TÜBİTAK

TÜBİTAK ARAŞTIRMA PROJESİ
PROJE SONUÇ RAPORU

Sonlu Metrik Uzayların Gromov Çarpımları İle İncelenmesi Ve Filogenetik Uygulamaları

Program Kodu: 1001

Proje No: 118F412

Proje Yürütücüsü:
Prof. Dr. Ayşe Hümeyra Bilge

Araştırmacı(lar):
Prof. Dr. Ebru Demet AKDOĞAN
Prof. Dr. Mehmet Şahin KOÇAK
Dr. Derya ÇELİK

Bursiyer(ler):
Arash Mohammadian REZAEİNAZHAD

Mart 2021
İstanbul

Önsöz

Sonlu metrik uzaylar, sonlu bir X kümesi üzerinde tanımlı bir d uzaklık fonksiyonu ile karakterize edilen uzaylardır. Bu uzayların sınıflandırılması gerek matematiksel gerekse özellikle filogenetikte ortaya çıkan uygulamaları açısından da ilginç bulunmaktadır. Projede sonlu metrik uzayların sınıflandırılması, çizge temsilleri ve filogenetik sınıflandırma problemlerine "Gromov çarpımları" vasıtasıyla yeni bir yaklaşım getirilmiştir.

Sonlu metrik uzayların sınıflandırılması, bu kümenin, "hipersimpleks bölümlemesi" olarak adlandırılan kanonik bir bölümlemesi ile yapılmaktadır. Literatürde n>6 elemanlı uzaylar için herhangi bir sonuç bulunmamaktadır. Çalışmamızda, Gromov çarpımı yapısı ve dörtgen yapısı ile tanımlanan sınıflamaların sınıflamasının hipersimpleks bölümlemesi ile ilintisi incelenmiş, 8 elemanlı uzaylar için Gromov çarpımı sınıfları ve 6 elemanlı uzaylar için dörtgen yapıları elde edilmiştir.

Filogenetik uygulamalar konusunda da, verilen bir ağaç yapısı ile temsil edilip edilmeyeceği/ne ölçekte temsil edilebileceği konusunda kriterler getirilerek incelenen örneklerde ağaç yapısından ayrılan döngülerin varlığı gösterilmiştir.

Bu proje Mart 2019 ve Mart 2021 tarihleri arasında TÜBİTAK ARDEB, MFAG 118F412 numaralı proje desteği ile gerçekleştirilmiştir. Proje yürütücüsü, araştırmacıları ve bursiyerleri katkıları için TÜBİTAK'a teşekkür ederler.

Özet

Sonlu metrik uzaylar, sonlu bir X kümesi üzerinde tanımlı bir d uzaklık fonksiyonu ile karakterize edilen uzaylardır. Projede, sonlu metrik uzayların sınıflandırılması ve çizge temsilleri problemlerine “Gromov çarpımları” vasıtasıyla yeni bir yaklaşım getirilmiş ve filogenetik analiz hesaplarına uygulanmıştır.

Sonlu metrik uzaylar, üçgen eşitsizliklerini sağlayan ve negatif olmayan noktalar kümesi ile tanımlanır ve bir polihedral koni oluştururlar. Sonlu metrik uzayların sınıflandırılması, bu kümenin, “hipersimpleks bölümlemesi” olarak adlandırılan kanonik bir bölümlemesi ile yapılmaktadır. Literatürde, n=4,5,6 için hipersimpleks bölünlemesi bilinmektedir ancak n>6 elemanlı uzaylar için herhangi bir sonuç bulunmamaktadır. Çalışmamızda, n-elemanlı metrik uzayların, sınıflandırma problemine, Gromov çarpımı ve dörtgen yapısı yöntemleri ile yaklaşılarak $n \leq 6$ için dörtgen yapısı sınıflaması ve $n \leq 8$, için Gromov çarpımı sınıflamaları elde edilmiştir.

Biyolojik sistemlerin incelenmesinde bir araç olarak kullanılan filogenetik ağaçlar, bir türün varyantları arasındaki mesafelerden hareketle, birbirleri arasındaki geçişleri ve oluşum aşamalarını temsil eden çizgelerdir. Literatürde, bu çizgelerin ağaç yapısında olduğundan hareketle çeşitli yöntemler geliştirilmiş, ancak ağaç yapısı varsayımının geçerli olmadığı durumlarda zorluklarla karşılaşılmıştır. Projede, ağaç varsayımı yapılmadan, verilen bir ailenin gen/protein yaısındaki dizilimlerden hareketle hesaplanan mesafe fonksiyonları hesaplanmış, sonlu bir metrik uzay olarak Gromov çarpımı ve dörtgen yapıları çıkarılmış, ağaç yapısına sahip olma/olmama ve döngülerin varlığına karar verme konusunda yaklaşıklık kriterleri getirilmiştir.

Anahtar Kelimeler: Sonlu metrik uzaylar, Optimal temsiller, Filogenetik sınıflamalar

Abstract:

Finite metric spaces are characterized by a metric d defined on a finite set X. In the framework of the project, we propose a new approach to the problem of classification and graph representation of finite metric spaces in terms of “Gromov products” and apply these methods to phylogenetic analysis problems.

In a metric space with n elements, distances are positive numbers satisfying triangle inequalities; they form a pointed polyhedral cone in $R^{n(n-1)/2}$, defined by the triangle inequalities among the d(x,y). The “hypersimplex” classification of finite metric spacies is a canonical decomposition of the metric cone via an auxiliary polyhedron. In the literature, hypersimplex decomposition is given for n=4,5,6.but there is no systematic analysis for n>6. In the present work we approached the classification problem in terms of Gromov product structures and qquadrangle structures and we obtained the quadrangle structure decompositions for n≤6 and the Gromov product structure decomposition for n≤8.

Phylogenetic trees, as a tool for the study of bilogical systems, are graphs that aim to represent the evolution and interelation in biological systems, sytarting from their dissimilarities. In the literatürde, methods based on the assumption of the exsitence of a tree representation failed to explain certain evolutionary mechanisms and researchers were interested in more general network models. In the framwork of the project we started from gene/protein sequences to compute metric functions, we computed Gromov product and quadrangle structures and we obtained criteria for deciding on the existence/non-existence of tree structures and on the nıumbers and places of loops in the graph representation of these metrics.

Keywords: Finite metric spaces, Optimal representations, Phylogenetic classifications

*

Giriş

Sonlu metrik uzaylar, n elemanlı bir küme üzerinde mesafe fonksiyonları tanımlayarak oluşturulan yapılardır. Noktalar arasındaki mesafeler, üçgen eşitsizliklerini sağlayan pozitif sayılar olup N=n(n-1)/2 olmak üzere, R^N uzayında, “metrik fan” adı verilen polihedral bir koni oluştururlar. Metrik uzayların sınıflandırılması konusunda literatürdeki genel yaklaşım, üçgen eşitsizlikleri ile tanımlanan bu polihedral koninin bölümlenmesinin incelenmesidir. Projede, metrik konunun bölümlenemesi için Gromov çarpımı yapıları, dörtgen yapıları kullanılarak metrik sınıflarının parametrize edilmesi, İş Paketi 1 kapsamında incelenmiştir. Bu iş paketi çerçevesinde elde edilen sonuçların ayrıntıları Ek 1.1-7 de verilmiştir.

Sonlu metrik uzayların incelenmesinde, optimal temsil adı verilen, metrik uzayın toplam kenar uzunluğu minimum olan bir çizgeye gömülmesi problemi önemli bir yer tutmaktadır. Literatürde, optimal temsilin varlığı ve bazı özel durumlarda nasıl elde edileceği konusunda bazı sonuçlar bulunmaktadır ancak bu konu hala açık ve zor bir problem olarak görülmektedir. Projede, İş Paketi 2 kapsamında, verilen bir Gromov çarpımı veya dörtgen yapısı için, metriklerin “lokal optimal” olarak adlandırdığımız temsilleri ve benzer optimale yakın indirgeme yöntemleri incelenmiştir. Bu iş paketi çerçevesinde elde edilen sonuçlar Ek 2.1-3 de verilmiştir.

Biyolojik sistemlerde türlerin arasındaki farklılıklar çeşitli yöntemlerle bir “farklılıklar matrisi” (dissimilarity matrix) ile ifade edilmektedir. Bu farklılıkladan hareketle, türler arasındaki ilintiler bir hiyerarşik ağaç yapısına oturtulmaktadır. Bu amaçla, türlerin DNA veya protein dizilimlerinin uygun bölgelerinden alınan örnekler karşılaştırılarak çeşitli farklılık matrisleri oluşturulmaktadır. İş Paketi 3 kapsamında, çeşitli örnekler için literatürde kullanılan çeşitli farklılık matrisleri hesaplanmış, bunların üçgen eşitsizliği koşullarını sağlayıp sağlamadıkları kontror edilmiş, matematiksel anlamda metrik uzay oluşturanlar için Gromov çarpımı yapıları incelenmiştir. Bu iş paketi çerçevesinde elde edilen sonuçlar Ek 3.1-3 de verilmiştir.

Biyolojik sistemlerin farklılık matrislerinin bir çizge olarak yapılandırılmasında, literatürde kullanılan hiyerarşik ağaç yapısı oluşturma yerine, farklılık matrisinin tanımladığı metrik uzayın çizge temsillerinin indirgenmesi yöntemi kullanılmıştır. Ancak, ders kitaplarında verilen basit örnekler dışına çıkıldığında, incelenen türlerin genel olarak jenerik metrikler olduğu ve matematiksel kesinlik çerçevesinde ağaç benzeri yapılara indirgenmediği görülmüştür. Öte yandan, bu metriklerin belli bir yaklaşıklık çerçevesinde ağaç yapılarına yakın temsilleri olduğu görülmüştür. İş Paketi 4 çerçevesinde, filogenetik analizde karşımıza çıkan metriklerin hangi kriterlere göre, hangi yaklaşıklıklar dahilinde ağaç yapısında kabul edilebilecekleri konusunda somut ölçütler getirilmiştir. Bu iş paketi çerçevesinde elde edilmiş olan sonuçlar Ek 4.1 de verilmiştir.

Sonlu metrik uzayların Gromov çarpımları ve dörtgen yapıları ile incelenmesi yeni bir yaklaşımdır. Bu yaklaşımla, 6, 7 ve 8 noktalı uzaylar için bir eşdeğerlik sınıfı dekompozisyonu elde edilmiştir. Literatürde n>6 için de sonuçlar olmaması nedeniyle sınıflandırmaya yönelik genel sonuçlar verebilecek etkin yöntemlerin geliştirilmesi özgün bir katkı olmuştur. Gromov çarpımı matris temsilleri ve çeşitli invaryantları teknik bir kolaylık olarak işlev görmüştür. Gromov çarpımları vasıtasıyla indirgemeler yolu ile optimal/suboptimal temsiller elde edilebileceği gösterilmiştir. Filogenetik sınıflamaların ağaç varsayımından bağımsız olarak yapılması incelenmiştir.

Proje çalışmalarına ilişkin literatür 2. Bölümde tartışılmıştır. 3. Bölümde, projede kullanılan yöntemler tanıtılmıştır. 4. Bölümde bulgular sunulmuş, 5. Bölümde tartışılmıştır. Sonuç ve öneriler 6. Bölümde verilmiştir. Proje süresince 1 makale yayına sunulmuş ve kabul almıştır. Diğer bir makale halen hakem değerlendirmesi sürecindedir.

Hazırlanmakta olan makaleler ve ayrıntılı matematisel metinler ekler halinde sunulmuştur. Proje çerçevesinde elde edilen sonuçlar proje grubunun uzun süreli çalışmalarının son aşamasını oluşturmaktadır. Bu nedenle, teorik çerçevenin bütünlük içinde sunulması amacı ile, gerekli yerlerde referanslar verilerek, ön çalışmalara elde edilmiş olan sonuçlar da rapora dahil edilmiştir. Proje eklerinin bir bölümü içerik olarak birbirleri ile ilintilidir ancak bağımsız olarak hazırlandıkları için bir takım örtüşmeler içermektedirler. Proje ekleri listesi aşağıdadır.

Ek-1.1 Establishing the connection between the hypersimplex classification, Gromov product structures and quadrangle structures

Ek-1.2 Five point metric spaces, Gromov product structures, quadrangle structures and explicit parametrizations

Ek-1.3 Parametrization of 6-point Metric Space Type I17 using Quadrangle Structure Parameters

Ek-1.4 Gromov product decomposition of 7-point metric spaces

Ek-1.5 Gromov product structures for eight point metric spaces

Ek-1.6 Quadrangle structures for 6-point spaces

Ek 2.1 Gromov product structures, quadrangle structures and split metric decompositions for finite metric spaces

Ek 2.2 Optimal realizations, h-optimal realizations and tight spans of metric spaces

Ek-2.3 Tight Reductions of Finite Metric Spaces

Ek-3.1 Examples of Distance Matrices

Ek-3.2, Non-generic metrics in phylogenetics, the Bee family

Ek-3.3 A New Method for Constructing Phylogenetic Trees

Ek-4.1 Phylogenetic Data for Biological Organisms

Literatür Özeti:

Literatürde, sonlu metrik uzaylar, sınıflandırma (Strumfels ve Yu, 2004), Euclid uzaylarına gömme, çizge temsilleri (Imrich et al. 1984) gibi. açılardan incelenmiş olup, filogenetik sınıflandırmalar, internet ağları vb. alanlara uygulamaları bulunmaktadır.

Metrik Uzaylar: Metrik uzaylar bir X kümesi üzerinde tanımlı d uzaklık reel fonksiyonu ile belirlenir ve (X,d) şeklinde gösterilir. Metrik uzaylarda uzaklık fonksiyonu, (i) $d(x,y)\geq 0$, $d(x,y)=0$ ancak ve ancak $x=y$, (ii) $d(x,y)=d(y,x)$; (iii) $d(x,y)+d(y,z)\geq d(x,z)$ koşullarını sağlar. Koşul (iii) üçgen eşitsizliği olarak da isimlendirilir. Eğer $d(x,y)=0$ olduğu halde x ve y farklı olabiliyorsa, d "yalancı (pseudo)-metrik" olarak adlandırılır. Sonlu metrik uzayların incelenmesinde önemli bir yer tutan split metrikler bu yapıdadır. Split metrikler sonlu X uzayının iki parçaya bölümlenip, farklı parçalarda yer alan noktaların uzaklıklarının 1 ve aynı parçada olanların uzaklıklarının 0 tanımlanmasından ortaya çıkmaktadır. Üçgen eşitsizliğinden daha güçlü olan "$\max(d(x,y),d(y,z)) \geq d(x,z)$" koşulunu sağlanıyorsa d, "ultra-metrik" olarak adlandırılır. Ultra metrikler filogenetik uygulamalarında kullanılmaktadır. Bizim yaklaşımımız, metrik uzayları, $d(x,y)$ uzunlukları yanısıra, Gromov çarpımı olarak adlandırılan $\Delta(x,y,z) = (1/2)(d(x,y)+d(x,z)-d(y,z))$ büyüklüklerinden hareketle tanımlamak ve sınıflandırmaktır. Bir diğer yaklaşım da, uzayın $\{x,y,z,w\}$ gibi dörtlü alt kümelerine, diagonal olarak tanımlanan $\max\{d(x,y)+d(w,z),\ d(x,z)+d(w,y),\ d(x,w)+d(y,z)\}$ büyüklüklerinden dörtgenler yapısı belirleyip sınıflandırmaktır.

Metrik koni sınıflaması: X kümesi eğer n elemanlı ise, metrik yapısı, bu noktaların birbirine uzaklıkları olan $N=n(n-1)/2$ adet pozitif sayı ile belirlenir. Ancak bu sayılar, $\Delta(x,y,z)=(1/2)(d(x,y)+d(x,z)-d(y,z))\geq 0$ olacak şekilde üçgen eşitsizliklerini sağlamaktadır. Dolayısıyla, n elemanlı metrik uzaylar, R^N boyutlu bir uzayda $d(x,y)\geq 0$ ve $\Delta(x,y,z)\geq 0$ ile verilen eşitsizlikleri ile belirlenen bir polihedral koni oluşturmaktadır. Her n için, bu koninin yapısının belirlenmesi, n elemanlı metrik uzayların sınıflandırılması olarak adlandırılmaktadır. Metrik koninin farklı açılardan sınıflandırılmaları mümkündür. Literatürde "hipersimpleks" sınıflaması n=6 noktalı metrik uzayların sınıflamasına uygulanmıştır. Proje ekibinin çalışmalarında ise Gromov çarpımları ve dörtgen yapıları eşdeğerlik sınıfları tanımlanmıştır. Her üç yapı açısından da 3 ve 4 elemanlı metrik uzayların tek bir eşdeğerlik sınıfı, 5 elemanlı metrik uzayların ise 3 eşdeğerlik sınıfı olduğu bilinmektedir. 6 elemanlı uzaylar ise hipersimplex sınıflandırılmasına göre 339, Gromov çarpımları sınıflandırılmasına göre 26 ve dörtgen yapısı sınıflandırılmasına göre 210 sınıftan oluşmaktadır. n=7 için Gromov çarpımları sınıflamasına göre 431 sınıf tarafımızdan elde edilmiştir. n=8 için ise 11470 Gromov çarpımı sınıfı bulunmuştur. Literatürde, bunların dışında sınıflandırmaya yönelik genel sonuçlar yoktur.

Metrik uzayların çizge temsilleri: Sonlu metrik uzaylar, (G,w) ile gösterilen ağırlıklı bir tam çizge olarak da düşünülebilir. Bu yapıda (X,d) uzayın elemanları çizgenin noktaları, kenarların ağırlıkları da bu noktalar arasındaki uzaklıklardır. Metriklerin çizge olarak realizasyonları literatürde çeşitli yönleri ile incelenmiş bir konudur. Ağaçlar halka içermeyen çizgeler olup, bir çizgenin ağaç yapısında olması için gerek ve yeter koşul, her 4 elemanlı kümesinin ağaç yapısında olmasıdır. 4-nokta koşulu dediğimiz başka bir koşulda, bir çizgenin ağaç olup olmadığınl şöyle vermektedir: Bir çizge ağaçtır ancak ve ancak uzaydaki her nokta için $\max\{d(x,z)+d(w,y),\ d(x,w)+d(y,z)\} \geq d(x,y)+d(w,z)$ (Semple ve Steel, 2003). Bu koşul dörtgen yapısındaki bütün dörtgenlerin dejenere, yani dörtgenlerin birden fazla diagonale sahip olmaları ile aynı almanda ve dörtgen yapısının önemini, çizgelerin ağaç olup olmamalarında ortaya çıkarmaktadır.

Optimal temsiller: Tam çizgede, bütün köşeleri birbirine bağlayan kenarlar olduğu için, kenarların ağırlıklarının toplamı, metrik uzayın noktaları arasındaki mesafelerin toplamına eşittir. Ancak, tam çizgede, bir noktadan çıkan kenarları birleştirmek ve dolaylı olarak realize olan mesafeler varsa doğrudan bağlantı sağlayan kenarları atarak çizgenin toplam ağırlığını azaltmak mümkündür (Bilge vd., 2015). Bu şekilde, yeni noktalar eklenmiş, noktalar arasındaki mesafeleri koruyan, tam olmayan bu çizge, metrik uzayın bir temsilidir. Bu temsiller arasında, toplam ağırlığı en az olan, “optimal temsil” olarak adlandırılmaktadır. Yukarıda bahsedilen ve “Δ-hamleleri” olarak adlandırılan bu işlemlerle her adımda çizgenin kenarlarının ağırlıkları toplamı azalacaktır ancak bu işlem sonucuna “optimal” olarak adlandırılan, kenar ağırlıkları toplamı en az olan çizgeye varılıp varılmayacağı bilinmemektedir. Optimal temsiller ilk önce Hakimi ve Yau (Hakimi, 1964, Hakimi ve Yau, 1964) tarafından tanımlanmıştır. Daha sonraları filogenetik (Landry vd., 1996), psikoloji (Cunningham, 1978), compression software (Li vd.,2004), internet tomografisi (Chung vd., 2001) gibi alanlarda uygulanmıştır. Optimal realizasyonların inşası konusundaki çalışmalar çok fazla değildir (Imrich vd., 1984; Koolen vd., 2009; Koolen vd., 2012; Hertz ve Varone, 2008; Dkatohress vd., 2010). Her sonlu metrik uzayın en az bir optimal realizasyonu olduğu gösterilmiştir (Dress, 1984; Simres-Pereira, 1984). Ancak bu optimal realizasyonu inşa etmenin NP-zor bir problem olduğu bilinmektedir (Althöfer, 1988; Winkler, 1988). Ayrıca, optimal temsillerin bulunması konusunda bazı sezgisel algoritmalar geliştirilmiştir (Simres-Pereira, 1984; Dress vd., 2010; Varone, 2006).

Split metrikler: X metrik uzayı A ve B şeklinde iki ayrık kümenin bileşimi olarak yazıldığında, bu kümelere karşı gelen $\delta_{(A,B)}$ (yalancı) metriği eğer x ve y aynı küme içinde ise 0, farklı kümelerde ise 1 olarak tanımlıdır (Dress, 1984). Split metrikler, sonlu metrik uzayların incelenmesinde önemli bir yer tutmaktadır. Split metrikler, kombinatoryel anlamda sınıflamada, üçgen eşitsizlikleri ile tanımlanan polihedral koninin metrik sınıfını tanımlayan bölümlerinin ayrıtları arasındadır. Ancak polihedral konininin tüm ayrıtları split metrik şeklinde değildir. Bunlar ise split-asal (split-prime) şeklinde yazılabilmektedir. Verilen bir metriğin split ve split-asal metriklerin toplamı şeklinde yazılması tek türlü olmayabilir. Proje kapsamında, split metrik dekompozisyonu dörtgen yapısı ve Gromov çarpımları arasındaki sıralama bağıntıları kullanılarak incelenmiştir. Bu doğrultuda yine dörtgen yapılarını kullanarak, metriğin dekompozisyonda hangi splitlerin olup olmayacağı tesbit edilebilir ancak 1-splitler (yani A yada B kümesinin tek elemanlı olduğu durumdan ortaya çıkan split) dışındakılerin katsayılarını bulmak ancak hesaplamalar yapıldıktan sonra belirlenmektedirler. Ancak 1-splitlerin katsayıları minimal Gromov çarpımlar olarak belirlenmektedir (Bilge vd., 2021).

Filogenetik yapılar: Biyolojik canlılardaki evrimsel süreç çoğu zaman filogenetik bir ağaç topolojisi ile tanımlansa da, evrimsel değişimin daha karmaşık olduğu durumlarda, örneğin hibridizasyon, horizontal gen transferi (HGT), rekombinasyon, gen düplikasyonu ya da kaybı gibi biyolojik olaylar, türler arasındaki değişim ifadesini sağlayan iki dallı (bifurcate) ağaç modellerini geçersiz ya da az güvenilir yapmaktadır. Çözüm olarak filogenetik ağ yapı modelleri ortaya atılmıştır (Posada ve Crandall*.,* 2001; Morrison, 2005). 2018 yılında, Science Advances dergisinde yayınlanan bir çalışma, balinaların evrimsel değişiminin bir filogenetik ağ yapısı ile açıklanabileceğini ortaya koymuştur (Arnason vd*.,* 2018). Çubuk balinalar (baleen whale) da denilen bu türün evrimsel değişimi hem moleküler düzeyde hem de morfolojik düzeyde yapılan analizlere rağmen açıklığa kavuşturulamamıştır. Altı farklı balina türünden alınan farklı gen bölgelerinin karşılaştırılması sonucu elde edilen uzaklık verilerinin farklı filogenetik ağaçlar doğurması, ağaç topolojisinin balinalardaki evrimsel değişimi açıklamada yetersiz kaldığını göstermiştir. Genetik çalışmaların sonucunda, buna sebep olan kritik olay, türlerden bazılarının ataları da dahil olmak üzere, türler arasında çok sayıda gerçekleşmiş olan gen transferleri olmuştur.

Filogenetik ağaç yapısına uymayan türlerin başında bakteriler gelmektedir. Bakterilerdeki genetik çeşitliliğin büyük bir bölümü horizontal gen transferi ile ortaya çıkmıştır (Kurland vd., 2003; Nakamura vd., 2004; Ochman vd., 2000). Bunun yanı sıra, bitkilerde de çok sayıda horizontal gen transferine rastlanmıştır (Bergthorsson vd., 2003, 2004; Mower vd., 2004). Yaygın görülen bir diğer biyolojik olay virüslerdeki türler arası rekombinasyondur (Posada ve Crandall, 2002). Son olarak, bazı balık ve kurbağa türlerindeki en temel evrimsel mekanizmanın melez türleşme olduğu belirlenmiştir (Mallet, 2007). Tüm bu karmaşık mekanizmalar, retikülat, diğer bir deyişle ağ şeklinde evrimsel olaylar olarak adlandırılmıştır. Retikülasyon gerçekleştiğinde, iki veya daha fazla birbirinden bağımsız türlerin arasında topolojik bir bifürkat ağacın tanımlayamayacağı karmaşık bir ilişki ortaya çıkmaktadır. Bu durum, filogenetik ağ yapı modellerini kaçınılmaz yapmıştır ve bu konuda çok sayıda farklı ağ yapı modelleri geliştirilmiştir (Gemeinholzer, 2008; Huson ve Bryant, 2006). Bunlardan biri de, “split ağ” modelidir. Bu model, DNA ya da protein dizilerinin hizalanması sonucunda ortaya çıkan veri kaynağındaki uyuşmazlıkların sonucunda birden fazla olası ağaç modelinin ortaya çıktığı durumlar için geliştirilmiş bir ağ yapı modelidir. Bir split ağ yapı modelinde ara nodlar her durumda ortak ataları ifade etmeyebilir.

Gereç ve yöntem

Metrik uzaylar: Sonlu metrik uzaylar, n elemanlı bir X kümesi üzerinde tanımlı bir d uzaklık fonksiyonu ile karakterize edilen uzaylardır. Bu tür bir uzay (X,n,d) ile gösterilmekte olup d fonksiyonu aşağıdaki koşullar sağlamaktadır.

X kümesinin her P_i, P_j, P_k elemanları için

1) $d(Pi,Pj) \geq 0$, ve $d(Pi,Pj)=0$ ancak ve ancak $P_i=P_j$
2) $d(P_i,P_j)=d(P_j,P_i)$, (simetri özelliği)
3) $d(P_i,P_j)+d(P_i,P_k) \geq d(P_i,P_k)$, (üçgen eşitsizliği)

Eğer (1) numaralı özellikte, $P_i \neq P_j$ için, $d(Pi,Pj)=0$ olabiliyorsa d "sözde-metrik" (pseudo-metric) adını alır. Aşağıda tanımlanan metrik koninin iç noktaları yukarıdaki tanım anlamında metriklerdir. Metrik koninin sınır noktalar ise sözde-metriklerdir. Ancak çalışmamızda sözde-metrikler önemli bir yer tuttuğu ve yaygın olarak kullanışdığı için metrik ve sözde-metrik ayrımı yapılmayacak, tümü "metrik" olarak isimlendirilecektir.

Metrik koni: Yukarıdaki koşullardan görüleceği gibi, n-elemanlı bir (X,n,d) metrik uzayı, N=n(n-1)/2 olmak üzere, R^N uzayında negatif olmayan ve üçgen eşitsizliklerini sağlayan noktadan oluşmaktadır. Yukarıda verilen lineer eşitsizliklerin çözüm kümesi "metrik koni" olarak adlandırılan bir polihedral koni oluşturmaktadır. Yarı doğrular tarafından üretilen koniler "pointed cone" olarak adlandırılmaktadır. Ancak çalışmamızda başka tür koniler gözönüne alınmadığı için sadece "koni" terimi kullanılacaktır.

Metrik koninin ayrık polihedral alt-konilere dekompozisyonu eşdeğerlik sınıfları oluşturmaktadır. Bu polihedral alt koniler, üçgen eşitsizliklerine kanonik bir şekilde yeni lineer eşitsizlikler ilave edilmesi ile oluşturulmaktadır. Bu alt-konilerin R^N de açık kümeler olması gereklidir. X'in elemanlarının permütasyonu altında birbirine gönderilen alt-koniler eşdeğer olarak tanımlanmaktadır. Bizim yaklaşımımızda, Gromov çarpımı yapıları, dörtgen yapıları (quadrangle structures), beşgen yapıları (pentacle structures) ve daha genel olarak k<n boyutlu alt uzayların yapılarından oluşan k-gen yapıları ile eşdeğerlik sınıfları tanımlanmıştır.

Ekstrem metrikler ve metrik koninin dekompozisyonu: Metrik koninin ayrıtları olan yarı doğrular, "ekstrem metrikler" olarak adlandırılmaktadır. Ekstrem metrikler arasında "split-metrikler" adı verilen bir sınıf özel bir yer tutmaktadır. Çalışmamızda, Gromov çarpımı ve dörtgen yapıları kullanılarak split-metriklerin elde edilebileceği gösterilmiştir.

Split-metrik yapısında olmayan ekstrem metrikler is "split-asal" metrikler olarak adlandırılmaktadır. Split-asal metriklerin geometrik yöntemlerle elde edilmesi konusunda literatürde n=5,6,7 için çalışmalar mevcuttur. Çalışmamızda split-asal metriklerin elde edilmesi için bir algortima tasarlanmış ve n=5 ve n=6 durumlarına uygulanmıştır.

n-elemanlı bir metrik uzay üzerinde tanımlı herhangi bir eşdeğerlik sınıfı, yani bir takım eşitsizliklerle tanımlanan bir alt koni, ayrıtlarınını oluşturan vektörlerin pozitif sayılarla gerdiği uzay olarak düşünülebilir. Eğer herhangi bir eşdeğerlik sınıfı N sayıda ayrıtı olan bir koni ile tanımlanıyorsa o zaman ayrıtlar bir baz takımı oluşturur. Bu durumda eşdeğerlik sınıfının her elemanı, yani alt-koninin her noktası N sayıda pozitif parametre cinsinden ifade edilmiş olur. Genel olarak M>N sayıda ayrıtla tanımlanan bir alt-koninin veya eşdeğerlik sınıfının parametrize edilmesi, bu alt koninin N tane ayrıt tarafından üretilen alt-konilere ayrıştırılmasıdır. Verilen bir sınıf için split metrikler ve split asal metriklerin bulunması bu anlamda verilen sınıfı parametrize etmek anlamına gelir. Çalışmamızda, Gromov çarpımı, dörtgen yapısı ve beşgen yapısı kullanılarak metriklerin parametrize edilmesi için bir yöntem geliştirilmiş ve n=5 durumuna ve n=6 için bazı sınıflara uygulanmıştır.

Metrik eşdeğerlik sınıflarının parametrizasyonu: Metriklerin parametrizasyonu için kullandığımız yöntem, Gromov çarpımı yapıları, dörtgen yapıları, beşgen yapıları ve genel olarak n-k elemanlı alt uzayların yapıları kullanılarak elde edilecek sınıflamanın halen literatürde verilmiş olan ve aşağıda tanımlanacak olan hipersimpleks sınıflamasından farklı olabileceğini göstermiştir. Çalışmamızda, Gromov çarpımı sınıflarının hipersimpleks sınıfları ile uyumlu olduğu yani hipersimpleks eşdeğerlik sınıflarının Gromov çarpımı eşdeğerlik sınıflarının inceltmesi (refinement) olduğu gösterilmiştir. Öte yandan, dörtgen sınıfları ile hipesimpleks sınıfları örtüşebilmektedir. Çalışmamızda önerilen yöntemin hipersimpleks sınıflamasından farkı, sınıflamanın doğrudan doğruya o sınıftaki metriklerin kanonik bir parametrizasyonunun vermesidir.

Çalışmalarımızda, n≤8 elemanlı metrik uzayların Gromov çarpımı sınıfları elde edilmiştir. n=5 için 3, n=6 için 26, n=7 için 431 ve n=8 için 11470 eşdeğerlik sınıfı bulunmuştur. Bu sınıfların nasıl elde edildiği ve eşdeğerlikleri aşağıda ayrıntılı olarak verilecektir. Bu sınıflandırmanın elde edilmesinde, Gromov çarpımı zincirleri, Gromov yapısını matrisi gibi kavramlar önem kazanmaktadır. Proje kapsamında, n=6 için dörtgen yapıları belirlenmiş ve seçilmiş örnekler için parametrizasyonlar elde edilmiştir.

n=5 için bilinen 3 sınıf önerdiğimiz yöntemlerle incelenmiş, parametrizasyonları, aşağıda açıklanacak olan optimal temsilleri elde edilmiş ve eşdeğerlik sınıflarının hacimsel büyüklükleri Monte Carlo simülasyonu yöntemi ile hesaplanmıştır. n=6 için ise, daha önce elde edilmiş olan 26 Gromov çarpımı sınıfınındörtgen yapıları çıkarılmış bu şekilde 210 adet dörtgen yapısı sınıfı bulunmuştur. Açık parametrizasyon elde edilen örnklerden biri farklı hipersimpleks tiplerinin aynı dörtgen yapısı ile temsil edilip parametrize edilebileceğini göstermiştir. Diğer örnek ise, dörtgen yapısının 6 elemanlı metrikleri parametrize etmek için yeterli olmadığını, beşgen yapılarına ihtiyaç duyulduğunu göstermiştir. Bu örnekten hareketle, n elemanlı bir metriğin parametrize edilmesinde n-k elemanlı alt kümelerinin yapılarına dayalı bir algortima önerilmiştir.

Optimal ve lokal optimal temsiller: Sonlu metrik uzaylar, ağırlıklandıralmış çizgeler olarak da düşünülebilir. Bu bağlamda, n elemanlı bir metrik uzak, n düğümlü bir tam çizge ile temsil edilir. Metrik uzayın elemanları arasındaki uzaklıklar da, çizgenin kenarlarının ağırlıklarını oluşturur. Verilen metrik uzay n+k düğümlü herhangi bir çizge ile de temsil edilebilir. Bu durumda, çizgenin düğümleri arasındaki en kısa yol metrik uzayda o iki nokta arasındaki mesafeye eşit olmalıdır. Verilen metriğin temisil edildiği kenarların ağırlıkları toplamı en küçük olan çizge, o metriğin optimal temsili olarak adlandırılır. Optimal temsil problemi literatürde sadece n=3,4,5 elemanlı metrik uzaylar için tam olarak çözülmüştür. Projede ele aldığımız “lokal optimal temsil” kavramı ise, (Bilge vd, 2015) de tanımladığımız “hamle” kavramı ile ilintilidir. Verilen metriğin herhangi bir temsili üzerinde, mesafeleri koruyacak şekilde yapılan kenarları birleştirme ve başka dolaylı yollar tarafından gerçeklenen mesafeler için gereksiz kenarları atma işlemleri "hamle" olarak adlandırılmaktadır. Bir takım hamle dizileri sonunda elde edilen temsilde çizgenin kenarlarının toplam ağırlığı daha fazla azaltılamıyorsa, elde edilen temsil "lokal optimal" olarak adlandırılmaktadır. Çalışmamızda, verilen metriğin parametrizasyonundan hareketle lokal optimal temsilinin elde edilmesi algoritmik olarak verilmiş ve n=5 durumundaki 3 tip ile, n=6 için seçilmiş örneklere uygulanmıştır.

Filogenetik problemlerinde ortaya çıkan metriklerin incelenmesi: Filogenetikte türler arasındaki farklar bir "benzemezlik matrisi" halinde ifade edilerek türler arasındaki geçişler hiyerarşik bir ağaç yapısı halinde temsil edilmektedir. Her ne kadar bir mütasyonlarla oluçan bir ailenin elemanlarının ağaç yapısı içine yerleştirilebilmesi beklense de, "yatay gen transferi" durumunda aileyi temsil eden çizgede döngüler olmaktadır.

Güncel literatürde bu farklılıklar türlerin DNA veya protein dizilimlerinin belli bölgelerinden alınan diziler karşılaştırılarak çeşitli benzemezlik matrisleri oluşturulmaktadır. Bu benzemezlik matrisleri bazı durumlarda üçgen eşitsizliklerini sağlamamaktadırlar. Proje kapsamında, çeşitli örnekler için doğrudan gen bankalarından alınan diziler işlenerek literatürde kullanılan yöntemlerdle benzemezlik matrisleri oluşturulmuştur. Bu yöntemlerle hesaplanan matrislerin hangilerinin üçgen eşisizliklerini sağladığı kontrol edilmiştir. Aynı aileden farklı yöntemlerle elde edilen benzemezlik matrislerinin yaklaşık %10 oranında farklılık gösterdiği gözlenmiş, ve bu oran aşağıda açıklanacak olan modelleme aşamasında da yaklaşıklık kriteri olarak esas alınmıştır.

Biyolojik canlılardaki evrimsel değişimin filogenetik modellenmesi: Yukarıda açıklanan yöntemlerle elde edilen benzemezlik matrisleri sonlu bir metrik uzayın mesafe fonksiyonları olarak alınarak seçilen örnekler için Gromov çarpımı yapıları ve dörtgen yapıları hesaplanmıştır. Bu yapılardan hareketle türlerin ağaç veya şebeke yapısını çıkarmak için 2 yöntem benimsenmiştir. Her iki yöntemde de hesaplamalar %10 mertebesinde bir yaklaşıklıkla yapılmıştır. İlk yöntem ise, daha çok elemanlı ailelerin incelenmesinde kullanılmıştır. Sonlu bir metrik uzayın ağaç yapısında bir metrikle temsil edilebilmesi için gerek ve yeter şartın "4-nokta koşulu" olarak adlandırılan ve aşağıda ayrıntıları verilecek olan bir koşul olduğu bilinmektedir. Seçilen örneklerde bu koşulun %10 mertebesinde sağlanıp sağlanmadığı kontrol edilmiştir. Filogenetik örneklerde çok az sayıda dörtgen haricinde 4-nokta koşulunun %10 mertebesinde sağlandığı gözlenmiştir. Bu koşulun sağlanmadığı dörtgenler ise o türler arasında bir yatay gen transferi olasılığını gündeme getirmektedir. Proje çerçevesinde seçilen örneklerin ağ yapıları elde edimiştir.

Bulgular ve Tartışma

Proje kapsamında sonlu metrik uzaylar “Gromov çarpımı yapısı” olarak adlandırdığımız bir yöntemle incelenerek eşdeğerlik sınıflarına bölünmüştür. Bu eşdeğerlik sınıfları “dörtgen yapıları” kurularak daha ince (fine) bölümlere ayrılmıştır. Dörtgen yapıları kullanılarak incelenen sınıfın split-metrik yapısı elde edilmiştir. İncelenen sınıfın parametrizasyonu için ise n-elemanlı metrik uzayın n-k elemanlı alt uzaylarının yapısının bilinmesine gerek olduğu gösterilmiştir.

Gromov çarpımı ve dörtgen yapıları “jenerik” metrikler için tanımlıdır. Öte yandan filogenetik uygulamalarında karşılaşılan metrikler jenerik olmaktan çok uzaktır. Bu nedenle her iki grup için farklı yöntemlerin uygulanması gereklidir. Çalışmalarımızda kullanılan yöntemler ve elde edilen sonuçlar aşağıda açıklanmıştır.

Gromov çarpımı yapıları: Bir (X,d) metrik uzayı verildiğinde, X’in her P_i, P_j, P_k elemanları için Gromov çarpımları

$$\Delta_{ijk}=(1/2)\ [d(P_i,P_j)+d(P_i,\ P_k)-d(P_j,\ P_k)]$$

şeklinde tanımlanır. Eğer bu sayılar kümesinin tek bir minimum elemanı varsa, o uzay, “Δ-jenerik” olarak adlandırılır. Bu şekilde metrik uzayın her elemanı için o noktadaki minimal Gromov çarpımının seçimi, o uzay üzerinde bir Gromov çarpımı yapısı ismini alır.

Verilen bir P_i noktası için hangi Δ_{ijk} sayısının minimum olduğu, diğer noktalardaki minimum Gromov çarpanlarının neler olabileceğini kısıtlamaktadır (Bilge vd., 2017).

Önerme: Eğer P_i noktasında Δ_{ijk} minimal ise, P_j noktasında Δ_{jkl}, P_k noktasında Δ_{kjl} ve P_l noktasında Δ_{lij}, Δ_{lik} minimal olamaz.

Bu önerme ile verilen kısıtlamalardan hareketle, n elemanlı bir metrik uzay için Gromov çarpımı yapılarının listesini elde etmek mümkündür. Ancak bu şekilde elde edilen listede, birbirlerine permütasyonla dönüştürülebilecek çok sayıda yapı bulunmaktadır. Bu şekilde birbirlerine permütasyonla dönüşebilecek Gromov çarpımı yapılarının eşdeğerlik sınıfları Δ-eşdeğerlik sınıfları olarak adlandırılmaktadır.

Δ-eşdeğerlik sınıflarını bulma probleminde n elemanlı permütasyon grubunun etkisi altında eşdeğer olan sınıfların belirlenmesi önemli bir yer tutmaktadır. Gromov çarpımı yapılarının matris temsili bu problemin çözümünde önemli bir rol oynamaktadır (Bilge ve İncegül, 2018a).

Yukarıdaki önerme ile elde edilmiş olan eşdeğerlik sınıflarının sınıfları Δ-jenerik olmayabilmektedir. Δ-jenerik olmayan sınıfların belirlenmesi için bir yöntem (Bilge ve İncegül, 2018c)’de verilmiştir. Ayrıca, aşağıda açıklanacağı gibi lineer programlama kullanılarak da Gromov çarpımı yapısı ile verilen eşitsizliklerin çözüm kümesinin açık küme olup olmadığı dolayısıyla verilen sınıfın jenerik olup olmadığı kontrol edilebilmektedir.

Örnek olarak n=5 için, uzayın elemanlarını $P_1,..,P_5$ ile gösterirsek, aşağıdaki A,B, ve C tipi olarak adlandırılan 3 Gromov çarpımı yapısı elde edilmektedir. İndirgenebilirlik/indirgenemezlik kavramı, Gromov çarpımı yapılarının matris temsilleri ile tanımlanacaktır.

(A) $\{\Delta_{125},\ \Delta_{213},\ \Delta_{324},\ \Delta_{435},\ \Delta_{514}\}$, (indirgenemez yapı)

(B) $\{\Delta_{125},\ \Delta_{213},\ \Delta_{325},\ \Delta_{425},\ \Delta_{514}\}$, (indirgenemez yapı)

(C) $\{\Delta_{125},\ \Delta_{213},\ \Delta_{325},\ \Delta_{425},\ \Delta_{513}\}$, (indirgenebilir yapı)

Gromov çarpımı yapıları ve hipersimpleks sınıflaması: Literatürde metrik uzayların sınıflaması için, yukarıda da belirtildiği gibi hipersimpleks sınıflaması kullanılmaktadır. Proje çalışmalarında, hipersimpleks sınıflaması ve önermiş olduğumuz Gromov çarpımları yapısının ilintisi nin kurulmuştur. Bu konu ile ilgili ayrıntılı matematiksel açıklamalar **Ek-1.1, "Establishing the connection between the hypersimplex classification, Gromov product structures and quadrangle structures"** isimli metinde verilmiştir. Kısaca özetlemek gerekirse, hipersimpleks sınıflamasının "thracle" adı verilen, düğüm sayısı kenar sayısına eşit olan çizgeler kümesi verdiği, ve "thracle"lardan bir 3 elemanlı döngü ve n-2 dereceli bir düğümden oluşan özel grubunun Gromov çarpımı yapısına eşdeğer olduğu gösterilmiştir. Burada hareketle, hipersimpleks sınıflamasının Gromov çarpımı eşdeğerlik sınıflarının bir incelmesi (refinement) olduğu gösterilmiştir.

Hipersimpleks yapısı n elemanlı bir metrik uzayın n+1 elemanlı bir metrik uzayın alt uzayı olarak inşa edilmesi olarak da düşünülebilir. Elemanları $P_1,,..,P_n$ olan X uzayına yeni bir P_0 noktası ekleyelim ve $d(P_0,P_i)=x_i$ diyelim. Bu yeni uzayda üçgen eşitsizlikleri

$$x_i + x_j - d_{ij} \geq 0$$

koşullarını oluşturur. Bu sayılar ise P_0 noktasındaki Gromov çarpımlarıdır. P_0 noktasındaki Gromov çarpımlarının maksimal olarak jeneriklikten uzak olmasını istersek, tüm x_i lerin pozitif (özellikler sıfırdan farklı) sayılar olacak şekilde belirlenmesi gerekir. Bu ise, P_0 noktasındaki üçgen eşitsizliklerini $AX \geq d$ şeklinde yazacak olursak A matrisinin $A=[A_1\ A_2]^t$, d vektörünün $d=[d_1\ d_2]^t$ şeklinde $A_1X=d_1$, $A_2X \geq d_2$ olacak şekilde pozitif çözümlerinin bulunmasından ibarettir. Bu koşul ise, yukarıda bahsedilen "thracle" olarak adlandırılan çizgeler kümesinin varlığını gerektirmektedir.

Hipersimpleks sınıflaması görüldüğü gibi n elemanlı metrik uzayın n+1 elemanlı bir uzaya belli koşulları sağlayan gömülmesi olarak ortaya çıkmaktadır. Bizim yaklaşımımızda ise, metrik uzayın içsel (intrinsic) özellikleri kullanılmaktadır. Her iki yaklaşımda da problem birtakım lineer eşitsizlik sistemlerinin pozitif çözümlerinin bulunmasına indirgenmektedir. Hipersimpleks sınıflamasını tanımlayan eşitsizlik sitemlerinin daha kıstlayıcı olması beklenebilir. n=5 ve bazı n=6 örnekleri için hipersimpleks sınıflaması ve bizim yaklaşımımızda ortaya çıkan eşitsizlik sistemleri karşılaştırılmıştır. Eşitsizlik denklemleri farklı olmakla birlikte çözümlerin örtüştüğü veya hipersimpleks sınıflamasında daha kısıtlı olduğu görülmüştür. Ancak bu konuda kesin yargılara varmak için daha ayrıntılı incelemeler gereklidir.

5 noktalı uzayların sınıflaması: 5 noktalı uzayların sınıflaması literatürde iyi bilinmektedir. Bu uzayların 102 sınıftan oluştuğu ve permütasyonlar altında birbirine denk olmayan, yukarıda A,B ve C tipleri ile adlandırılan 3 sınıf olduğu bilinmektedir. Bu 3 sınıfın split ve split-asal metrikler cinsinden parametrizasyonu da verilmiştir. Bu sınıfların optimal temsilleri da elde edilmiş, A ve C tiplerinin tek bir optimal temsili, B tipini ise, biri C tipi ile ortak olan 2 optimal temsili olduğu gösterilmiştir. **Ek-1.2 de verilen, "Five point metric spaces, Gromov product structures, quadrangle structures and explicit parametrizations"** isimli yayına sunulmuş olan metinde, 5 noktalı uzaylar için bilinen sonuçların Gromov çarpımı ve dörtgen yapıları ile elde edilmesi ayrıntılı olarak verilmiştir. Bu çalışmada katkımız metriğin parametrizasyonda hareketle, (Bilge vd 2015) de tanımlanan "hamleler" kullanılarak optimal temsilinin elde edilmesidir. Bu şekilde B tipi metriklerin hangi durumlarda hangi optimal temsile sahip olacağı belirlenmektedir. Makalede ayrıca, 3 tipin oransal hacimleri Monte-Carlo simülasyonları ile bulunmuş, her bir eşdeğerlik sınıfının sınıfın permütasyonlar altında yörüngeleleri belirlenmiştir.

Gromov çarpımı yapılarının matris temsili: Bir Δ-eşdeğerlik sınıfına karşı gelen M_Δ matrisi, eğer herhangi bir k için:

$$\Delta_{ijk} \text{ minimal ise, } M(i,j)=1,$$

diğer durumlarda

$$M(i,j)=0$$

olarak tanımlanır.

Bu matrisin rankı, özdeğerleri, indirgenemez olup olmadığı, vb. özellikler o eşdeğerlik sınıfının invaryantları olarak ortaya çıkmaktadır. Bu invaryantların aynı olması hatta matrislerin benzer olması Gromov çarpımı yapılarının eşdeğer olduğu anlamına gelmemektedir. Çünkü iki Gromov çarpımı yapısının eşdeğer olması demek bunlara karşı gelen matrislerin arasındaki benzerlik dönüşümünün bir permütasyon matrisi olması demektir (Bilge ve Incegül, 2018b).

İndirgenebilir Gromov çarpımı yapıları: Eğer bir Gromov çarpımı yapısının matrisi indirgenebilir bir matris ise yani M_Δ =[A_{11} A_{12} ;A_{21} 0] şeklinde ise o Gromov çarpımı yapısı indirgenebilir olarak tanımlanmaktadır. Gromov çarpımı yapılarının indirgenebilir ve indirgenemez olarak ayrışımı sınıflandırmada büyük kolaylık sağlamaktadır.

Gromov çarpımı zincir ve halkaları: Bir Gromov çarpımı yapısı verildiğinde, eğer Δ_{ijk} Δ_{kil} minimal ise, P_i ve P_k bir Gromov çarpımı zincirinin parçasını oluşturmakta ve bu durum aşağıdaki diagramla temsil edilmektedir.

P_j P_i P_k P_l
○———●———●———○

n-elemanlı bir metrik uzayda Gromov çarpımı zincirleri n tamsayısının bölümlemeleri ile sınıflanabilir. Örneğin, n=5 için,

5, 4+1, 3+2, 2+2+1, 1+1+1+1+1

şeklinde Gromov çarpımı zincirleri olabilir. Eğer P_i noktasında Δ_{ijk} minimal iken P_j ve P_k noktalarında hiçbir Δ_{jil} ve Δ_{kil} minimal değilse, bu durumda 1+1+...+1 şeklinde bir dekompozisyon söz konusudur.

P_j P_i P_k

○———●———○

Genel olarak l uzunluğunda bir zincir

P_j P_{i1} P_{i2} .... P_{il} P_k

○———●———● .... ●———○

şeklindedir. Eğer bir Gromov çarpımı zincirinin açık uçları birleşiyorsa yani, j=k ise, l+1 uzunluğunda bir Gromov çarpımı halkası oluşur.

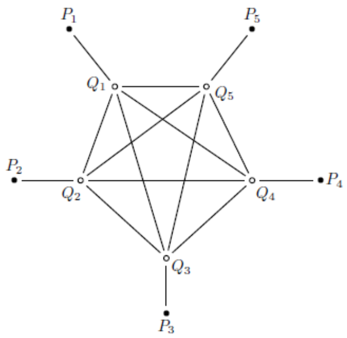
Gromov çarpımı zincirleri ve halkalarının sayıları büyük ölçüde M_Δ matrisi kullanılarak belirlenmektedir. Bu özellikler sınıflandırmanın elde edilebilmesinde önemli bir rol oynamış olup özellikle n=8 için Gromov çarpımı sınıfların elde edilebilmesi büyük ölçüde bu eşdeğerlikler kullanılması sayesine mümkün olmuştur.

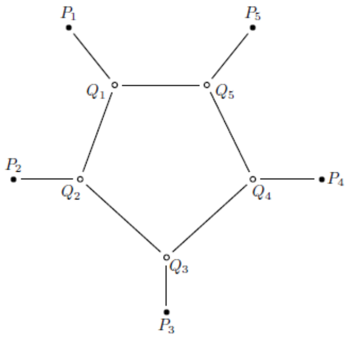
6 noktalı uzayların sınıflaması: 6 noktalı uzayların hipersimpleks sınıflaması (Sturmfels- ve Yu, 2004) makalesinde verilmiştir. Bu makalede 339 adet sınıf elde edilmiştir. Bu sınıfların çeşitli özellikleri ve sıkı-germe (tight-span)lerinin grafikleri, yazarların önerisi üzerine, güncellenerek proje web sitesine konmuştur. 6 noktalı uzayların Gromov çarpımı sınıflaması ve bir Gromov çarpımı sınıfında hangi Sturmfels-Yu metriklerinin bulunduğu (Bilge vd, 2018) de verilmiştir. n=6 için 17 adet indirgenemez, 9 adet indirgenebilir olmak üzer toplam 26 adet Gromov çarpımı sınıfı bulunmuştur. **Ek-1.3 Parametrization of 6-point Metric Space Type I17 using Quadrangle Structure Parameters** metninde bu sınıfların elde edilmesini ayrıntıları ve I17 olarak adlandırdığımız ve tamamen split-ayrışabilir bir metriği de içeren sınıf için, parametrizasyonlar açık olarak elde edilmiştir. I17 Gromov çarpımı yaıpısı içinde SSS, SST, STT ve TTT olarak adlandırdığımız 4 adet dörtgen yapısı sınıfı olduğu gösterilmiştir. Bu sınıflardam SSS, SST ve STT nin her birine birer Sturmfels-Yu hipersimpleks sınıfı tekabül etmektedir. TTT sınıfına ise 2 adet hipersimpleks sınıfı tekabül etmektedir.

7 noktalı uzayların sınıflaması: 7 noktalı metriklerin sınıflaması 6 noktalı metriklerin sınıflamasına benzer şekilde proje öncesi çalışmalarda tamamlanmış ancak yayımlanmamıştır. 7 noltalı uzaylar için, Gromov çarpımı matrisi ve Gromov çarpımı zincirleri kullanılarak toplamda 431 adet Gromov çarpımı sınıfı elde edilmiştir. **Ek-1.4 Gromov product decomposition of 7-point metric spaces'**de verilen metinde 7 noktalı uzayların sınıflandırılmasına ilişkin sonuçlar sunulmuştur. Bu metin proje öncesinde hazırlanmış ancak yayımlanmamış bir makale olup konunun bütünlüğü açısından rapora dahil edilmiştir.

8 noktalı uzayların sınıflaması: 8 noktalı metriklerin sınıflaması yöntem olarak 6 ve 7 noktalı metriklerin sınıflamasına benzerdir. Ancak seçenekler çok fazlalaştığı için matris temsillerinin kullanılarak eşdeğer sınıfların önceden elimine edilmesi önem kazanmıştır. Bu sadeleştirmelere rağmen sınıflamanın elde edilebilmesi ancak TRUBA bilgisayar sisteminin kullanılabilmesi ile mümkün olmuştur toplamda 11470 sınıf bulunmuştur. Elde edilen sonuçlar **Ek-1.5 "Gromov product structures for eight point metric spaces"** isimli metinde sunulmuştur.

Metrik uzayların çizge temsilleri, saçaksız (pendant-free) temsiller: Her sonlu metrik uzay ağırlıklandırılmış bir tam çizge ile temsil edilir. Eğer P_i noktasında Δ_{ijk} minimal ise, P_i noktasından çıkan kenarların Δ_{ijk} uzunluğundaki bölümünün Q_i noktalarına kadar birleştirlebileceği bilinmektedir. Bütün noktalarda bu birleştirmeler yapıldıktan sonra, jenerik durumda ortaya 2n noktalı bir çizge çıkmaktadır. Bu yöntemle elde edilen çizgede P_i noktalarının derecesi (bu düğümlere bağlı kenar sayısı) 1'dir. Q_i düğümleri ise yine bir tam çizge oluşturmaktadır. Ancak Δ_{ijk}.ve Δ_{jil}.minimal ise Q_j ve Q_k düğümlerini birleştiren kenat çizgeden atılabilir, çünkü P_i ve P_j noktaları arasındaki uzaklık, $\Delta_{ijk}+d(Q_j,Q_k)+\Delta_{jil}$ ile realize edilmektedir. Bu şekilde elde edilen çizgeden (Pi,Qi) kenarları da atılırsa elde edilen çizge literatürde saçaksız temsil veya indirgeme (pendant-free reduction) olarak adlandırılmaktadır. Aşağıda, n=5 A tipi metrik için bu indirgemeler örnek olarak verilmiştir.





Saçaksız temsiller aşağıda anlatılacak olan lokal optimal temsillerin hareket noktasını oluşturmaktadır. Ayrıca, eğer herhangi bir i,j çifti için $d(Q_i,Q_j)=0$ ise, saçaksız temsilde efektif olarak metriğin eleman sayısı azalmaktadır. Bu ise, filogenetik problemlerinde ağaç yapısının oluşmasını sağlayan bir özelliktir.

Dörtgen yapıları: (X,d) uzayında verilen her $\{P_i,P_j,P_k,P_l\}$ nokta dörtlüsü için eğer,

$$d(P_i,P_j)+d(P_k,P_l), \qquad d(P_i,P_k)+d(P_j,P_l), \qquad d(P_i,P_l)+d(P_j,P_k)$$

sayılarının maksimal elemanı tek ise (X,d) uzayı dörtgen yapısı anlamında lenerik (Q-jenerik) olarak adlandırılır. Örneğin $d(P_i,P_l)+d(P_j,P_k)$ maksimal ise, $\{P_i,P_j,P_k,P_l\}$ noktaları (P_i,P_l) ve (P_j,P_k) kenarlarının köşegen olduğu bir dörtgene izometrik olarak gönderilebilir. Q-jenerik bir uzayda, her $\{P_i,P_j,P_k,P_l\}$ dörtlüsü için yukarıdaki toplamlardan maksimal olanının seçimi (X,d) uzayında dörtgen yapısı olarak tanımlanmıştır.

Gromov çarpımı yapısı yani her noktada hangi Gromov çarpımının minimum olduğu verildiğinde, metrik uzayın 4 elemanlı kümelerindeki Gromov çarpımları arasındaki bazı sıralama bağıntıları da belirlenir. Örneğin, P_i noktadında Δijk minimal ise, her l için,

$$\Delta_{ijl} \geq \Delta_{ijk}, \qquad \Delta_{ikl} \geq \Delta_{ijk},$$

eşitsizlikleri sağlanır. Buradan hareketle ayrıca

$$\Delta_{jik} \geq \Delta_{jil}, \qquad \Delta_{jkl} \geq \Delta_{jil},$$

$$\Delta_{kij} \geq \Delta_{kil}, \qquad \Delta_{kjl} \geq \Delta_{kil},$$

$$\Delta_{lij} \geq \Delta_{ljk}, \qquad \Delta_{lik} \geq \Delta_{ljk},$$

eşitsizlikleri elde edilir. Bu eşitsizlikler kullanılarak, Ancak {Pi,Pj,Pk,Pl} elemanlarının hiç bir üçlü kombinasyonu minimal Gromov çarpımlarından biri ile eşleşmiyorsa, bu tür eşitsizlikleri yazma imkanı yoktur. Dolayısıyla, dörtgen yapısı içinde bazı 4 elemanlı kümelerin bir dörtgene nasıl gönderilmesi gerektiği belirlenemeyecektir. n=5 için Gromov çarpımı yapısı tüm dörtgenlerin yapısını belirlemektedir. n>5 için ise verilen bir Δ-eşdeğerlik sınıfına karşılık birden çok dörtgen yapısı bulunabilmektedir.

Gromov çarpımı yapısından hareketle veya ondan bağımsız olarak tanımlanmış dörtgen yapılarının jenerik olması için yukarıda verilen eşitsizliklerin kesin büyük olarak sağlanması gereklidir. Çalışmamızda, bu koşulun sağlanıp sağlanmadığı lineer programlama yöntemleri ile kontrol edilmiştir.

Verilen bir (jenerik) dörtgen yapısına sahip birden fazla Gromov çarpımı yapısı olması da mümkündür. Bunun en basit örneği n=5 de C tipi için karşımıza çıkmaktadır. Aşağıda ayrıntılı olarak tartışıldığı gibi, A ve B tiplerinde dörtgen yapısı Gromov çarpımı yapısını belirlemektedir. C tipi için ise minimal Gromov çarpımları $\{\Delta_{125}, \Delta_{213}, \Delta_{325}, \Delta_{425}, \Delta_{513}\}$ şeklindedir. Bu Gromov çarpımı yapısından hareketle kurulan dörtgen yapısı ise $\{\Delta_{125}, \Delta_{213}, \Delta_{325}, \Delta_{413}, \Delta_{513}\}$ ile de uyumludur.

Proje çerçevesinde, n=6 için, 26 adet Gromov çarpımı yapısı ile uyumlu 210 adet dörtgen yapısı bulumuş ve bu sonuçlar **Ek-1.6 Quadrangle structures for 6-point spaces'**de verilmiştir. Bu dörtgen sınıflarının belirlenmesi için, verilen Gromov çarpımı yapısının hangi dörtgenleri belirlediği hesaplanmış, yapısı belirlenmeyen denklem sayısı k ise, bunların her biri için 3 seçenek olmasından hareketle 3^k adet dörtgen yapısını belirleyen eşitsizlikler belirlenmiştir. Bu eşitisizliklerin çözüm kümesinin boş olup olmadığı MATLAB yazılımı ile, lineer programlama kullanılarak kontrol edilmiştir. Lineer programa algoritması kullanılırken, maliyet fonksiyonu sabit bir fonksiyon olarak alınmıştır. Lineer programlama probleminin kısıtları ise dörtgen yapısını tanımlayan eşitsizliklerden oluşmaktadır. Eğer eşitsizliklerin çözüm kümesi açık bir küme ise algoritma bu küme içinde herhangi bir eleman vermektedir.

Beşgen yapıları ve (n-k) elemanlı alt uzay yapıları: Çalışmamızda Gromov çarpımı yapıları ve dörtgen yapılarının tanımlanma amacı metriklerin parametrize edilmesidir. n=6 için dörtgen yapısının metriği parametrize etmek için yeterli olmadığı I_{01} Gromov çarpımı tipi içindeki bir dörtgen yapısındada gözlenmiştir. I1 Gromov çarpımı tipinin minimal gromov çarpımları aşağıdaır.

$$\text{Tip } I_{01} : \{\Delta_{124},\ \Delta_{213}, \Delta_{324}, \Delta_{456}, \Delta_{524}, \Delta_{624}\}.$$

Bu minimal Gromov çarpımları kullanılarak dörtgen yapısı belirlenmeye çalışıldığında, 5 adet dörtgenin yapısının serbest kaldığı görülmüştür. Buradan hareketle 3^5 olası dörtgen yapısını tanıımlayan eşitsizliklerin tutarlı olup olmadığı kontrol edilmiş, permütasyon altında eşdeğer olan sınıflar birleştirilerek 18 adet dörtgen yapısı sınıfı bulunmuştur. Bu sınıfların parametrize edilmesi çalışılırken, I_{01}-Q_{06} sınıfının dörtgen yapısı kullanılarak parametrize edilemeyeceği görülmüştür. Bu sınıfın dörtgen yapısı, dörtgenlerin köşegenleri ile aşağıdaki gibi belirlenmiştir. (ik,jl) notasyonu, $Q(P_i,P_j,P_k,P_l)$ dörtgeninde, (P_i,P_k) ve (P_j,P_l) kenarlarının köşegen olduğu anlamına gelmektedir.

(13,24), (13,25), (13,26), (15,24), (16,24),

(12,56), (14,35), (14,36), (13,56), (14,56),

(24,35), (24,36), (23,56), (24,56), (34,56).

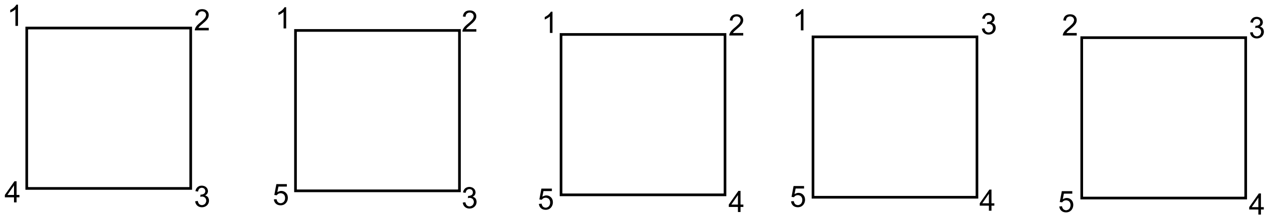
Bu sınıf Sturmfels-Yu sınıflamasında 260 ve 321 sayılı metrikleri içermekte olup, bu iki metriği birbirinden ayıran Δ_{314} ve , Δ_{356} Gromov çarpımlarının hangisinin daha büyük olduğudur:

Sturmfels-Yu 260 tipi: $\Delta_{314} < \Delta_{356}$ Sturmfels-Yu 321 tipi: $\Delta_{356} < \Delta_{314}$

Bu durum 6 elemanlı bir metriğin 5 elemanlı alt kümelerinin yapısı ile şu şekilde açıklanır. 5 elemanlı bir metriği yukarıda verilen A, B, C tiplerininden birindedir. A tipi bir metriğin minimal Gromov çarpımları

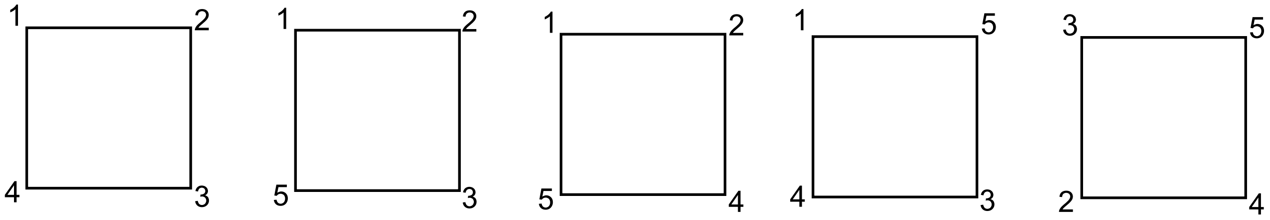
Tip A: $\{\Delta_{125}, \Delta_{213}, \Delta_{324}, \Delta_{435}, \Delta_{514}\}$,

şeklindedir. Bir (P_i,P_j,P_k,P_l) dörtgeninde eğer Δ_{ijl} minimal ise, (P_i,P_k) ve (P_j,P_l) köşegendir. Buradan hareketle dörtgenlerin yapısı aşağıdaki gibi belirlenir.



Şimdi bu dörtgen yapısının verildiğini ama minimal Gromov çarpımlarının neler olduğunu bilmediğimizi varsyalım ve bu dörtgenlerin yapısından minimal Gromov çarpımlarını belirlemeye çalışalım. İlk dörtgene baktığımızda, $\Delta_{124} < \Delta_{123}$ ve $\Delta_{124} < \Delta_{134}$ olduğunu görürüz, yani, Δ_{123} ve Δ_{134} minimal olamazlar. Benzer şekilde, Δ_{124}, Δ_{135} ve Δ_{145}'in de minimal olamayacağı görülür. Dolayısıyla bu dörtgen yapısı verildiğinde P_1 noktasındaki minimal Gromov çarpımı Δ_{125} olmak zorundadır. Benzer şekilde diğer noktalarda da A tipini belirleyen Gromov çarpımı yapılarının minimal olmaz zorunda olduğu görülebilir. Yani A tipi metrikler için dörtgen yapısı Gromov çarpımı yapısını belirler. Benzer durumun B tipi metrikler için de geçerli olduğu görülebilir. C tipi metrikler için ise durum farklıdır. C tipi bir metriği belirleyen minimal Gromov çarpımları aşağıda verildiği gibidiryukarıda

Tip C: $\{\Delta_{125}, \Delta_{213}, \Delta_{325}, \Delta_{425}, \Delta_{513}\}$,



Bu dörtgen yapısına bakarak minimal olamyacak Gromov çarpımlarını üstünü çizerek gösterdiğmizda aşağıdaki durum ortaya çıkar

~~Δ_{123}~~, ~~Δ_{124}~~, Δ_{125}, ~~Δ_{134}~~, ~~Δ_{135}~~, ~~Δ_{145}~~

Δ_{213}, ~~Δ_{214}~~, ~~Δ_{215}~~, ~~Δ_{234}~~, ~~Δ_{235}~~, ~~Δ_{245}~~

~~Δ_{312}~~, ~~Δ_{314}~~, ~~Δ_{315}~~, ~~Δ_{324}~~, Δ_{325}, ~~Δ_{345}~~

~~Δ_{412}~~, Δ_{413}, ~~Δ_{415}~~, ~~Δ_{423}~~, Δ_{425}, ~~Δ_{435}~~

~~Δ_{512}~~, Δ_{513}, ~~Δ_{514}~~, ~~Δ_{523}~~, ~~Δ_{524}~~, ~~Δ_{534}.~~

Buradan da, C tipi metrikte, (yukarıdaki sayma konvansiyonu çerçevesinde) dörtgen yapısnın P_1, P_2, P_3,ve P_5 noktalarında minimal Gromv çarpımlarını belirlediği ancak, Δ_{413} ve Δ_{425} değerlerinin karşılaştırılamaz olduğu görülmektedir.

6 noktalı uzay örneğinde, 5 adet, 5 elemanlı alt uzay vardır. 6 noktalı uzayın dörtgen yapısından başlayarak bu 5 elemanlı alt uzayların dörtgen yapıları çıkarılır. Eğer bu 5 elemanlı alt uzayların dörtgen yapısı A veya B tipinde ise, Gromov çarpımları arasında yeni bir bağıntı ortaya çıkmaz. Eğer C tipinde ise ve (yukarıda incelenen sayma konvansiyonu ile), Δ_{413} veya Δ_{425}'in minimal olduğu biliniyorsa yine yeni bir eşitsizlik bağıntısı elde edilemez. Ancak, C tipi metrikteki Δ_{413} veya Δ_{425} P_4 noktasında minimal değilse, $\Delta_{413} < \Delta_{425}$ veya $\Delta_{413} < \Delta_{425}$ durumlarının her ikisi de olabilir ve bunlar dörtgen tipi verilen metriğin beşgen tipini (pentacle structure) belirlerler.

I_{01}-Q_{06} tipini 5 elemanlı alt kümelerinin dörtgen yapıları incelenip buradan 5 elemanlı uzaylar olarak Gromov çarpımı yapıları oluşturulduğunda aşağıdaki 5 elemanlı kümeler için dörtgen yapıları elde edilir.

(P_6 hariç): (13,24), (13,25), (15,24), (14,35), (24,35),

(P_5 hariç): (13,24), (13,26), (16,24), (14,36), (24,36),

(P_4 hariç): (13,25), (13,26), (12,56), (13,56), (23,56),

(P_3 hariç): (15,24), (16,24),(12,56), (14,56), (24,56),

(P_2 hariç): (14,35), (14,36), (13,56), (14,56), (34,56),

(P_1 hariç): (24,35), (24,36), (23,56), (24,56), (34,56).

Bu dörtgen yapılarının Gromov çarpımı tipleri belirlenmeye çalışıldığında ise, P2 noktasının hariç tutulduğu 5 elemanlı küme için 2 durumun mümkün olduğu görülür. Aşağıda, 5 elemanlı alt kümeler için I_{01} tipinin minimal Gromov çarpımları ile aynı olanlar italik olarak, P_2 hariç için iki farklı seçenek koyu ile belirtilmiştir.

Tip I_{01} : $\{\Delta_{124}, \Delta_{213}, \Delta_{324}, \Delta_{456}, \Delta_{524}, \Delta_{624}\}$.

(P_6 hariç): (13,24), (13,25), (15,24), (14,35), (24,35), {*Δ_{124}*, *Δ_{213}*, *Δ_{324}*, Δ_{435}, *Δ_{524}*}.

(P_5 hariç): (13,24), (13,26), (16,24), (14,36), (24,36), {*Δ_{124}*, *Δ_{213}*, *Δ_{324}*, Δ_{436}, *Δ_{624}*}.

(P_4 hariç): (13,25), (13,26), (12,56), (13,56), (23,56), {Δ_{156}, *Δ_{213}*, Δ_{356}, Δ_{513}, Δ_{613}}.

(P_3 hariç): (15,24), (16,24),(12,56), (14,56), (24,56), {*Δ_{124}*, Δ_{256}, *Δ_{456}*, *Δ_{524}*, *Δ_{624}*}.

(P_2 hariç): (14,35), (14,36), (13,56), (14,56), (34,56), {Δ_{156}, **Δ_{314},** *Δ_{456}*, Δ_{514}, Δ_{614}}.

(P_2 hariç): (14,35), (14,36), (13,56), (14,56), (34,56), {Δ_{156}, **Δ_{356},** *Δ_{456}*, Δ_{514}, Δ_{614}}

(P_1 hariç): (24,35), (24,36), (23,56), (24,56), (34,56). {Δ_{256}, *Δ_{324}*, *Δ_{456}*, *Δ_{524}*, *Δ_{624}*}.

Görüldüğü gibi, verilen dörtgen yapısı sınıfı için, 5 elemanlı alt kümelerin yapısı bu dörtgen yapısı tarafından belirlenmiyorsa, yeni eşitsizlikler elde edilmekte ve çözüm kümeleri farklılaşmaktadır. İncelediğimiz örnekte, yeni elde edilen eşitsizlik bu dörtgen sınıfının 2 farklı parametrizasyonunu vermiştir. n=6 için beşgen yapılarının çıkarılması ve parametrizasyon çalışmaları devam etmektedir. n>6 için ise, benzer şekilde n-k elemanlı alt uzayların yapılarının parametrizasyon için gerekli olacağı düşünülmektedir, ancak bu konuda henüz bir çalışma yapılmamıştır

Lineer programlama: Sonlu metrik uzayların sınıflamasında yaklaşımımız, her sonlu metrik uzay için sağlanması gereken üçgen eşitsizlikleri kümesini aşama aşama genişletmektir. Her aşamada, elde edilen eşitsizlikler sisteminin çözüm kümesinin R^N de açık bir küme olup olmadığının kontrol edilmesi gereklidir. Lineer programlama, lineer bir amaç fonksiyonunun lineer eşitlik ve eşitsizlikler tanımlı bir küme üzerinde minimize edilmesi problemidir. Çalışmamızda, Gromov çarpımı yapılarının jenerik olup olmadığını belirlemek için (Incegül, Bilge 2018) de verilen bir algoritma kullanılmıştır. n=6 için elde edilmiş olan Gromov çarpımı yapılarının belirlenmesi için ise lineer programlama yöntemi kullanılmıştır.

Split metrikler: Metrik uzayı oluşturan X kümesini X=AUB şeklinde bölümlenmiş olsun. X üzerinde δ_A split (sözde) metriği,

$\delta_A(P_i,P_j)=1$ P_i ve P_j farklı A ve B kümelerinde ise,

$\delta_A(P_i,P_j)=0$ P_i ve P_j aynı A veya B kümelerinde ise,

kuralı ile tanımlanır. Projenin ön çalışmalarında da görüldüğü gibi, dörtgen yapısı, literatür özeti kısmında ayrıntıları verilen split-metriklerin belirlenmesinde önemli bir rol oynamaktadır. δ_A şeklinde bir split metriğin izolasyon indisinin (Bandelt ve Dress, 1992; Koolen vd., 2009; Celik, 2018) 0 olmaması için gerekli koşul, A'nın her (P_i,P_j) elemanı ve A'da olmayan her (P_k,P_l) elemanı için, P_i ve P_j'yi birleştiren kenarın, (P_i,P_j,P_k,P_l) dörtgeninde köşegen olmamasına eşdeğerdir. Bu gözlemden hareketle, dörtgen yapısı oluşturulduğu zaman, herhangi bir d metriğinin split metrik dekompozisyonu kolayca yazılabilmekte, split-kısım ise kısım ise δ_P ile gösterilmektedir. Örneğin (B) tipindeki metrik için (P_1P_2), (P_2P_3), (P_3P_5) ve (P_1P_5) in her dörtgende kenar olarak bulunduğu gözlenerek, split dekompozisyon hemen yazılabilir.

$$d=m_{12}\,\delta_{12}+m_{23}\,\delta_{23}+m_{45}\,\delta_{45}+m_{15}\,\delta_{15}+m_p\,\delta_p$$

Yukarıda δij terimleri split metrikleri temsil etmektedir ancak, split-asal ismi verilen δ_p metriğinin bulunması ve split metriklerin izolasyon katsayıları olan m_{ij} parametrelerinin (d_{ij}' ler cinsinden) bulunabilmesi kolay bir problem değildir. Split metriklerin izolasyon indisleri ve split asal kısmın aynı anda belirlenmesi gerekmektedir. Bu ise zaten metriğin parametrize edilmiş olması anlamına gelmektedir. Proje kapsamında n=6 için tüm metriklerin parametrizasyonu hedeflenmiş olsa da, kısmi ancak farklı olasılıkları yansıtacak şekilde seçilmiş örnekler çalışılmıştır. Split metrikler ve dörtgen yapılarının ilişkileri Discrete mathematics dergisinde yayımlanmış **olan Ek 2.1 "Gromov product structures, quadrangle structures and split metric decompositions for finite metric spaces"** makalesinde verilmiştir.

Split-asal metrikler: Split-asal metrikler, metrik koninin, split metrik formunda olmayan ayrıtları olarak tanımlanmaktadır. Bir polihedral koninin ayrıtlarının bulunması prensip olarak basit gözükse de pratikte zor bir problemdir (Kölbig ve Schwarz, 1986) makalesinde bir eşitsizlik kümesi ile tanımlanan polihedral koninin tüm ayrıtlarını verecek bir algoritma öne sürülmüştür. Proje çerçevesinde bu algoritmanın uyarlanması denese de başarılı olmamıştır. n=6 ve n=7 için split asal metriklerin hesabı sırasıyla (Koolen, 2000, Grishljkhin, 1992) makalelerinde hesaplanmıştır. (Koolen 200) makalesinde, n=6 için 11 adet split asal metrik verilmiştir. (Grishljkhin, 1992) makalesinde ise n=7 için metrik koninin toplamda 60 000 den fazla ayrıtı olduğu belirtilmi ve permutasyon altında eşdeğer olmanlar doğrudan hesaplanarak 643 tanesi listelenmiştir.

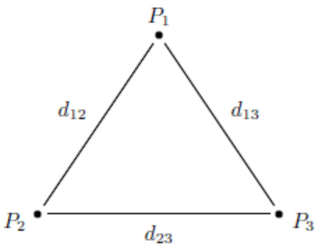
Proje çerçevesinde, split-asal metriklerin belirlenmesi için bir algoritma geliştirilerek n=5 durumuna ve n=6 için bazı örneklere uygulanmıştır. Metrik koninin ayrıları, üçgen eşitsizlikleri, yani Gromov çarpımlarının bir bölümünün sıfırlanması ile elde edilcek lineer sistemin 1 boyutlu pozitif çözümüdür. n elemanlı uzay için M=n(n-1)(n-2)/2 sayıda üçgen eşitsizliği bulunmaktadır. Dolayısıyla, N=n(n-1)/2 sayıda bilinmeyen için M sayıda denklem söz konusudur. Bu denklem sistemini AX=0 şeklinde temsil edersek, problem bu sistemin

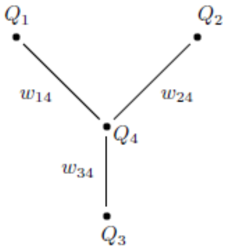
$$[A_1\ A_2]\ [X]^t \geq 0, \quad A_1X^t=0, \quad \mathrm{rank}(A_1)=1, \quad X>0 \text{ ve } A_2X^t>0$$

Şeklinde çözümleri aranmaktadır. Bu problem çok fazla sayıda kombinasyonun denemesini içermektedir. Ancak X vektörünün hiçbir elemanının sıfır olmaması istenirse, $\Delta_{ijk}=0$, $\Delta_{jik}=0$, $\Delta_{kij}=0$ denklemlerinden sadece birinin sağlanması gerektiği kolayca görülebilir. Bu gözlemden hareketle denenmesi gereken kombinasyonlar problemi çözülebilir bir seviyeye indirgemektedir. n=5 için bu yöntem kullanılarak bilinin split asal metrik bulunmuştur. n=6 için Gromov çarpımı yapıları kullanılarak denklem sistemi biraz daha basitleştirilmiş ve incelenen örnekler için split asal metrikler bulunmuştur, ancak bu yöntemin n>6 için uygulanması zordur. Split-asal metriklerin, metriğin parametrizasyonundan bağımımsız olarak elde edilmesi yönünde çalışmalar devam etmektedir.

Lokal optimal (sıkı, tight) temsiller: Bir sonlu metrik uzayın optimal temsili, bu uzayın resmedilmiş olduğu toplam kenar uzunluğu en az olan ağırlıklandırılmış çizgedir. Optimal temsilin NP zor bir problem olduğu ve optimal temsilin tek olmadığı bilinmektedir. Verilen bir temsilden hareketle, (Bilge, Çelik, Koçak, 2015) de tanımlanmış olan “hamle”ler ile temsilin ağırlığının azaltılması için bir yöntem önerilmiştir. Jenerik metrikler için, uygulanacak hamle dizisi temsil çizgesinin düğümlerinin derecelerinin 1 veya 3 olduğu noktada sonlanmaktadır. Varılan nokta toplam uzunluğun daha fazla azaltılamaması açısından, “lokal optimum” olarak adlandırılmıştır.

Bu temsillerin en temel örneği 3 noktalı bir uzayın üçgen şeklinde bir çizge yerine, 4 noktalı Y şeklinde bir çizge ile temsilidir. 3-noktalı uzayın bu gömülmesi Δ-Y dönüşümü olarak adlandırılmaktadır.





Yukarıda, w_{i4} ağırlıkları Qi noktasındaki (minimal) Gromov çarpımı olup, $w_{i4}+w_{j4}=d_{ij}$ olduğu görülebilir.

4 veya daha fazla noktalı uzaylarda ise, her P_i noktasından çıkan kenarlar o noktada minimal olan $\Delta(P_iP_jP_k)$ uzunluğunda bir kenar boyunca bir Q_i noktasına kadar birleştirilebilir. Bu işlemler tamamlandıktan sonra, örneğin $\Delta(P_iP_jP_k)$ minimal ise, Q_jQ_k kenarı gereksiz olduğu için atılabilir. Bu aşamadan sonra Q_i noktalarından çıkan kenarların tümünün birleştirilmesi söz konusu değildir; ancak bir alt grubu seçilerek yine bu gruba ilişkin minimal Gromov çarpımı kadar birleştirmeler yapılabilir ve bazı kenarlar atılabilir. Bu işlemler Δ-hamleleri olarak adlandırılmaktadır. Δ-hamlelerinin peşpeşe uygulanması ile, daha fazla indirgeme yapılamayan, “lokal optimal” olarak adlandıracağımız bir çizge edilir; ancak bu çizge optimal olmak zorunda değildir.

n=6 için ise, optimal temsillerin tek olmadığı bilinmektedir. **Ek 2.2 “Optimal realizations, h-optimal realizations and tight spans of metric spaces”** metinde bu durumlar incelenmiştir. ,(Koolen XXX) makalesinde n=6 için optimal temsilin tek olmadığı bir örnek verilmiştir. Ek 2.2’de bu örneğin I17 SSS sınıfından bir örnek olduğu gösterilerek bu sınıf için optimal temsilin algoritmik olarak elde edilmesi gösterilmiş, ayrıca SST tipi için aynı yöntemin neden geçerli olmadığı açıklanmıştır. I17 tipinde SSS, SST, STT ve TTT tipleri için ise, lokal optimal indirgemeler verilmiştir.

Optimal temsillerin elde edilmesine yönelik, global olarak aldandırabileceiğimiz bir yaklaşım, verilen metriği ayrık üçgenlere ayırarak herbirine Delta-Y dönüşümü uygulamaktır. Eğer verilen uzay (n=7 durumunda olduğu gibi) ayrık üçgenlere dekompoze edilebiliyorsa bu şekilde Delta-Y dönüşümü uygulanarak toplam ağırlık yarıya indirilebilir. Ama genelde, optimal temsillerde bu oran daha düşüktür. Proje kapsamında uygulanan bir diğer yaklaşım, saçaksız (pendant-free) temsili ayrık üçgenlere ayırarak Delta-Y dönüşümleri uygulamaktır. Bu yöndeki çalışmalar **Ek-2.3 Tight Reductions of Finite Metric Spaces**’de sunulmuştur. Bu yöntemde karşımıza şu durumlar çıkmaktadır: İlk durumda, saçaksız çizge tam olarak ayrık üçgenlerin bileşiminden oluşmaktadır. Toplam uzunluğun azalması açısından en elverişli olan durum budur. İkinci durumda ise, saçaksız çizgeden çıkarılmış olan kenarlar tekrar eklenerek ayrık üçgenler oluşturulmakta ve Delta-Y dönüşümü uygulanmaktadır. Bazı örneklerde, her 2 durum için de ayrık üçgen oluşturulamayacak şekilde kenarlar kalmaktadır. Bu yöntem optimale yakın temsillerin elde edilmesinin yanı sıra Gromov çarpımları yapısından daha kaba sınıfları oluşturma açısından da önemli olup bu konuda çalışmalar devam etmektedir.

Jenerik olamayan metrikler: Filogenetik uygulamalarında genelde daha fazla elemanlı ve ağaç yapısında ya da az sayıda döngü içeren metrik uzayların ortaya çıkması nedeniyle, jenerik olmayan metriklerin incelenmesi önem kazanmaktadır. Ağaç yapısındaki olan metriklerin incelenmesi matematik literatüründe geniş bir yer tutmaktadır. Literatürde, tam olarak ağaç yapısında olmayan, ancak optimal indirgemeleri ağaç yapısına çeşitli anlamlarda yakın olan metrikler de incelenmiştir. Bu yönde bir ölçüt de, Gromov hiperbolisite adı verilen, ve her dörtgen için tight-span’daki 2 boyutlu elemanın “eni” üzerinde sınırlama getiren kavramdır.

Jenerik olmayan metriklerin incelenmesinin öne çıkaran bir durum da, hipersimpleks sınıflamasına göre çok daha az sınıf olmasını beklediğimiz Gromov çarpımı sınıflamasında bile n=8 için 11 470 sınıfın ortaya çıkması daha kaba eşdeğerlik sınıflarının aranmasını gündeme getirmiştir. Bu çerçevede, verilen bir Gromov çarpımı sınıfı için jenerik olmaktan en uzak olan temsilin bulunması konusu incelenmiştir. Bir Gromov çarpımı yapısı verildiği zaman, bu yapıda bulunan her Δ_{ijk} için, jenerik olsun olmasın, saçaksız temsilde (pendant free reduction) (Q_i,Q_j) ve (Q_i,Q_k) kenarları çizgede kalacak, (Q_j,Q_k) kenarı ise çizgeden atılacaktır. Eğer metrik jenerik değilse, saçaksız temsilde, $\Delta_{ijk}= \Delta_{imn}$ olacak şekilde başka Gromov çarpımları da olacaktır ve bu durumda (Q_m,Q_n) kenarı atılacaktır. Verilen bir Gromov çarpımı yapısı için jenerik olmaktan maksimal derecede uzak olan temsil (maximally non-generic representation) o Gromov çarpımı yapısını “çatısı” (frame of the Gromov product structure) olarak adlandırılmıştır. n=5,6,7,8 için Gromov çarpımı yapılarının çatıları hesaplanmıştır. Bu çatılar bilinen örneklerde, çizge yapısındaki split-asal metriklerle örtüşmektedir. Çatı metriklerinin özelliklerinin incelenmesi çalışmaları devam etmektedir.

Filogenetik problemlerde ortaya çıkan metriklerin incelenmesi: Filogenetik sınıflandırma problemlerinde ortaya çıkan metriklerin incelenmesi için veri bankalarından ve literatürden elde edilen DNA/RNA ve/ya protein sekans dizilerini birbiriyle hizalayarak türler arası farklı benzemezlik (dissimilarity) matrisleri elde edilmiştir. **Ek-3.1 Examples of Distance Matrices'**de dizilerden elde edilen matrislerin yanısıra literatürden alınan çeşitli örnekler de verilmiştir. Bu metinde, "Arı (Bee family)", "Yosun (Algea family)" ve "Primat (Primate family)" aileleri için literatürde kullanılan çeşitli benzemenzlik matrisleri çıkarılmıştır. Referans olarak kullanılacak diğer aileler için tek bir benzemezlik matrisi verilmiştir. Tüm örnekler için Gromov çarpımları hesaplanarak üçgen eşitsizliklerinin sağlanıp sağlanmadığı kontrol edilmiştir. "Base Frequency" mesafe hesaplama yöntemi dıuşında hemen her durumda üçgen eşitsizliklerinin sağlandığı görülmüştür.

Ağaç yapısında örnekler: Tam olarak ağaç yapısına sahip bir örnek aşağıda incelenerek yöntem tanıtılacaktır. Bu örnek 6 elemanlı bir aile olup mesafeler:

$d_{12}=6$, $d_{13}=8$, $d_{14}=1$, $d_{15}=2$, $d_{16}=6$, $d_{23}=8$, $d_{24}=6$, $d_{25}=6$, $d_{26}=4$, $d_{34}=8$, $d_{35}=8$, $d_{36}=8$, $d_{45}=2$, $d_{46}=6$, $d_{56}=6$

şeklindedir. Bu metrik için Gromov çarpımları hesaplandığında,

$$\Delta_{124} = \Delta_{134} = \Delta_{145} = \Delta_{146} = 0.5,$$

$$\Delta_{216} = \Delta_{236} = \Delta_{246} = \Delta_{256} = 2.0,$$

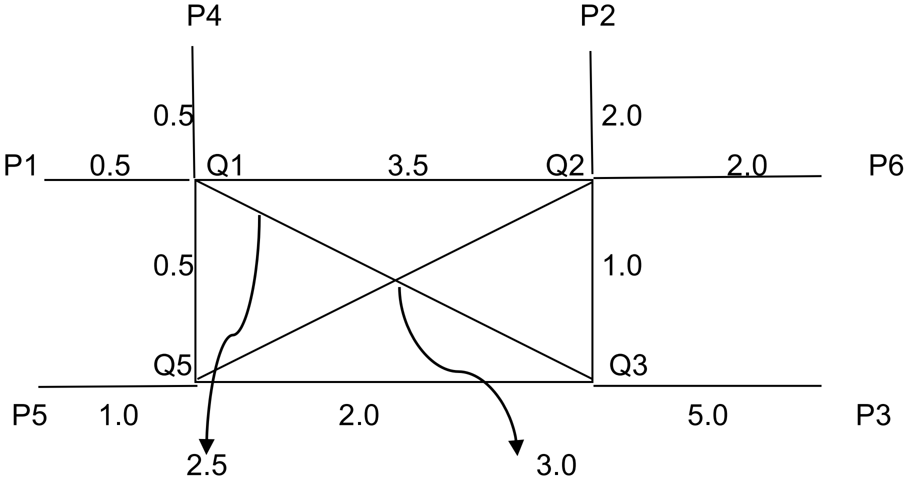
$$\Delta_{312} = \Delta_{316} = \Delta_{324} = \Delta_{325} = \Delta_{346} = \Delta_{356} = 5.0,$$

$$\Delta_{412} = \Delta_{413} = \Delta_{415} = \Delta_{416} = 0.5,$$

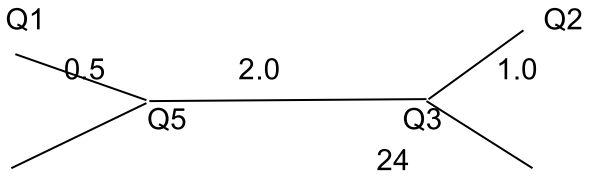
$$\Delta_{512} = \Delta_{513} = \Delta_{516} = \Delta_{524} = \Delta_{534} = \Delta_{546} = 1.0,$$

$$\Delta_{612} = \Delta_{623} = \Delta_{624} = \Delta_{625} = 2.0$$

olduğu görülür. Dikkat edilirse, P_1 ve P_4 noktaları arasındaki mesafe 1, bu noktalardaki Gromov çarpımları 0.5 olduğu için saçaksız temsilde Q_1 ve Q_4 noktaları birleşmiştir. Benzer şekilde Q_2 ve Q_6 noktaları da birleşip saçaksız temsil 4 noktalı bir uzaya indirgenmiştir.



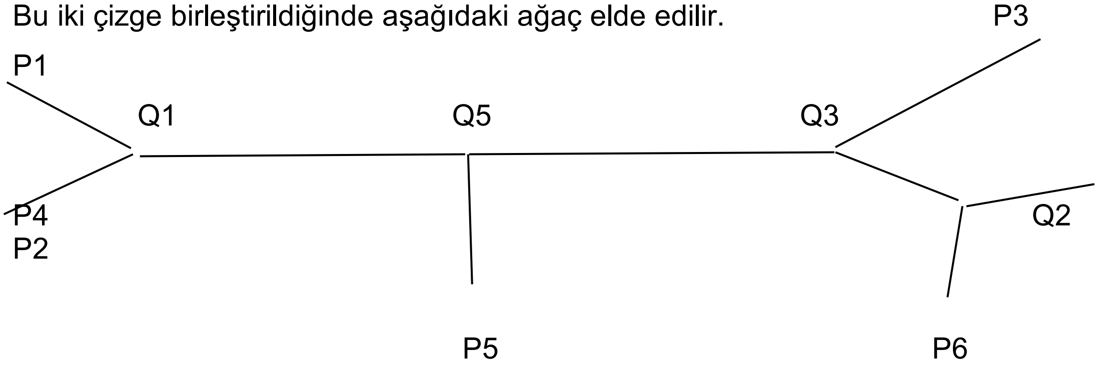
Yeni durumda, Q1, Q2, Q3, Q5 noktalarından oluşan dörtgende, köşegenler toplamı ile iki uzun kenar toplamı birbirine eşittir. Bu durumda dörtgenin dejenere olacağı bilinmektedir. Gromov çarpımları hesaplanarak indirgemeler yapıldığında iç dörtgenin yapısı aşağıdaki gibi bulunur.



0 0

Q5 Q3

Bu iki çizge birleştirildiğinde aşağıdaki ağaç elde edilir.



Benzer yöntemin 6 elemamlı Arı ailesine uygulanması ise ağaç metriği vermemiştir. Bu örnek **Ek-3.2, “Non-generic metrics in phylogenetics, the Bee family”** metninde ayrıntılı olarak incelenmiştir.

Ek-3.2, “Non-generic metrics in phylogenetics, the Bee family” metninde, 6 elemanlı Arı ailesi için önce çeşitli metriklerle çalışılmış, bu metriklerdeki çok küçük pertürbasyonların bile metriğin Gromov çarpımı tipini değiştirdiği gözlenmiştir. Doğrudan dizilerden elde edilmiş olan Hamming mesafe fonksiyonu ile çalışıldığında ise, metriğin yukarıdaki örneğe benzer şekilde, 5 elemanlı bir aileye indirgendiği görülmüştür. Gromov çarpımları hesaplandığında metriğin B tipinde olduğu görülmüş ve bu tip için bilinen optimal ve lokal-optimal (tight) temsiller elde edilmiştir. Her 2 durum da bu Arı ailesi için olası konfigürasyonlardır. Her iki durumda da cizge temsilinde 2 döngü bulunmaktadır. Bu nedenle, bu türler arasında gen transferinden bahsedilebilir.

Yaklaşık ağaç yapıları: Bir metriğin ağaç yapısınıa sahip olması için tüm dörtgenlerin dejenere yapıda olması gerektiği bilinmektedir. Bu koşulun tam olarak sağlandığı örnek yukarıda verilmiştir. Proje çerçevesinde çalışılan gerçek örneklerin hiçbirinde bu koşul tam olarak sağlanmamıştır. Bu nedenle verilen bir ailenin ağaç yapısı ile temsil edilip edilemeyeceğini anlamak için bir krited verilmeye çalışılmıştır.

Öncelikle, çalışılan örneklerde, dizinlerde küçük pertürbasyonlar yapılmış metriğin normunun bu pertürbasyonlar altında değişimi Arı, Yosun, Yarasa, Zebra balığı aileleri için incelenmiş, %10 ölçeğinde bir yaklaşıklığın kobul edilebilir olduğu sonucuna varılmıştır.

Verilen bir metrik için sayısal bir kriter getirme amacıyla her Q(i,j,k,l) dörtlüsü için,

$$a=d_{ij}+d_{kl}, \qquad b=d_{ik}+d_{jl}, \qquad c=d_{il}+d_{jk}$$

büyüklükleri hesaplanmış, ve

$$a \geq b \geq c$$

olacak şekilde adlandırılmışlardır. Ardından a-b ve a-c farkları hesaplanarak, aşağıdaki r_{max} ve r_{min} oranları hesaplanmışır.

$$r_{max}=\max\{ (a-b)/a, (a-c)/a)\}, \qquad r_{min}=\min\{(a-b)/a, (a-c)/a\}.$$

Daha sonra bir α yaklaşıklık seviyesi seçilmiş ve eğer verilen bir dörtgen için $r_{min}<\alpha$ ise o dörtgen dejenere kabul edilmiştir. Benzer şekşlde bir metriğin α seviyesinde ağaç temsilinin olması ise

$$\max (r_{min}) < \alpha$$

koşulu ile verilmiştir. Burada kritik problem, istenen koşulları sağlayan bir metriğin varlığının garanti edilmesi ve bu metriğin orjinal metriğe yeterince yakın olmasıdır. Bu problem **Ek-3.3 "A New Method for Constructing Phylogenetic Trees"** metninde tartışılmış ve örnekler verilmiştir. Yukarıda verilen kriterin orjinal mesafe fonksiyonlarına ve saçaksız temsil metriğine uygulanması da farklı sonuçlar vermektedir. Genel olarak saçaksız temsil metriğinde mesafeler daha küçük olduğu için makul bir yaklaşıklıkta ağaç temsili kriterinin sağlanması daha zordur. Örneğin, Ek 3.3 de orjinal metrik kullanılırsa %12 seviyesinde ağaç temsili olabileceği sonucu çıkarılmıştır ancak Ek 3.2 de, saçaksız temsilede %10 yaklaşıklığı ile bir döngü yok edilebilmekte anca diğer döngünün yok edilebilmesi için %30 seviyesinde yaklaşıklık gerekmektedir ki bu makul seviyenin üstündedir. Öte yandan daha çok elemanlı ailelerde, hangi temsil kullanılırsa kullanılsın, dörtgenlerin büyük çoğunluğunun dejenere olduğu sadece çok az sayıda dörtgen için makul yaklaşıklık seviyelerinde, dejenere olayan az sayıda dörtgenin olduğu gözlenmiştir. Önerdiğimiz yöntem filogenetik örnekleri dışındaki metriklere uygulanmış, bu örneklerde, yukarıda tanımlanan yaklaşıklık kriterinin çok daha az dörtgen için sağlandığı ve α seviyesinin %30 civarında olduğu görülmüştür. Dolayısıyla, filogenetik örneklerin, büyük çoğunluğu dejenere olan, az sayıda döngü dışında ağaca yakın yapılarda olduğu söylenebilmektedir. Sonuç olarak, yukarıda verilen kriter, metriğin ağaç temsili olup olmadığı, hangi döngülerin var olduğu konusunda kullanışlı bir gösterge olsa da, kesin karara varmak için daha ayrıntılı inceleme gereklidir.

Biyolojik canlılardaki evrimsel değişimin filogenetik modellemesi

1'inci ara rapor döneminde tamamlanan bu iş paketi için detaylı açıklamalar 1'inci ara raporda verilmiştir. Sonraki dönemlerde bu iş paketindeki çalışmalar sona erdiğinden, bu iş paketi ile ilgili çalışma son dönemde yapılmamıştır. Bu iş paketinde elde edilen diziler özetle aşağıdaki gibi verilmiştir. Dizin vb ayrıntılar **Ek-4.1 Phylogenetic Data for Biological Organisms**

Arı ailesi: Projenin ilk aşamalarında altı farklı arı türüne ait 677 nükleotid molekülünden oluşan DNA dizileri üzerinden hesaplanmış uzaklık matrisi kullanılmıştır. Bu veri seti Huson ve Bryant tarafından yazılmış SpitsTree programı için test amaçlı kullanılmıştır (Huson ve Bryant, 2006). (http://page.math.tu- berlin.de/~joswig/tightspans/index.html).

Balık ailesi: Luciobarbus ve barbus balık türleri için sitokrom oksidaz enzimini kodlayan 33 adet gen dizisi belirlendikten sonra hizalanmıştır. Kimura uzaklık yöntemi kullanılarak uzaklık matrisleri elde edilmiştir (Khaefi *vça.,* 2018). 10 farklı balık türü için hesaplanmış uzaklık matrisi verilmiştir.

Nehirde yaşayan misk sıçan türleri arasındaki kafatasına ait morfolojik farklılıklar: Eric Le Boulangé tarafından gerekleştirilen 144 farklı kafatası ölçümleri sonucunda türler arası farklılıkları ifade eden bir uzaklık matrisi elde edilmiştir (Le Boulange *vça.,* 1996). Bu örnekte, misk sıçanları arasındaki evrimsel değişimin bir ağaç topolojisiyle açıklanamadığı, ancak bir ağ yapı modelinin çözümleyebildiği gen transferleri söz konusudur. (Legendre ve Makarenkov, 2002)

AIDS hastalığına yol açan HIV virüsünün DNA ve protein dizilimleri: Hakkında kesin bir sonuca varılamayan primatlardan insanlara geçiş yolunu belirleyecek mutasyonları barındıran HIV virüsüne ait 16 farklı türe bakılmıştır. AIDS hastalığına yol açan HIV retrovirüsünün iki farklı türü mevcuttur; HIV-1 ve HIV-2. Bir başka retrovirüs te sadece primatlarda görülen SIV'dir ve HIV-2'nin SIV'den dönüşerek insanlara geçtiği düşünülmektedir, ancak hala tartışılan bir konudur. Virüse yapısını kazandıran kiritik öneme sahip üç gen (*gag*, *pol* ve *env*) bölgesine ait DNA dizilimleri NCBI veri bankasından (http://www.ncbi.nlm.nih.gov/genbank) elde edilmiştir. Daha sonra MATLAB® Bioinformatik modülü (*n2aa* fonksiyonu) yardımıyla protein dizisine dönüştürülmüş, ardından çoğul dizilim tekniğiyle hizalanmıştır. Uzaklık değerleri, hizalanan diziler üzerinden farklı yöntemler kullanılarak elde edilmiştir. DNA ve protein dizileri, uzaklık matrisleri (farklılıklar) ve kullanılan yöntemler verilmiştir.

Proje web sitesi: Proje kapsamında kullanılan programlar, veri setleri ve elde edilen sonuçların yayınlandığı http://finitemetricspaces.khas.edu.tr/ adresli bir internet sitesi hazırlanmıştır. Proje konusuna ilişkin yeni çalışmalar bu sitede sunulmaya devam edecektir.

6 noktalı uzayların kataloğu: (Sturmfels-Yu, 2004) makalesinde elde edilmiş 6 noktalı uzayların hipersimpleks sınıflamaları ve çeşitli özellikleri http://sixpointmetrics.khas.edu.tr adresli internet sitesinde sunulmuştur. Bu site anılan çalışmanın yazarların önerisi ile oluşturulmuş görseller güncel yazılımlarla uyumlu, JavaView isimli bir programla görselleştirilmiştir. Toplam 339 adet metrik sınıfı için tüm sonuçlar kullanıcının farklı açılar ve uzaklıklarda incelemesine olanak veren interaktif resimlerinin oluşturulmasını sağlayacak şekilde görselleştirilmiştir. JavaScript programlama dilinin kullanılmasıyla internet sitelerinde güncel olarak kullanılan HTML-5 diline uygun görseller oluşturulmuş ve sonuçlar ve görsellerin aynı anda incelenmesine olanak sunan bir önyüz tasarlanmıştır. Bu işlem için JavaScript programlama dilinin "Tree" isimli özel bir kütüphanesi ile kullanılmıştır.

Proje ekibinin katkıları: Proje yürütücüsü ve araştırıcılar, proje önerisinde belirtilen plana uygun olarak çalışmışlardır. Projede görev alan lisansüstü bursiyerlerden Arash M. Rezaeinazhad sonlu metrik uzayların yapısal özelliklerinin teorik incelemesi, Burak Erkan Kaya ise filogenetik uygulamalar, bilgisayar programları ve web sistesinin kurulması alanlarında çalışmışlardır. Lisans bursiyerleri ise, yoğun emek gerektiren hesaplarda ve hazırlanan programların işletilmesi alanlarında katkıda bulunmuşlardır.

Tartışma

Proje uzun süreli devam eden bir çalışmanın son aşaması olmuştur. Proje desteği sayesinde n=8 sınıflaması yapılabilmiş, filogenetik örnekler incelenebilmiş, 6 noktalı uzayların hipersimpleks sınıflamalarının sunulduğu web sitesi düzenlenebilmiştir. Covid-19 salgını süresince bilimsel çalışmalara devam edilebilmiş ancak nedeniyle yaygınlaştırma çalışmaları yeterince yapılamamıştır.

Projede elde edilen sonuçlar, ana hatları ile, bir sonraki bölümde de değerlendirilen 4 grupta toplanmıştır. “Tartışma” bölümünde bu sonuçların içerikleri, bir sonraki “Sonuçlar ve Öneriler” bölümünde ise ileride bu konuda yapılabilecek çalışmalar verilecektir.

- Sonlu metrik uzayların Gromov çarpımı sınıflaması konusunda, Gromov çarpımı zincirleri ve Gromov çarpımlarının matris temsilleri kullanılarak hesap yöntemleri geliştirilmiş ve 8 elemanlı metriklerin sınıflamasına uygulanmıştır.

- Metriklerin Gromov çarpımı ve dörtgen yapılarından hareketle parametrizasyonların elde edilmesi için n=6 için dörtgen yapılarının yetersiz olduğu, “beşgen-yapıları” olarak adlandırdığımız, 5 elemanlı alt uzaylarının yapılarının bilinmesi gerektiği gösterilmiştir. Buradan hareketle, n-elemanlı bir metriğin n-k elemanlı alt kümelerinin yapısı ile parametrize edilebileceği savlanmıştır.

- Optimal ve lokal-optimal olarak adlandırdığımız temsillerin elde edilmesinde parametrizasyonun nasıl kullanılacağı gösterilmiştir.

- Filogenetik yapıların incelenmesinde yaklaşık hesap yapılması gerektiği ortaya konmuştur.

Sonuç ve Öneriler

Proje çerçevesinde elde edilen sonuçlar ve ilerideki çalışmalar için öneriler aşağıda özetlenmiştir.

- Proje kapsamında 8 elemanlı metrik uzayların Gromov çarpımı sınıflaması elde edilmiş ve 11470 sınıf elde edilmiştir. Benzer problem literatürde kullanılan benzer yöntemler için de geçerlidir. Bu bakış açısı ile, verilen bir n sayısı için tüm metrik uzayların sınıflandırılması teknik olarak mümkün olsa bile çok anlamlı değildir.

- Öte yandan, proje çerçevesinde incelenen, Gromov çarpımları, dörtgen yapıları, beşgen yapıları, n-k yapılar şeklinde iteratif olarak ilerleyen dizinler kullnarak metrik sınıflarını kullanarak kanonik bir parametrizasyon elde etme problemi daha anlamlı gözükmektedir. Bu yaklaşımda, n elemanlı bir metrik uzayda herhangi bir metrik verildiğinde o metriğin ait olduğu Gromov çarpımı dörtgen vb. yapılar aşama aşama çıkarılarak o metriğin ait olduğu N=n(-1)/2 parametreli bir metrik ailesi inşa edilecektir. Bu ise verilen metriğin hangi özelliklerin pertürbasyonlar altında korunduğu gibi özelliklerinin incelenmesinde yararlı olacaktır.

- Bir metrik uzayın toplam uzunluğu miniman olan bir çizge ile temsili olan “optimal temsil” problemi örneğin bir şebekenin tasarımı aşamasında anlamlı ve gereklidir. Öte yandan, filogenetikte olduğu gibi, çeşitli aşamalarla dallanan bir yapı için varılan son noktada ortaya çıkan çizgenin optimal olması gerekmez, sadece her aşamanın toplam ağırlığı azaltacak şekilde oluşması beklenir. Bizim “lokal-optimal” veya “tight” olarak adlandırdığımız bu temsiller, metriklerin parametrizasyonu kullanılarak sistematik olarak incelenebilmektedir. Bu yöntemle verilen bir metrik yapısı için birden fazla lokal-optimal temsil bulunabilir ve bunlar belki de eşit olasılıkta, farklı oluşum seçeneklerini yansıtabilirler. Bu anlamda spantan oluşumlar içeren bazı problemler için lokal-optimal temsillerin optimal temsillerden daha anlamlı olduğu düşünülebilir.

- Proje çerçevesinde yapılan çalışmalar, filogenetik analiz örneklerine %10 ölçeğinde bir yaklaşıklık çerçevesinde bakılması gerektiğini göstermiştir. Çalışılan örnekler için ağaç ve döngü yapıları bu yaklaşıklık içerisinde ortaya çıkarılmış, verilen bir yaklaşıklık oranı için istenen metrik özellikler korunarak ortaya çıkacak yapılar belirlenmiştir. Ancak, verilen bir metriğin “temel” özelliklerini koruyarak yaklaşık olarak temsil edilmesi durumunu tasvir edecek kuramsal bir çerçeveye ihtiyaç vardır.

Kaynaklar

Alizon, M., et al., "Genetic variability of the AIDS virus: nucleotide sequence analysis of two isolates from African patients", Cell, 46(1):63-74, 1986.

Althöfer, I. 1988. “On optimal realizations of finite metric spaces by graphs”, Discrete & Computational Geometry, 3, 103-122.

Arnason U, Lammers F, Kumar V, Nilsson MA, Janke A, 2018. “Whole-genome sequencing of the blue whale and other rorquals finds signatures for introgressive gene flow”, Evolutionary Biology, Science Advances, 4(4): eaap9873.

Bandelt, H. J. & A.W.M. Dress, 1992, “A canonical decomposition theory for metrics on a finite set”, Advances in Mathematics, Vol. 92, No 1, pp.47-102.

Bergthorsson, U., Adams, K., Thomason, B., Palmer, J. 2003. “Widespread horizontal transfer of mitochondrial genes in flowering plants”, Nature 424, 197–201.

Bergthorsson, U., Richardson, A., Young, G., Goertzen, L., Palmer, J. 2004. “Massive horizontal transfer of mitochondrial genes from diverse land plant donors to basal angiosperm Amborella”, Proc. Nat’l Acad. Sci., USA 101, 17,747-752

Bilge, A. H., Çelik, D., Koçak, Ş. 2015. “Optimal embeddings of finite metric spaces into graphs”, Anadolu Univ. J. Sci. Technol. B- Theor. Sci.. 3 (2), 133–147.

Bilge, A. H., Çelik, D., Koçak, Ş. 2017. “An equivalence class decomposition of finite metric spaces via Gromov products”, Discrete Mathematics, 340, 1928–1932.

Bilge, A. H. ve M. İncegül, 2018a, “A matrix representation for the Gromov product decomposition of finite metric spaces”,preprint.

Bilge, A. H. ve M. İncegül, 2018b, “Gromov product decomposition of 7-point metric spaces”, arXiv, 1804.03051 (math.co) preprint.

Bilge, Çelik, Koçak, Rezaeinazhad, 2021a, Gromov product structures, quadrangle structures and split metric decompositions for finite metric spaces,

Bilge, Çelik, Koçak, Rezaeinazhad, 2021b Five-point metric spaces, Gromov product structures, quadrangle structures and explicit parametrizations (yayına sunuldu)

Boc, A., Diallo, Alpha B. and Makarenkov, V. 2012. “T-REX: a web server for inferring, validating and visualizing phylogenetic trees and networks”, Nucleic Acids Research, 40(W1), W573-W579.

Boesch, F.T., 1969. “Properties of the distance matrix of a tree”. Quarterly of applied mathematics, 26(4), pp.607-609.

Buneman, P. 1974. “A note on the metric properties of trees”, J. Combin. Theory Ser. B 17, 48-50.

Chaiken, S., Dewdney, A. K., Slater, P. S. 1983. “An optimal diagonal tree code”, SIAM J. Algebraic Discrete Methods 4, 42-49.

Chung, F., Garrett, M., Graham, R., Shallcross, D. 2001. “Distance realizations problems with applications to Internet tomography”, J. Comput. System Sci., 64, 432–448.

Cunningham, J. 1978. “Free trees and bidirectional trees as representations of psychological distance”, J. Math. Psych,. 17, 165–188.

Çelik, D., 2018, Explicit determination of Gromov product types of five-point metric spaces, Anadolu Univ. J. Sci. Technol. B- Theor. Sci., basım aşamasında.

Dress, A. W. M. 1984. “Trees, tight extensions of metric spaces, and the cohomological dimension of certain groups: a note on combinatorial properties of metric spaces”, Adv. in Math., 53, 321 – 402.

Dress, A., Huber, K.T., Koolen, J., Moulton, V., Spillner, A. 2010. “An algorithm for computing cutpoints in finite metric spaces”, J. Classification, 27 (2), 158-172.

Dress, A., Huber, K.T., Koolen, J., Moulton, V., Spillner, A. 2011. “Basic phylogenetic combinatorics”, Cambridge University Press, Cambridge.

Edgar, R.C. 2004. “MUSCLE: multiple sequence alignment with high accuracy and high throughput”, Nucleic Acids Research, 32(5):1792–97.

Felsenstein, J. 1981. “Evolutionary trees from DNA sequences: a maximum likelihood approach”, J. Mol. Evol., 17, 368–376.

Gao, F., et al., "Origin of HIV-1 in the chimpanzee Pan troglodytes troglodytes", Nature, 397(6718):436-41, 1999.

Gemeinholzer, B. 2008. “Phylogenetic networks”. In: B.H. Junker, F. Schreiber (eds.) Analysis of Biological Networks, pp. 255–282. John Wiley and Sons Ltd.

Grishljkhin, V.P., 1992. “Computing extreme rays of the metric cone for seven points”. European Journal of Combinatorics, 13, pp. 153,165

Hakimi, S. L. 1964. “Optimum locations of switching centers and medians of a graph”, Oper. Res., 12, 450-459.

Hakimi, S. L., Yau, S. S. 1964. “Distance matrix of a graph and its realizability”, Quart. AppL Math., 22, 305-317.

Hertz, A., Varone, S. 2008. “The metric cutpoint partition problem”, J. Classification, 25, 159–175.

Huson, D.H. and Bryant, D. 2006. “Application of Phylogenetic Networks in Evolutionary Studies”, Molecular Biology and Evolution, 23(2):254-267.

Imrich, W. 1977. “On metric properties of tree like spaces”, Contributions to Graph Theory and Its Applications, 129-156.

Imrich, W., Simres-Pereira, J. M. S., Zamfirescu, C. M. 1984. “On optimal embeddings of metrics in graphs”, J. Combin. Theory Ser. B, 36, 1-15.

M. İncegül, ve Bilge, A. H. ve 2018, Algorithms for Gromov product structures, preprint.

Jin, L. and Nei, M. 1990. “Limitations of the evolutionary parsimony method of phylogenetic analysis”, Mol. Biol. Evol., 7, 82-102.

Katoh K, Kuma K, Toh H, Miyata T. 2005. “MAFFT version 5: improvement in accuracy of multiple sequence alignment”, Nucleic Acids Res. Jan 20;33(2):511-8. (doi: 10.1093/nar/gki198)

Kestler, H.W., et al., "Comparison of simian immunodeficiency virus isolates", Nature, 331(6157):619-22, 1998.

Khaefi, R., Esmaeili, H.R., Chermahini, M.A. 2018. "Natural Hybridization of *Luciobarbus barbulus* x *Luciobarbus kersin* and *Luciobarbus barbulus* x *Luciobarbus xanthopterus* in the Persian Gulf Basin", Turk. J. Fish. Aquat. Sci. 18, 1399-1407.

Kilic, M. ve S. Kocak, 2016, "Tight span of subsets of the plane with the maximum metric", Advances in Mathematics, Vol. 301 pp.693-710

Kimura, M. 1983. "The neutral theory of molecular evolution", Cambridge University Press, Cambridge.

Kolbig Schwarz;

Koolen, J. H., Lesser, A., Moulton, V., Wu, T. 2012. "Injective optimal realizations of finite metric spaces", Discrete Mathematics, 312, 1602–1610.

Koolen, J., Lesser, A., Moulton, V. 2009. "Optimal realization of generic 5-point metrics", European J. Combin., 30 (5), 1164–1171.

Koolen, J., Moulton, V., Tonges, U., 2000. "A classification of six-point prime metrics", European J. Combin., 21, 815-829.

Kurland, C., Canback, B., Berg, O. 2003. "Horizontal gene transfer: A critical view", Proc. Nat'l Acad. Sci., USA (17), 9658–9662.

Landry, P., Lapointe, F., Kirsch, J. 1996. "Estimating phylogenies from distance matrices: additive is superior to ultrametric estimation", Mol. Biol. Evol., 13, 818-823.

Le, S. Q. and Gascuel, O. 2008. "LG: an improved, general amino-acid replacement matrix", Mol. Biol. Evol., 25, 1307-1320.

Le Boulange, E., Legendre, C., De Le Court, P, Le Boulange, N, Languy M. 1996. "Microgeographic morphological differentiation in muskrats. J. Mammal., 77, 684–701.

Legendre, P. ve Makarenkov, V. 2002. "Reconstruction of Biogeographic and Evolutionary Networks Using Reticulograms", *Systematic Biology,* 51(2), 199-216.

Li, M., Chen, X., Li, X., Ma, B., Vitanyi, P. 2004. "The similarity metric", IEEE Trans. Inf. Theory, 50, 3250–3264.

Morrison D. 2005. "Networks in phylogenetic analysis: new tools for population biology", International J for Parasitology, 35:567–582.

Mower, J., Stefanovic, S., Young, G., Palmer, J. 2004. "Gene transfer from parasitic to host plants". Nature, 165–166.

Mallet, J. 2007. "Hybrid speciation", Nature, 279–283.

Nakamura, Y., Itoh, T., Matsuda, H. ,Gojobori, T. 2004. "Biased biological functions of horizontally transferred genes in prokaryotic genomes", Nature Genetics (7), 760–766.

Nei, M. and Miller, J.C. 1990. "A simple method for estimating average number of nucleotide substitutions within and between populations from restriction data", Genetics, 125:873–879.

Ochman, H., Lawrence, J., Groisman, E. 2000. "Lateral gene transfer and the nature of bacterial innovation", Nature 405(6784), 299–304.

Patrinos, A. N., Hakimi, S. L. 1972. "The distance matrix of a graph and its tree realization", Quart. AppL Math., 30, 255-269.

Posada D, Crandall K. 2001. “Intraspecific gene genealogies: trees grafting into networks”, Trends in Ecology and Evolution, 16:37–45

Posada, D., Crandall, K. 2002. “The effect of recombination on the accuracy of phylogeny estimation”, J. Mol. Evol. 54(3), 396–402

Semple, C. and Steel, M., 2003. “*Phylogenetics”* (Vol. 24). Oxford University Press on Demand. (tree metrics)

Simfes-Pereira, J. M. S. 1969. “A note on the tree realizability of a distance matrix”, JL Combin. Theory, 6, 303-310.

Sturmfels, B., Yu, J. 2004. “Classification of six-point metrics”, Electron. J. Combin., 11, R44.

Steel, M.A. 1994. “Recovering a tree from the leaf colorations it generates under a Markov model”, Appl. Math. Lett., 7(2):19–24.

Swofford, D.L., Olsen, G.J., Waddell, P.J. and Hillis, D.M. 1996. Chapter 11: Phylogenetic inference. In D. M. Hillis, C. Moritz, and B. K. Mable, editors, Molecular Systematics, pages 407–514. Sinauer Associates, Inc., 2nd edition.

Tajima, F. and Nei, M. 1984. “Estimation of evolutionary distance between nucleotide sequences”, Mol. Biol. Evol., 1, 269-285.

Tamura, K. 1992. “Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G+C content biases”, Mol. Biol. Evol., 9, 678–687.

Thompson, J.D., Higgins, D.G. and Gibson, T.J. 1994. “CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice”, Nucl. Acids Res, 22:4673-4680.

Varone, S. 2006. “A constructive algorithm for realizing a distance matrix”, European J. Oper. Res., 174, 102–111.

Whelan, S. and Goldman, N. 2001. “A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach”, Mol. Biol. Evol., 18, 691-699.

Winkler, P. 1988. “The complexity of metric realization”, SIAM J. Discrete Math., 1, 552–559.

Establishing the Connection between the Hypersimplex Classification, Gromov Product Structures and Quadrangle Structures

Ayşe H. Bilge, Şahin Koçak, Derya Çelik, Arash M. Rezaeinazhad

Introduction

The set C_n of all pseudo-metrics $d = (d_{ij}) \in \mathbb{R}^{\binom{n}{2}}$ on a given n-point set X, is called the metric cone and consists of all the n-point finite metric spaces. The problem of classification of these finite metric spaces according to a meaningful and manageable approach is a problem still under investigation. Several attempts have been done for this aim. Hypersimplex classification [5], Gromov product structure [1], and quadrangle structures [2] are tools to decompose the metric cone. In this paper this concepts will be reviewed and their relations to each other will be demonstrated.

Let (X, d) be a finite metric space with n elements P_i, $i = 1, \ldots, n$ $(n \geq 3)$ and let d_{ij} be the distance between P_i and P_j. The elements of X are also referred to as "vertices" or "nodes". E_{ij} and T_{ijk} denote respectively an edge and a triangle with corresponding vertices.

Gromov products The quantity Δ_{ijk}, defined as

$$\Delta_{ijk} = \Delta_{ikj} = \tfrac{1}{2}(d_{ij} + d_{ik} - d_{jk}) \tag{1}$$

is called the Gromov product of the triangle T_{ijk} at the vertex P_i [4].

Δ-generic metrics A metric space is called Δ-generic, if for each P_i the set of Gromov products Δ_{ijk} has a unique smallest element.

Gromov product structures Let (X, d) be a Δ-generic finite metric space. Let $P_i \in X$, and let Δ_{ijk} be the minimal Gromov product at P_i, $(i = 1, \ldots, n)$. The function that assigns the edge E_{jk} to the vertex P_i is called the Gromov product structure on X. Two Δ-generic metric spaces (X, d) and (X, d') are Δ-equivalent, if the corresponding Gromov product structures are the same up to a permutation of X.

Quadrangle generic metric spaces An n-point finite metric space X is called "quadrangle generic", or Q-generic, if for every 4-point subset $\{P_a, P_b, P_c, P_d\} \subseteq X$, the set of distances

$$\{d_{ab} + d_{cd}, d_{ac} + d_{bd}, d_{ad} + d_{bc}\}$$

has a unique maximal element.

Quadrangle Structures A quadrangle structure on a Q-generic finite metric space (X, d) is a map which assigns to any 4-point subset $\{P_a, P_b, P_c, P_d\}$ of X, the pair of edges corresponding to the maximal element of the set $\{d_{ab} + d_{cd}, d_{ac} + d_{bd}, d_{ad} + d_{bc}\}$. We denote the 4-point subset $\{P_a, P_b, P_c, P_d\}$ without any restriction on the sides by $Q(a, b, c, d)$ in which the ordering of the indices is irrelevant. If $d_{ac} + d_{bd}$ is maximal, the vertices are ordered as (P_a, P_b, P_c, P_d) and we denote this structured quadrangle by $Q(abcd)$ in which the cyclic permutations and reversal of the order of the indices give equivalent quadrangles. Two Q-generic metric spaces (X, d) and (X, d') are called Q-equivalent, if the corresponding quadrangle structures are the same up to a permutation of X.

Hypersimplex Decomposition Let $\mathcal{A}_n = \{\mathbf{e}_i + \mathbf{e}_j : 1 \leq i < j \leq n\} \subseteq \mathbb{R}^n$. This is the set of column vectors of the vertex-edge incidence matrix of the complete graph K_n. The convex hull of $\mathcal{A}_n$ is called the **second hypersimplex** of order n and is denoted by $\Delta(n, 2)$. In general, the **r-th hypersimplex** in $\mathbb{R}^n$ is the convex hull $\Delta(n, r)$ of the configuration

$$\{\mathbf{e}_{i_1} + \mathbf{e}_{i_2} + \ldots + \mathbf{e}_{i_r} : 1 \leq i_1 < i_2 < \ldots < i_r \leq n\}$$

Two metrics from the metric cone will be called equivalent if they give rise to the same triangulation of this second hypersimplex, thus the name Hypersimplex classification. We will first give the necessary definitions, then will give a second equivalent definition which uses much less definitions from algebraic geometry and is easy to work with algorithmically. For more on the following material the reader is referred to [6].

Fan, Normal Cone and Normal Fan Let P be any polyhedron in $\mathbb{R}^n$ and $w \in \mathbb{R}^n$, viewed as a linear functional. Define

$$face_w(P) := \{\mathbf{u} \in P \mid w.\mathbf{u} \geq w.\mathbf{v} \; for \; all \; \mathbf{v} \in P\}.$$

Every subset F of P which has this form, that is, which maximizes some linear functional, is called a face of P. Note that $P = face_0(P)$ is a face of itself.

A (polyhedral) complex Δ is a finite collection of polyhedra in $\mathbb{R}^n$ such that

1. if $P \in \Delta$ and F is a face of P, then $F \in \Delta$;
2. if $P_1, P_2 \in \Delta$, then $P_1 \bigcap P_2$ is a face of P_1 and of P_2.

The support of a complex Δ is $|\Delta| := \bigcup_{P \in \Delta} P$. A complex Δ which consists of cones is called a **fan**. A fan Δ is complete if $|\Delta| = \mathbb{R}^n$. If $P \subset \mathbb{R}^n$ is a polyhedron and F a face of P, then the **normal cone** of F at P is

$$\mathcal{N}_P(F) := \{w \in \mathbb{R}^n \mid face_w(P) = F\}.$$

If F and F' are faces of P, then F' is a face of F if and only if $\mathcal{N}_P(F)$ is a face of $\mathcal{N}_P(F')$. Hence the collection of normal cones $\mathcal{N}_P(F)$, where F ranges over the faces of P, is a fan. This fan is denoted by $\mathcal{N}(P)$ and called the **normal fan** of P.

Triangulations A triangulation of a convex polytope P is a decomposition of P into a finite number of simplices such that the intersection of any two of these simplices is a common face of them both (maybe empty). We regard a triangulation as a collection of its simplices of maximal dimension. All the lower-dimensional simplices are just faces of the maximal ones. By a triangulation of (P, A), we mean simply a triangulation of P into simplices with vertices in A. Note that it is not required that every element of A to appear as a vertex of a simplex. Let T be a triangulation of (P, A) and fix a translation invariant volume form Vol on $\mathbb{R}^n$. By the characteristic function of T we mean the function $\phi_T : A \longrightarrow \mathbb{R}$ defined as:

$$\phi_T(w) = \sum_{\sigma \,:\, \text{w is a vertex of } \sigma} Vol(\sigma)$$

where the summation is over all (maximal) simplices of T for which w is a vertex. In particular, $\phi_T(w) = 0$ if w is not a vertex of any simplex of T.

Fix a positive vector called weight vector $w \in \mathbb{R}^n_{\geq 0}$. Given any term $c_{\mathbf{u}}\mathbf{x}^{\mathbf{u}}$ of a polynomial $f \in k[\mathbf{x}]$, where $c_{\mathbf{u}}$ is a nonzero scalar in k, the weight of $c_{\mathbf{u}}\mathbf{x}^{\mathbf{u}}$ is the dot product $w.u = w_1u_1 + \ldots + w_nu_n$. For a polynomial $f \in k[\mathbf{x}]$, the sum of all terms of f having maximal weight under w is called the initial form of f and is denoted by $in_w(f)$. If I is any ideal inside $k[\mathbf{x}]$, then we express the initial ideal of I under the weight order defined by w as the ideal generated by initial forms of all polynomials in I. So

$$in_w(I) :=< in_w(f) \; : \; f \in I > .$$

We say that w is generic for I if $in_w(I)$ is a monomial ideal. A finite subset $\mathcal{G}$ of I is a **Gröbner basis** for I with respect to w if $in_w(I)$ is generated by $\{in_w(g) \; : \; g \in \mathcal{G}\}$. If w is generic for I and every element of $\mathcal{G}$ is monic, then $\mathcal{G}$ is a reduced Gröbner basis if the initial term of each element in $\mathcal{G}$ does not divide any term of any other element in $\mathcal{G}$. We can assume that $\prec$ is any term order in $k[\mathbf{x}]$. $in_{\prec}(f)$ and $in_{\prec}(I)$ are defined similarly like the weigh order. It is convenient to identify the zero set of $in_w(I)$ with a simplicial complex.

The initial complex $\Delta_w(I)$ of I is the simplicial complex on the vertex set $\{1, 2, \ldots, n\}$ defined by the following rule: A subset $F \subset \{1, 2, \ldots, n\}$ is a face of $\Delta_w(I)$ if there is no polynomial $f \in I$ whose initial monomial $in_w(f)$ uses only the variables $\{x_i \ : \ i \in F\}$ (or whose initial monomial $in_w(f)$ has support F).

Identify a given set of latice points $\mathcal{A} = \{\mathbf{a}_i, ..., \mathbf{a}_n\} \subset \mathbb{Z}^d$ with index set $\{1, 2, \ldots, n\}$. If σ is a subset of $\mathcal{A}$ then we write $pos(\sigma)$ for the cone spanned by σ. A triangulation of $\mathcal{A}$ is a collection Δ of subsets of $\mathcal{A}$ such that $\{pos(\sigma) \ : \ \sigma \in \Delta\}$ is the set of cones in a simplicial fan whose support equals $pos(\mathcal{A})$. Every sufficiently generic vector $w \in \mathbb{R}^n$ defines a triangulation Δ_w as follows:
A subset $\{i_1, i_2, \ldots, i_r\}$ is a face of Δ_w if there exists a vector $\mathbf{c} = (c_1, \ldots, c_d) \in \mathbb{R}^d$ such that

$$\begin{aligned} &\mathbf{a}_j.\mathbf{c} = w_j \ \ if \ j \in \{i_1, i_2, \ldots, i_r\} \ and \\ &\mathbf{a}_j.\mathbf{c} < w_j \ \ if \ j \in \{1, 2, \ldots, n\} \setminus \{i_1, i_2, \ldots, i_r\} \end{aligned}$$

A triangulation Δ of $\mathcal{A}$ is called **regular** (or coherent) if $\Delta = \Delta_w$ for some $w \in \mathbb{R}^n$. Regular triangulations can be defined geometrically too.

1. Use the coordinates of w as "heights", and lift the configuration $\mathcal{A}$ into the next dimension. The result is the configuration $\hat{\mathcal{A}} = \{(\mathbf{a}_1, w_1), (\mathbf{a}_2, w_2), \ldots, (\mathbf{a}_n, w_n)\} \subset \mathbb{R}^{d+1}$.
2. The "lower faces" of the cone $pos(\hat{\mathcal{A}})$ form a d-dimensional polyhedral complex. (A face is "lower" if it has a normal vector with negative last coordinate). The triangulation Δ_w is the image of this complex under projection onto the first d coordinates.

The reader is highly advised to not confuse the notion of Δ_w for **triangulation with a generic vector** w with the similar notation of $\Delta_w(I)$ for **initial complex** and also the notion of Δ for **triangulation**s and **polyhedral complexes** defined above. The usage of last one must be clear from the context.

Polyhedral Fan With every regular triangulation Δ of a set $\mathcal{A}$ we associate the polyhedral cone

$$\mathcal{C}_{\Delta} := \{w \in \mathbb{R}^n \ : \ \Delta_w = \Delta\}.$$

The cone $\mathcal{C}_{\Delta}$ consists of all lifting functions which induce the triangulation Δ. The collection of these cones together with their faces is a **polyhedral fan**.

Theorem 1. [Gel'fand, Kapranov and Zelevinsky] There exists an $(n-d)$-dimensional polytope $\Sigma(A)$, such that the normal cones at the vertices of $\Sigma(A)$ are precisely the cones $\mathcal{C}_{\Delta}$ [3].

Secondary Polytope The polytope $\Sigma(A)$ in Theorem 1 above is called the **secondary polytope**. It can be shown that $\Sigma(A)$ is the convex hull in the space $\mathbb{R}^n$ of the vectors $\phi_T(w)$ for all the triangulations T of (P, A). In other terms:

$$\Sigma(A) := conv\{\phi_T \ |\text{T is a triangulation of (P,A)}\}$$

Secondary Fan The normal fan of secondary polytope $\mathcal{N}(\Sigma(A))$ is called the **secondary fan** and

$$\mathcal{N}(\Sigma(A)) = \{\mathcal{C}_{\Delta} \ : \ \Delta \text{ is a regular triangulation of } \mathcal{A} \ \}.$$

Hypersimplex Decomposition of Metric Cone: First Description The subject of study in [5] is the canonical subdivision of metric cone C_n which is called the metric fan and is denoted by MF_n. The metric fan is the secondary fan of the second hypersimplex. Every metric d defines a regular polyhedral subdivision Δ_d of $\Delta(n, 2)$ in the following way: The vertices of $\Delta(n, 2)$ are identified with the edges of the complete graph K_n, and subpolytopes of $\Delta(n, 2)$ correspond to arbitrary subgraphs of K_n. A subgraph G is a cell of Δ_d if there exists an $x \in \mathbb{R}^n$ such that

$$x_i + x_j = d_{ij} \ \text{ if } \{i, j\} \in G \text{ and } x_i + x_j > d_{ij} \ \text{ if } \{i, j\} \notin G$$

Two metrics d and d' lie in the same cone of the metric fan MF_n if they induce the same subdivision $\Delta_d = \Delta_{d'}$ of the second hypersimplex $\Delta(n, 2)$ or in other words give rise to the same triangulations of $\Delta(n, 2)$. A metric is called generic in this sense if the corresponding triangulation is regular.

Hypersimplex Decomposition of Metric Cone: Second Description Consider the $\binom{n}{2} \times n$ matrix $\mathcal{A}$ where the rows are labeled by

$$(1,2),(1,3),\ldots,(1,n),(2,3),(2,4),\ldots,(2,n),\ldots,(n-1,n)$$

and the (i,j)-row $(i<j)$ is given by $e_i + e_j = (0,\ldots,1,\ldots,1,\ldots,0) \in \mathbb{R}^n$.

Let $\mathcal{B}$ be an invertible $n \times n$ submatrix of $\mathcal{A}$ and denote the $\left[\binom{n}{2} - n\right] \times n$ matrix obtained by deleting $\mathcal{B}$ from $\mathcal{A}$ by $\mathcal{B}'$. Likewise, define $d_{\mathcal{B}} \in \mathbb{R}^n$ by choosing the components of $d \in \mathbb{R}^{\binom{n}{2}}$ corresponding to $\mathcal{B}$ and $d_{\mathcal{B}'} \in \mathbb{R}^{\binom{n}{2}-n}$ corresponding to $\mathcal{B}'$. Now consider the following system of equations and inequalities for $x \in \mathbb{R}^n$:

$$\mathcal{B}x = d_{\mathcal{B}} \text{ and } \mathcal{B}'x > d_{\mathcal{B}'}. \tag{2}$$

If this system has a solution we say that the matrix $\mathcal{B}$ is a "cell" or a "thrackle" for the metric d. We denote the collection of cells of a metric d by $Cell(d)$. This terminology stems from the fact that the row vectors of $\mathcal{B}$ can be viewed as the vertices of an $(n-1)$-simplex in $\mathbb{R}^n$. Still another interpretation is that, a row vector $e_i + e_j$ can be viewed as an edge of the complete graph K_n with n nodes so that a cell $\mathcal{B}$ can be viewed as a sub-graph of K_n.

Now we define two metrics d and d' on an n-point set X to be equivalent if they have the same collection of cells, i.e. $Cell(d) = Cell(d')$ (or what amounts to the same, the same collection of sub-graphs). The equivalence class of a metric d is a sub-cone of the metric cone and these sub-cones constitute altogether the metric fan.

The variables x_{ij} are indexed by the edges in the complete graph K_n. For the cases $n = 4, 5$, the hypersimplex $\Delta(4,2)$ is a regular octahedron in $\mathbb{R}^4$. It has three distinct regular triangulations. Therefore the secondary polytope, is a triangle in $\mathbb{R}^6$. The hypersimplex $\Delta(5,2)$ has dimension four with 10 vertices and 10 facets (5 tetrahedra and 5 octahedra). Its secondary polytope $\Sigma(\mathcal{A}_5)$ is five-dimensional and has 102 vertices, 255 edges, 240 two-faces, 105 three-faces and 20 facets. Under the natural S_5-action the 102 regular triangulations of $\Delta(5,2)$ fall into three distinct orbits.

In Table 1 the numerical instances for 5-point metric spaces is given.

Table 1: Numeric examples of 5-point metric spaces

	d_{12}	d_{13}	d_{14}	d_{15}	d_{23}	d_{24}	d_{25}	d_{34}	d_{35}	d_{45}
Type A	*10*	*11*	*10*	*7*	*8*	*10*	*10*	*6*	*10*	*8*
Type B	*5*	*6*	*6*	*5*	*3*	*5*	*6*	*5*	*5*	*4*
Type C	*6*	*10*	*6*	*8*	*8*	*6*	*10*	*9*	*6*	*8*

For $n = 6$, since $\Delta(6,2)$ has 339 regular triangulations, by means of hypersimplex decomposition there 339 classes of 6-point metrics. These types can also be studied in terms of "cells" or "thrackles" that we have introduced. First we will give two results:

Proposition 1. Each cell, is not a tree and also does not include any cycle of even length.

Proof. Considering that each cell has n vertices and n edges, since it has more edges than a tree can have, it is not a tree (Note that if there was an isolated vertex in the cell, then there was a row consisting of zeroth which would cause the matrix to be singular). For proving that each cell does not include any cycle of even length, consider that it has an even cycle of length $2k$ $(2 \leq k \leq \frac{n}{2})$ and assume that $i_1, i_2, \ldots, i_{2k-1}, i_{2k}$ are the vertices of this cycle labeled counterclockwise. By the definition, the edges of this cycle are $e_{i_1} + e_{i_2}, e_{i_2} + e_{i_3}, \ldots, e_{i_{2k-1}} + e_{i_{2k}}, e_{i_{2k}} + e_{i_1}$. Since

$$(e_{i_1} + e_{i_2}) - (e_{i_2} + e_{i_3}) + \cdots + (e_{i_{2k-1}} + e_{i_{2k}}) - (e_{i_{2k}} + e_{i_1}) = 0$$

a subset of rows of corresponding matrix is not linearly independent and hence the whole matrix is not invertible, a clear contradiction with our assumption about the invertible matrix. This completes the proof. □

Proposition 2. Consider that for a given finite metric space (X,d) all the triangle inequalities $d_{ij} + d_{jk} > d_{ik}$ hold strictly for all distinct $1 \leq i,j,k \leq n$. Then the x_i's in (2), for each cell are positive numbers.

Proof. Our proof will be in 3 steps.

1. If a triangle with edges d_{ij}, d_{jk} and d_{ik} is recognized, then x_i, x_j and x_k (which can be seen as the vertices of this triangle) are positive. For this claim we should keep in mind that since the edges are realized, by (1) the followings hold:

$$x_i + x_j = d_{ij},\ x_i + x_k = d_{ik},\ x_j + x_k = d_{jk}.$$

   By adding the two first equations and using the third one, one has $x_i = \frac{1}{2}(d_{ij} + d_{ik} - d_{jk}) > 0$. By symmetry x_j and x_k are also positive.

2. Assume that a cycle of odd length (proposition 1) is recognized in a cell. First we must mention that a cell can have several odd cycles (which cannot be connected together, since the number of edges must match the number of vertices). Consider one of these cycles of length $2k+1$ and assume that its vertices are labeled clockwise as $x_{i_1}, x_{i_2}, \ldots, x_{i_{2k}}, x_{i_{2k+1}}$. Since these edges are realized, by (1) we have:

$$\begin{aligned} & x_{i_1} + x_{i_2} = d_{i_1 i_2},\ x_{i_2} + x_{i_3} = d_{i_2 i_3}, \\ & \ldots \\ & x_{i_{2k}} + x_{i_{2k+1}} = d_{i_{2k} i_{2k+1}},\ x_{i_{2k+1}} + x_{i_1} = d_{i_1 i_{2k+1}}. \end{aligned}$$

   Consider a vertex x_{i_j}. Since x_{i_j} is not connected to $x_{i_{j+2}}$ and $x_{i_{j+3}}$, by (1) $x_{i_j} + x_{i_{j+2}} > d_{i_j i_{j+2}}$ and $x_{i_j} + x_{i_{j+3}} > d_{i_j i_{j+3}}$. Adding these two inequalities will result in $2x_{i_j} + x_{i_{j+2}} + x_{i_{j+3}} > d_{i_j i_{j+2}} + d_{i_j i_{j+3}}$. Now since $x_{i_{j+2}}$ and $x_{i_{j+3}}$ are neighbor vertices, by (1) we have $x_{i_{j+2}} + x_{i_{j+3}} = d_{i_{j+2} i_{j+3}}$ and thus $2x_{i_j} + d_{i_{j+2} i_{j+3}} > d_{i_j i_{j+2}} + d_{i_j i_{j+3}}$. So $x_{i_j} > \frac{1}{2}(d_{i_j i_{j+2}} + d_{i_j i_{j+3}} - d_{i_{j+2} i_{j+3}}) > 0$.

3. Assume that the vertex x_k is not one of the vertices in the odd cycle of step 2 but lies in the same component, that is it's connected with a path to this cycle at vetrex x_{i_j}. We claim that x_k is not connected to $x_{i_{j+1}}$ and $x_{i_{j+2}}$. For proving this claim first let us assume that x_k is connected to $x_{i_{j+1}}$ via an edge. Depending on the parity of length of the path connecting x_k to x_{i_j} two scenarios are possible; This path is of odd length, then the cycle $[(x_k, \ldots, x_{i_j}), (x_{i_{j-1}}, x_{i_{j-2}}, \ldots, x_{i_{j+1}}), x_k]$ is of even length and this contradicts proposition 1. If the path connecting x_k to x_{i_j} is of even length then the cycle $[(x_k, \ldots, x_{i_j}), (x_{i_{j+1}}, x_k)]$ would also have even length which again contradicts proposition 1. This proofs that x_k is not connected to $x_{i_{j+1}}$. A similar argument will show that x_k is not connected to $x_{i_{j+2}}$ too. One should keep in mind that if the main cycle in step 2 is not of odd length then these arguments would have been false.

   Now since there are no edges between $x_k, x_{i_{j+1}}$ and $x_k, x_{i_{j+2}}$ by (1), $x_k + x_{i_{j+1}} > d_{ki_{j+1}}$ and $x_k + x_{i_{j+2}} > d_{ki_{j+2}}$. But since $x_{i_{j+1}}$ and $x_{i_{j+2}}$ are neighbors we have $x_{i_{j+1}} + x_{i_{j+2}} = d_{i_{j+1} i_{j+2}}$. Now by adding the two last inequalities we have $2x_k + x_{i_{j+1}} + x_{i_{j+2}} > d_{ki_{j+1}} + d_{ki_{j+2}}$, thus $x_k = \frac{1}{2}(d_{ki_{j+1}} + d_{ki_{j+2}} - d_{i_{j+1} i_{j+2}}) > 0$.

This completes the proof. □

We can obtain all graphs for cells by counting arguments. For example, for $n = 5$, if there is a 3 cycle, the remaining 2 edges can be attached to the same vertex or to different vertices etc. In Figures 1 and 2 all the cell types for 5-point and 6-point metric spaces are shown.

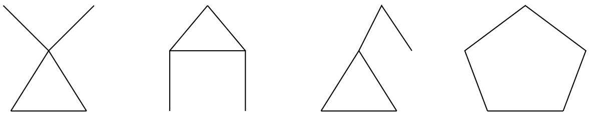


Figure 1: All possible unlabeled cells for a 5-point metric space.

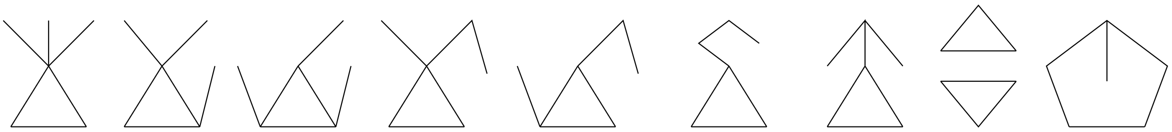


Figure 2: All possible unlabeled cells for a 6-point metric space.

Before we continue, there is a useful corollary about Gromov products that we will use later. The proofs are given in [1].

Lemma 1. Let (X,d) be a finite metric space with n elements P_i, $i=1,\dots,n$. Then the following equations hold.

$$\Delta_{ijl}-\Delta_{ijk}=\Delta_{kjl}-\Delta_{kil}=\Delta_{lik}-\Delta_{ljk}=\Delta_{jik}-\Delta_{jil}.$$

where $i,j,k,l=1,2,\dots,n$.

Corollary 1. Let (X,d) be a Δ-generic finite metric space and let Δ_{ijk} be the minimal Gromov product at node P_i. Then,

1. Δ_{jkl} $(=\Delta_{jlk})$ cannot be minimal at node P_j, where $l\neq j,k$,
2. Δ_{kjl} $(=\Delta_{klj})$ cannot be minimal at node P_k, where $l\neq j,k$,
3. Δ_{lij} $(=\Delta_{lji})$ and Δ_{lik} $(=\Delta_{lki})$ cannot be minimal at node P_l, where $l\neq i,j,k$.

It is shown in [1], that any 5-point metric space (X,d), is equivalent to a metric space (X,d') with one of the following Gromov product structures:

$$S(\mathcal{E}_1)=\{\Delta_{125},\Delta_{213},\Delta_{324},\Delta_{435},\Delta_{514}\},\ \text{type } A$$
$$S(\mathcal{E}_2)=\{\Delta_{125},\Delta_{213},\Delta_{325},\Delta_{425},\Delta_{514}\},\ \text{type } B$$
$$S(\mathcal{E}_3)=\{\Delta_{125},\Delta_{213},\Delta_{325},\Delta_{425},\Delta_{513}\}.\ \text{type } C$$

On decomposition of $\Delta(5,2)$ and classification of five-point metric spaces

For the case $n=5$, $\Delta(5,2)$ is the hypersimplex generated by columns of matrix

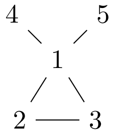
$$A^t=\begin{bmatrix}1&1&1&1&0&0&0&0&0&0\\1&0&0&0&1&1&1&0&0&0\\0&1&0&0&1&0&0&1&1&0\\0&0&1&0&0&1&0&1&0&1\\0&0&0&1&0&0&1&0&1&1\end{bmatrix}$$

From A we can extract 252 system of five equations in five variables and five inequalities. Only 162 of such submatrices have rank 5. Let us assume that A_1 is the matrix of equalities. According to how many of x_i's we include at most into system of equations we can categorize three general types of systems:

(a) system type (I): At most four x_i for each $i=1,\dots,5$ can appear in the equations, the typical instance of such a system is (columns (1,2,3,4,5) of A^t):

$$x_1+x_2=d_{12}\ ,\ x_1+x_3=d_{13}\ ,\ x_1+x_4=d_{14}\ ,\ x_1+x_5=d_{15}\ ,\ x_2+x_3=d_{23},$$
$$x_2+x_4>d_{24}\ ,\ x_2+x_5>d_{25}\ ,\ x_3+x_4>d_{34}\ ,\ x_3+x_5>d_{35}\ ,\ x_4+x_5>d_{45}.$$

For this choice of equalities, one will have a graph like



By solving the equations, and then replacing the x_i in the inequalities, the following inequalities are drived:

$$d_{12}+d_{34}<d_{14}+d_{23}\ (I)\ ,\ d_{12}+d_{35}<d_{15}+d_{23}\ (III),$$
$$d_{13}+d_{24}<d_{14}+d_{23}\ (II)\ ,\ d_{13}+d_{25}<d_{15}+d_{23}\ (IV),$$
$$\Delta_{123}<\Delta_{145}\ (V)$$

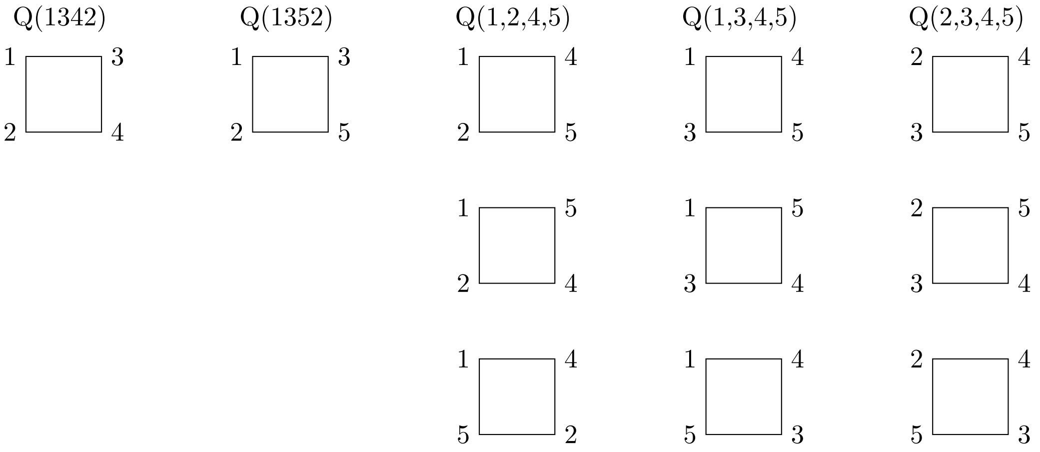


Figure 3: Allowable quadrangles for a 5-point metric space with a graph of type I.

We will interpret these inequalities on quadrangle structures. There are 15 different quadrangles for a five point metric space, which can be grouped to 5 sets of 3 quadrangles according to which vertex x_i they do not contain and the relative settlement of diagonals (Figure 3). From (I) and (II) one can conclude that in a metric space which does have such a graph, the only legitimate quadrangle is $Q(1342)$ and from (III) and (IV) it is concluded that the only legitimate quadrangle is $Q(1352)$. We cannot conclude anything about Q(1,2,4,5), Q(1,3,4,5) and Q(2,3,4,5) (Figure 3). We will ask this question: Given these possible quadrangles which Gromov product structures cannot be put on this metric space? For answering this question it should be noted that each inequality $d_{12} + d_{34} < d_{14} + d_{23}$ gives four inequalities among Gromov products (one for each vertex of quadrangle):

$$
\begin{aligned}
d_{12} + d_{34} < d_{14} + d_{23} &\Leftrightarrow d_{12} + d_{13} - d_{23} < d_{13} + d_{14} - d_{34} \Leftrightarrow \Delta_{123} < \Delta_{134},\\
d_{12} + d_{34} < d_{14} + d_{23} &\Leftrightarrow d_{12} + d_{24} - d_{14} < d_{23} + d_{24} - d_{34} \Leftrightarrow \Delta_{214} < \Delta_{234},\\
d_{12} + d_{34} < d_{14} + d_{23} &\Leftrightarrow d_{13} + d_{34} - d_{14} < d_{13} + d_{23} - d_{12} \Leftrightarrow \Delta_{314} < \Delta_{312},\\
d_{12} + d_{34} < d_{14} + d_{23} &\Leftrightarrow d_{24} + d_{34} - d_{23} < d_{14} + d_{24} - d_{12} \Leftrightarrow \Delta_{423} < \Delta_{412}.
\end{aligned}
$$

For this case, using the inequalities, the allowable quadrangles are depicted in Figure 3. Considering Corollary 1, one can write the list of all Gromov product structures and eliminate the impossible ones.

For the given graph, the possible Gromov structures are the followings (considering the symmetry for 4 and 5, the case when Δ_{215} is minimal is omitted):

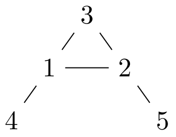
123, 214, 314, 423, 514.
123, 214, 314, 423, 523.
123, 214, 314, 425, 514.
123, 214, 314, 435, 514.
123, 214, 315, 423, 523.
123, 214, 315, 425, 534.
123, 214, 345, 423, 523.
123, 245, 314, 423, 523.
123, 245, 315, 423, 523.
123, 245, 345, 423, 523.

(b) system type (II): At most three x_i for each $i = 1, \ldots, 5$ can appear in the equations, this systems can be put into four subtypes:

  (i) (II)-1 : the typical instance of such a system is (columnss (1,2,3,5,7) of A^t):

$$
\begin{aligned}
&x_1 + x_2 = d_{12}\ ,\ x_1 + x_3 = d_{13}\ ,\ x_1 + x_4 = d_{14}\ ,\ x_2 + x_3 = d_{23}\ ,\ x_2 + x_5 = d_{25},\\
&x_1 + x_5 > d_{15}\ ,\ x_2 + x_4 > d_{24}\ ,\ x_3 + x_4 > d_{34}\ ,\ x_3 + x_5 > d_{35}\ ,\ x_4 + x_5 > d_{45}.
\end{aligned}
$$

For this choice of equalities, one will have a graph like



Solving the equations, and then replacing the x_i in the inequalities, the following formulas can be obtained:

$$d_{12} + d_{35} < d_{13} + d_{25} \ (I) \ , \ d_{12} + d_{34} < d_{14} + d_{23} \ (III),$$
$$d_{15} + d_{23} < d_{13} + d_{25} \ (II) \ , \ d_{13} + d_{24} < d_{14} + d_{23} \ (IV),$$
$$d_{12} + d_{45} < d_{14} + d_{25} \ (V).$$

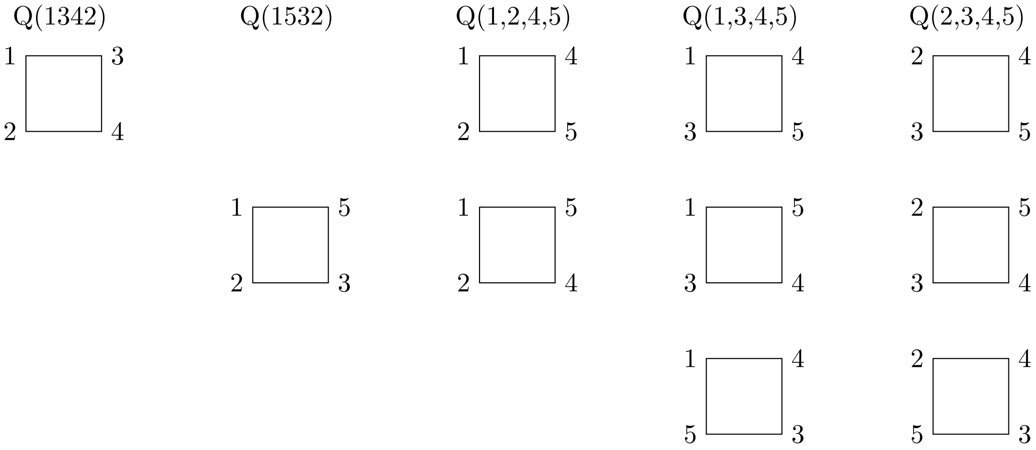


Figure 4: Allowable quadrangles for a 5-point metric space with a graph of type II-1.

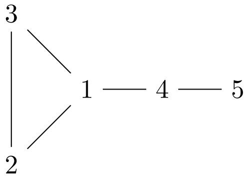
Using these inequalities, the allowable quadrangles are depicted in Figure 4 for this case. By Corollary 1, for this given graph one will have the following possible Gromov structures:

125, 214, 314, 423, 514.
125, 214, 314, 425, 514.
125, 214, 314, 435, 514.
125, 214, 325, 425, 513.
125, 214, 325, 425, 514.
125, 214, 325, 425, 534.
125, 214, 345, 423, 513.

(ii) (II)-2 : the typical instance of such a system is (columns (1,2,3,5,10) of A^t):

$$x_1 + x_2 = d_{12} \ , \ x_1 + x_3 = d_{13} \ , \ x_1 + x_4 = d_{14} \ , \ x_2 + x_3 = d_{23} \ , \ x_4 + x_5 = d_{45},$$
$$x_1 + x_5 > d_{15} \ , \ x_2 + x_4 > d_{24} \ , \ x_2 + x_5 > d_{25} \ , \ x_3 + x_4 > d_{34} \ , \ x_3 + x_5 > d_{35}.$$

For this choice of equalities, a graph like



is obtained. Solving the equations, and then replacing the x_i in the inequalities, one derives these formulas:

$$d_{12} + d_{34} < d_{14} + d_{23}\ (I)\ ,\ d_{14} + d_{25} < d_{12} + d_{45}\ (III),$$
$$d_{13} + d_{24} < d_{14} + d_{23}\ (II)\ ,\ d_{14} + d_{35} < d_{13} + d_{45}\ (IV),$$
$$\Delta_{145} < \Delta_{123}\ (V).$$

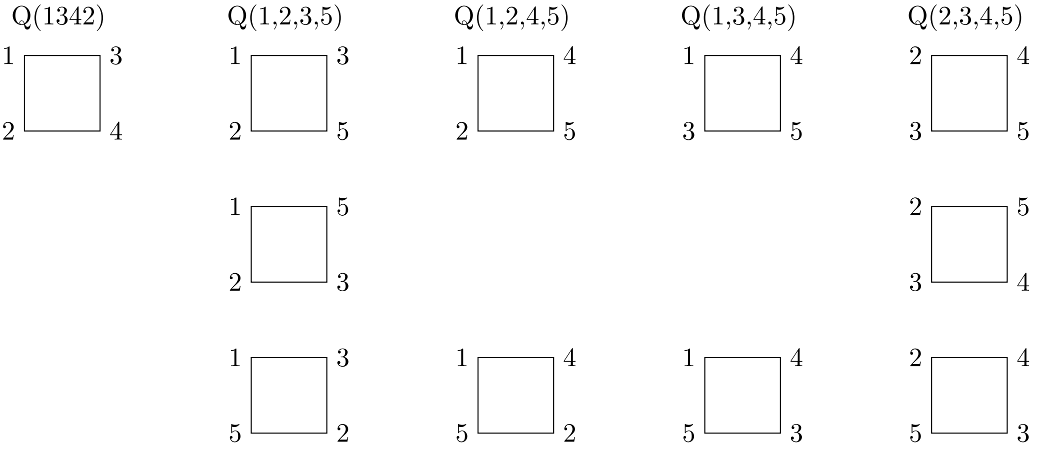


Figure 5: Allowable quadrangles for a 5-point metric space with a graph of type II-2.

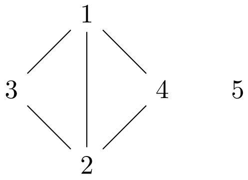
Using these inequalities, the allowable quadrangles are depicted in Figure 5 for this case. By Corollary 1, for this given graph one will have the following possible Gromov structures:

145, 245, 345, 423, 512.
145, 245, 345, 423, 513.
145, 245, 345, 423, 523.

(iii) (II)-3 : the typical instance of such a system is (columns (1,2,3,5,6) of A^t):

$$x_1 + x_2 = d_{12}\ ,\ x_1 + x_3 = d_{13}\ ,\ x_1 + x_4 = d_{14}\ ,\ x_2 + x_3 = d_{23}\ ,\ x_2 + x_4 = d_{24},$$
$$x_1 + x_5 > d_{15}\ ,\ x_2 + x_5 > d_{25}\ ,\ x_3 + x_4 > d_{34}\ ,\ x_3 + x_5 > d_{35}\ ,\ x_4 + x_5 > d_{45}.$$

For this choice of equalities, a graph like

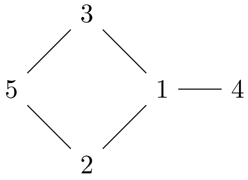


is obtained. (There is an even cycle in the graph.) In this case, the matrix of equations system is not invertible.

(iv) (II)-4 : the typical instance of such a system is (columns (1,2,3,7,9) of A^t):

$$x_1 + x_2 = d_{12}\ ,\ x_1 + x_3 = d_{13}\ ,\ x_1 + x_4 = d_{14}\ ,\ x_2 + x_5 = d_{25}\ ,\ x_3 + x_5 = d_{35},$$
$$x_1 + x_5 > d_{15}\ ,\ x_2 + x_3 > d_{23}\ ,\ x_2 + x_4 > d_{24}\ ,\ x_3 + x_4 > d_{34}\ ,\ x_4 + x_5 > d_{45}.$$

For this choice of equalities, a graph like

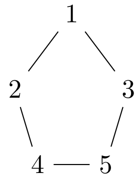


is obtained. (There is an even cycle in the graph.) In this case, the matrix of equations system is not invertible.

(c) system type (III): At most two x_i for each $i = 1, \dots, 5$ can appear in the equations, the typical instance of such a system is (columns (1,2,6,9,10) of A^t):

$$x_1 + x_2 = d_{12}\ ,\ x_1 + x_3 = d_{13}\ ,\ x_2 + x_4 = d_{24}\ ,\ x_3 + x_5 = d_{35}\ ,\ x_4 + x_5 = d_{45},$$
$$x_1 + x_4 > d_{14}\ ,\ x_1 + x_5 > d_{15}\ ,\ x_2 + x_3 > d_{23}\ ,\ x_2 + x_5 > d_{25}\ ,\ x_3 + x_4 > d_{34}.$$

For this choice of equalities, one has a graph like



Solving the equations, and then replacing the x_i in the inequalities, one derives these formulas:

$$d_{12} + d_{34} < d_{13} + d_{24}\ (I)\ ,\ d_{15} + d_{24} < d_{12} + d_{45}\ (III),$$
$$d_{13} + d_{25} < d_{12} + d_{35}\ (II)\ ,\ d_{14} + d_{35} < d_{13} + d_{45}\ (IV),$$
$$d_{23} + d_{45} < d_{24} + d_{35}\ (V).$$

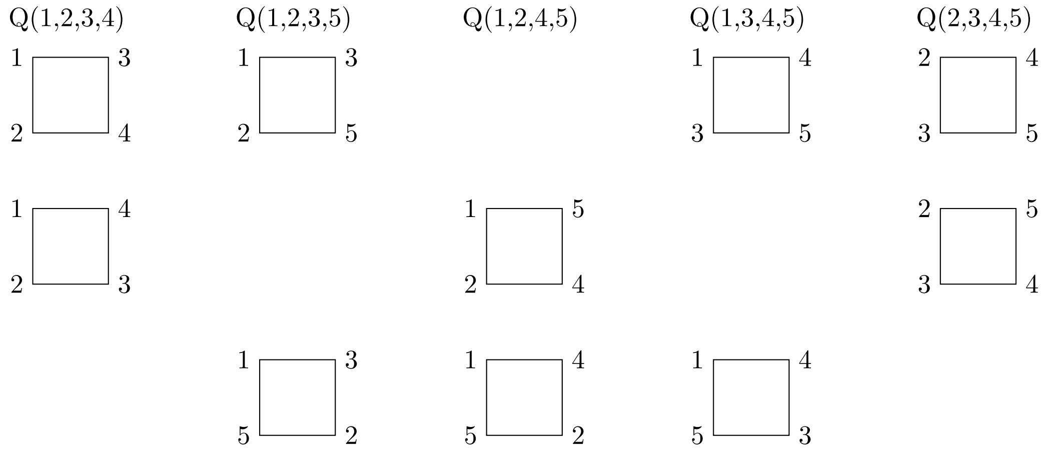


Figure 6: Allowable quadrangles for a 5-point metric space with a graph of type III.

Using these inequalities, allowable quadrangle structures are shown in Figure 6. By Corollary 1, for this given graph one will have the following possible Gromov structures:

123, 214, 315, 425, 534.

145, 235, 324, 413, 512.

5-point metric spaces and related graphs

Remember that every metric d defines a regular polyhedral subdivision Δ_d of $\Delta(n,2)$. For each subsystem of equations and inequalities, we let A_1 and A_2 denote the rows of A related with equalities and (remaining) inequalities respectively and D_1 and D_2 to denote the vector d_{ij} whose related variables x_i and x_j have appeared in system of equations and inequalities respectively. What we will consider here, is to look at compatible systems for three different metrics of a 5-point metric spaces. Numerical instances for three types have been given before in Table 1.

5-point metric type A ($S(\mathcal{E}_1)$)

Let A be the matrix whose rows are generators of $\Delta(5,2)$ as before and assume that d is the metric type $A1$ ($D = [10, 11, 10, 7, 8, 10, 10, 6, 10, 8]^t$). The only rows of matrix A which give compatible solutions with the metric d satisfying (2) are the following rows:

$$(1,2,3,4,7)\ ,\ (1,2,3,6,7)\ ,\ (1,2,5,6,7)\ ,\ (2,3,4,7,9)\ ,\ (2,3,6,7,9)\ ,\ (2,3,6,8,9)\ ,$$
$$(2,5,6,7,9)\ ,\ (2,5,6,8,9)\ ,\ (3,4,7,9,10)\ ,\ (3,6,7,9,10)\ ,\ (3,6,8,9,10)$$

Since a system type (III) (rows (2,3,6,7,9)) is among the systems, the complete structure is determined.

The cell families for these type and quadrangle structure is given below:

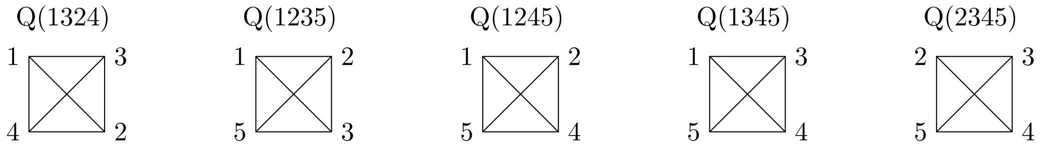


Figure 7: Quadrangles structure of type A 5-point metric space.

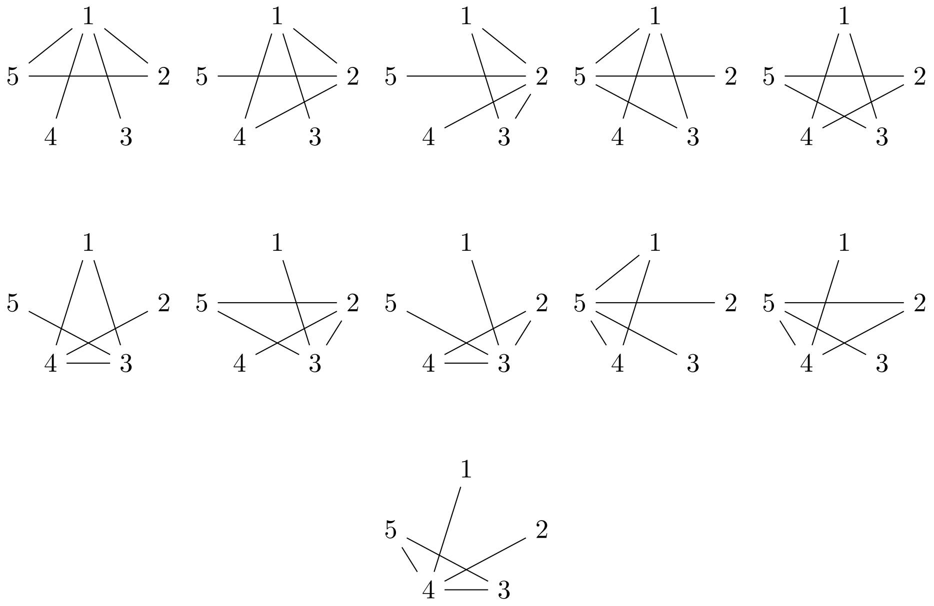


Figure 8: Cells of metric type A.

5-point metric type B ($S(\mathcal{E}_2)$)

Let A be the matrix whose rows are generators of $\Delta(5,2)$ as before and assume that d is the metric type $A2$ ($D = [5,6,6,5,3,5,6,5,5,4]^t$). The only rows of matrix A which give compatible solutions with the metric d satisfying (2) are the following rows:

$$(1,2,3,4,7)\ ,\ (1,2,3,6,7)\ ,\ (1,2,5,6,7)\ ,\ (2,3,4,7,9)\ ,\ (2,3,6,7,8)\ ,\ (2,3,7,8,9)\ ,$$
$$(2,5,6,7,8)\ ,\ (2,5,7,8,9)\ ,\ (3,4,7,9,10)\ ,\ (3,6,7,8,10)\ ,\ (3,7,8,9,10)$$

The cell families for these type and quadrangle structure is given below:

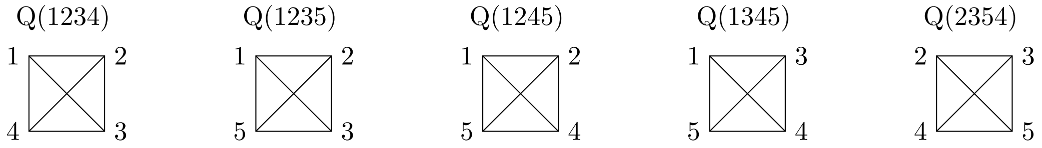


Figure 9: Quadrangles structure of type B 5-point metric space.

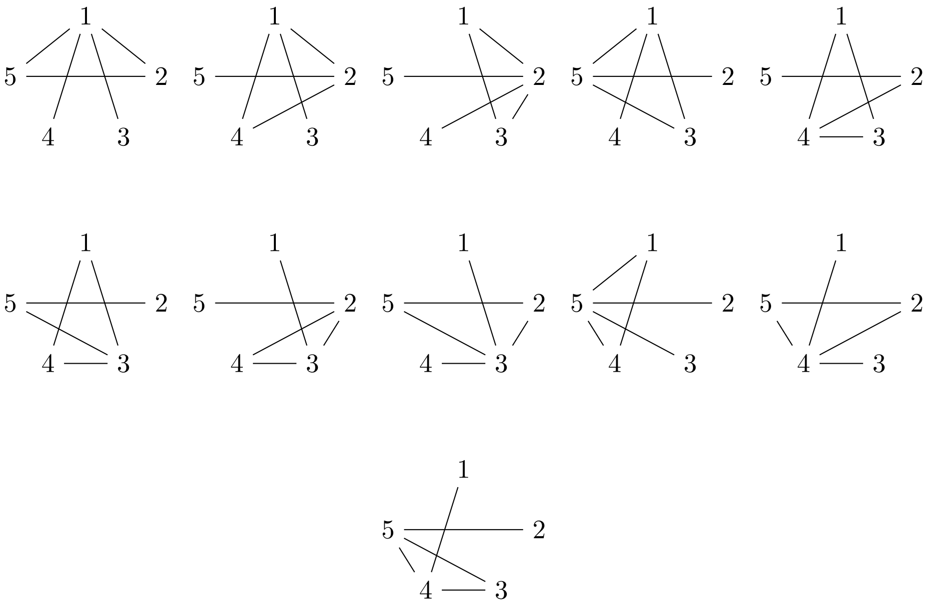


Figure 10: Cells of metric type B.

5-point metric type C ($S(\mathcal{E}_3)$)

Let A be the matrix whose rows are generators of $\Delta(5,2)$ as before and assume that d is the metric type C ($D = [6,10,6,8,8,6,10,9,6,8]^t$). The only rows of matrix A which give compatible solutions with the metric d satisfying (2) are the following rows:

$$(1,2,3,4,7)\ ,\ (1,2,3,6,7)\ ,\ (1,2,5,6,7)\ ,\ (2,3,4,7,10)\ ,\ (2,3,6,7,8)\ ,\ (2,3,7,8,10)\ ,$$
$$(2,4,7,9,10)\ ,\ (2,5,6,7,8)\ ,\ (2,5,7,8,9)\ ,\ (2,7,8,9,10)\ ,\ (3,6,7,8,10)$$

The cell families for these type and quadrangle structure is given below:

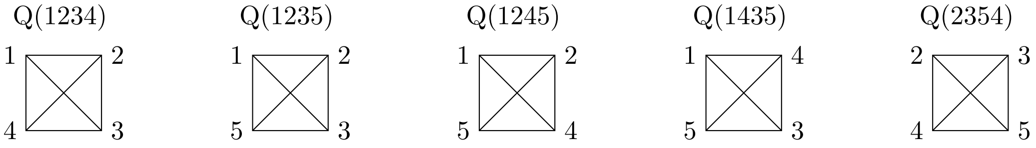


Figure 11: Quadrangles structure of type C 5-point metric space.

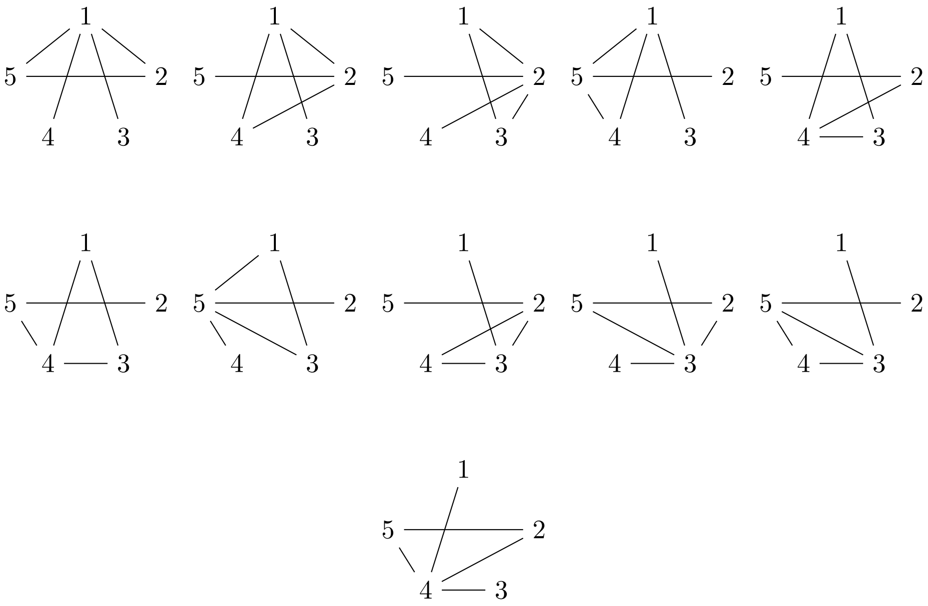
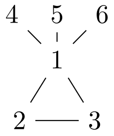


Figure 12: Cells of metric type C.

6-point Metric spaces and compatible cell types

For a 6-point metric space, the different cell types, which correspond to different choice of set of equations from the rows of generators for $\Delta(6,2)$, can be put into 9 different types. Each of these types come with an initial set of equation, but one can add up the equations to derive new inequalities from those initials. These different nine graph types and all the inequalities derived from them are as following:

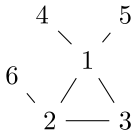
1. type I :



This type corresponds to the choice of rows (1,2,3,4,5,6) from the matrix of $\Delta(6,2)$. Inequalities are as following:

$$\begin{aligned}
&d_{12}+d_{13}+d_{56}<d_{15}+d_{16}+d_{23},\\
&d_{12}+d_{13}+d_{46}<d_{14}+d_{16}+d_{23},\\
&d_{12}+d_{13}+d_{45}<d_{14}+d_{15}+d_{23},\\
&d_{12}+d_{36}<d_{16}+d_{23},\quad d_{12}+d_{35}<d_{15}+d_{23},\quad d_{12}+d_{34}<d_{14}+d_{23},\\
&d_{13}+d_{26}<d_{16}+d_{23},\quad d_{13}+d_{25}<d_{15}+d_{23},\quad d_{13}+d_{24}<d_{14}+d_{23}.
\end{aligned}$$

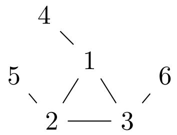
2. type II :



This type corresponds to the choice of rows (1,2,3,4,6,9) from the matrix of $\Delta(6,2)$. Inequalities are as following:

$$\begin{aligned}&d_{12}+d_{13}+d_{45}<d_{14}+d_{15}+d_{23},\\ &d_{12}+d_{56}<d_{15}+d_{26},\quad d_{12}+d_{35}<d_{15}+d_{23},\quad d_{12}+d_{46}<d_{14}+d_{26},\\ &d_{12}+d_{34}<d_{14}+d_{23},\quad d_{12}+d_{36}<d_{13}+d_{26},\quad d_{13}+d_{25}<d_{15}+d_{23},\\ &d_{13}+d_{24}<d_{14}+d_{23},\quad d_{16}+d_{25}<d_{15}+d_{26},\quad d_{16}+d_{24}<d_{14}+d_{26},\\ &d_{16}+d_{23}<d_{13}+d_{26}.\end{aligned}$$

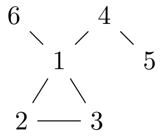
3. type III :



This type corresponds to the choice of rows (1,2,3,6,8,12) from the matrix of $\Delta(6,2)$. Inequalities are as following:

$$\begin{aligned}&d_{12}+d_{45}<d_{14}+d_{25},\quad d_{12}+d_{34}<d_{14}+d_{23},\quad d_{12}+d_{35}<d_{13}+d_{25},\\ &d_{13}+d_{46}<d_{14}+d_{36},\quad d_{13}+d_{24}<d_{14}+d_{23},\quad d_{23}+d_{56}<d_{25}+d_{36},\\ &d_{26}+d_{35}<d_{25}+d_{36},\quad d_{15}+d_{24}<d_{14}+d_{25},\quad d_{16}+d_{34}<d_{14}+d_{36},\\ &d_{15}+d_{23}<d_{13}+d_{25},\quad d_{13}+d_{26}<d_{12}+d_{36},\quad d_{16}+d_{23}<d_{12}+d_{36}.\end{aligned}$$

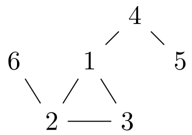
4. type IV :



This type corresponds to the choice of rows (1,2,3,5,6,13) from the matrix of $\Delta(6,2)$. Inequalities are as following:

$$\begin{aligned}&d_{12}+d_{13}+d_{46}<d_{14}+d_{16}+d_{23},\\ &d_{14}+d_{15}+d_{23}<d_{12}+d_{13}+d_{45},\\ &d_{12}+d_{36}<d_{16}+d_{23},\quad d_{12}+d_{34}<d_{14}+d_{23},\quad d_{13}+d_{26}<d_{16}+d_{23},\\ &d_{13}+d_{24}<d_{14}+d_{23},\quad d_{14}+d_{56}<d_{16}+d_{45},\quad d_{15}+d_{46}<d_{16}+d_{45},\\ &d_{24}+d_{35}<d_{23}+d_{45},\quad d_{25}+d_{34}<d_{23}+d_{45},\quad d_{14}+d_{35}<d_{13}+d_{45},\\ &d_{15}+d_{34}<d_{13}+d_{45},\quad d_{14}+d_{25}<d_{12}+d_{45},\quad d_{15}+d_{24}<d_{12}+d_{45}.\end{aligned}$$

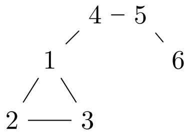
5. type V :



This type corresponds to the choice of rows (1,2,3,6,9,13) from the matrix of $\Delta(6,2)$. Inequalities are as following:

$$\begin{aligned}
&d_{14}+d_{23}+d_{56}<d_{13}+d_{26}+d_{45},\\
&d_{14}+d_{15}+d_{23}<d_{12}+d_{13}+d_{45},\\
&d_{12}+d_{46}<d_{14}+d_{26},\quad d_{12}+d_{34}<d_{14}+d_{23},\quad d_{12}+d_{36}<d_{13}+d_{26},\\
&d_{13}+d_{24}<d_{14}+d_{23},\quad d_{24}+d_{56}<d_{26}+d_{45},\quad d_{25}+d_{46}<d_{26}+d_{45},\\
&d_{24}+d_{35}<d_{23}+d_{45},\quad d_{25}+d_{34}<d_{23}+d_{45},\quad d_{16}+d_{24}<d_{14}+d_{26},\\
&d_{14}+d_{35}<d_{13}+d_{45},\quad d_{15}+d_{34}<d_{13}+d_{45},\quad d_{16}+d_{23}<d_{13}+d_{26},\\
&d_{14}+d_{25}<d_{12}+d_{45},\quad d_{15}+d_{24}<d_{12}+d_{45}.
\end{aligned}$$

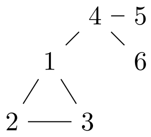
6. type VI :



This type corresponds to the choice of rows (1,2,3,6,13,15) from the matrix of $\Delta(6,2)$. Inequalities are as following:

$$\begin{aligned}
&d_{12}+d_{13}+d_{45}+d_{46}<2d_{14}+d_{23}+d_{56},\\
&d_{12}+d_{36}+d_{45}<d_{14}+d_{23}+d_{56},\\
&d_{13}+d_{26}+d_{45}<d_{14}+d_{23}+d_{56},\\
&d_{14}+d_{15}+d_{23}<d_{12}+d_{13}+d_{45},\\
&d_{12}+d_{34}<d_{14}+d_{23},\quad d_{13}+d_{24}<d_{14}+d_{23},\quad d_{24}+d_{35}<d_{23}+d_{45},\\
&d_{25}+d_{34}<d_{23}+d_{45},\quad d_{25}+d_{36}<d_{23}+d_{56},\quad d_{26}+d_{35}<d_{23}+d_{56},\\
&d_{15}+d_{46}<d_{14}+d_{56},\quad d_{16}+d_{45}<d_{14}+d_{56},\quad d_{25}+d_{34}<d_{23}+d_{45},\\
&d_{14}+d_{35}<d_{13}+d_{45},\quad d_{15}+d_{34}<d_{13}+d_{45},\quad d_{15}+d_{36}<d_{13}+d_{56},\\
&d_{16}+d_{35}<d_{13}+d_{56},\quad d_{14}+d_{25}<d_{12}+d_{45},\quad d_{15}+d_{24}<d_{12}+d_{45},\\
&d_{15}+d_{26}<d_{12}+d_{56},\quad d_{16}+d_{25}<d_{12}+d_{56}.
\end{aligned}$$

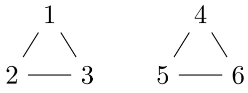
7. type VII :



This type corresponds to the choice of rows (1,2,3,6,13,14) from the matrix of $\Delta(6,2)$. Inequalities are as following:

$$\begin{aligned}
&2d_{14}+d_{23}+d_{56}<d_{12}+d_{13}+d_{45}+d_{46},\\
&d_{14}+d_{15}+d_{23}<d_{12}+d_{13}+d_{45},\\
&d_{14}+d_{16}+d_{23}<d_{12}+d_{13}+d_{46},\\
&d_{12}+d_{34}<d_{14}+d_{23},\quad d_{13}+d_{24}<d_{14}+d_{23},\quad d_{24}+d_{35}<d_{23}+d_{45},\\
&d_{24}+d_{36}<d_{23}+d_{46},\quad d_{25}+d_{34}<d_{23}+d_{45},\quad d_{26}+d_{34}<d_{23}+d_{46},\\
&d_{14}+d_{35}<d_{13}+d_{45},\quad d_{14}+d_{36}<d_{13}+d_{46},\quad d_{15}+d_{34}<d_{13}+d_{45},\\
&d_{16}+d_{34}<d_{13}+d_{46},\quad d_{14}+d_{25}<d_{12}+d_{45},\quad d_{14}+d_{26}<d_{12}+d_{46},\\
&d_{15}+d_{24}<d_{12}+d_{45},\quad d_{16}+d_{24}<d_{12}+d_{46}.
\end{aligned}$$

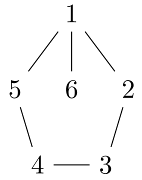
8. type VIII :



This type corresponds to the choice of rows (1,2,6,13,14,15) from the matrix of $\Delta(6,2)$. Inequalities are as following:

$$
\begin{aligned}
&d_{12} + 2d_{34} + d_{56} < d_{13} + d_{23} + d_{45} + d_{46},\\
&d_{12} + 2d_{35} + d_{46} < d_{13} + d_{23} + d_{45} + d_{46},\\
&d_{12} + 2d_{36} + d_{45} < d_{13} + d_{23} + d_{46} + d_{56},\\
&d_{13} + 2d_{24} + d_{56} < d_{12} + d_{23} + d_{45} + d_{46},\\
&d_{13} + 2d_{25} + d_{46} < d_{12} + d_{23} + d_{45} + d_{56},\\
&d_{13} + 2d_{26} + d_{45} < d_{12} + d_{23} + d_{46} + d_{56},\\
&2d_{14} + d_{23} + d_{56} < d_{12} + d_{13} + d_{45} + d_{46},\\
&2d_{15} + d_{23} + d_{46} < d_{12} + d_{13} + d_{45} + d_{56},\\
&2d_{16} + d_{23} + d_{45} < d_{12} + d_{13} + d_{46} + d_{56},\\
&d_{24} + d_{35} < d_{23} + d_{45}, \quad d_{24} + d_{36} < d_{23} + d_{46}, \quad d_{25} + d_{34} < d_{23} + d_{45},\\
&d_{25} + d_{36} < d_{23} + d_{56}, \quad d_{26} + d_{34} < d_{23} + d_{46}, \quad d_{26} + d_{35} < d_{23} + d_{56},\\
&d_{14} + d_{35} < d_{13} + d_{45}, \quad d_{14} + d_{36} < d_{13} + d_{46}, \quad d_{15} + d_{34} < d_{13} + d_{45},\\
&d_{15} + d_{36} < d_{13} + d_{56}, \quad d_{16} + d_{34} < d_{13} + d_{46}, \quad d_{16} + d_{35} < d_{13} + d_{56},\\
&d_{14} + d_{25} < d_{12} + d_{45}, \quad d_{14} + d_{26} < d_{12} + d_{46}, \quad d_{15} + d_{24} < d_{12} + d_{45},\\
&d_{15} + d_{26} < d_{12} + d_{56}, \quad d_{16} + d_{24} < d_{12} + d_{46}, \quad d_{16} + d_{25} < d_{12} + d_{56}.
\end{aligned}
$$

9. type IX :



This type corresponds to the choice of rows (1,4,5,6,10,13) from the matrix of $\Delta(6,2)$. Inequalities are as following:

$$
\begin{aligned}
&d_{12} + d_{34} + d_{56} < d_{16} + d_{23} + d_{45},\\
&d_{15} + d_{26} + d_{34} < d_{16} + d_{23} + d_{45},\\
&d_{12} + d_{36} < d_{16} + d_{23}, \quad d_{12} + d_{35} < d_{15} + d_{23}, \quad d_{13} + d_{46} < d_{16} + d_{34},\\
&d_{13} + d_{26} < d_{16} + d_{23}, \quad d_{13} + d_{45} < d_{15} + d_{34}, \quad d_{13} + d_{25} < d_{15} + d_{23},\\
&d_{14} + d_{56} < d_{16} + d_{45}, \quad d_{14} + d_{36} < d_{16} + d_{34}, \quad d_{14} + d_{35} < d_{15} + d_{34},\\
&d_{15} + d_{46} < d_{16} + d_{45}, \quad d_{24} + d_{35} < d_{23} + d_{45}, \quad d_{25} + d_{34} < d_{23} + d_{45},\\
&d_{13} + d_{24} < d_{12} + d_{34}, \quad d_{14} + d_{23} < d_{12} + d_{34}, \quad d_{14} + d_{25} < d_{12} + d_{45},\\
&d_{15} + d_{24} < d_{12} + d_{45}.
\end{aligned}
$$

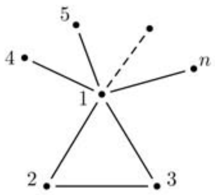
1 Derivation of Gromov product structure from cell types

Before discussing the derivation of Gromov product structure from the cells of a hypersimplex metric type, we need to examine the properties of a particular cell. This cell type consists of a triangle and several other edges individually connected to a single vertex (Figure **??**). Typical examples of such a cell under discussion can also be seen as the left most cell in Figures 1 and 2.

Proposition 3. *Let $(X = \{P_1, P_2, \ldots, P_n\}, d)$ be an n-point metric space. Then, (X, d) is Δ-generic with the Gromov product structure $P_i \mapsto E_{jk}, (i = 1, \ldots, n)$ if and only if the following sub-graphs $G_{i,jk}$ of (the complete graph) K_n belong to the cell-collection $Cell(d)$ of the metric d.*

Proof. For the ease of notation let $i = 1,\ j = 2,\ k = 3$. Then, the graph $G_{1,23}$ corresponds to the matrix $\mathcal{B}$ below:

$G_{1,23}$:



$$\longleftrightarrow \qquad \mathcal{B}=\begin{bmatrix}1&1&0&\cdots&0\\1&0&1&\cdots&0\\\vdots&\vdots&\vdots&\ddots&\vdots\\1&0&0&\cdots&1\\0&1&1&\cdots&0\end{bmatrix}$$

The system $\mathcal{B}x = d_{\mathcal{B}}$ and $\mathcal{B}'x > d_{\mathcal{B}'}$ reads as follows:

$$\begin{array}{rcl@{\qquad}rcl}
x_1+x_2 &=& d_{12} & x_2+x_4 &>& d_{24}\\
x_1+x_3 &=& d_{13} & &\vdots& \\
x_1+x_4 &=& d_{14} & x_2+x_n &>& d_{2n}\\
&\vdots& & &\vdots& \\
x_1+x_n &=& d_{1n} & x_r+x_s &>& d_{rs}\ \ (r\geq 2,\ s\geq 4,\ r<s)\\
x_2+x_3 &=& d_{23} & &\vdots& \\
& & & x_{n-1}+x_n &>& d_{n-1\ n}.
\end{array}$$

If the system has a solution in terms of x_i's, then Δ_{123} is uniquely minimal among Δ_{1rs} for $1 \neq r \neq s \neq 1$ and vice versa. To show this, first consider the first 2 equations and the last one from left column to see that $x_1 = \Delta_{123}$, $x_2 = \Delta_{213}$ and $x_3 = \Delta_{312}$. From other equations it follows that $x_k = d_{1k} - \Delta_{123}$ for $4 \leq k \leq n$. By applying algebraic manipulations, it can be seen that the inequalities $x_2 + x_k > d_{2k}$ $(4 \leq k \leq n)$ are equivalent to $\Delta_{123} < \Delta_{12k}$, the inequalities $x_3 + x_k > d_{3k}$ $(4 \leq k \leq n)$ are equivalent to $\Delta_{123} < \Delta_{13k}$ and finally the inequalities $x_r + x_s > d_{rs}$ $(4 \leq r, s$ and $r < s)$ are equivalent to $\Delta_{123} < \Delta_{1rs}$. Applying this to every node of X, the proposition follows. □

This proposition shows that Gromov product equivalence is weaker than the metric fan equivalence:

Proposition 4. Let d and d' be two Δ-generic metrics on an n-point set $X = \{P_1, P_2, \ldots, P_n\}$. If d and d' are equivalent in the metric fan sense, i.e. $Cell(d) = Cell(d')$, then they have the same Gromov product structure (and so, a fortiori, they are equivalent in the Gromov product sense).

Proof. If the Δ-generic metric d has the Gromov product structure $P_i \to E_{jk}$, for $i = 1, 2, \ldots, n$, then by Proposition 2 the sub-graphs $G_{i,jk}$ of (the complete graph) K_n belong to the cell-collection $Cell(d)$ of the metric d. Since by assumption $Cell(d) = Cell(d')$, the sub-graphs $G_{i,jk}$ belong also to $Cell(d')$. This means, again by Proposition 2, that the metric d' has the same Gromov product structure. □

2 Derivation of Quadrangle structures from cell types

The hypersimplex classification of finite metric spaces is to use triangulations of $\Delta(n, 2)$ to find out a 'compatible' set of cells. Assume that we have a class of cells which represents a hypersimplex metric type. Using this set of compatible cells, one can derive the quadrangle structure of this metric type. We will need the following proposition for this aim.

Proposition 5. For a Q-generic metric space, if $\{P_iP_j\}$ and $\{P_kP_l\}$ are realized as two edges of a cell of a class, then they are diagonals of $Q(P_i, P_j, P_k, P_l)$.

Proof. Consider four vertices of a cell P_i, P_j, P_k and P_l in the given class of hypersimplex metric type. We write the equation (1) in the following form:

$$\begin{aligned}
d_{ij} + d_{kl} &= x_i + x_j + x_k + x_l - \epsilon_{(ij)(kl)},\\
d_{ik} + d_{jl} &= x_i + x_k + x_j + x_l - \epsilon_{(ik)(jl)},\\
d_{il} + d_{jk} &= x_i + x_l + x_j + x_k - \epsilon_{(il)(jk)}.
\end{aligned}$$

Here for instance $\epsilon_{(ik)(jl)}$ is a non-negative number which is zero if and only if $\{ik\}$ and $\{jl\} \in G$ which is if E_{ik} and E_{jl} are two recognized edges of a cell. We are looking for $max\{d_{ij} + d_{kl}, d_{ik} + d_{jl}, d_{il} + d_{jk}\} = max\{x_i + x_j + x_k + x_l - \epsilon_{(ij)(kl)}, x_i + x_k + x_j + x_l - \epsilon_{(ik)(jl)}, x_i + x_l + x_j + x_k - \epsilon_{(il)(jk)}\}$. Since we have assumed that the metric space is Q-generic then only one of the three epsilons can be zero which will result in the only maximum (the diagonals). This means if E_{ik} and E_{jl} are recognized edges of a cell then $max\{d_{ij} + d_{kl}, d_{ik} + d_{jl}, d_{il} + d_{jk}\} = d_{ik} + d_{jl}$ and hence E_{ik} and E_{jl} are diagonals of $Q(i, j, k, l)$. □

Using Proposition 5, the quadrangle structure of a hypersimplex metric type can be simply determined as following: For quadrangle $Q(P_1, P_2, P_3, P_4)$, look at the cells in the metric type, if for instance $\{P_1P_2\}$ and $\{P_3P_4\}$ are realized as two edges of a cell (These edges can be realized in multiple cells of a same class) then they are diagonals and hence the quadrangle structure for $\{P_1, P_2, P_3, P_4\}$ is $Q(P_1P_3P_2P_4)$. The same can be done for other quadrangles till the diagonals of all quadrangles have been found.

Corollary 2. Two metric spaces with the same hypersimplex class have the same quadrangle structure, in other words quadrangle structure classification is coarser than hypersimplex classification.

References

[1] Bilge, A.H., Çelik, D. and Koçak, Ş., 2017. An equivalence class decomposition of finite metric spaces via Gromov products. Discrete Mathematics, 340(8), pp.1928-1932.

[2] Bilge, A.H., Çelik, D., Koçak, Ş. and Rezaeinazhad, A.M., 2021. Gromov product structures, quadrangle structures and split metric decompositions for finite metric spaces. Discrete Mathematics, 344(6), p.112358.

[3] De Loera, J.A., Rambau, J. and Santos, F., 2010. Triangulations, volume 25 of Algorithms and Computation in Mathematics.

[4] Ghys, É. and De La Harpe, P., 1990. Espaces métriques hyperboliques. In Sur les groupes hyperboliques d'après Mikhael Gromov (pp. 27-45). Birkhäuser, Boston, MA.

[5] Sturmfels, B. and Yu, J., 2004. Classification of six-point metrics. arXiv preprint math/0403147.

[6] Sturmfels, B., 1996. Grobner bases and convex polytopes (Vol. 8). American Mathematical Soc.

Five Point Metric Spaces: Gromov Product Structures, Quadrangle Structures and Explicit Parameterizations

Ayşe Hümeyra Bilge [*], Derya Çelik [†], Şahin Koçak [‡], Arash M. Rezaeinazhad [§]

Abstract

Let (X, d) be a finite metric space with elements $P_i, i = 1, \dots, n$ and with distances $d_{ij} := d(P_i, P_j)$ for $i, j = 1, \dots, n$. The "Gromov product" Δ_{ijk}, is defined as $\Delta_{ijk} = \frac{1}{2}(d_{ij} + d_{ik} - d_{jk})$. (X, d) is called Δ-generic, if for each fixed i, the set of Gromov products has a unique least element, $\Delta_{ij_ik_i}$. The Gromov product structure on a Δ-generic finite metric space (X, d) is the map that assigns the edge $E_{j_ik_i}$ to P_i. A finite metric space is called "quadrangle generic", if for all 4-point subsets $\{P_i, P_j, P_k, P_l\}$, the set $\{d_{ij}+d_{kl}, d_{ik}+d_{jl}, d_{il}+d_{jk}\}$ has a unique maximal element. We define the "quadrangle structure" on a quadrangle generic finite metric space (X, d) as the map that assigns to each 4-point subset of X, the pair of edges corresponding to the maximal element of the sums of the distances. Two metric spaces (X, d) and (X, d') are said to be Δ-equivalent (Q-equivalent), if the corresponding Gromov product (quadrangle) structures are the same, up to a permutation of X.

In this paper, Gromov product structures, quadrangle structures, optimal reductions and explicit parameterizations for 5-point spaces are obtained and compared with previous results in the literature. In the final part of this review paper, we have used the Monte Carlo method to obtain the relative volume of each of the 5-point metric types inside the corresponding metric cone for 5-point spaces, meanwhile 102 different partitions of metric cone for 5-point spaces are derived, considering Gromov product structures. These 102 partitions, come in three symmetric classes forming three types of metrics for 5-point spaces. Thus one can say that all the methods of classification given here or given before in the literature of finite metric spaces, give 3 types of metrics for 5-point spaces.

Keywords: Finite metric spaces, Split metric decompositions, Gromov products, Quadrangle structures

1 Introduction

The notions of Gromov product structures, Δ-equivalence, quadrangle structures and Q-equivalence have been defined in previous work [3]. Here, we present the applications of these notions to 5-point spaces. Basic definitions are quoted from [3].

Notation Let (X, d) be a finite metric space with n elements P_i, $i = 1, \dots, n$ $(n \geq 3)$ and let d_{ij} be the distance between P_i and P_j. The elements of X are also referred to as "vertices" or "nodes". E_{ij} and T_{ijk} denote respectively an edge and a triangle with corresponding vertices.

---

[*] Faculty of Engineering and Natural Sciences, Department of Industrial Engineering, Kadir Has University, Istanbul, Turkey.

[†] (Corresponding author) Eskişehir Technical University, Department of Mathematics, 26470, Eskişehir, Turkey. E-mail: deryacelik@eskisehir.edu.tr

[‡] Anadolu University, Department of Mathematics, 26470, Eskişehir, Turkey.

[§] Faculty of Management, Management Information Systems Department, Kadir Has University, Istanbul, Turkey.

Gromov products The quantity Δ_{ijk}, defined as

$$\Delta_{ijk} = \Delta_{ikj} = \tfrac{1}{2}(d_{ij} + d_{ik} - d_{jk}) \tag{1}$$

is called the Gromov product of the triangle T_{ijk} at the vertex P_i [6].

Δ-generic metrics A metric space is called Δ-generic, if for each P_i the set of Gromov products Δ_{ijk} has a unique smallest element.

Gromov product structures Let (X, d) be a Δ-generic finite metric space. Let $P_i \in X$, and let Δ_{ijk} be the minimal Gromov product at P_i, $(i = 1, \ldots, n)$. The function that assigns the edge E_{jk} to the vertex P_i is called the Gromov product structure on X. Two Δ-generic metric spaces (X, d) and (X, d') are Δ-equivalent, if the corresponding Gromov product structures are the same up to a permutation of X.

The metric cone The set C_n of all pseudo-metrics $d = (d_{ij}) \in \mathbb{R}^{\binom{n}{2}}$ on a given n-point set X, is called the metric cone.

The metric fan A decomposition of metric cone C_n into some sub-cones defined as below is called the metric fan [8]. Consider the $\binom{n}{2} \times n$ matrix $\mathcal{A}$ where the rows are numbered by the edges as

$$(1,2), (1,3), \ldots, (1,n), (2,3), (2,4), \ldots, (2,n), \ldots, (n-1,n)$$

and the (i,j)-row $(i < j)$ is given by $e_i + e_j = (0, \ldots, 1, \ldots, 1, \ldots, 0) \in \mathbb{R}^n$. Let $\mathcal{B}$ be an invertible $n \times n$ submatrix of $\mathcal{A}$ and denote the $\left[\binom{n}{2} - n\right] \times n$ matrix obtained by deleting $\mathcal{B}$ from $\mathcal{A}$ by $\mathcal{B}'$. Likewise, define $d_{\mathcal{B}} \in \mathbb{R}^n$ by choosing the components of $d \in \mathbb{R}^{\binom{n}{2}}$ corresponding to $\mathcal{B}$ and $d_{\mathcal{B}'} \in \mathbb{R}^{\binom{n}{2}-n}$ corresponding to $\mathcal{B}'$. Now consider the following system of equations and inequalities for $x \in \mathbb{R}^n$:

$$\mathcal{B}x = d_{\mathcal{B}} \quad \text{and} \quad \mathcal{B}'x > d_{\mathcal{B}'}.$$

If this system has a solution we say that the matrix $\mathcal{B}$ is a "cell" or a "thrackle" for the metric d. The collection of cells of a metric d is denoted by $Cell(d)$. Two metrics d and d' on an n-point metric space X are said to be equivalent in the metric-fan sense, if they have the same collection of cells or what amounts to the same collection of sub-graphs, i.e. $Cell(d) = Cell(d')$. The equivalence class of a metric d is a sub-cone of the metric cone and these sub-cones constitute altogether the metric fan.

Quadrangle generic metric spaces An n-point finite metric space X is called "quadrangle generic", or Q-generic, if for every 4-point subset $\{P_a, P_b, P_c, P_d\} \subseteq X$, the set of distances

$$\{d_{ab} + d_{cd}, d_{ac} + d_{bd}, d_{ad} + d_{bc}\}$$

has a unique maximal element.

Quadrangle Structures A quadrangle structure on a Q-generic finite metric space (X, d) is a map which assigns to any 4-point subset $\{P_a, P_b, P_c, P_d\}$ of X, the pair of edges corresponding to the maximal element of the set $\{d_{ab} + d_{cd}, d_{ac} + d_{bd}, d_{ad} + d_{bc}\}$. We denote the 4-point subset $\{P_a, P_b, P_c, P_d\}$ without any restriction on the sides by $Q(a, b, c, d)$ in which the ordering of the indices is irrelevant. If $d_{ac} + d_{bd}$ is maximal, the vertices are ordered as (P_a, P_b, P_c, P_d) and we denote this structured quadrangle by $Q(abcd)$ in which the cyclic permutations and reversal of the order of the indices give equivalent quadrangles.

Q-equivalence Two Q-generic metric spaces (X, d) and (X, d') are called Q-equivalent, if the corresponding quadrangle structures are the same up to a permutation of X.

Parameterization of 4-point spaces Let the set of minimal Gromov products of the quadrangle $Q(abcd)$ be $\{\Delta_{abd}, \Delta_{bac}, \Delta_{cbd}, \Delta_{dac}\}$ and let α and β be defined as

$$\alpha = \Delta_{abc} - \Delta_{abd}, \qquad \beta = \Delta_{adc} - \Delta_{adb}, \tag{2}$$

then, one has the following equalities between Gromov products

$$\begin{aligned} \alpha &= \Delta_{abc} - \Delta_{abd} = \Delta_{bad} - \Delta_{bac} = \Delta_{cda} - \Delta_{cdb} = \Delta_{dcb} - \Delta_{dca}, \\ \beta &= \Delta_{adc} - \Delta_{adb} = \Delta_{bcd} - \Delta_{bca} = \Delta_{cba} - \Delta_{cbd} = \Delta_{dab} - \Delta_{dac}, \end{aligned} \tag{3}$$

and the distances are expressed as

$$\begin{aligned} d_{ab} &= \Delta_{abd} + \Delta_{bac} + \alpha, & d_{cd} &= \Delta_{cbd} + \Delta_{dac} + \alpha, \\ d_{bc} &= \Delta_{bac} + \Delta_{cbd} + \beta, & d_{ad} &= \Delta_{abd} + \Delta_{dac} + \beta, \\ d_{ac} &= \Delta_{abd} + \Delta_{cbd} + \alpha + \beta, & d_{bd} &= \Delta_{bac} + \Delta_{dac} + \alpha + \beta. \end{aligned} \tag{4}$$

This is shown in Figure **??** below.

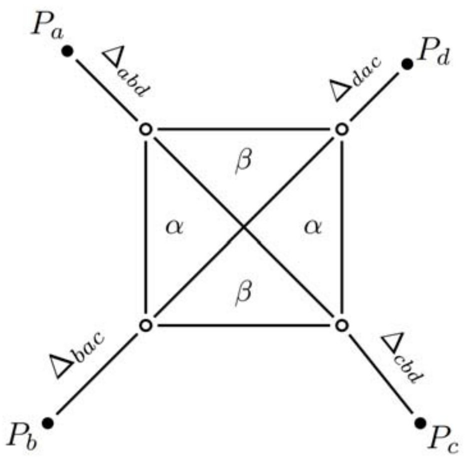


Figure 1.1: A quadrangle with the set of minimal Gromov products $\{\Delta_{abd}, \Delta_{bac}, \Delta_{cbd}, \Delta_{dac}\}$.

??

Matrix representation of Gromov product structures Gromov product structures on an n-point space are represented by the $n \times n$ matrix M_Δ defined by $M_\Delta(i, j) = 1$ and $M_\Delta(i, k) = 1$ if Δ_{ijk} is the minimal Gromov product at P_i and 0 otherwise [4].

Matrix representation of quadrangle structures The matrix of a quadrangle structure Q, M_Q on an n-point space is an $n_d \times n_d$ matrix ($n_d = n(n-1)/2$) such that $M_Q(ab, cd) = 1$ if the edges E_{ab} and E_{cd} are diagonals in $\{P_a, P_b, P_c, P_d\}$, and $M_Q(ab, cd) = 0$ otherwise.

Split pseudo-metrics A "split" $S = \{A, B\}$ of a finite set X is a partition of X into two non-empty subsets A and B. For simplicity we often identify the set of points of A with its index set. For each $P_a \in X$, we denote by $S(a)$ the subset A or B that contains P_a. Corresponding to each split S we define the pseudo-metric δ_S by

$$\delta_S(a, a') = \begin{cases} 1 & \text{if } S(a) \neq S(a'), \\ 0 & \text{if } S(a) = S(a'). \end{cases}$$

If the number of elements of A or B is equal to k, the split is referred to as a k-split.

Totally split decomposable metrics A metric on X is called totally split decomposable if it can be expressed as a linear combination (with non-negative coefficients) of the split metrics [1].

The isolation index of a split The isolation index of a split $S = \{A, B\}$ is defined as

$$\alpha_{A,B} = \frac{1}{2} min_{a,a' \in A, b,b' \in B}\{max\{d_{ab} + d_{a'b'}, d_{ab'} + d_{a'b}, d_{aa'} + d_{bb'}\} - (d_{aa'} + d_{bb'})\}.$$

Split prime A pseudo-metric is called a split prime if all of its isolation indices are equal to zero [1].

Lemma 1.1. *Let (X, d) be a finite metric space with n elements P_i $(i = 1, \ldots, n)$ and let $S = \{A, B\}$ be a split for X. Then,*

i. The isolation index for the 1-split with $A = \{P_a\}$ is the minimal Gromov product at P_a,

ii. If (X, d) is Q-generic, then the isolation index for the k-split with $A = \{P_{i_1}, \ldots, P_{i_k}\}$ is non-zero if and only if for no pair of indices $a, a' \in A$, $E_{aa'}$ is a diagonal of the quadrangles $Q(a, a', b, b')$ where $b, b' \in B$.

Proof. See [3]. $\square$

2 Parameterization of 5-Point Metric Spaces

In this section we will give an explicit parameterization of 5-point spaces using Gromov product structures, quadrangle structures and partial orders on Gromov products at each P_a. This parameterization coincides with the parameterization given in [7].

It is known that the Gromov product equivalence gives the known classification of 5-point Δ-generic metric spaces [2].

$$\begin{aligned} A: &\quad \{\Delta_{125}, \Delta_{213}, \Delta_{324}, \Delta_{435}, \Delta_{514}\}, \\ B: &\quad \{\Delta_{125}, \Delta_{213}, \Delta_{325}, \Delta_{425}, \Delta_{514}\}, \\ C: &\quad \{\Delta_{125}, \Delta_{213}, \Delta_{325}, \Delta_{425}, \Delta_{513}\}. \end{aligned}$$

Note that, if say Δ_{ijk} is minimal in the metric space X, then it is also minimal in every quadrangle $Q = \{P_i, P_j, P_k, P_l\}$. In a graphical presentation we indicate this by marking the corresponding angle by a filled arc as shown in Figure 2.2. For a 5-point metric space X, at least one of the Gromov products in any quadrangle belongs to the list of minimal Gromov products. It follows that for a 5-point space, the Gromov product structure determines the quadrangle structure. The determination of the parameters displayed in the quadrangles will be explained below.

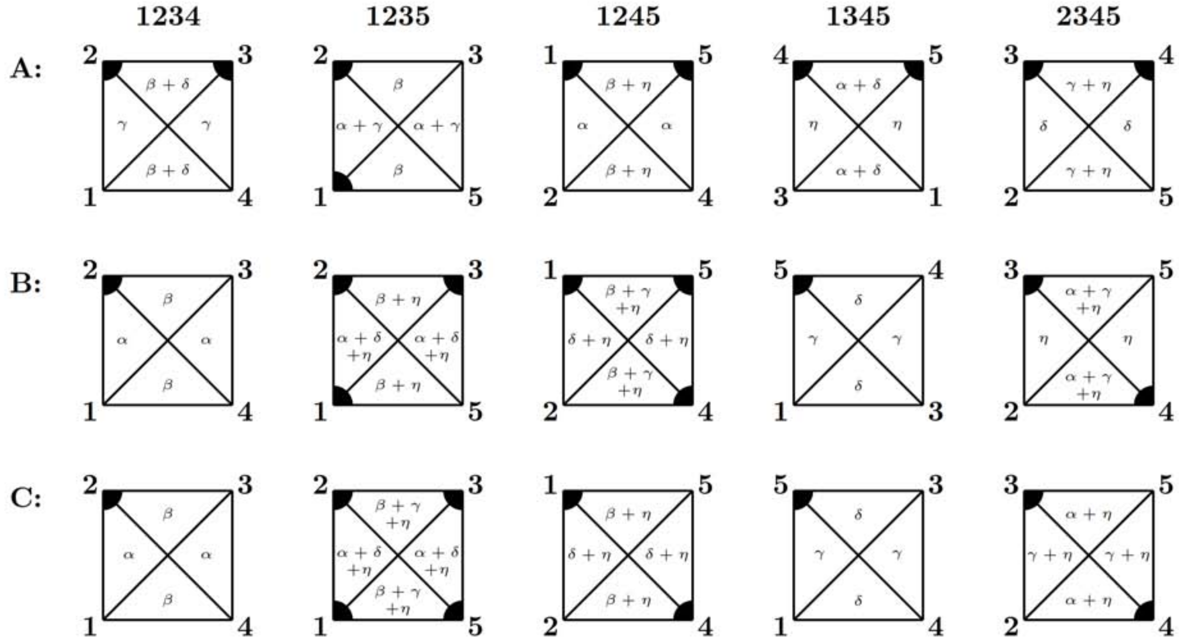


Figure 2.2: The structure of the 4-point subsets for the three types of 5-point metric spaces.

From Figure 2.2, we can see that, in Type A, the edges E_{12}, E_{23}, E_{34}, E_{45} and E_{15} are “sides” in all quadrangles, hence Type A metrics are totally split-decomposable by Lemma 1.1. For Type B, there are 4 edges E_{45}, E_{15}, E_{12} and E_{23} that occur as sides in all quadrangles. Therefore it is not totally split-decomposable. Similarly for Type C, the edges that occur as “sides” in all quadrangles are E_{12}, E_{23}, E_{34} and E_{45}, hence it is not totally split-decomposable.

In order to obtain an explicit parameterization of these metrics, we will use the quadrangle structure to obtain partial order relations among the Gromov products, then use the relations $d_{ij} = \Delta_{ijk} + \Delta_{jik}$. The structure of the quadrangles in Figure 2.2 lead to the following order relations for each of the types A, B, C in the following way: Take quadrangle $Q(1234)$ of Type A for instance. Since $d_{12} + d_{34} < d_{13} + d_{24}$, equivalently $\frac{1}{2}(d_{12}+d_{14}-d_{24}) < \frac{1}{2}(d_{13}+d_{14}-d_{34})$ which is to say $\Delta_{124} < \Delta_{134}$; we can also say that since $d_{14} + d_{23} < d_{13} + d_{24}$ is equivalent to $\frac{1}{2}(d_{12} + d_{14} - d_{24}) < \frac{1}{2}(d_{12} + d_{13} - d_{23})$ which means $\Delta_{124} < \Delta_{123}$. Thus for each vertex of a quadrangle, two inequalities among three Gromov products could be derived by similar algebraic manipulations. The list of these inequalities for each type is given below. These order relations are used to determine isolation indices for 2-splits and the split primes.

From quadrangles of Type A, we have the following relations among Gromov products:

$Q(1234) : \Delta_{124} < \Delta_{123}, \Delta_{134}, \quad \Delta_{213} < \Delta_{214}, \Delta_{234}, \quad \Delta_{324} < \Delta_{312}, \Delta_{314}, \quad \Delta_{413} < \Delta_{412}, \Delta_{423},$

$Q(1235) : \Delta_{125} < \Delta_{123}, \Delta_{135}, \quad \Delta_{213} < \Delta_{215}, \Delta_{235}, \quad \Delta_{325} < \Delta_{312}, \Delta_{315}, \quad \Delta_{513} < \Delta_{512}, \Delta_{523},$

$Q(1245) : \Delta_{125} < \Delta_{124}, \Delta_{145}, \quad \Delta_{214} < \Delta_{215}, \Delta_{245}, \quad \Delta_{425} < \Delta_{412}, \Delta_{415}, \quad \Delta_{514} < \Delta_{512}, \Delta_{524},$

$Q(1345) : \Delta_{135} < \Delta_{134}, \Delta_{145}, \quad \Delta_{314} \leq \Delta_{315}, \Delta_{345}, \quad \Delta_{435} < \Delta_{413}, \Delta_{415}, \quad \Delta_{514} < \Delta_{513}, \Delta_{534},$

$Q(2345) : \Delta_{235} < \Delta_{234}, \Delta_{245}, \quad \Delta_{324} < \Delta_{325}, \Delta_{345}, \quad \Delta_{435} < \Delta_{423}, \Delta_{425}, \quad \Delta_{524} < \Delta_{523}, \Delta_{534}.$

which lead to the following Hasse diagrams given in Figure 2.3.

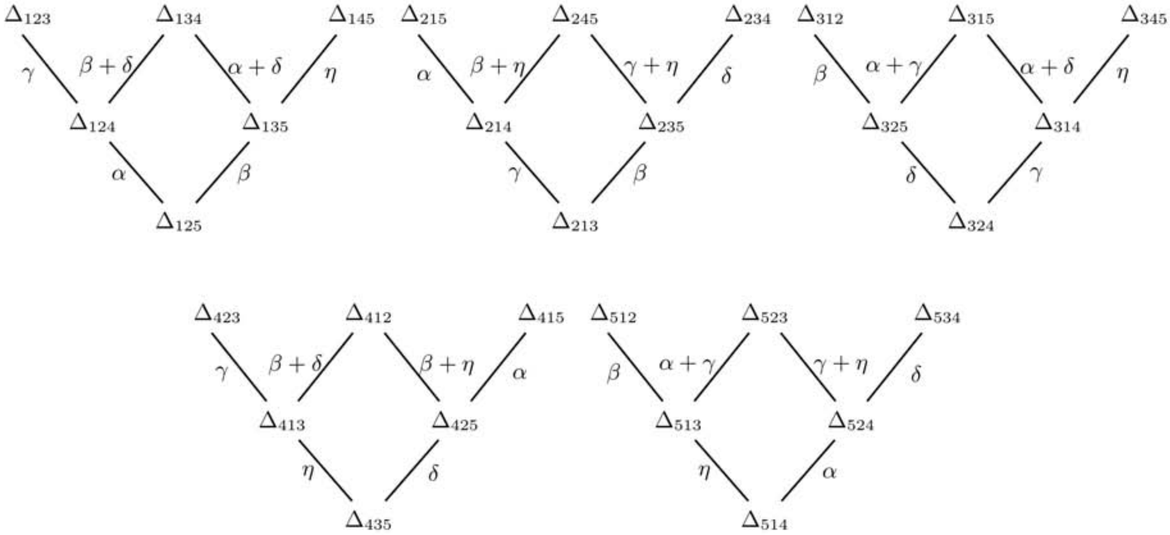


Figure 2.3: The partial order diagrams for the Type A.

For Type B, we have:

$Q(1234) : \Delta_{124} < \Delta_{123}, \Delta_{134}, \quad \Delta_{213} < \Delta_{214}, \Delta_{234}, \quad \Delta_{324} < \Delta_{312}, \Delta_{314}, \quad \Delta_{413} < \Delta_{412}, \Delta_{423},$

$Q(1235) : \Delta_{125} < \Delta_{123}, \Delta_{135}, \quad \Delta_{213} < \Delta_{215}, \Delta_{235}, \quad \Delta_{325} < \Delta_{312}, \Delta_{315}, \quad \Delta_{513} < \Delta_{512}, \Delta_{523},$

$Q(1245) : \Delta_{125} < \Delta_{124}, \Delta_{145}, \quad \Delta_{214} < \Delta_{215}, \Delta_{245}, \quad \Delta_{425} < \Delta_{412}, \Delta_{415}, \quad \Delta_{514} < \Delta_{512}, \Delta_{524},$

$Q(1345) : \Delta_{135} < \Delta_{134}, \Delta_{145}, \quad \Delta_{314} < \Delta_{315}, \Delta_{345}, \quad \Delta_{435} < \Delta_{413}, \Delta_{415}, \quad \Delta_{514} < \Delta_{513}, \Delta_{534},$

$Q(2354) : \Delta_{234} < \Delta_{235}, \Delta_{245}, \quad \Delta_{325} < \Delta_{324}, \Delta_{345}, \quad \Delta_{425} < \Delta_{423}, \Delta_{435}, \quad \Delta_{534} < \Delta_{523}, \Delta_{524}.$

which give the following Hasse diagrams given in Figure 2.4.

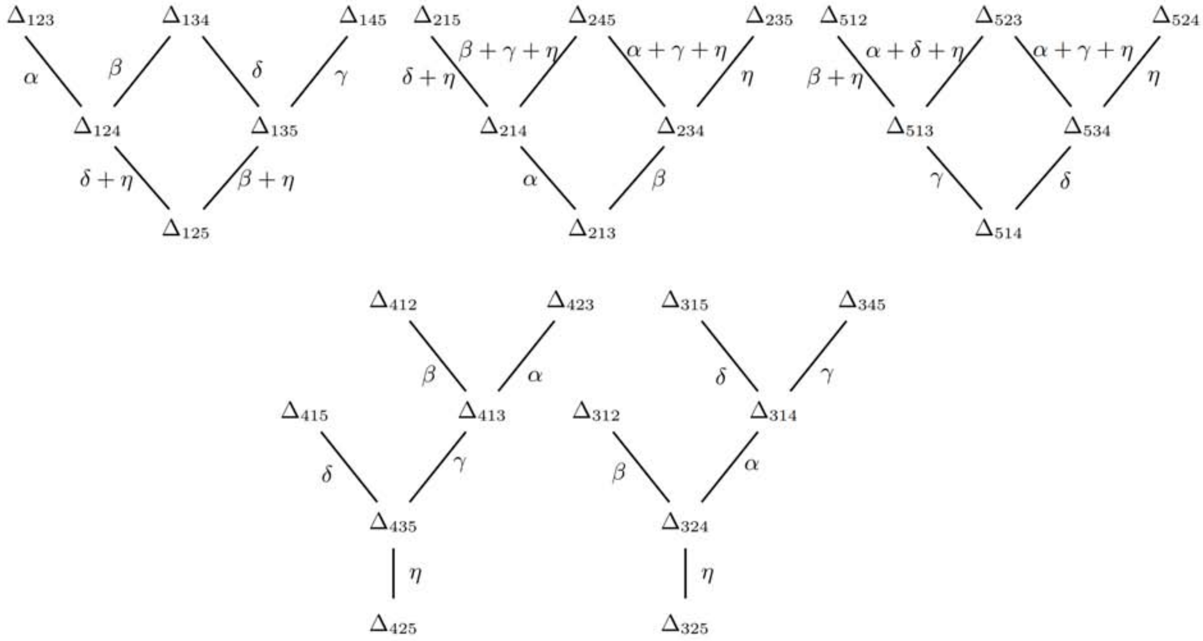


Figure 2.4: The partial order diagrams for the Type B.

For Type C, the quadrangles give the following relations:

$$\begin{aligned}
&Q(1234): \Delta_{124} < \Delta_{123}, \Delta_{134}, \quad \Delta_{213} < \Delta_{214}, \Delta_{234}, \quad \Delta_{324} < \Delta_{312}, \Delta_{314}, \quad \Delta_{413} < \Delta_{412}, \Delta_{423}, \\
&Q(1235): \Delta_{125} < \Delta_{123}, \Delta_{135}, \quad \Delta_{213} < \Delta_{215}, \Delta_{235}, \quad \Delta_{325} < \Delta_{312}, \Delta_{315}, \quad \Delta_{513} < \Delta_{512}, \Delta_{523}, \\
&Q(1245): \Delta_{125} < \Delta_{124}, \Delta_{145}, \quad \Delta_{214} < \Delta_{215}, \Delta_{245}, \quad \Delta_{425} < \Delta_{412}, \Delta_{415}, \quad \Delta_{514} < \Delta_{512}, \Delta_{524}, \\
&Q(1435): \Delta_{145} < \Delta_{134}, \Delta_{135}, \quad \Delta_{345} < \Delta_{314}, \Delta_{315}, \quad \Delta_{413} < \Delta_{415}, \Delta_{435}, \quad \Delta_{513} < \Delta_{514}, \Delta_{534}, \\
&Q(2354): \Delta_{234} < \Delta_{235}, \Delta_{245}, \quad \Delta_{325} < \Delta_{324}, \Delta_{345}, \quad \Delta_{425} < \Delta_{423}, \Delta_{435}, \quad \Delta_{534} < \Delta_{523}, \Delta_{524}.
\end{aligned}$$

which lead to the Hasse diagrams given in Figure 2.5.

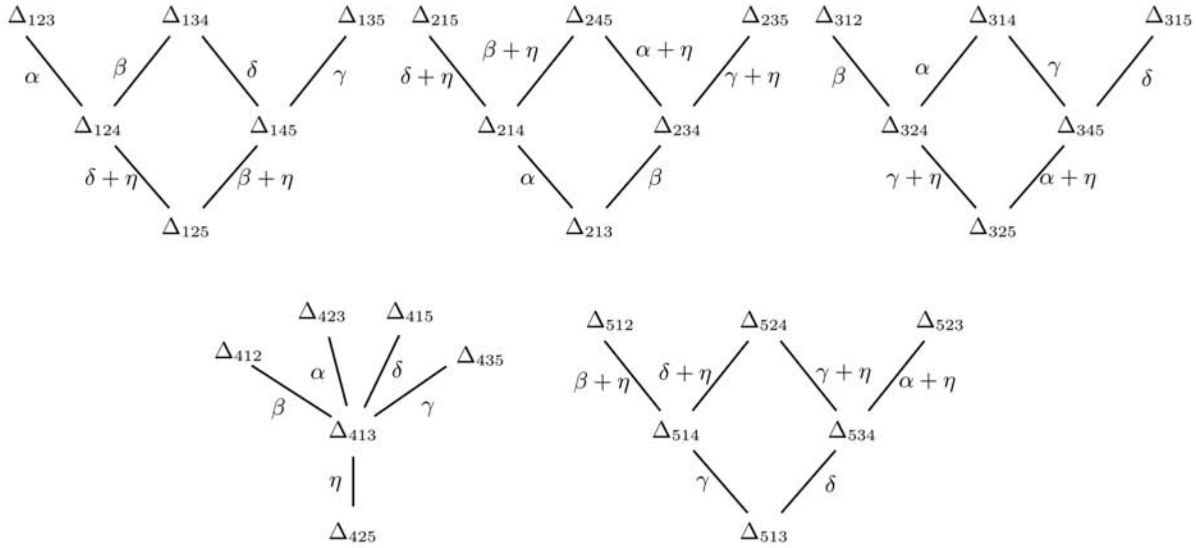


Figure 2.5: The partial order diagrams for the Type C.

Remark For Types A and B, the quadrangle structure determines the Gromov product structure, in the sense that the partial order relations deduced from the quadrangle structure determine the smallest Gromov product at each P_i. On the other hand, for Type C, the partial order relations imply that both Δ_{413} and Δ_{425} are smaller than $\Delta_{412}, \Delta_{415}, \Delta_{423}$, and Δ_{435}, but the order relation between Δ_{413} and Δ_{425} is not determined by the quadrangle structure. This is an example for the case where the quadrangle structure does not determine the Gromov product structure.

Recall that the minimal Gromov products at each P_a are the isolation indices of 1-splits. In what follows, we assume that minimal Gromov products are zero. The isolation indices for 2-splits will serve as free variables for the parameterization of the distances. For example, for Type A,

$$\begin{aligned}\alpha_{12} = \frac{1}{2} min\{ & max\{d_{13}+d_{24}, d_{14}+d_{23}, d_{12}+d_{34}\} - (d_{12}+d_{34}),\\ & max\{d_{13}+d_{25}, d_{15}+d_{23}, d_{12}+d_{35}\} - (d_{12}+d_{35}),\\ & max\{d_{14}+d_{25}, d_{15}+d_{24}, d_{12}+d_{45}\} - (d_{12}+d_{45})\}\\ = & \frac{1}{2} min\{d_{13}+d_{24}-d_{12}-d_{34}, d_{13}+d_{25}-d_{12}-d_{35}, d_{14}+d_{25}-d_{12}-d_{45}\}.\end{aligned}$$

Which reformulating by using Gromov products gives:

$$\begin{aligned}\alpha_{12} = min\{ & \Delta_{134} - \Delta_{124} = \Delta_{234} - \Delta_{213},\\ & \Delta_{135} - \Delta_{125} = \Delta_{235} - \Delta_{213},\\ & \Delta_{145} - \Delta_{125} = \Delta_{245} - \Delta_{214}\}.\end{aligned}$$

Finally since $\Delta_{125} = \Delta_{213} = 0$ we may write it as:

$$\alpha_{12} = min\{\Delta_{234}, \Delta_{135} = \Delta_{235}, \Delta_{145}\}.$$

From the partial order relations it is clear that α_{12} cannot be equal to Δ_{145}. Similarly, as $\Delta_{234} > \Delta_{235}$, we choose Δ_{135} as a free variable for the parameterization. By similar arguments and what is given when discussing "Parameterization of 4-point spaces" and Figure **??** in the introduction, the parameterization of the Gromov products and of the distance functions can be obtained as given below.

Type A: $\Delta_{125} = \Delta_{213} = \Delta_{324} = \Delta_{435} = \Delta_{514} = 0$.

$$\begin{array}{lllll}
\Delta_{124} = \alpha, & \Delta_{135} = \beta, & \Delta_{123} = \alpha + \gamma, & \Delta_{145} = \beta + \eta, & \Delta_{134} = \alpha + \beta + \delta, \\
\Delta_{214} = \gamma, & \Delta_{235} = \beta, & \Delta_{215} = \alpha + \gamma, & \Delta_{234} = \beta + \delta, & \Delta_{245} = \beta + \gamma + \eta, \\
\Delta_{314} = \gamma, & \Delta_{325} = \delta, & \Delta_{312} = \beta + \delta, & \Delta_{345} = \gamma + \eta, & \Delta_{315} = \alpha + \delta + \gamma, \\
\Delta_{413} = \eta, & \Delta_{425} = \delta, & \Delta_{415} = \alpha + \delta, & \Delta_{423} = \gamma + \eta, & \Delta_{412} = \beta + \delta + \eta, \\
\Delta_{513} = \eta, & \Delta_{524} = \alpha, & \Delta_{512} = \beta + \eta, & \Delta_{534} = \alpha + \delta, & \Delta_{523} = \alpha + \gamma + \eta,
\end{array}$$

$$\begin{array}{ll}
d_{12} = \alpha + \gamma, & d_{13} = \alpha + \beta + \delta + \gamma, \\
d_{23} = \beta + \delta, & d_{24} = \beta + \delta + \gamma + \eta, \\
d_{34} = \gamma + \eta, & d_{35} = \alpha + \delta + \gamma + \eta, \\
d_{45} = \alpha + \delta, & d_{14} = \alpha + \beta + \delta + \eta, \\
d_{15} = \beta + \eta, & d_{25} = \alpha + \beta + \gamma + \eta.
\end{array}$$

Type B: $\Delta_{125} = \Delta_{213} = \Delta_{325} = \Delta_{425} = \Delta_{514} = 0$.

$$\begin{array}{lllll}
\Delta_{124} = \delta + \eta, & \Delta_{135} = \beta + \eta, & \Delta_{134} = \beta + \delta + \eta, & \Delta_{123} = \alpha + \delta + \eta, & \Delta_{145} = \beta + \gamma + \eta, \\
\Delta_{214} = \alpha, & \Delta_{234} = \beta, & \Delta_{235} = \beta + \eta, & \Delta_{215} = \alpha + \delta + \eta, & \Delta_{245} = \alpha + \beta + \gamma + \eta, \\
\Delta_{324} = \eta, & \Delta_{312} = \beta + \eta, & \Delta_{314} = \alpha + \eta, & \Delta_{315} = \alpha + \delta + \eta, & \Delta_{345} = \alpha + \gamma + \eta, \\
\Delta_{435} = \eta, & \Delta_{413} = \gamma + \eta, & \Delta_{415} = \delta + \eta, & \Delta_{423} = \alpha + \gamma + \eta, & \Delta_{412} = \beta + \gamma + \eta, \\
\Delta_{534} = \delta. & \Delta_{513} = \gamma, & \Delta_{524} = \delta + \eta, & \Delta_{512} = \beta + \gamma + \eta, & \Delta_{523} = \alpha + \delta + \gamma + \eta,
\end{array}$$

$$\begin{array}{ll}
d_{12} = \alpha + \delta + \eta, & d_{13} = \alpha + \beta + \delta + 2\eta, \\
d_{14} = \beta + \delta + \gamma + 2\eta, & d_{15} = \beta + \gamma + \eta, \\
d_{23} = \beta + \eta, & d_{24} = \alpha + \beta + \gamma + \eta, \\
d_{25} = \alpha + \beta + \delta + \gamma + 2\eta, & d_{34} = \alpha + \gamma + 2\eta, \\
d_{35} = \alpha + \delta + \gamma + \eta, & d_{45} = \delta + \eta.
\end{array}$$

Type C: $\Delta_{125} = \Delta_{213} = \Delta_{325} = \Delta_{425} = \Delta_{513} = 0$.

$$\begin{array}{lllll}
\Delta_{124} = \delta + \eta, & \Delta_{145} = \beta + \eta, & \Delta_{123} = \alpha + \delta + \eta, & \Delta_{134} = \beta + \delta + \eta, & \Delta_{135} = \beta + \gamma + \eta, \\
\Delta_{214} = \alpha, & \Delta_{234} = \beta, & \Delta_{215} = \alpha + \delta + \eta, & \Delta_{235} = \beta + \gamma + \eta, & \Delta_{245} = \alpha + \beta + \eta, \\
\Delta_{324} = \gamma + \eta, & \Delta_{345} = \alpha + \eta, & \Delta_{312} = \beta + \gamma + \eta, & \Delta_{314} = \alpha + \gamma + \eta, & \Delta_{315} = \alpha + \delta + \eta, \\
\Delta_{413} = \eta, & \Delta_{412} = \beta + \eta, & \Delta_{415} = \delta + \eta, & \Delta_{423} = \alpha + \eta, & \Delta_{435} = \gamma + \eta, \\
\Delta_{514} = \gamma, & \Delta_{534} = \delta, & \Delta_{512} = \beta + \gamma + \eta, & \Delta_{523} = \alpha + \delta + \eta, & \Delta_{524} = \delta + \gamma + \eta.
\end{array}$$

$$\begin{array}{ll}
d_{12} = \alpha + \delta + \eta, & d_{13} = \alpha + \beta + \delta + \gamma + 2\eta, \\
d_{14} = \beta + \delta + 2\eta, & d_{15} = \beta + \gamma + \eta, \\
d_{23} = \beta + \gamma + \eta, & d_{24} = \alpha + \beta + \eta, \\
d_{25} = \alpha + \beta + \delta + \gamma + 2\eta, & d_{34} = \alpha + \gamma + 2\eta, \\
d_{35} = \alpha + \delta + \eta, & d_{45} = \delta + \gamma + \eta.
\end{array}$$

These parameterizations are exactly the ones given by Koolen, Lesser and Moulton [7]. In the paper

[7], the classes obtained via the decomposition of the metric cone are denoted as Type I, Type II and Type III. These correspond respectively to our equivalence classes denoted by Type A, Type C and Type B. The metrics of Type I, II and III are defined by their split decompositions, given as below. For simplicity we consider the pendant free case, i.e, we take the coefficients of the 1-splits as zero, equivalently the minimal Gromov products at each node are zero.

We use the labeling of the nodes by $\{x, y, u, v, w\}$.

(Type I): $d = \alpha_{xy}\delta_{xy} + \alpha_{yu}\delta_{yu} + \alpha_{uv}\delta_{uv} + \alpha_{vw}\delta_{vw} + \alpha_{wx}\delta_{wx}$,

(Type II): $d = \alpha_{xu}\delta_{xu} + \alpha_{xv}\delta_{xv} + \alpha_{uy}\delta_{uy} + \alpha_{vy}\delta_{vy} + cd'$,

(Type III): $d = \alpha_{xu}\delta_{xu} + \alpha_{xv}\delta_{xv} + \alpha_{wy}\delta_{wy} + \alpha_{vy}\delta_{vy} + cd'$,

where $d'(a,b) = 0$ if $a = b$, $d'(x,y) = d'(u,v) = d'(u,w) = d'(v,w) = 2$ and $d'(a,b) = 1$ for all other cases.

We identify the indices x, y, u, v, w with our notation. For example, for Type I, i.e, our Type A, x, y, u, v, w correspond to $1, 2, 3, 4, 5$ respectively and the correspondence of the parameters are

$$\alpha_{xy} = \beta, \quad \alpha_{yu} = \gamma, \quad \alpha_{uv} = \delta, \quad \alpha_{vw} = \eta, \quad \alpha_{wx} = \alpha.$$

For Type II, i.e, our Type C, x, y, u, v, w correspond to $5, 2, 1, 3, 4$ respectively and the correspondence of the parameters are

$$\alpha_{xu} = \delta, \quad \alpha_{xv} = \gamma, \quad \alpha_{uy} = \beta, \quad \alpha_{vy} = \alpha, \quad c = \eta.$$

For Type III, i.e, our Type B, x, y, u, v, w correspond to $2, 5, 3, 1, 4$ respectively and the correspondence of the parameters are

$$\alpha_{xu} = \alpha, \quad \alpha_{xv} = \beta, \quad \alpha_{wy} = \gamma, \quad \alpha_{vy} = \delta, \quad c = \eta.$$

Optimal Reductions of 5-Point Metric Spaces

Optimal realizations of 5-point metric spaces for three types are given in [7], in what follows we will give underlying graphs for each metric type and will drive their optimal reductions.

The weighted graph $G = (V, E, w)$ is called a realization of the finite metric space (X, d) if there is a labeling function $\phi\colon X \to V$ such that for all x, $y \in X$ the weight of any path between $\phi(x)$ and $\phi(y)$ is equal to $d(x, y)$. Any such realization is called optimal if $||G||$, the total edge weight of the graph G , is minimal among all realizations of the metric space (X, d) [7].

As it is clear from the definition above that a finite metric space can have many realizations. In the following, we will start with the pendant free reductions and use certain "moves" as defined in [5] to reduce the total weight and reach the optimal representation. This kind of operations are generally done by adjoining new vertices to the original graph, which in this case the added vertices are called secondary vertices and the original vertices as primary, discarding some edges or adding new edges between the enlarged set of vertices and assigning weights to the new edges in a way that the distance between primary nodes are unchanged but the weight of the graph, namely $||G||$, is reduced.

The first move, which is called *joining edges*, is done in the following way: Consider a vertex u and all (or some) of the other nodes $v_1, v_2, \ldots, v_k$ of G, which are neighbors of u. Calculate the Gromov products of all triangles $T_{uv_iv_j}$ with $1 \leq i, j \leq k$ at vertex u and call the minimum m_u. Now delete all the edges between u and v_i's, introduce a new vertex v and connect v_i's to v by edges of weight $w_{uv_i} - m_u$ for $1 \leq i \leq k$ and also u to v by an edge of weight m_u; hence the nodes v_i become connected to u by two edges through v and the total weight of the graph is reduced by an amount of $(k-1)m_u$.

The second move, which is called *edge removing*, is done by deleting the edge between two nodes u and v if it can be avoided by a shortest path. This move reduces $||G||$ by an amount of the weight of the deleted edge.

The "$\Delta - Y$" transform is a consequence of the above moves and can be applied to any triangle with 1-connected vertices in G. It is called a $\Delta - Y$ transform, because a triangle shape (Δ) turns to a Y shape after the operation.

We should also note that what we mean by *underlying graph of a metric*, is the complete graph with the same set of vertices as the metric space and all the edges with weight d_{ij} removed for which there is a point in space p_k such that $d_{ij} = d_{ik} + d_{kj}$.

For Type A with the Gromov product structure as $\{\Delta_{125}, \Delta_{213}, \Delta_{324}, \Delta_{435}, \Delta_{514}\}$, when edge removing operations are applied and passed to pendant-free reduction, a 5-cycle given as below is obtained. The optimal realization given in [7] is a 5-cycle with edges connected to each of its nodes (Type (a) of [7]).

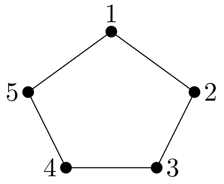


Figure 2.6: Optimal reduction of metric Type A

For Type B with the Gromov product structure as $\{\Delta_{125}, \Delta_{213}, \Delta_{325}, \Delta_{425}, \Delta_{514}\}$, the underlying graph is given below:

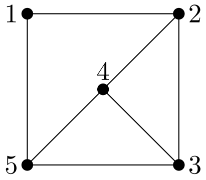


Figure 2.7: Underlying graph of metric Type B

By applying a $\Delta - Y$ transform to T_{345} we have the following:

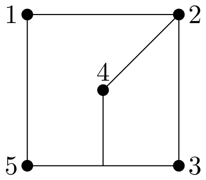


Figure 2.8: Graph with $\Delta - Y$ transformed.

In this step, one can follow two different approaches which reduce the metric to Type (b) or (c) of [7]. To observe the process closely we need to point out that the parameterization of Type B is given as follows:

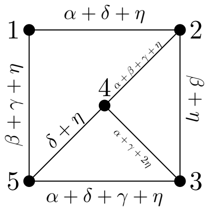


Figure 2.9: Underlying graph of metric Type B with distances parameterized

Here we have $\Delta_{345} = \alpha+\gamma+\eta$, $\Delta_{435} = \eta$ and $\Delta_{534} = \delta$, and applying a $\Delta - Y$ transform to T_{345} will result in the following:

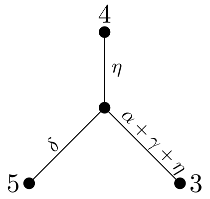


Figure 2.10: T_{345} of Type B after $\Delta - Y$ transform

So the Type B with parameters are as following:

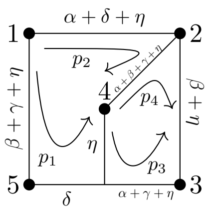


Figure 2.11: Metric Type B with the parameters

Now according to the graph above, we have d_{14} equal to $\beta+\delta+\gamma+2\eta$ (path p_1) or equal to $2\alpha+\beta+\gamma+\delta+2\eta$ (path p_2). Path p_2 is longer than path p_1 by an amount of 2α. Likewise d_{34} is equal to $\alpha+\gamma+2\eta$ (path p_3) or equal to $\alpha+2\beta+\gamma+2\eta$ (path p_4). Here path p_4 is longer than path p_3 by a difference of 2β. It should be noted that $\alpha = \Delta_{214}$ and $\beta = \Delta_{234}$ and two scenarios are possible: either $\alpha > \beta$ or $\beta > \alpha$. If $\alpha > \beta$, in order to decrease the total weight of the graph, we will introduce a new node called v on the edge joining 1 to 2 as shown below:

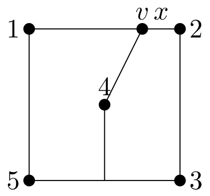


Figure 2.12: Reduction of Type B to (b)

This will reduce the total weight as $x = \Delta_{214}$ and that results the Type B to reduced into (b) of [7] and the metric will be as following:

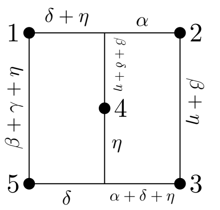


Figure 2.13: Reduction of Type B to (b) when $\alpha > \beta$ (parameters given)

In the other case, when $\beta > \alpha$, if we do the same operation as before, but this time for the edge joining 2 to 3 we will have the following reduction:

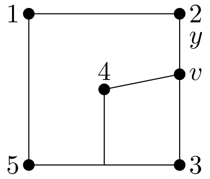


Figure 2.14: Reduction of Type B to (c)

This reduces the weight of graph as $y = \Delta_{234}$ and turns it into Type (c) given as below:

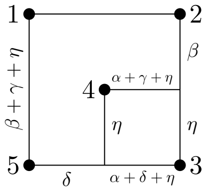


Figure 2.15: Reduction of Type B to (c) when $\beta > \alpha$ (parameters given)

For Type C which the underlying graph with the parameters given is depicted below, the following can be done:

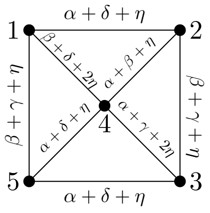


Figure 2.16: Underlying graph of metric Type C with the metric parameterized

Since $\Delta_{124} = \delta + \eta$, $\Delta_{214} = \alpha$ and $\Delta_{412} = \beta + \eta$, applying a $\Delta - Y$ transform to T_{124} will result in the following:

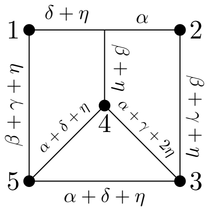


Figure 2.17: Type C with a $\Delta - Y$ transform applied to T_{124}

Finally considering that $\Delta_{345} = \alpha+\eta$, $\Delta_{435} = \gamma+\eta$ and $\Delta_{534} = \delta$, applying another $\Delta - Y$ transform to T_{345} will result in the following:

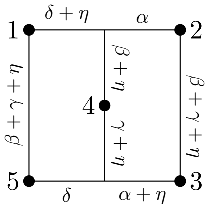


Figure 2.18: Type C with a second $\Delta - Y$ transform applied to T_{345} and reduced to (b)

Volumes of Gromov metric types

One of the ways to study the stability of Δ-equivalence classes under small numerical perturbations on components of metric represented as the vector $d = (d_{ij})$, is to consider the relative volume of each class inside the metric cone. To estimate these relative sizes of Δ-equivalence classes in an n-point space, we generate random points that lie in the intersection of the metric cone with unit ball in $\mathbb{R}^{n(n-1)/2}$ and then count the occurrence of points in each class.

We note that the volume of unit ball in $\mathbb{R}^N$ is equal to $V_N = \frac{\pi^{\frac{N}{2}}}{\Gamma(\frac{N}{2}+1)}$, where Γ is the Gamma function. It should also be noted that since the rate of growth of Gamma function is greater than the exponentials, as the dimension of space increases this volume decreases. It is known that the maximum volume is obtained for $N = 6$ and for the values of N greater than 6, V_N starts to decrease. On the other hand by keeping in mind that a metric d on an n-point space can be shown by a vector of positive coordinates in $\mathbb{R}^N$ where $N = n(n-1)/2$, we need to work with the intersection of unit ball with the orthant in which all the coordinates are positive (the first orthant in higher dimensions). Both of these issues leave us with only a few samples to work with.

To deal with the problem of generating a statistically significant number of points in the metric cone in $\mathbb{R}^{10}$ (since every metric on a 5-point space can be shown by a vector in $\mathbb{R}^{10}$) on a standard computer, we generate 10^7 random points $P = (x_1, x_2, \dots, x_{10})$, $0 < x_i < 1$ and accumulate these points from 10 such runs to get 10^8 points. Each of these points has 10 positive coordinates that are uniformly distributed random numbers in the range $(0, 1)$. Then the points that fall inside the unit ball are chosen and in the next step by checking which points satisfy the triangle inequalities, we select the points inside the metric cone. Finally for each of these points (metrics) we calculate the Gromov product structure in order to determine the metric type. This process is repeated 30 times and some of the results are given in Table 2.1 below. The Matlab code for this program is available at `http://finitemetricspaces.khas.edu.tr/Volume_of_Metric_Cone_n=5.m`.

points in unit ball	points in metric cone	Type A	Type B	Type C
274578	705	142	360	203
273136	735	186	351	198
273891	716	161	362	193
273426	733	170	376	187
272959	721	167	363	191

Table 2.1: Sample results of accumulating 10^8 points in $\mathbb{R}^{10}$. Each row is a single run of the program and shows how many points fall inside the unit ball, metric cone, and each type.

As shown in Table 2.1, from 10^8 points in the cube, around 2.7×10^5 points (0.275%) fall inside the unit ball and around 0.25% of these points fall inside the metric cone. To understand why these small amounts of points in unit cube of $\mathbb{R}^{10}$ fall inside the unit ball, it should be noted that the volume of unit ball V_{10} in $\mathbb{R}^{10}$ is equal to $\frac{\pi^5}{120}$ and we work only with the portion of unit ball intersecting the first orthant. This volume is approximately 0.00249 which is 0.24% of the volume of the unit cube.

In order to interpret the data given in Table 2.1, some clarifications must be made. 5-point metrics inside the metric cone in $\mathbb{R}^{10}$, when the Gromov product structure is considered, fall into 102 classes. Under permutation of the points of underlying metric space, these 102 classes form 3 families. In a family which is the orbit of the Gromov product structure $\{\Delta_{125}, \Delta_{213}, \Delta_{324}, \Delta_{435}, \Delta_{514}\}$ under the action of the permutation group S_5, there are 12 elements. The metrics that have a Gromov product structure in this family are called Type A metrics. Furthermore, the orbit of the Gromov product structure $\{\Delta_{125}, \Delta_{213}, \Delta_{325}, \Delta_{425}, \Delta_{514}\}$ and $\{\Delta_{125}, \Delta_{213}, \Delta_{325}, \Delta_{425}, \Delta_{513}\}$ have 60 and 30 elements respectively and the metrics of these families are called Type B and Type C in this order.

For calculating the type of a metric inside the metric cone to obtain the results given in Table 2.1, these 102 classes are taken into consideration. With this view in hand, the volume of Type A, Type B and Type C metrics on average are 22.07%, 51.02% and 26.26% of the metric cone (within a standard deviation of 21.1 for points inside the metric cone, 10.83 for Type A metrics, 17.03 for Type B metrics and 12.43 for Type C metrics in our runs to obtain the data given in Table 2.1). If we take the other view, without considering the permutations, results of Type A, B and C should be divided by 12, 60 and 30 respectively to obtain the volume of a single representative of each class. This means that within error bounds, the volumes of a single representative of Type A, B and C are respectively 1.84%, 0.85% and 0.87% of the metric cone.

The results above, give us the following intuitive conclusions: first that the volume of a single representative of Type B and Type C metrics are almost equal and Type A is "thicker" than these two types. Second, although a single representative of metric Type A is thicker than other types, these representatives are small in number (12 among 102 classes) with respect to Type B (60 among 102) and Type C (30 among 102) inside the metric cone.

Acknowledgment This work has been supported by the Scientific and Technological Research Council of Turkey (TUBITAK) under the project number 118F412 titled "Analysis of Finite Metric Spaces via Gromov Products and their Applications to Phylogenetics".

References

[1] Bandelt, H.J. and Dress, A.W., 1992. A canonical decomposition theory for metrics on a finite set. Advances in mathematics, 92(1), pp.47-105.

[2] Bilge, A.H., Çelik, D. and Koçak, Ş., 2017. An equivalence class decomposition of finite metric spaces via Gromov products. Discrete Mathematics, 340(8), pp.1928-1932.

[3] Bilge, A.H., Çelik, D., Koçak, Ş. and Rezaeinazhad, A.M., 2021. Gromov product structures, quadrangle structures and split metric decompositions for finite metric spaces. Discrete Mathematics, 344(6), p.112358.

[4] Bilge, A.H. and Incegul, M., 2020. Matrix Invariants of Finite Metric Spaces. arXiv preprint arXiv:2003.03335.

[5] Çelik, D., Bilge, A.H. and Koçak, Ş., 2015. Optimal embeddings of finite metric spaces into graphs. Anadolu Üniversitesi Bilim Ve Teknoloji Dergisi-B Teorik Bilimler, 3(2), pp.133-147.

[6] Ghys, É. and De La Harpe, P., 1990. Espaces métriques hyperboliques. In Sur les groupes hyperboliques d'après Mikhael Gromov (pp. 27-45). Birkhäuser, Boston, MA.

[7] Koolen, J., Lesser, A. and Moulton, V., 2009. Optimal realizations of generic five-point metrics. European Journal of Combinatorics, 30(5), pp.1164-1171.

[8] Sturmfels, B., 1996. Grobner bases and convex polytopes (Vol. 8). American Mathematical Soc.

Parametrization of 6-point Metric Space Type I_{17} using Quadrangle Structure Parameters

Ayşe Hümeyra Bilge[*], **Derya Çelik**[†], **Şahin Koçak**[‡],
Arash M. Rezaeinazhad[§]

Abstract

Let (X,d) be a finite metric space with elements $P_i, i=1,\ldots,n$ and with $d(P_i,P_j)=: d_{ij}$ for $i,j=1,\ldots,n$. The Gromov product is defined as $\Delta_{ijk} = 1/2(d_{ij}+d_{ik}-d_{jk})$ and (X,d) is called Δ-generic, if, for each fixed i, the set of Gromov products has a unique smallest element, $\Delta_{ij_ik_i}$. The Gromov product structure on a Δ-generic finite metric space (X,d) is the map that assigns the edge $\{P_{j_i},P_{k_i}\}$ to P_i. Two metric spaces (X,d) and (X,d') are said to be Δ-equivalent, if the corresponding Gromov product structures are the same up to a permutation of X. Matrix representation of the Gromov product structure allows to define various invariants of finite metric spaces. It can be easily seen that a 4-point metric space has, up to permutations, a unique Gromov product structure. Given a $\Delta-$generic four point metric space, then the set $\{d_{ij}+d_{kl}, d_{ik}+d_{jl}, d_{il}+d_{jk}\}$ has a unique largest element. A finite metric space is called "quadrangle generic", if the restrictions to all 4-point subsets with the restriction of the metric d are $\Delta-$generic 4-point metric spaces. A finite metric space is called "quadrangle generic", if for all 4-point subsets $\{P_i,P_j,P_k,P_l\}$, the set $\{d_{ij}+d_{kl}, d_{ik}+d_{jl}, d_{il}+d_{jk}\}$ has a unique largest element. We define the "quadrangle structure" on a quadrangle generic finite metric space (X,d) as the map that assigns to each 4-point subset of X the pair of edges corresponding to the maximal element of the sums of distances. Two metric spaces (X,d) and (X,d') are said to be Q-equivalent, if the corresponding quadrangle structures are the same up to a permutation of X. The quadrangle structure has a representation by a symmetrical matrix that can be used to define certain invariants. For $n=4$ and $n=5$, Δ-equivalence coincides with Q-equivalence. For $n\geq 6$, there are in general more than one quadrangle

---

[*] Faculty of Engineering and Natural Sciences, Department of Industrial Engineering, Kadir Has University, Istanbul, Turkey (ayse.bilge@khas.edu.tr)

[†] Anadolu University, Dept.of Mathematics, Eskişehir, Turkey (deryacelik@anadolu.edu.tr)

[‡] Anadolu University, Department of Mathematics, 26470 Eskişehir, Turkey (skocak@anadolu.edu.tr)

[§] Faculty of Management, Management Information Systems Department, Kadir Has University, Istanbul, Turkey, (arash.rezaeinazhad@khas.edu.tr)

structures compatible with a given Gromov product structure. Furthermore, a given quadrangle structure may be compatible with more than one Gromov product structure. The quadrangle structure compatible with a given Gromov product structure provides an algorithmic method for the determination of the partial order relations among the Gromov products at each P_i. These partial order relations in turn determine the for the parametrization of distance functions d_{ij}. We illustrate this procedure for each of the three types of 5 point spaces and for a specific Gromov product type of 6 point metric spaces.

Keywords: Finite metric spaces, Gromov product, Weighted graphs, Quadrangle structure.

1 Introduction

In a previous paper [2] we presented an equivalence class decomposition of finite metric spaces using the set of minimal Gromov products at each point of that space. In this section, we recall definitions and results concerning Gromov product structures.

Let (X, d) be a finite metric space with n elements P_i, $i = 1, \ldots, n$ $(n \geq 3)$ and let d_{ij} be the distance between P_i and P_j. Since a finite metric space can be considered as a weighted complete graph, we will also denote the elements of X as "vertices" or "nodes". The set $\{P_i, P_j\}$, $i \neq j$, denotes an "edge" of the metric space and the set $\{P_i, P_j, P_k\}$, $i \neq j \neq k \neq i$, denotes a "triangle" with vertices P_i, P_j and P_k. We denote the set of edges of (X, d) by $E(X)$. For simplicity we write P_iP_j for $\{P_i, P_j\}$ and $P_iP_jP_k$ for $\{P_i, P_j, P_k\}$. We will also sometimes shorten those as ij and ijk.

Definition 1 *The quantity Δ_{ijk} defined as*

$$\Delta_{ijk} = \tfrac{1}{2}(d_{ij} + d_{ik} - d_{jk}) \tag{1}$$

is called the Gromov product of the triangle $P_iP_jP_k$ at the vertex P_i [6]. We call a metric space Δ-generic, if for each P_i the set of Gromov products Δ_{ijk} has a unique, smallest element.

By the triangle inequality, the Gromov products Δ_{ijk} are non-negative numbers. It is easy to see that the distances d_{ij} can be expressed in terms of the Gromov products

$$d_{ij} = \Delta_{ijk_1} + \Delta_{jik_1} = \Delta_{ijk_2} + \Delta_{jik_2} = \cdots = \Delta_{ijk_{n-2}} + \Delta_{jik_{n-2}}, \tag{2}$$

where the indices k_l run from 1 to n, excluding i and j, leading to a total of $n-3$ equalities. In [2], we proved that a metric space can be defined using the Gromov products Δ_{ijk} as the primary ingredients. If X is an n-point metric space, the number of d_{ij}'s and Δ_{ijk}'s are respectively $n_d = n(n-1)/2$ and $n_\Delta = n(n-1)(n-2)/2$. The number of equality relations in Eq.(2) are $n_e = n_d(n-3) = n(n-1)(n-3)/2$, hence, $n_e + n_d = n_\Delta$. In this characterization, an n-point metric space is determined by n_Δ positive numbers subject to $n_\Delta - n_d$ equalities. The problem is thus converted from the search of positive solutions of a set of linear inequalities [7], to the search of positive solutions of a set of linear equalities [8]. The aim of the present work is to give an algorithmic procedure for the parametrization of the distance functions in terms of (unconstrained) positive parameters. We call a metric space Δ-generic, if for each P_i the set of Gromov products Δ_{ijk} has a unique, smallest element.

We recall the definition of the Gromov product structure for a Δ-generic finite metric space [2].

Definition 2 *Let (X, d) be a Δ-generic finite metric space. Let $P_i \in X$, and let Δ_{ijk} be the minimal Gromov product at P_i, $(i = 1, \dots, n)$. The function that assigns the edge P_jP_k to the vertex P_i is called the Gromov product structure on X. Two Δ-generic metric spaces (X, d) and (X, d') are Δ-equivalent, if the corresponding Gromov product structures are same up to a permutation of X. Two Δ-generic metric spaces (X, d) and $(X, \tilde{d})$ is Δ-equivalent, if the corresponding Gromov product structures are the same up to a permutation of X.*

In [2] and [5] we had given all allowable Gromov product structures on 6 and 7-point spaces, respectively, hence obtained a decomposition of these spaces into Δ-equivalence classes. The algorithms for these decompositions is based on the results below.

Proposition 1 *Let (X, d) be a finite metric space with n elements $P_i, i = 1, \dots, n$. Then the following equations hold*

$$\begin{aligned} \Delta_{ijl} - \Delta_{ijk} &= \Delta_{kjl} - \Delta_{kil} = \Delta_{lik} - \Delta_{ljk} = \Delta_{jik} - \Delta_{jil}, \\ \Delta_{ikl} - \Delta_{ijk} &= \Delta_{jkl} - \Delta_{jil} = \Delta_{lij} - \Delta_{ljk} = \Delta_{kij} - \Delta_{kil} \end{aligned} \tag{3}$$

where $i, j, k, l = 1, 2, \dots, n$.

Corollary 1 *Let (X, d) be a $\Delta-$generic finite metric space and let Δ_{ijk} be the minimal Gromov product at node P_i. Then,*

(a) Δ_{jkl} cannot be minimal at node P_j, where $l \neq j, k$
(b) Δ_{kjl} cannot be minimal at node P_k, where $l \neq j, k$
(c) Δ_{lij} and Δ_{lik} cannot be minimal at node P_l, where $l \neq i, j, k$

In order to obtain the decomposition of a finite metric space into Δ-equivalence classes, we start by the Cartesian product of the sets of Gromov products at each P_i, then we use Corollary 1 to eliminate the ones that are not allowable. This list may contain non-generic Gromov products that are eliminated by an algorithm presented in [3]. Finally, the full permutation group on n elements is applied to the list of allowable Gromov product to form group orbits and representatives from each Δ-equivalence class are selected. Computationally, the action of the permutation group is the most time-consuming part of the algorithm; in [5] the group action has been reduced to a manageable size by decomposing the set of allowable Gromov product structures into equivalence classes given by a matrix invariants [4].

In Section 2, we define quadrangle structure and discuss their relation to Gromov product structures. In Section 3, we review split metric decompositions and we present the expression of the of the isolation indices of 2-splits in terms of quadrangle structures. In Section 4, we reproduce the well known classification of 5-point spaces by using quadrangle structures. The Section 5, we consider the problem of the parametrization for 6-point spaces. We recall that there are 339 combinatorial types of 6-point spaces, as given in [13]. In [2], we have shown that 6-point metrics fall in 26 Δ-equivalence class and we indicated the corresponding combinatorial types. Here we study the Δ-equivalence class I_{17}, containing the combinatorial types $\{12, 13, 39, 65, 66\}$ given in [13]. We show that the Δ-equivalence class I_{17} splits in 4 quadrangle equivalence classes, denoted as SSS, SST, STT and TTT and we obtain their explicit parametrizations. Representatives of the combinatorial types 66, 65 and 39 fall in the Q-equivalence classes SSS, SST and STT respectively. The Q-equivalence class TTT contains representatives of the combinatorial types 12 and 13, hence, Q-equivalence is strictly coarser than the combinatorial equivalence of finite metric spaces.

2 Quadrangle Structures

In this section we define quadrangle structures on a finite metric space, as the collection of compatible embeddings of 4-point sets into 4 point metric spaces. In the generic case, for each 4-point set, a unique pair of edges is identified as diagonals and this information is coded into a symmetrical matrix. It is shown that those edges that never appear as diagonals correspond to 2-splits, as discussed in the next section.

Definition 3 *We call an n-point finite metric space X, "quadrangle generic", or Q-generic, if for every 4-point subset $\{P_i, P_j, P_k, P_l\} \subseteq X$, the set of distances*

$$\{d_{ij} + d_{kl}, d_{ik} + d_{jl}, d_{il} + d_{jk}\}$$

has a unique largest element.

We remark that for a $4-$point metric space the notions of $\Delta-$genericness and $Q-$genericness coincide. Let $X = \{P_a, P_b, P_c, P_d\}$ be a 4-point metric space. If X is $\Delta-$generic and if we assume without loss of generality that the minimal Gromov product at P_a is Δ_{abd}, then we have the relations

$$\Delta_{abc} - \Delta_{abd} = \frac{1}{2}(d_{ab} + d_{ac} - d_{bc}) - \frac{1}{2}(d_{ab} + d_{ad} - d_{bd}) = \frac{1}{2}(d_{ac} + d_{bd} - d_{ad} - d_{bc}) > 0$$

$$\Delta_{acd} - \Delta_{abd} = \frac{1}{2}(d_{ac} + d_{ad} - d_{cd}) - \frac{1}{2}(d_{ab} + d_{ad} - d_{bd}) = \frac{1}{2}(d_{ac} + d_{bd} - d_{ab} - d_{cd}) > 0,$$

which show that $d_{ac} + d_{bd}$ is the unique maximal element of the set $\{d_{ab} + d_{cd}, d_{ac} + d_{bd}, d_{ad} + d_{bc}\}$. Hence X is $Q-$generic. Conversely if $X = \{P_a, P_b, P_c, P_d\}$ is $Q-$generic with $d_{ac} + d_{bd}$ the maximal element of the set $\{d_{ab} + d_{cd}, d_{ac} + d_{bd}, d_{ad} + d_{bc}\}$, then X is $\Delta-$generic with Δ_{abd} the minimal Gromov product at P_a. Hence X is $\Delta-$generic.

We now define the notion of a quadrangle structure:

Definition 4 *A quadrangle structure on a Q-generic finite metric space (X, d) is a map which assigns to any 4-point subset $\{P_a, P_b, P_c, P_d\}$ of X the pair of edges corresponding to the maximal element of the set $\{d_{ab} + d_{cd}, d_{ac} + d_{bd}, d_{ad} + d_{bc}\}$.*

If, for example, $d_{ac} + d_{bd}$ is the maximal element of the set $\{d_{ab} + d_{cd}, d_{ac} + d_{bd}, d_{ad} + d_{bc}\}$, then the pair $\{P_aP_c, P_bP_d\}$ is assigned to the $4-$point subset $\{P_a, P_b, P_c, P_d\}$. We will say that the edges P_aP_c and P_bP_d are "diagonals" of the quadrangle $\{P_a, P_b, P_c, P_d\}$. The other edges P_aP_b, P_aP_d, P_bP_c, P_cP_d will be called the "sides" of the quadrangle $\{P_a, P_b, P_c, P_d\}$.

Definition 5 *Two Q-generic metric spaces (X, d) and (X, d') are Q-equivalent, if the corresponding quadrangle structures are same up to a permutation of X.*

For $n = 5$, the Gromov product structures determine the quadrangle structure completely, as seen in Figure 2. For $n = 6$, the Gromov product structure determines only part of the quadrangles, hence Q-equivalence refines Δ-equivalence. In a 6-point space there are 15 quadrangles. For the Δ-equivalence class type I_{17} contains 4 Q-equivalence classes. Figure 3 displays the structure of the 12 quadrangles that are common to all Q-equivalence types, while Figure 4 shows the structure of the remaining 3 triangles that are different for each type.

If a quadrangle structure refines a Gromov product structure we say that it is compatible with Δ. The examples above show that for $n \geq 6$ there may be more than one quadrangle structures compatible with a given Gromov product structure. Also, a quadrangle structure may be compatible with inequivalent Gromov product structures, as illustrated by the example below for $n = 8$.

Example 1 *For $n = 8$ consider the Gromov product structures*

$$\Delta^{(A)} = \{\Delta_{124}, \Delta_{213}, \Delta_{324}, \Delta_{413}, \Delta_{513}, \Delta_{613}, \Delta_{713}, \Delta_{813}\},$$

$$\Delta^{(B)} = \{\Delta_{157}, \Delta_{257}, \Delta_{357}, \Delta_{468}, \Delta_{568}, \Delta_{657}, \Delta_{768}, \Delta_{857}\}.$$

Using the constructions given in [4], it can be seen that the matrix representation of these Gromov product structures are similar but the change of basis is not a permutation matrix, hence $\Delta^{(A)}$ and $\Delta^{(B)}$ are inequivalent. The number of quadrangles in an 8-point space is 70. It can be seen that $\Delta^{(A)}$ and $\Delta^{(B)}$ determine respectively the structure of 23 and 26 quadrangles, with the only common one being $Q(1537)$.

In the following we will work with quadrangle structures compatible with a given Gromov product structure, as a tool for obtaining explicit parametrizations of distance functions. The following parametrization of 4-point spaces will constitute the building blocks for the parametrization of metric spaces. In each quadrangle $\{P_a, P_b, P_c, P_d\}$ with the set of minimal Gromov products $\{\Delta_{abd}, \Delta_{bac}, \Delta_{cbd}, \Delta_{dac}\}$, we define

$$\alpha = \Delta_{abc} - \Delta_{abd}, \qquad \beta = \Delta_{adc} - \Delta_{adb}. \tag{4}$$

One can see that the following equalities hold

$$\begin{aligned}\alpha &= \Delta_{abc} - \Delta_{abd} = \Delta_{bad} - \Delta_{bac} = \Delta_{cda} - \Delta_{cdb} = \Delta_{dcb} - \Delta_{dac},\\ \beta &= \Delta_{adc} - \Delta_{adb} = \Delta_{bcd} - \Delta_{bca} = \Delta_{cba} - \Delta_{cbd} = \Delta_{dab} - \Delta_{dac},\end{aligned} \tag{5}$$

and check that

$$\begin{aligned}d_{ab} &= \Delta_{abd} + \Delta_{bac} + \alpha,\\ d_{bc} &= \Delta_{bac} + \Delta_{cbd} + \beta,\\ d_{cd} &= \Delta_{cbd} + \Delta_{dac} + \alpha,\\ d_{ad} &= \Delta_{abd} + \Delta_{dac} + \beta,\\ d_{ac} &= \Delta_{abd} + \Delta_{cbd} + \alpha + \beta,\\ d_{bd} &= \Delta_{bac} + \Delta_{dac} + \alpha + \beta.\end{aligned} \tag{6}$$

This is shown in Figure 1 below.

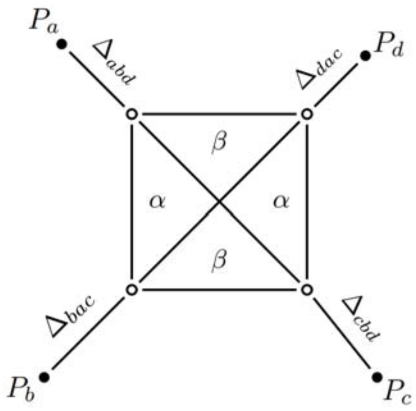


Figure 1: A quadrangle with the set of minimal Gromov products $\{\Delta_{abd}, \Delta_{bac}, \Delta_{cbd}, \Delta_{dac}\}$.

Gromov product structures on an n-point space have a convenient representation by an $n \times n$ matrix as defined in [4]. At present, the most useful feature of this matrix representation turned out to be the possibility of splitting the (huge) problem of finding structures that are equivalent under permutations to reasonable sizes, as used in [5] to obtain the Gromov product decomposition of 7-point spaces. We have a matrix representation for quadrangle structures as given below. We call a Gromov product structure "reducible", if there is a non-empty subset $X' \neq X$ of the metric space X such that its restriction to X' is a Gromov product structure on X'. Otherwise it is called "irreducible". In terms of matrices this corresponds to the reducibility of the associated matrix. Similarly, we define

Definition 6 *The matrix of a quadrangle structure Q, $M(Q)$ on an n-point space is an $n_d \times n_d$ matrix ($n_d = n(n-1)/2$) such that $M(Q)_{ab,cd} = 1$ if the edges P_aP_b and P_cP_d are diagonals in $\{P_a, P_b, P_c, P_d\}$, and $M(Q)_{ab,cd} = 0$ otherwise.*

The advantage of a matrix representation is the possibility of using matrix invariants to split the problem of finding structures equivalent under permutations into smaller subproblems, based on the invariants of these matrices. As a simple example, the nullity of $M(Q)$ is the number of edges that are never diagonals in any of the quadrangles they occur. This number is equal to the number of 2 splits in the prime metric decomposition, as shown below.

3 Split Metric Decompositions

"Split Metrics" are defined as follows. A "split" $S = \{A, B\}$ of a finite set X is a partition of X into two non-empty subsets A and B. For $x \in X$, we denote by $S(x)$ the subset(A or B) that contains x. Corresponding to each split S we define a metric δ_S by

$$\delta_S(x,y) = \begin{cases} 1 & \text{if } S(x) \neq S(y), \\ 0 & \text{if } S(x) = S(y). \end{cases} \tag{7}$$

The isolation index of the split $\{A, B\}$ is defined as

$$\alpha_{A,B} = \frac{1}{2} min_{a,a' \in A, b, b' \in B} \{ max\{d_{ab} + d_{a'b'}, d_{ab'} + d_{a'b}, d_{aa'} + d_{bb'}\} \tag{8}$$

$$- (d_{aa'} + d_{bb'})\}. \tag{9}$$

One can easily see that the isolation index for a 1-split is just the minimal Gromov product. If $A = \{a\}$ and $B = \{b_1, \dots, b_{n-1}\}$, then since $a = a'$, $d_{aa'} = 0$, the expression of the isolation index is

$$\alpha_{\{a\},B} = \frac{1}{2} min_{b,b' \in B} \left\{ max\{d_{ab} + d_{ab'}, d_{ab'} + d_{ab}, d_{bb'}\} - d_{bb'} \right\}.$$

Since the first two elements in the braces are equal,

$$\alpha_{\{a\},B} = \frac{1}{2} min_{b,b' \in B} \left\{ max\{d_{ab} + d_{ab'}, d_{bb'}\} - d_{bb'} \right\}.$$

By triangle inequality, $d_{ab} + d_{ab'} \geq d_{bb'}$, thus

$$\alpha_{\{a\},B} = \frac{1}{2} min_{b,b' \in B} \left\{ d_{ab} + d_{ab'} - d_{bb'} \right\}. \tag{10}$$

This means that the isolation index for the case $A = \{a\}$ is the minimal Gromov product at a.

Let's now compute the isolation index for the case where A is a 2-point set. If a, a' is a diagonal of the quadrangle $\{a, a', b, b'\}$, then

$$max\{d_{ab} + d_{a'b'}, d_{ab'} + d_{a'b}, d_{aa'} + d_{bb'}\} = d_{aa'} + d_{bb'}$$

and the isolation index of the split $\{\{a, a'\}, B\}$ is

$$\alpha_{\{a,a'\},B} = \frac{1}{2} min_{b,b' \in B} \left\{ d_{aa'} + d(b, b') - (d_{aa'} + d_{bb'}) \right\} = 0.$$

Therefore, if the points a, a' are on the diagonals of at least one quadrangle, the isolation index is zero. Thus, the isolation indices of the form $\alpha_{(\{a,a'\},B)}$ are nonzero only for the pairs a, a' that appear as sides in every quadrangle they belong to. In this case, without loss of generality, we may assume that

$$max\{d_{ab} + d_{a'b'}, d_{ab'} + d_{a'b}, d_{aa'} + d_{bb'}\} = d_{ab'} + d_{a'b}$$

and the isolation index is

$$\alpha_{\{a,a'\},B} = \frac{1}{2} min_{b,b' \in B} \left\{ d_{ab'} + d_{a'b} - (d_{aa'} + d_{bb'}) \right\}.$$

Note that

$$\begin{aligned} d_{ab'} + d_{a'b} - (d_{aa'} + d_{bb'}) &= [d_{ab'} + d_{ab} - d_{bb'}] - [d_{aa'} + (d_{ab} - d_{a'b}] \\ &= 2[\Delta_{abb'} - \Delta_{aa'b}]. \end{aligned}$$

It follows that

$$\alpha_{\{a,a'\},B} = min_{b,b' \in B}\{\Delta_{abb'} - \Delta_{aa'b}\}. \tag{11}$$

We can summarize these results below.

Proposition 2 *Let (X, d) be a Δ-generic finite metric space with n elements $P_i, i = 1, \ldots, n$, let Δ be its Gromov product structure, $Q(\Delta)$ be a compatible quadrangle structure and let $S = \{A, B\}$ be a split decomposition for X. Then,*

i. *The isolation index for the 1-split with $A = \{P_a\}$ is the minimal Gromov product at P_a,*

ii. *The isolation index for the 2-split with $A = \{P_a, P_{a'}\}$ is non zero if and only if the edge $P_a P_{a'}$ occurs as a "side" in each quadrangle $\{P_a, P_{a'}, P_b, P_{b'}\}$ of the quadrangle structure $Q(\Delta)$ and the isolation index is given by*

$$\alpha_{\{a,a'\},B} = min_{b,b' \in B}\{\Delta_{abb'} - \Delta_{aa'b}\}$$

.

4 Parametrization of 6-Point Metric Spaces: Examples

The classification of 6-point metric spaces with respect to the metric fan decomposition is given in [13] where it is shown that there are 339 combinatorial types. In previous work we presented the Gromov product decomposition of 6-point metric spaces and we gave a list of the corresponding types in [2]. Here we will give explicit parametrization of the Δ-equivalence class I_{17}, characterized by the minimal Gromov products below, that are assumed to be zero.

$$\Delta_{126} = \Delta_{213} = \Delta_{324} = \Delta_{435} = \Delta_{546} = \Delta_{615} = 0. \tag{12}$$

It follows that for all cases, we have the equalities

$$d_{13} = d_{12} + d_{23}, \quad d_{24} = d_{23} + d_{34}, \quad d_{35} = d_{34} + d_{45},$$

$$d_{46} = d_{45} + d_{56}, \quad d_{15} = d_{56} + d_{16}, \quad d_{26} = d_{16} + d_{12}.$$

For $n = 6$ there are 15 quadrangles. For the class I_{17} the structures of 12 of these are determined by the minimal Gromov products, as shown below in Figure 2.

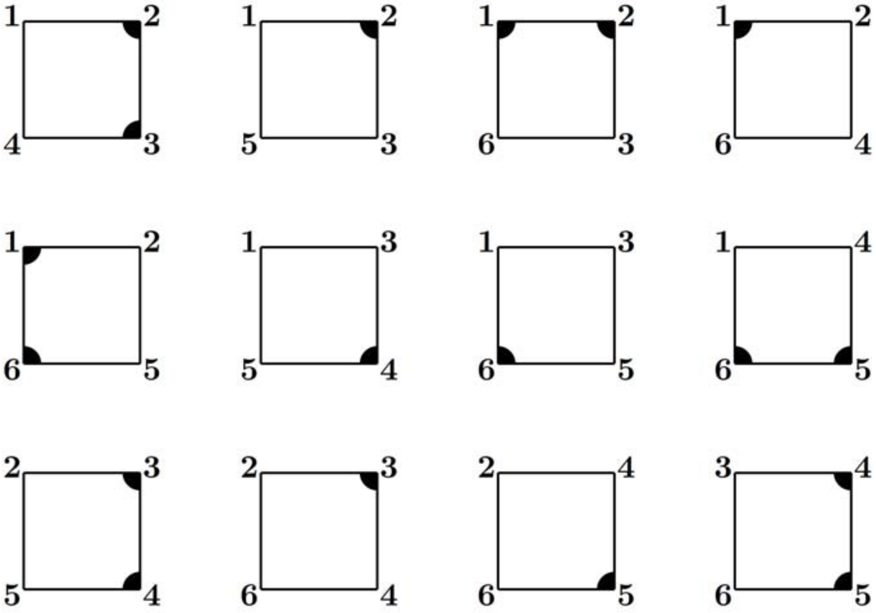


Figure 2: The structure of the 4-point subsets determined by the Gromov product structure, for the 6-point space of type I_{17}.

From this figure, on can see that the edges that occur as sides in all quadrangles are

$$P_1P_2,\ P_2P_3,\ P_3P_4,\ P_4P_5,\ P_5P_6,\ P_1P_6.$$

Therefore, there will be at least 6 2-splits in the decomposition. We then consider possible structures for the remaining 3 quadrangles. We use the notation $Q(a, b, c, d)$ to denote a quadrangle with these vertices. On the other hand $Q(abcd)$ denotes the quadrangle with vertices ordered as indicated. With this notation, the quadrangles whose structures are not determined by the Gromov product structure are $Q(1, 2, 4, 5)$, $Q(2, 3, 5, 6)$ and $Q(1, 3, 4, 6)$. As an example consider the quadrangle $Q(1, 2, 4, 5)$. The structure of this quadrangle depends on the choice of a minimal Gromov product at any of its vertices. At P_1, these are Δ_{124}, Δ_{125} and Δ_{145}. We first show that Δ_{145}

cannot be minimal among these three. From Figure 2, one can see that in quadrangle $Q(1,2,3,5)$, $\Delta_{135} > \Delta_{125}$ and in quadrangle $Q(1,3,4,5)$, $\Delta_{145} > \Delta_{135}$, it follows that $\Delta_{145} > \Delta_{125}$, hence it cannot be minimal. This corresponds to the fact that the edge P_1P_2 cannot be diagonal. By similar considerations, one can check that the edges listed above cannot be diagonals in the remaining three quadrangles too. It follows that the isolation indices corresponding to these 2-splits corresponding to the edges given above, are non-zero.

For each of $Q(1,2,4,5)$, $Q(2,3,5,6)$ and $Q(1,3,4,6)$, there are 2 possibilities. If the ordering of the vertices agrees with the ordering $(1,2,3,4,5,6)$ we call the quadrangle of S type, otherwise we call it of T type. For example, $Q(1245)$ is of S type, while $Q(1254)$ is of T type. It can be seen that quadrangle structures with only T type are all equivalent. Similarly the ones with two T types are also equivalent. It follows that there are only 4 inequivalent quadrangle structures:

$$\begin{aligned}
\text{Type}(SSS) :& Q(1245), Q(2356), Q(3461), \\
\text{Type}(SST) :& Q(1245), Q(2356), Q(3416), \\
\text{Type}(STT) :& Q(1245), Q(2365), Q(3416), \\
\text{Type}(TTT) :& Q(1254), Q(2365), Q(3416).
\end{aligned}$$

The structure of the 4-point subsets $Q(1,2,4,5)$, $Q(2,3,5,6)$ and $Q(1,3,4,6)$ for each type is shown below in Figure 3.

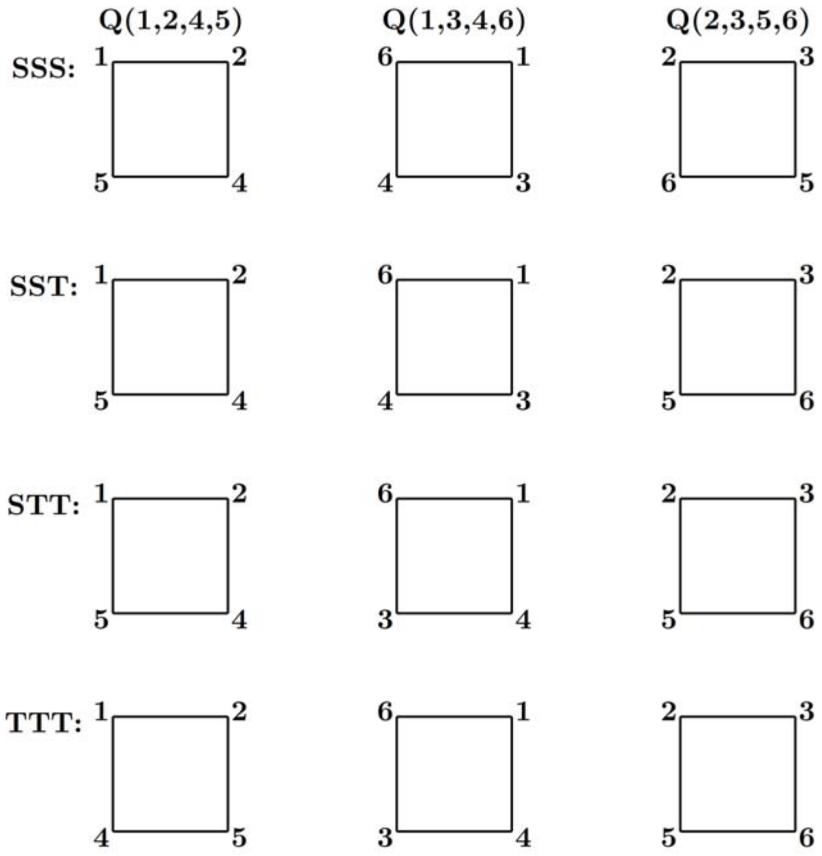


Figure 3: The structure of the 4-point subsets $Q(1,2,4,5)$, $Q(1,3,4,6)$ and $Q(1,3,5,6)$ for the SSS, SST, STT and TTT types.

These are the metrics numbered as 12, 13, 39, 65, 66 in the Sturmfels-Yu classification. The labels of vertices in these types, as compared with our notation, are as follows.

$$\text{Types } 12, 13: (1,2,5,4,6,3) \to (1,2,3,4,5,6),$$
$$\text{Types } 39, 65, 66: (1,2,5,4,6,3) \to (1,2,3,4,5,6). \tag{13}$$

We give below the complete graphs for Type 12 and Type 66, in Figure 4.

Table 1:

Edges (St.)	d_{12}	d_{13}	d_{14}	d_{15}	d_{16}	d_{23}	d_{24}	d_{25}	d_{26}	d_{34}	d_{35}	d_{36}	d_{45}	d_{46}	d_{56}
Type 12(TTT)	3	3	5	6	6	6	6	3	5	6	5	3	3	3	6
Type 13(TTT)	5	5	7	10	10	10	10	5	7	10	11	5	5	5	10
Edges (ours)	d_{12}	d_{16}	d_{14}	d_{13}	d_{15}	d_{26}	d_{24}	d_{23}	d_{25}	d_{46}	d_{36}	d_{56}	d_{34}	d_{45}	d_{35}

Table 2:

Edges (St.)	d_{12}	d_{13}	d_{14}	d_{15}	d_{16}	d_{23}	d_{24}	d_{25}	d_{26}	d_{34}	d_{35}	d_{36}	d_{45}	d_{46}	d_{56}
Type 39(TTS)	2	2	5	5	4	4	3	4	5	5	3	5	4	2	2
Type 65(STS)	4	3	8	7	7	7	4	9	7	7	4	8	7	3	4
Type 66(SSS)	3	3	6	6	7	6	3	7	6	7	3	6	6	3	3
Edges (ours)	d_{12}	d_{16}	d_{13}	d_{15}	d_{14}	d_{26}	d_{23}	d_{25}	d_{24}	d_{36}	d_{56}	d_{46}	d_{35}	d_{34}	d_{45}

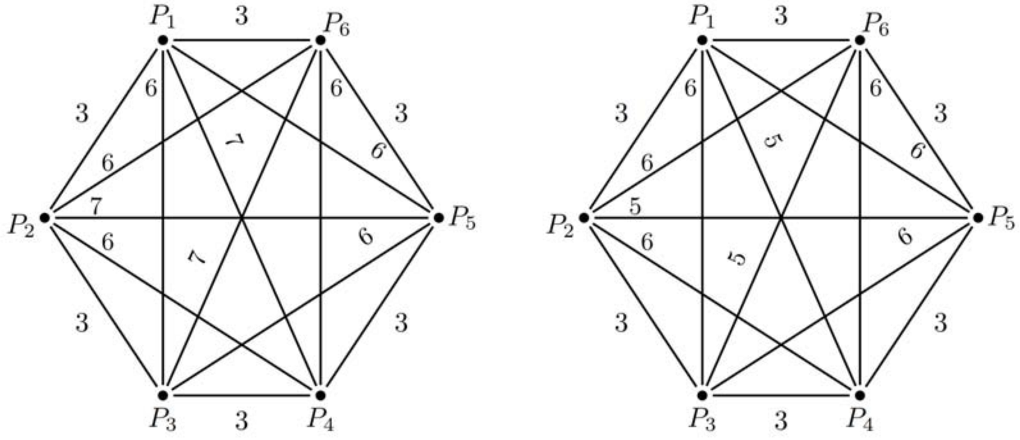


Figure 4: The first figure has Gromov product type I_{17}, Quadrangle Type SSS, Sturmfels-Yu Type 66 and the second has Gromov product type I_{17}, Quadrangle Type TTT, Sturmfels-Yu Type 12 (See Table 2).

We will explain how the parameterization for a 6-point space is done. For this aim we first work with the metric in which the type of all free quadrangles is S and we also write the parameters of all quadrangles. This is shown in Figure 5 below.

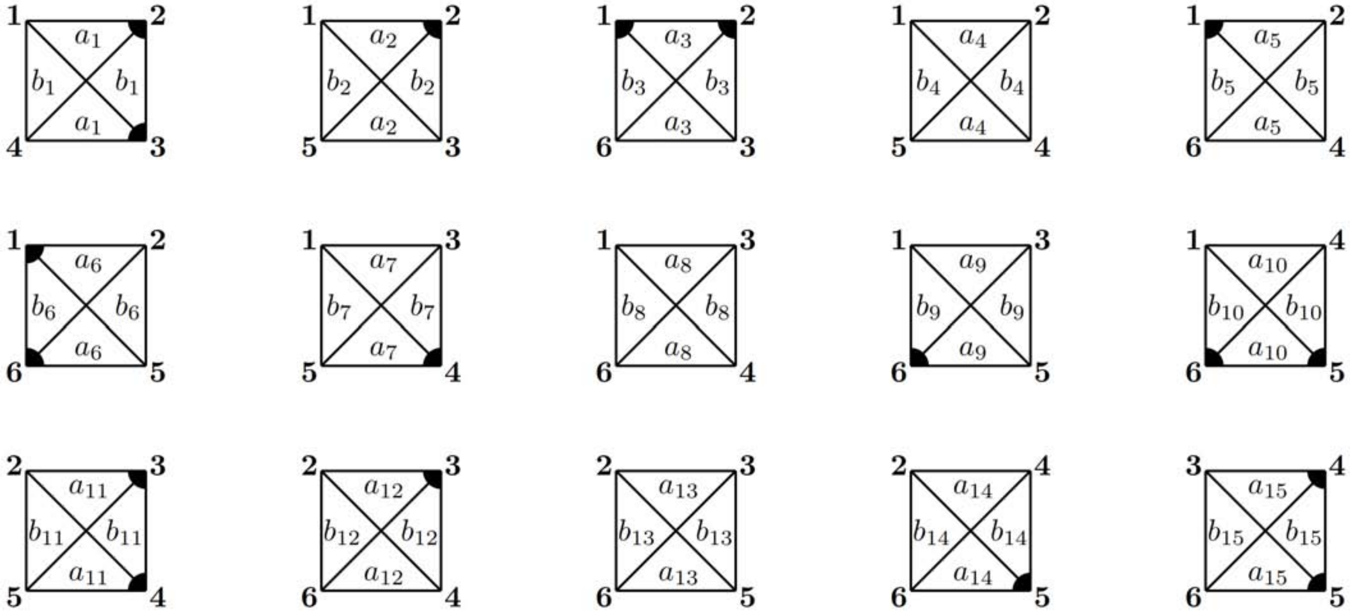


Figure 5: All quadrangles of Gromov product type I_{17}, Quadrangle Type SSS, Sturmfels-Yu Type 66 with quadrangle parameters.

What we need is first is to use quadrangle structure to find out about the isolation indices of the splits of a metric. As proved before, the isolation indices for 1-splits are Gromov products. Let's do the calculations for 2-splits. We will start by calculating the isolation index of δ_{12} which we show by m_{12}. By definition we have

$$m_{12} = \frac{1}{2} \min_{a,a' \in \{1,2\}, b,b' \in \{1,2\}^c} \{max\{d_{ab} + d_{a'b'}, d_{ab'} + d_{a'b}, d_{aa'} + d_{bb'}\} - (d_{aa'} + d_{bb'})\}.$$

To find out the maximum that has appeared here we look at all the quadrangles in Figure 5 which contain 1 and 2 and then select the diagonal pair as maximum, hence we will have:

$$\begin{aligned} m_{12} = \frac{1}{2} \min\{ & d_{13} + d_{24} - d_{12} - d_{34}, d_{13} + d_{25} - d_{12} - d_{35} \\ & d_{13} + d_{26} - d_{12} - d_{36}, d_{14} + d_{25} - d_{12} - d_{45} \\ & d_{14} + d_{26} - d_{12} - d_{46}, d_{15} + d_{26} - d_{12} - d_{56}\}. \end{aligned}$$

To be able to go on, one should keep in mind that the quadrangles of Figure 5 give the distances as a sum of some parameters and to make this clear we should note that the quadrangles can be drawn in a different way that help us achieve our aim. For instance $Q(1234)$ could be drawn as the following figure:

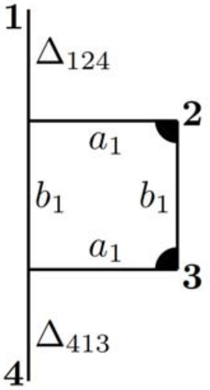


Figure 6: Another representation for quadrangle $Q(1234)$ with quadrangle parameters

Considering this way of representing a quadrangle, then we have $d_{12} = \Delta_{124} + a_1$ when quadrangle $Q(1234)$ is investigated. d_{12} comes in 5 other quadrangles also and thus we have:

$$\begin{aligned} d_{12} &= \Delta_{124} + a_1 = \Delta_{125} + a_2 = a_3 \\ &= \Delta_{124} + a_4 + \Delta_{214} = a_5 + \Delta_{214} = a_6 + \Delta_{215}. \end{aligned}$$

This second representation for all quadrangles are given below in Figure 7:

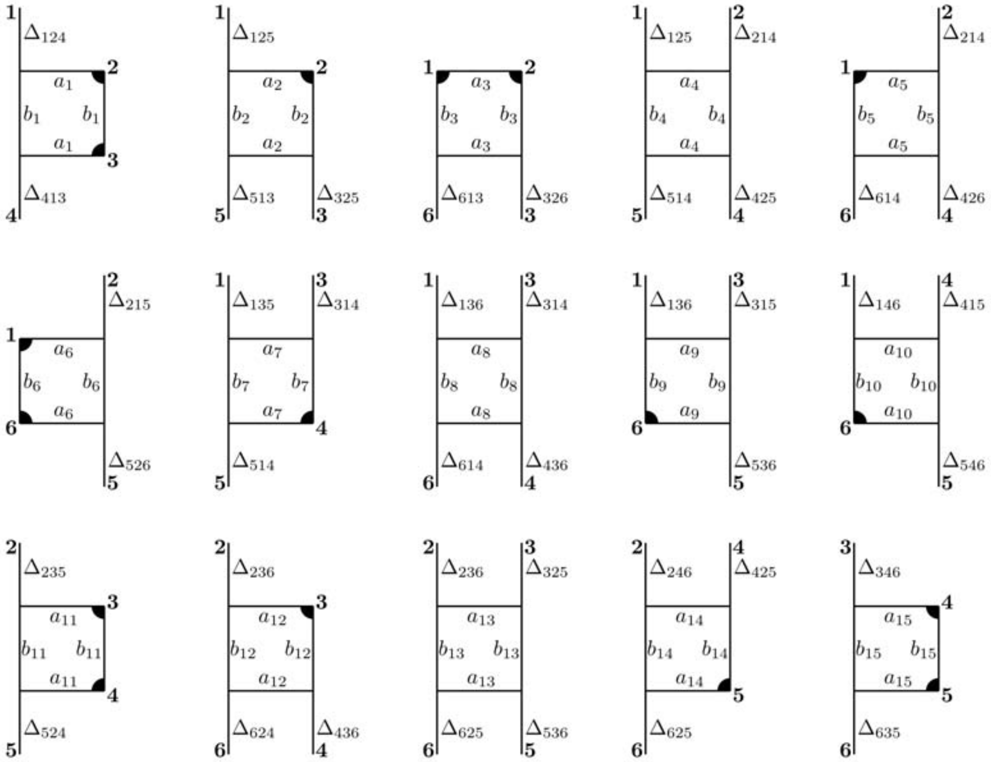


Figure 7: Second representation for quadrangles of type SSS with quadrangle parameters.

We have the following list for the other distances according to Figure 7:

$$\begin{aligned}
d_{13} &= \Delta_{124} + b_1 + a_1 = \Delta_{125} + b_2 + a_2 + \Delta_{325} = b_3 + a_3 + \Delta_{326} \\
&= \Delta_{135} + a_7 + \Delta_{314} = \Delta_{136} + a_8 + \Delta_{314} = \Delta_{136} + a_9 + \Delta_{315}. \\
d_{14} &= \Delta_{124} + b_1 + \Delta_{413} = \Delta_{125} + b_4 + a_4 + \Delta_{425} = b_5 + a_5 + \Delta_{426} \\
&= \Delta_{135} + b_7 + a_7 = \Delta_{136} + b_8 + a_8 + \Delta_{436} = \Delta_{146} + a_{10} + \Delta_{415}. \\
d_{15} &= \Delta_{125} + b_2 + \Delta_{513} = \Delta_{125} + b_4 + \Delta_{514} = a_6 + b_6 + \Delta_{526} \\
&= \Delta_{135} + b_7 + \Delta_{514} = \Delta_{136} + a_9 + b_9 + \Delta_{536} = \Delta_{146} + a_{10} + b_{10} + \Delta_{546}. \\
d_{16} &= b_3 + \Delta_{613} = b_5 + \Delta_{614} = b_6 \\
&= \Delta_{136} + b_8 + \Delta_{614} = \Delta_{136} + b_9 = \Delta_{146} + b_{10}.
\end{aligned}$$

$$
\begin{aligned}
d_{23} &= b_1 = b_2 + \Delta_{325} = b_3 + \Delta_{326}\\
&= \Delta_{235} + a_{11} = \Delta_{236} + a_{12} = \Delta_{236} + a_{13} + \Delta_{325}.\\
d_{24} &= a_1 + b_1 + \Delta_{413} = \Delta_{214} + b_4 + \Delta_{425} = \Delta_{214} + b_5 + \Delta_{426}\\
&= \Delta_{235} + a_{11} + b_{11} = \Delta_{236} + a_{12} + b_{12} + \Delta_{436} = \Delta_{246} + a_{14} + \Delta_{425}.\\
d_{25} &= a_2 + b_2 + \Delta_{513} = \Delta_{214} + a_4 + b_4 + \Delta_{514} = \Delta_{215} + b_6 + \Delta_{526}\\
&= \Delta_{235} + b_{11} + \Delta_{524} = \Delta_{236} + a_{13} + b_{13} + \Delta_{536} = \Delta_{246} + a_{14} + b_{14}.\\
d_{26} &= a_3 + b_3 + \Delta_{613} = \Delta_{214} + a_5 + b_5 + \Delta_{614} = \Delta_{215} + a_6 + b_6\\
&= \Delta_{236} + b_{12} + \Delta_{624} = \Delta_{236} + b_{13} + \Delta_{625} = \Delta_{246} + b_{14} + \Delta_{625}.
\end{aligned}
$$

$$
\begin{aligned}
d_{34} &= a_1 + \Delta_{413} = \Delta_{314} + b_7 = \Delta_{314} + b_8 + \Delta_{436}\\
&= b_{11} = b_{12} + \Delta_{436} = \Delta_{346} + a_{15}.\\
d_{35} &= \Delta_{325} + a_2 + \Delta_{513} = \Delta_{314} + a_7 + b_7 + \Delta_{514} = \Delta_{315} + b_9 + \Delta_{536}\\
&= a_{11} + b_{11} + \Delta_{524} = \Delta_{325} + b_{13} + \Delta_{536} = \Delta_{346} + a_{15} + b_{15}.\\
d_{36} &= \Delta_{326} + a_3 + \Delta_{613} = \Delta_{314} + a_8 + b_8 + \Delta_{614} = \Delta_{315} + a_9 + b_9\\
&= a_{12} + b_{12} + \Delta_{624} = \Delta_{325} + a_{13} + b_{13} + \Delta_{625} = \Delta_{346} + b_{15} + \Delta_{635}.
\end{aligned}
$$

$$
\begin{aligned}
d_{45} &= \Delta_{425} + a_4 + \Delta_{514} = a_7 + \Delta_{514} = \Delta_{415} + b_{10} + \Delta_{546}\\
&= a_{11} + \Delta_{524} = \Delta_{425} + b_{14} = b_{15}.\\
d_{46} &= \Delta_{426} + a_5 + \Delta_{614} = \Delta_{436} + a_8 + \Delta_{614} = \Delta_{415} + a_{10} + b_{10}\\
&= \Delta_{436} + a_{12} + \Delta_{624} = \Delta_{425} + a_{14} + b_{14} + \Delta_{625} = a_{15} + b_{15} + \Delta_{635}.\\
d_{56} &= \Delta_{526} + a_6 = \Delta_{536} + a_9 = \Delta_{546} + a_{10}\\
&= \Delta_{536} + a_{13} + \Delta_{625} = a_{14} + \Delta_{625} = a_{15} + \Delta_{635}.
\end{aligned}
$$

Now in order to simplify the formula for m_{12}, the distances d_{ij} must be replaced by the parameters given above. To chose the best candidate among the equalities we should look to the quadrangle which is related to all terms of each expression inside the formula of m_{12}. This means that if we want to replace d_{13}, d_{24}, d_{12} and d_{34} by the quadrangle parameters, we should use ones that come from the quadrangle $Q(1234)$; hence

$$d_{13}+d_{24}-d_{12}-d_{34} = (\Delta_{124}+a_1+b_1)+(a_1+b_1+\Delta_{413})-(\Delta_{124}+a_1)-(a_1+\Delta_{413}) = 2b_1.$$

Likewise we have:

$$\begin{aligned}
d_{13}+d_{25}-d_{12}-d_{35} &= (\Delta_{125}+a_2+b_2+\Delta_{325})+(a_2+b_2+\Delta_{513}) \\
&\quad -(a_2+\Delta_{125})-(\Delta_{325}+a_2+\Delta_{513}) = 2b_2, \\
d_{13}+d_{26}-d_{12}-d_{36} &= (a_3+b_3+\Delta_{326})+(a_3+b_3+\Delta_{613}) \\
&\quad -a_3-(\Delta_{326}+a_3+\Delta_{613}) = 2b_3, \\
d_{14}+d_{25}-d_{12}-d_{45} &= (\Delta_{125}+a_4+b_4+\Delta_{425})+(\Delta_{214}+a_4+b_4+\Delta_{514}) \\
&\quad -(\Delta_{125}+a_4+\Delta_{214})-(\Delta_{425}+a_4+\Delta_{514}) = 2b_4, \\
d_{14}+d_{26}-d_{12}-d_{46} &= (a_5+b_5+\Delta_{426})+(\Delta_{214}+a_5+b_5+\Delta_{614}) \\
&\quad -(a_5+\Delta_{214})-(\Delta_{426}+a_5+\Delta_{614}) = 2b_5, \\
d_{15}+d_{26}-d_{12}-d_{56} &= (a_6+b_6+\Delta_{526})+(\Delta_{215}+a_6+b_6) \\
&\quad -(a_6+\Delta_{215})-(\Delta_{526}+a_6) = 2b_6.
\end{aligned}$$

And thus

$$m_{12} = \min\{b_1, b_2, b_3, b_4, b_5, b_6\}.$$

To find out what is this minimum, we will do the following: According to the equations given before for the distances with quadrangle parameters and Gromov products we have $d_{23} = b_1 = b_2 + \Delta_{325}$, hence $b_2 < b_1$. Also $d_{16} = b_6 = b_3 + \Delta_{613}$, hence $b_3 < b_6$. So $m_{12} = \min\{b_2, b_3, b_4, b_5\}$. On the other hand we have $d_{23} = b_2 + \Delta_{325} = b_3 + \Delta_{326}$ and if we look at the quadrangle with the vertices 2, 3, 5 and 6, which is $Q(2356)$ here, we see that E_{25} and E_{36} are diagonals which imply that $\Delta_{325} < \Delta_{326}$. Thus $b_3 < b_2$. Likewise we have $d_{16} = b_3 + \Delta_{613} = b_5 + \Delta_{614}$ and from quadrangle $Q(1346)$, since E_{14} and E_{36} are diagonals, we have $\Delta_{614} < \Delta_{613}$ and hence $b_3 < b_5$. So $m_{12} = \min\{b_3, b_4\}$. To find the order relation between b_3 and b_4, there are no simple equalities, but we have $d_{15} = \Delta_{125} + b_2 + \Delta_{513} = \Delta_{125} + b_4 + \Delta_{514}$. So $b_2 + \Delta_{513} = b_4 + \Delta_{514}$ and from quadrangle $Q(1345)$, we have $\Delta_{514} < \Delta_{513}$, so $b_2 < b_4$. We have shown before that $b_3 < b_2$ and thus $b_3 < b_4$. So $m_{12} = b_3$. According to $Q(1236)$ of Figure 5, $b_3 = \Delta_{136} - \Delta_{126} = \Delta_{136}$.

What we did for representing the isolation index of a 2-split in terms of quadrangle parameters a_i's and b_i's in the above example can be summarized in the proposition below:

Proposition 3 *For a Q-generic finite metric space, the isolation index of a 2-split $A = \{a, a'\}$ is zero if $E_{aa'}$ is a diagonal in at least one quadrangle or it is equal to the minimum among all of the quadrangle parameter assigned to the neighbor edge of $E_{aa'}$ in every quadrangle in which $E_{aa'}$ is a side.*

Proof. By definition of the isolation index we have

$$\alpha_A = \alpha_{aa'} = \frac{1}{2} \min_{b,b' \in A^c} \{\max\{d_{ab} + d_{a'b'}, d_{ab'} + d_{a'b}, d_{aa'} + d_{bb'}\} - (d_{aa'} + d_{bb'})\}.$$

If there is a quadrangle in which $E_{aa'}$ is a diagonal, then one of the terms in which minimum is calculated is zero, since for this quadrangle we have $\max\{d_{ab} + d_{a'b'}, d_{ab'} + d_{a'b}, d_{aa'} + d_{bb'}\} = d_{aa'} + d_{bb'}$. So we assume that $\alpha_{aa'}$ is not zero and hence $E_{aa'}$ can only appear in quadrangles of the form $Q(aa'bb')$ or $Q(aa'b'b)$ for $b, b' \in A^c$. If the i's quadrangle (the quadrangle that has parameters a_i and b_i in the list of quadrangles) is $Q(aa'bb')$ then the term inside the min expression for this quadrangle is equal to

$$\begin{aligned} d_{ab} + d_{a'b'} - d_{aa'} - d_{bb'} &= 2(\Delta_{abb'} - \Delta_{aa'b'}) = 2(\Delta_{a'bb'} - \Delta_{a'ab'}) \\ &= 2(\Delta_{baa'} - \Delta_{ba'b'}) = 2(\Delta_{b'aa'} - \Delta_{b'ab}) = 2b_i. \end{aligned}$$

On the other hand If the i's quadrangle is $Q(aa'bb')$ then the term inside the min expression for this quadrangle is equal to

$$\begin{aligned} d_{ab'} + d_{a'b} - d_{aa'} - d_{bb'} &= 2(\Delta_{abb'} - \Delta_{aa'b}) = 2(\Delta_{a'bb'} - \Delta_{a'ab'}) \\ &= 2(\Delta_{baa'} - \Delta_{bab'}) = 2(\Delta_{b'aa'} - \Delta_{b'a'b}) = 2b_i. \end{aligned}$$

So in either cases the terms are equal to the quadrangle parameter assigned to the neighbor edge of $E_{aa'}$ and hence

$$\alpha_{aa'} = \min_k \{b_k\},$$

where k runs over all quadrangles that contain $E_{aa'}$ as a side and b_k is the parameter assigned to the neighbor side of the same quadrangle. ■

This proposition which is in fact a restatement of part ii. of Proposition 2, will be used to calculate the isolation indices of all 2-splits and 3-splits of $I_{17}SSS$ metric. First of all, in Figure 5, E_{13}, E_{14}, E_{15}, E_{24}, E_{25}, E_{26}, E_{35}, E_{36} and E_{46} have appeared as diagonals and hence isolation indices of their related 2-splits are zero and non-zero isolation indices are α_{12}, α_{16}, α_{23}, α_{34}, α_{45} and α_{56}. We have calculated $\alpha_{12} = m_{12} = b_3 = \Delta_{136}$ before.

For $\alpha_{16} = m_{16}$ we have $m_{16} = \min\{a_3, a_5, a_6, a_8, a_9, a_{10}\}$ according to proposition and Figure 5. Since $d_{12} = a_3 = a_5 + \Delta_{214} = a_6 + \Delta_{215}$, according to $Q(1245)$ we have $\Delta_{214} < \Delta_{215}$ and thus $a_6 < a_5$. Clearly $a_5 < a_3$. Likewise $d_{13} = \Delta_{136} + a_8 + \Delta_{314} = \Delta_{136} + a_9 + \Delta_{315}$ and from $Q(1345)$, we have $\Delta_{314} < \Delta_{315}$, thus $a_9 < a_8$. So $m_{16} = \min\{a_6, a_9, a_{10}\}$. On the other hand $d_{56} = \Delta_{526} + a_6 = \Delta_{536} + a_9 = \Delta_{546} + a_{10}$. From $Q(3456)$ and $Q(2356)$, we have $\Delta_{546} < \Delta_{536}$ and $\Delta_{536} < \Delta_{526}$ respectively and hence $a_6 < a_9 < a_{10}$. This means that

$$\alpha_{16} = m_{16} = a_6 = \Delta_{125}$$

The final equality $a_6 = \Delta_{125}$ is because of $Q(1256)$.

In the above calculations, we have also used the partial order among Gromov products implicitly, for type SSS this partial orders are given below in Figure 8:

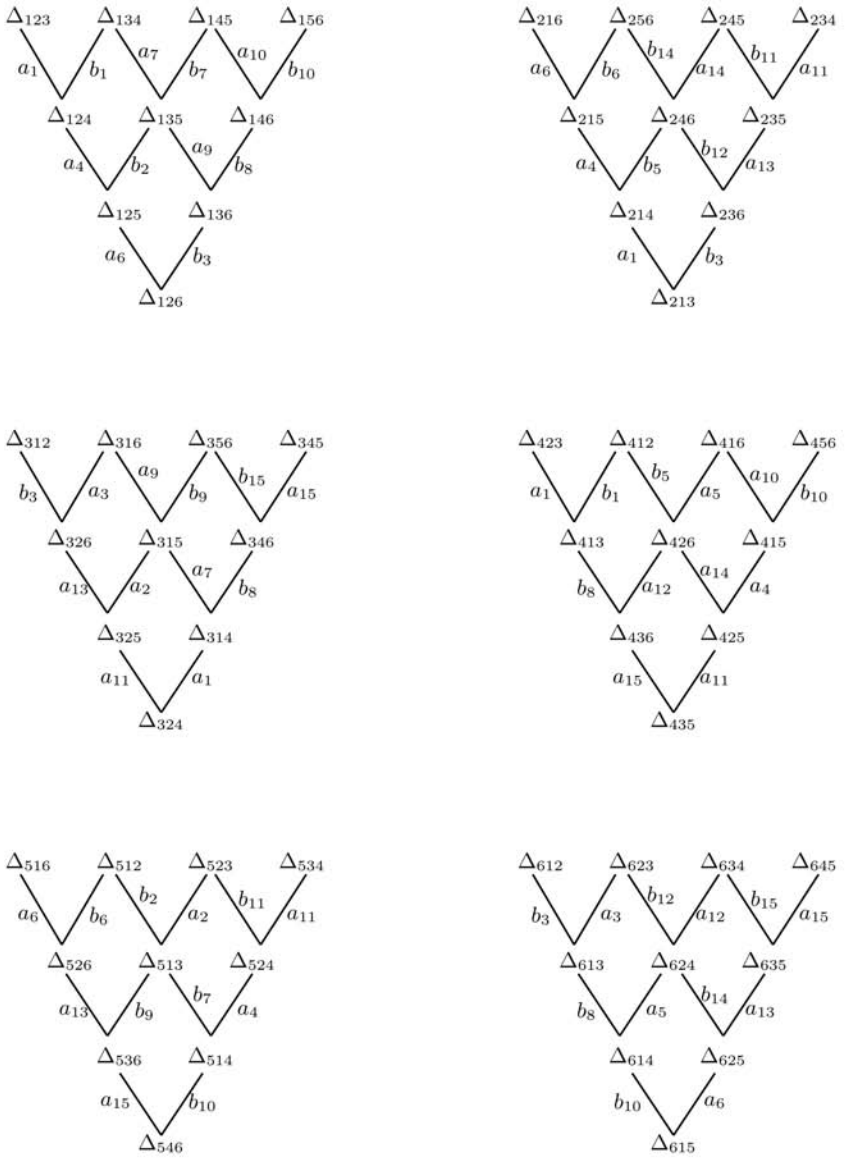


Figure 8: Partial order diagrams of Gromov product type I_{17}, Quadrangle Type SSS.

The rest is calculated likewise.

We give the first and second way of representing the quadrangles of type I_{17} in the following figures:

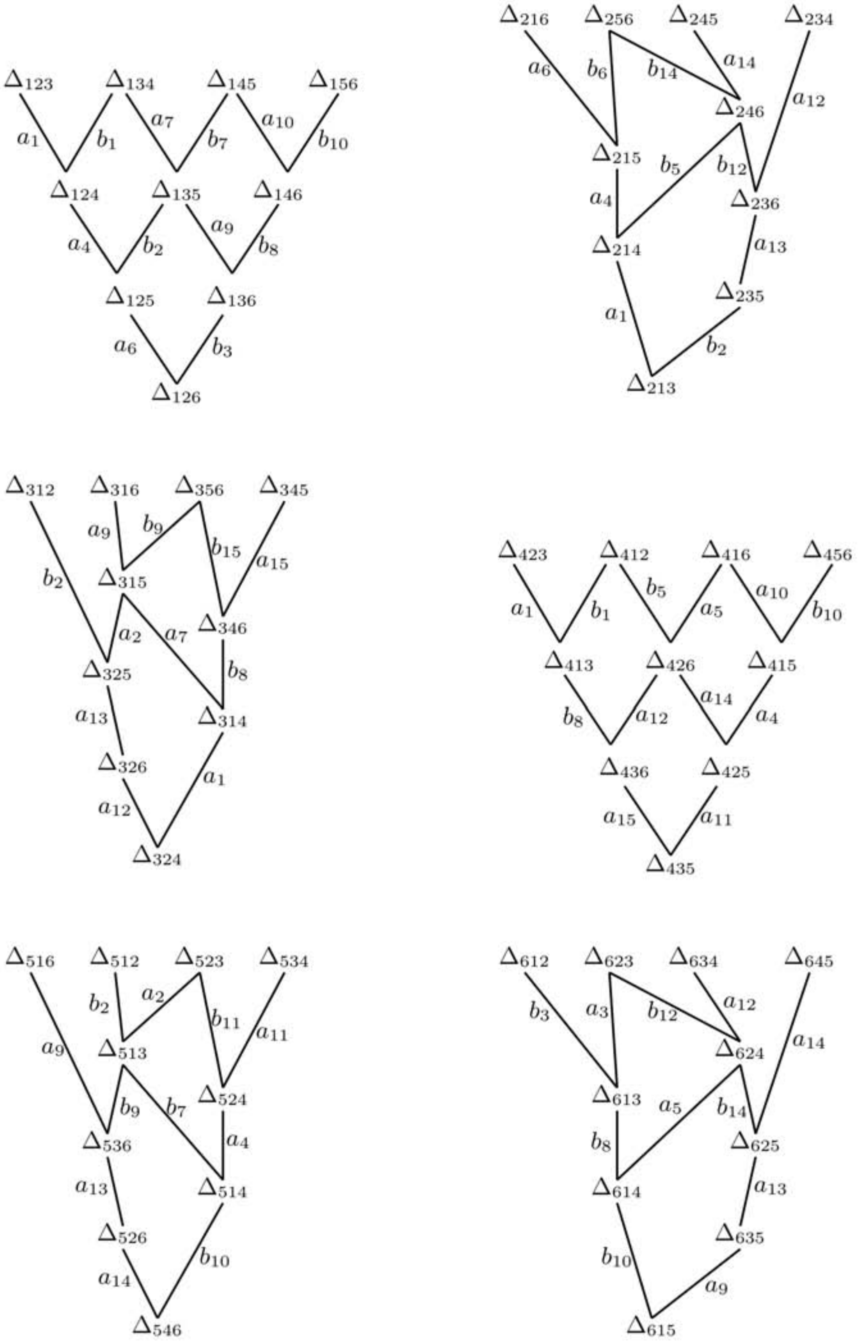


Figure 9: Partial order diagrams of Gromov product type I_{17}, Quadrangle Type SST.

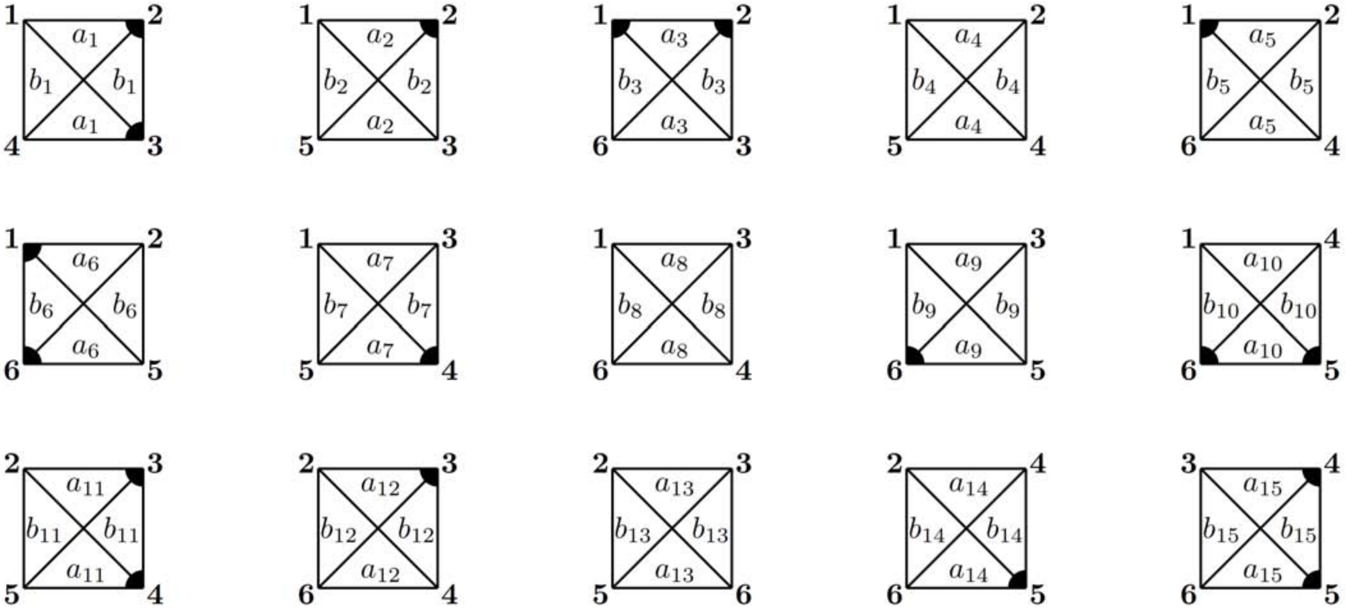


Figure 10: All quadrangles of Gromov product type I_{17}, Quadrangle Type SST.

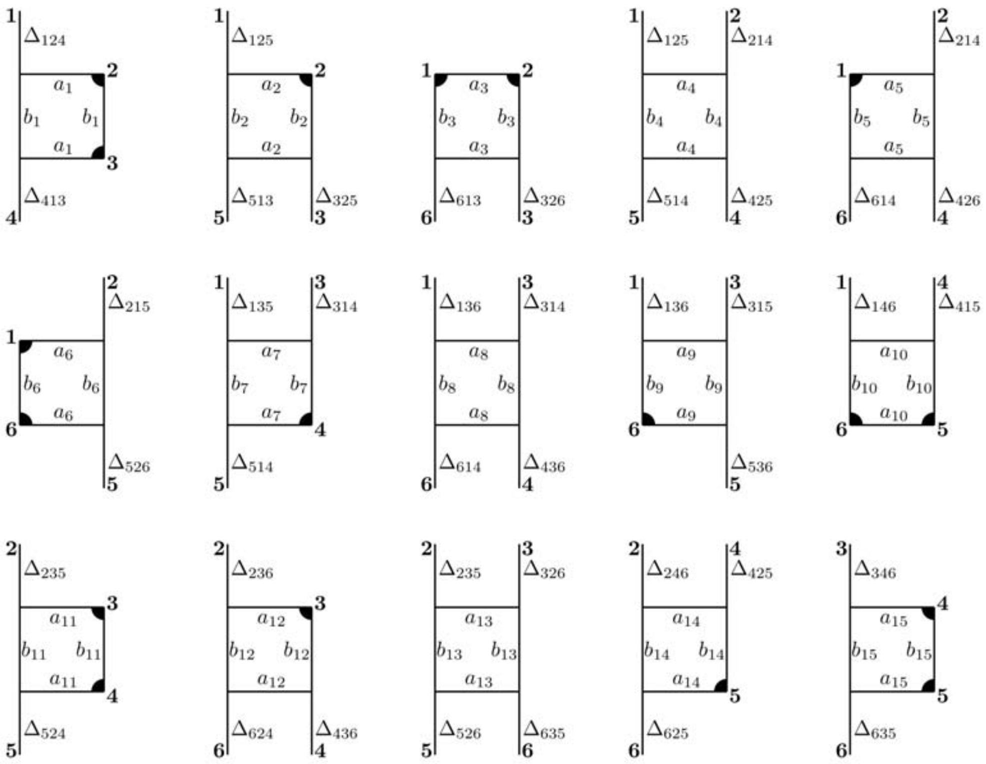


Figure 11: Second representation for quadrangles of type SST with quadrangle parameters.

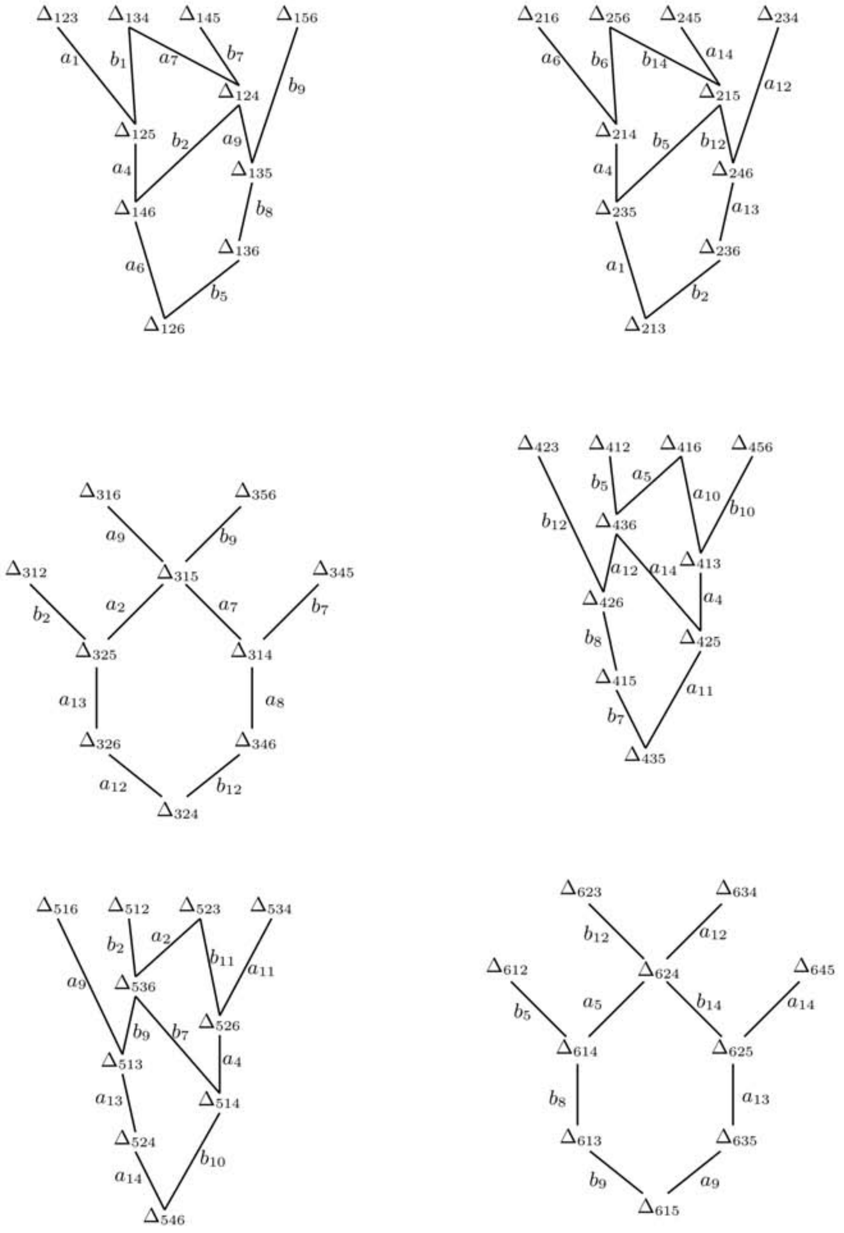


Figure 12: Partial order diagrams of Gromov product type I_{17}, Quadrangle Type STT.

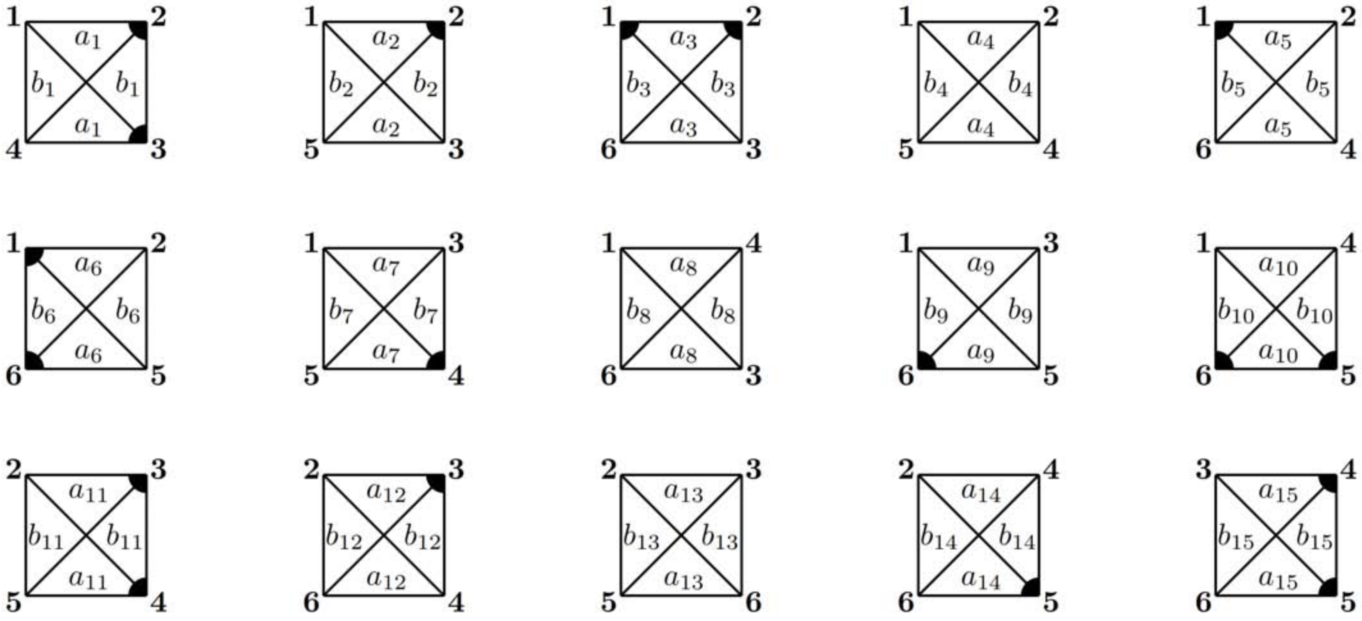


Figure 13: All quadrangles of Gromov product type I_{17}, Quadrangle Type STT.

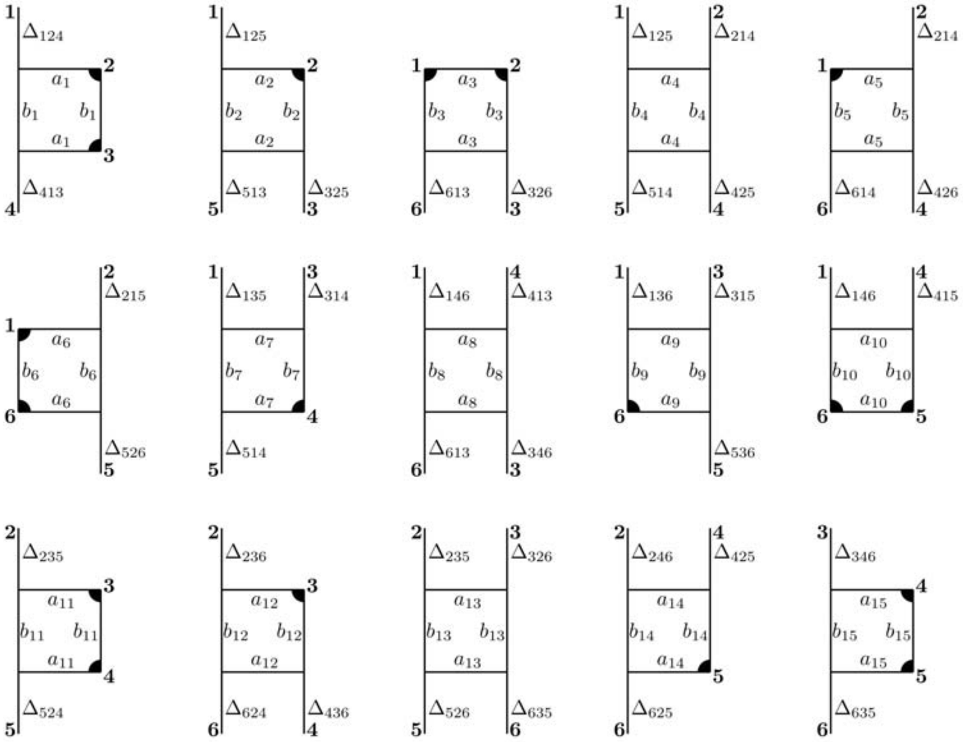


Figure 14: Second representation for quadrangles of type STT with quadrangle parameters.

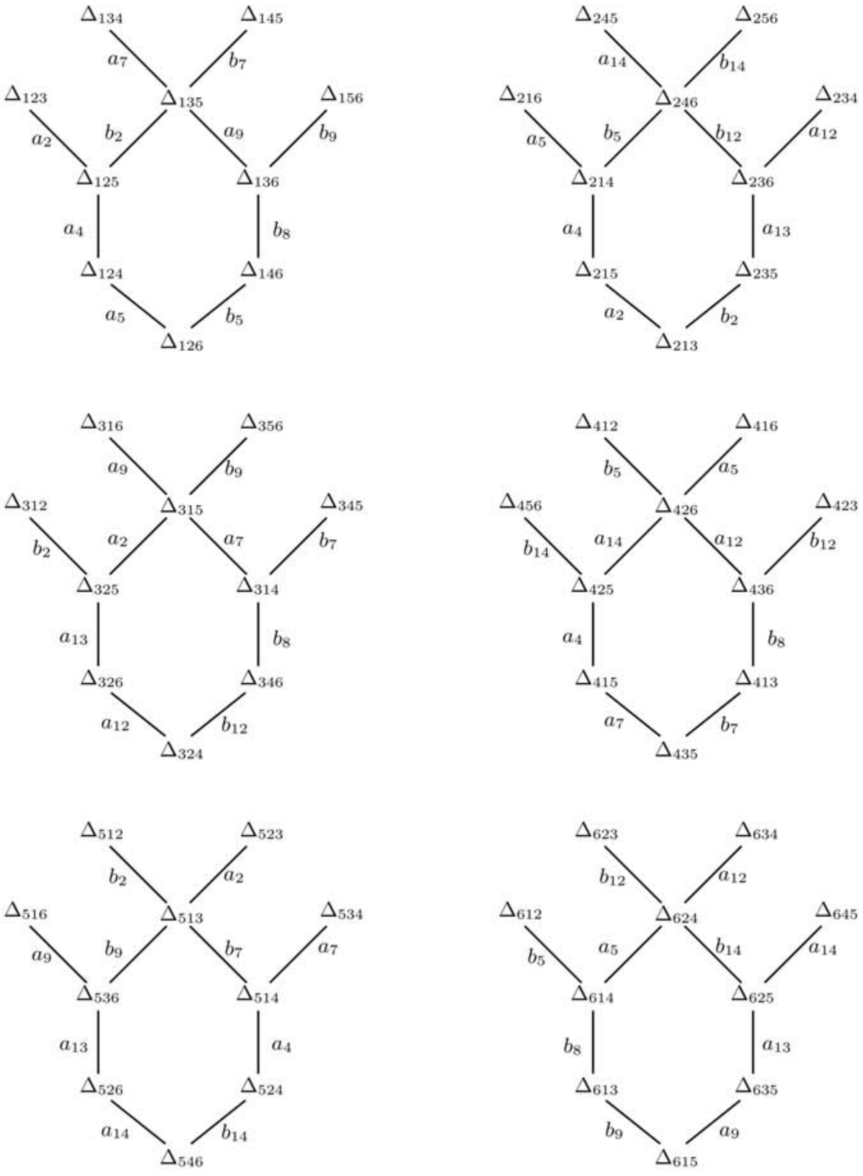


Figure 15: Partial order diagrams of Gromov product type I_{17}, Quadrangle Type TTT.

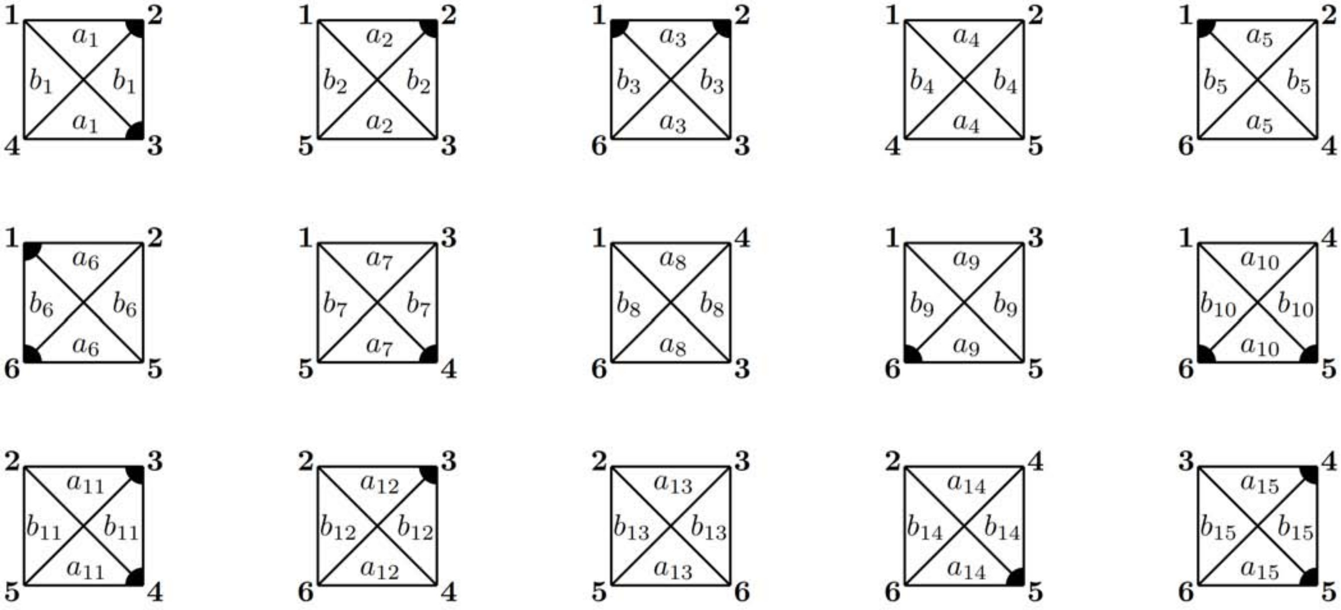


Figure 16: All quadrangles of Gromov product type I_{17}, Quadrangle Type TTT.

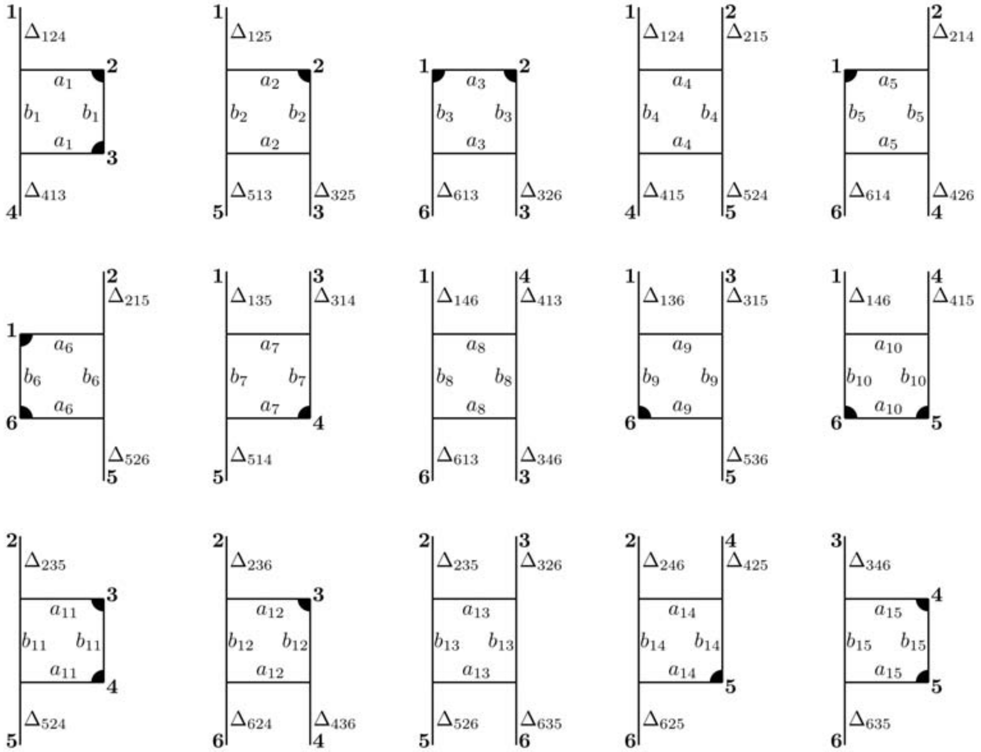


Figure 17: Second representation for quadrangles of type TTT with quadrangle parameters.

From Figures 3, one can see that all 4 quadrangle types have the same non zero isolation indices for 2-splits. Therefore each of the 4 quadrangle types are parameterized as below.

$$
\begin{aligned}
d_{12} &= m_{16} + m_{23} + \delta_{12}, & d_{23} &= m_{12} + m_{34} + \delta_{23},\\
d_{34} &= m_{23} + m_{45} + \delta_{34}, & d_{45} &= m_{34} + m_{56} + \delta_{45},\\
d_{56} &= m_{16} + m_{45} + \delta_{56}, & d_{16} &= m_{12} + m_{56} + \delta_{16},\\
d_{14} &= m_{12} + m_{16} + m_{34} + m_{45} + \delta_{14}, & &\\
d_{25} &= m_{12} + m_{23} + m_{45} + m_{56} + \delta_{25}, & &\\
d_{36} &= m_{16} + m_{23} + m_{34} + m_{56} + \delta_{36} & &. \qquad (14)
\end{aligned}
$$

It is then easy to compute the Gromov products for these type. The parameters δ_{ij}'s are to be determined from Eqns. (2), which consist of 60 equations to be solved for the 45 Δ_{ijk}'s. The passage to pendant free reductions reduces to number of parameters to 9. Furthermore, since each of the quadrangle types has exactly 6 nonzero isolation indices for 2-splits, we have only 9 (positive) parameters, the δ_{ij}'s to be determined from these set of equations. In order to ensure that the solutions of this linear system are positive, we have to use the structure of the quadrangles in Figure 3. By solving these linear equations using REDUCE, we obtain the following parameterizations for each case.

Type (SSS): p, q, r are isolation indices for 3 splits

$$\begin{aligned}
d_{12} &= m_{16} + m_{23} + p, & d_{23} &= m_{12} + m_{34} + r,\\
d_{34} &= m_{23} + m_{45} + q, & d_{45} &= m_{34} + m_{56} + p,\\
d_{56} &= m_{16} + m_{45} + r, & d_{16} &= m_{12} + m_{56} + q,\\
d_{14} &= m_{12} + m_{16} + m_{34} + m_{45} + p + q + r, &&\\
d_{25} &= m_{12} + m_{23} + m_{45} + m_{56} + p + q + r, &&\\
d_{36} &= m_{16} + m_{23} + m_{34} + m_{56} + p + q + r. &&
\end{aligned}$$

Type (SST): p,q are isolation indices for 3 splits

$$\begin{aligned}
d_{12} &= m_{16} + m_{23} + p + r, & d_{23} &= m_{12} + m_{34} + r,\\
d_{34} &= m_{23} + m_{45} + q + r, & d_{45} &= m_{34} + m_{56} + p + r,\\
d_{56} &= m_{16} + m_{45} + r, & d_{16} &= m_{12} + m_{56} + q + r,\\
d_{14} &= m_{12} + m_{16} + m_{34} + m_{45} + p + q + 3r, &&\\
d_{25} &= m_{12} + m_{23} + m_{45} + m_{56} + p + q + r, &&\\
d_{36} &= m_{16} + m_{23} + m_{34} + m_{56} + p + q + r. &&
\end{aligned}$$

Type (STT): p is the isolation index for a 3 split

$$\begin{aligned}
d_{12} &= m_{16} + m_{23} + p + 2q + v, & d_{23} &= m_{12} + m_{34} + q + v,\\
d_{34} &= m_{23} + m_{45} + q + v, & d_{45} &= m_{34} + m_{56} + p + 2q + v,\\
d_{56} &= m_{16} + m_{45} + q + v, & d_{16} &= m_{12} + m_{56} + q + v,\\
d_{14} &= m_{12} + m_{16} + m_{34} + m_{45} + p + 2q + 3v, &&\\
d_{25} &= m_{12} + m_{23} + m_{45} + m_{56} + p + 2q + v, &&\\
d_{36} &= m_{16} + m_{23} + m_{34} + m_{56} + p + 2q + v. &&
\end{aligned}$$

Type (TTT):

$$\begin{aligned}
d_{12} &= m_{16} + m_{23} + p + 2u + v, & d_{23} &= m_{12} + m_{34} + p + u + v,\\
d_{34} &= m_{23} + m_{45} + p + u + v, & d_{45} &= m_{34} + m_{56} + p + 2u + v,\\
d_{56} &= m_{16} + m_{45} + p + u + v, & d_{16} &= m_{12} + m_{56} + p + u + v,\\
d_{14} &= m_{12} + m_{16} + m_{34} + m_{45} + p + 2u + 3v, & &\\
d_{25} &= m_{12} + m_{23} + m_{45} + m_{56} + p + 2u + v, & &\\
d_{36} &= m_{16} + m_{23} + m_{34} + m_{56} + p + 2u + v. & &
\end{aligned}$$

The Sturmfels-Yu Type 12 metric is characterized by the inequalities given in Example 5 in [13]. In our notation (see Table 1), the last three of the inequalities characterizing this metric are

$$\begin{aligned}
d_{13} + d_{26} + d_{46} + d_{35} &\geq d_{15} + d_{24} + 2d_{36},\\
d_{15} + d_{26} + d_{24} + d_{35} &\geq d_{13} + 2d_{25} + d_{46},\\
d_{13} + d_{15} + d_{24} + d_{46} &\geq 2d_{14} + d_{26} + d_{35}.
\end{aligned}$$

After using the fact that minimal Gromov products are zero, these inequalities are reduced to

$$d_{12} + d_{45} \geq d_{36}, \quad d_{16} + d_{34} \geq d_{25}, \quad d_{23} + d_{56} \geq d_{14}.$$

When we substitute the expressions of the d_{ij}'s we obtain

$$p + 2u + v \geq 0, \quad p + v \geq 0, \quad p - v \geq 0.$$

Thus the only non-trivial condition distinguishing Sturmfels-Yu Type-12 in the quadrangle equivalence class TTT, is $p \geq v$. This shows that the quadrangle equivalence is coarser than the decomposition of the metric fan.

The matrix representation for these types is a 9×9 matrix with the following numbering of the columns.

$$\{P_1P_3, P_1P_4, P_1P_5, P_2P_4, P_2P_5, P_2P_6, P_3P_5, P_3P_6, P_4P_6\}.$$

Table 3:

Type	x	y	z	u	v	w
SSS	0	1	1	0	1	0
SST	0	1	1	0	0	1
STT	1	1	0	0	0	1
TTT	1	0	0	1	0	1

The (common) matrix representation for these 4 quadrangle structures is

$$Q = \begin{pmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & x \\ 0 & 0 & 0 & 0 & y & 1 & 1 & z & 0 \\ 0 & 0 & 0 & u & 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & u & 0 & 0 & 0 & 1 & 1 & 0 \\ 1 & y & 0 & 0 & 0 & 0 & 0 & v & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & w & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & w & 0 & 0 & 1 \\ 0 & z & 1 & 1 & v & 0 & 0 & 0 & 0 \\ x & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \end{pmatrix},$$

where the values of the indeterminates are as follows.

It has been checked that the matrices (SSS) and (SST) are similar but the change of basis matrix is not a permutation matrix. The matrices for the types (STT) and (TTT) are not similar.

5 Gromov products

$$
\begin{aligned}
&\Delta_{126} = \Delta_{213} = \Delta_{324} = \Delta_{435} = \Delta_{546} = \Delta_{615} = 0,\\
&d_{13} = d_{12} + d_{23},\\
&d_{24} = d_{23} + d_{34},\\
&d_{35} = d_{34} + d_{45},\\
&d_{46} = d_{45} + d_{56},\\
&d_{15} = d_{16} + d_{56},\\
&d_{26} = d_{12} + d_{16}.
\end{aligned}
$$

Type SSS distances can be parameterized as following:

$$
\begin{aligned}
&d_{12} = m_{16} + m_{23} + p,\\
&d_{23} = m_{12} + m_{34} + r,\\
&d_{34} = m_{23} + m_{45} + q,\\
&d_{45} = m_{34} + m_{56} + p,\\
&d_{56} = m_{16} + m_{45} + r,\\
&d_{16} = m_{12} + m_{56} + q,\\
&d_{14} = m_{12} + m_{16} + m_{34} + m_{45} + p + q + r,\\
&d_{25} = m_{12} + m_{23} + m_{45} + m_{56} + p + q + r,\\
&d_{36} = m_{16} + m_{23} + m_{34} + m_{56} + p + q + r.
\end{aligned}
$$

For this type Gromov products can be parameterized as following:

$\Delta_{123} = m_{16} + m_{23} + p,$
$\Delta_{124} = m_{16} + p,$
$\Delta_{125} = m_{16},$
$\Delta_{126} = 0,$
$\Delta_{134} = m_{12} + m_{16} + m_{34} + p + r,$
$\Delta_{135} = m_{12} + m_{16} + r,$
$\Delta_{136} = m_{12},$
$\Delta_{145} = m_{12} + m_{16} + m_{45} + q + r,$
$\Delta_{146} = m_{12} + q,$
$\Delta_{156} = m_{12} + m_{56} + q.$

$\Delta_{213} = 0,$
$\Delta_{214} = m_{23},$
$\Delta_{215} = m_{23} + p,$
$\Delta_{216} = m_{16} + m_{23} + p,$
$\Delta_{234} = m_{12} + m_{34} + r,$
$\Delta_{235} = m_{12} + r,$
$\Delta_{236} = m_{12},$
$\Delta_{245} = m_{12} + m_{23} + m_{45} + q + r,$
$\Delta_{246} = m_{12} + m_{23} + q,$
$\Delta_{256} = m_{12} + m_{23} + m_{56} + p + q.$

$\Delta_{312} = m_{12} + m_{34} + r,$
$\Delta_{314} = m_{23},$
$\Delta_{315} = m_{23} + m_{34} + p,$
$\Delta_{316} = m_{16} + m_{23} + m_{34} + p + r,$
$\Delta_{324} = 0,$
$\Delta_{325} = m_{34},$
$\Delta_{326} = m_{34} + r,$
$\Delta_{345} = m_{23} + m_{45} + q,$
$\Delta_{346} = m_{23} + q,$
$\Delta_{356} = m_{23} + m_{34} + m_{56} + p + q.$

$\Delta_{412} = m_{12} + m_{34} + m_{45} + q + r,$
$\Delta_{413} = m_{45} + q,$
$\Delta_{415} = m_{34} + p,$
$\Delta_{416} = m_{16} + m_{34} + m_{45} + p + r,$
$\Delta_{423} = m_{23} + m_{45} + q,$
$\Delta_{425} = m_{34},$
$\Delta_{426} = m_{34} + m_{45} + r,$
$\Delta_{435} = 0,$
$\Delta_{436} = m_{45},$
$\Delta_{456} = m_{34} + m_{56} + p.$

$\Delta_{512} = m_{12} + m_{45} + m_{56} + q + r,$
$\Delta_{513} = m_{45} + m_{56} + q,$
$\Delta_{514} = m_{56},$
$\Delta_{516} = m_{16} + m_{45} + r,$
$\Delta_{523} = m_{23} + m_{45} + m_{56} + p + q,$
$\Delta_{524} = m_{56} + p,$
$\Delta_{526} = m_{45} + r,$
$\Delta_{534} = m_{34} + m_{56} + p,$
$\Delta_{536} = m_{45},$
$\Delta_{546} = 0.$

$\Delta_{612} = m_{12} + m_{56} + q,$
$\Delta_{613} = m_{56} + q,$
$\Delta_{614} = m_{56},$
$\Delta_{615} = 0,$
$\Delta_{623} = m_{16} + m_{23} + m_{56} + p + q,$
$\Delta_{624} = m_{16} + m_{56} + p,$
$\Delta_{625} = m_{16},$
$\Delta_{634} = m_{16} + m_{34} + m_{56} + p + r,$
$\Delta_{635} = m_{16} + r,$
$\Delta_{645} = m_{16} + m_{45} + r.$

Type SST distances can be parameterized as following:

$$\begin{aligned}
d_{12} &= m_{16} + m_{23} + p + r,\\
d_{23} &= m_{12} + m_{34} + r,\\
d_{34} &= m_{23} + m_{45} + q + r,\\
d_{45} &= m_{34} + m_{56} + p + r,\\
d_{56} &= m_{16} + m_{45} + r,\\
d_{16} &= m_{12} + m_{56} + q + r,\\
d_{14} &= m_{12} + m_{16} + m_{34} + m_{45} + p + q + 3r,\\
d_{25} &= m_{12} + m_{23} + m_{45} + m_{56} + p + q + r,\\
d_{36} &= m_{16} + m_{23} + m_{34} + m_{56} + p + q + r.
\end{aligned}$$

For this type Gromov products can be parameterized as following:

$$\Delta_{123} = m_{16} + m_{23} + p + r,$$
$$\Delta_{124} = m_{16} + p + r,$$
$$\Delta_{125} = m_{16} + r,$$
$$\Delta_{126} = 0,$$
$$\Delta_{134} = m_{12} + m_{16} + m_{34} + p + 2r,$$
$$\Delta_{135} = m_{12} + m_{16} + r,$$
$$\Delta_{136} = m_{12} + r,$$
$$\Delta_{145} = m_{12} + m_{16} + m_{45} + q + 2r,$$
$$\Delta_{146} = m_{12} + q,$$
$$\Delta_{156} = m_{12} + m_{56} + q + r.$$

$$\Delta_{213} = 0,$$
$$\Delta_{214} = m_{23},$$
$$\Delta_{215} = m_{23} + p,$$
$$\Delta_{216} = m_{16} + m_{23} + p + r,$$
$$\Delta_{234} = m_{12} + m_{34} + r,$$
$$\Delta_{235} = m_{12},$$
$$\Delta_{236} = m_{12} + r,$$
$$\Delta_{245} = m_{12} + m_{23} + m_{45} + q + r,$$
$$\Delta_{246} = m_{12} + m_{23} + q + r,$$
$$\Delta_{256} = m_{12} + m_{23} + m_{56} + p + q + r.$$

$$\Delta_{312} = m_{12} + m_{34} + r,$$
$$\Delta_{314} = m_{23},$$
$$\Delta_{315} = m_{23} + m_{34} + p + r,$$
$$\Delta_{316} = m_{16} + m_{23} + m_{34} + p + r,$$
$$\Delta_{324} = 0,$$
$$\Delta_{325} = m_{34} + r,$$
$$\Delta_{326} = m_{34},$$
$$\Delta_{345} = m_{23} + m_{45} + q + r,$$
$$\Delta_{346} = m_{23} + q,$$
$$\Delta_{356} = m_{23} + m_{34} + m_{56} + p + q + r.$$

$$\Delta_{412} = m_{12} + m_{34} + m_{45} + q + 2r,$$
$$\Delta_{413} = m_{45} + q + r,$$
$$\Delta_{415} = m_{34} + p + r,$$
$$\Delta_{416} = m_{16} + m_{34} + m_{45} + p + 2r,$$
$$\Delta_{423} = m_{23} + m_{45} + q + r,$$
$$\Delta_{425} = m_{34} + r,$$
$$\Delta_{426} = m_{34} + m_{45} + r,$$
$$\Delta_{435} = 0,$$
$$\Delta_{436} = m_{45} + r,$$
$$\Delta_{456} = m_{34} + m_{56} + p + r.$$

$$\Delta_{512} = m_{12} + m_{45} + m_{56} + q + r,$$
$$\Delta_{513} = m_{45} + m_{56} + q + r,$$
$$\Delta_{514} = m_{56},$$
$$\Delta_{516} = m_{16} + m_{45} + r,$$
$$\Delta_{523} = m_{23} + m_{45} + m_{56} + p + q + r,$$
$$\Delta_{524} = m_{56} + p,$$
$$\Delta_{526} = m_{45},$$
$$\Delta_{534} = m_{34} + m_{56} + p + r,$$
$$\Delta_{536} = m_{45} + r,$$
$$\Delta_{546} = 0.$$

$$\Delta_{612} = m_{12} + m_{56} + q + r,$$
$$\Delta_{613} = m_{56} + q,$$
$$\Delta_{614} = m_{56},$$
$$\Delta_{615} = 0,$$
$$\Delta_{623} = m_{16} + m_{23} + m_{56} + p + q + r,$$
$$\Delta_{624} = m_{16} + m_{56} + p + r,$$
$$\Delta_{625} = m_{16} + r,$$
$$\Delta_{634} = m_{16} + m_{34} + m_{56} + p + r,$$
$$\Delta_{635} = m_{16},$$
$$\Delta_{645} = m_{16} + m_{45} + r.$$

Type STT distances can be parameterized as following:

$$\begin{aligned}
d_{12} &= m_{16} + m_{23} + p + 2q + r,\\
d_{23} &= m_{12} + m_{34} + q + r,\\
d_{34} &= m_{23} + m_{45} + q + r,\\
d_{45} &= m_{34} + m_{56} + p + 2q + r,\\
d_{56} &= m_{16} + m_{45} + q + r,\\
d_{16} &= m_{12} + m_{56} + q + r,\\
d_{14} &= m_{12} + m_{16} + m_{34} + m_{45} + p + 2q + 3r,\\
d_{25} &= m_{12} + m_{23} + m_{45} + m_{56} + p + 2q + r,\\
d_{36} &= m_{16} + m_{23} + m_{34} + m_{56} + p + 2q + r.
\end{aligned}$$

For this type Gromov products can be parameterized as following:

$\Delta_{123} = m_{16} + m_{23} + p + 2q + r,$
$\Delta_{124} = m_{16} + p + q + r,$
$\Delta_{125} = m_{16} + q + r,$
$\Delta_{126} = 0,$
$\Delta_{134} = m_{12} + m_{16} + m_{34} + p + 2q + 2r,$
$\Delta_{135} = m_{12} + m_{16} + q + r,$
$\Delta_{136} = m_{12} + q + r,$
$\Delta_{145} = m_{12} + m_{16} + m_{45} + q + 2r,$
$\Delta_{146} = m_{12} + r,$
$\Delta_{156} = m_{12} + m_{56} + q + r.$

$\Delta_{213} = 0,$
$\Delta_{214} = m_{23} + q,$
$\Delta_{215} = m_{23} + p + q,$
$\Delta_{216} = m_{16} + m_{23} + p + 2q + r,$
$\Delta_{234} = m_{12} + m_{34} + q + r,$
$\Delta_{235} = m_{12},$
$\Delta_{236} = m_{12} + q + r,$
$\Delta_{245} = m_{12} + m_{23} + m_{45} + q + r,$
$\Delta_{246} = m_{12} + m_{23} + q + r,$
$\Delta_{256} = m_{12} + m_{23} + m_{56} + p + 2q + r.$

$\Delta_{312} = m_{12} + m_{34} + q + r,$
$\Delta_{314} = m_{23} + q,$
$\Delta_{315} = m_{23} + m_{34} + p + 2q + r,$
$\Delta_{316} = m_{16} + m_{23} + m_{34} + p + 2q + r,$
$\Delta_{324} = 0,$
$\Delta_{325} = m_{34} + q + r,$
$\Delta_{326} = m_{34},$
$\Delta_{345} = m_{23} + m_{45} + q + r,$
$\Delta_{346} = m_{23},$
$\Delta_{356} = m_{23} + m_{34} + m_{56} + p + 2q + r.$

$\Delta_{412} = m_{12} + m_{34} + m_{45} + q + 2r,$
$\Delta_{413} = m_{45} + r,$
$\Delta_{415} = m_{34} + p + q + r,$
$\Delta_{416} = m_{16} + m_{34} + m_{45} + p + 2q + 2r,$
$\Delta_{423} = m_{23} + m_{45} + q + r,$
$\Delta_{425} = m_{34} + q + r,$
$\Delta_{426} = m_{34} + m_{45} + q + r,$
$\Delta_{435} = 0,$
$\Delta_{436} = m_{45} + q + r,$
$\Delta_{456} = m_{34} + m_{56} + p + 2q + r.$

$\Delta_{512} = m_{12} + m_{45} + m_{56} + q + r,$
$\Delta_{513} = m_{45} + m_{56} + q + r,$
$\Delta_{514} = m_{56} + q,$
$\Delta_{516} = m_{16} + m_{45} + q + r,$
$\Delta_{523} = m_{23} + m_{45} + m_{56} + p + 2q + r,$
$\Delta_{524} = m_{56} + p + q,$
$\Delta_{526} = m_{45},$
$\Delta_{534} = m_{34} + m_{56} + p + 2q + r,$
$\Delta_{536} = m_{45} + q + r,$
$\Delta_{546} = 0.$

$\Delta_{612} = m_{12} + m_{56} + q + r,$
$\Delta_{613} = m_{56},$
$\Delta_{614} = m_{56} + q,$
$\Delta_{615} = 0,$
$\Delta_{623} = m_{16} + m_{23} + m_{56} + p + 2q + r,$
$\Delta_{624} = m_{16} + m_{56} + p + 2q + r,$
$\Delta_{625} = m_{16} + q + r,$
$\Delta_{634} = m_{16} + m_{34} + m_{56} + p + 2q + r,$
$\Delta_{635} = m_{16},$
$\Delta_{645} = m_{16} + m_{45} + q + r.$

Type TTT distances can be parameterized as following:

$$
\begin{aligned}
d_{12} &= m_{16} + m_{23} + p + 2q + r, \\
d_{23} &= m_{12} + m_{34} + p + q + r, \\
d_{34} &= m_{23} + m_{45} + p + q + r, \\
d_{45} &= m_{34} + m_{56} + p + 2q + r, \\
d_{56} &= m_{16} + m_{45} + p + q + r, \\
d_{16} &= m_{12} + m_{56} + p + q + r, \\
d_{14} &= m_{12} + m_{16} + m_{34} + m_{45} + p + 2q + 3r, \\
d_{25} &= m_{12} + m_{23} + m_{45} + m_{56} + p + 2q + r, \\
d_{36} &= m_{16} + m_{23} + m_{34} + m_{56} + p + 2q + r.
\end{aligned}
$$

For this type Gromov products can be parameterized as following:

$\Delta_{123} = m_{16} + m_{23} + p + 2q + r,$
$\Delta_{124} = m_{16} + q + r,$
$\Delta_{125} = m_{16} + p + q + r,$
$\Delta_{126} = 0,$
$\Delta_{134} = m_{12} + m_{16} + m_{34} + p + 2q + 2r,$
$\Delta_{135} = m_{12} + m_{16} + p + q + r,$
$\Delta_{136} = m_{12} + p + q + r,$
$\Delta_{145} = m_{12} + m_{16} + m_{45} + p + q + 2r,$
$\Delta_{146} = m_{12} + r,$
$\Delta_{156} = m_{12} + m_{56} + p + q + r.$

$\Delta_{213} = 0,$
$\Delta_{214} = m_{23} + p + q,$
$\Delta_{215} = m_{23} + q,$
$\Delta_{216} = m_{16} + m_{23} + p + 2q + r,$
$\Delta_{234} = m_{12} + m_{34} + p + q + r,$
$\Delta_{235} = m_{12},$
$\Delta_{236} = m_{12} + p + q + r,$
$\Delta_{245} = m_{12} + m_{23} + m_{45} + p + q + r,$
$\Delta_{246} = m_{12} + m_{23} + p + q + r,$
$\Delta_{256} = m_{12} + m_{23} + m_{56} + p + 2q + r.$

$\Delta_{312} = m_{12} + m_{34} + p + q + r,$
$\Delta_{314} = m_{23} + p + q,$
$\Delta_{315} = m_{23} + m_{34} + p + 2q + r,$
$\Delta_{316} = m_{16} + m_{23} + m_{34} + p + 2q + r,$
$\Delta_{324} = 0,$
$\Delta_{325} = m_{34} + p + q + r,$
$\Delta_{326} = m_{34},$
$\Delta_{345} = m_{23} + m_{45} + p + q + r,$
$\Delta_{346} = m_{23},$
$\Delta_{356} = m_{23} + m_{34} + m_{56} + p + 2q + r.$

$\Delta_{412} = m_{12} + m_{34} + m_{45} + p + q + 2r,$
$\Delta_{413} = m_{45} + r,$
$\Delta_{415} = m_{34} + q + r,$
$\Delta_{416} = m_{16} + m_{34} + m_{45} + p + 2q + 2r,$
$\Delta_{423} = m_{23} + m_{45} + p + q + r,$
$\Delta_{425} = m_{34} + p + q + r,$
$\Delta_{426} = m_{34} + m_{45} + p + q + r,$
$\Delta_{435} = 0,$
$\Delta_{436} = m_{45} + p + q + r,$
$\Delta_{456} = m_{34} + m_{56} + p + 2q + r.$

$\Delta_{512} = m_{12} + m_{45} + m_{56} + p + q + r,$
$\Delta_{513} = m_{45} + m_{56} + p + q + r,$
$\Delta_{514} = m_{56} + p + q,$
$\Delta_{516} = m_{16} + m_{45} + p + q + r,$
$\Delta_{523} = m_{23} + m_{45} + m_{56} + p + 2q + r,$
$\Delta_{524} = m_{56} + q,$
$\Delta_{526} = m_{45},$
$\Delta_{534} = m_{34} + m_{56} + p + 2q + r,$
$\Delta_{536} = m_{45} + p + q + r,$
$\Delta_{546} = 0.$

$\Delta_{612} = m_{12} + m_{56} + p + q + r,$
$\Delta_{613} = m_{56},$
$\Delta_{614} = m_{56} + p + q,$
$\Delta_{615} = 0,$
$\Delta_{623} = m_{16} + m_{23} + m_{56} + p + 2q + r,$
$\Delta_{624} = m_{16} + m_{56} + p + 2q + r,$
$\Delta_{625} = m_{16} + p + q + r,$
$\Delta_{634} = m_{16} + m_{34} + m_{56} + p + 2q + r,$
$\Delta_{635} = m_{16},$
$\Delta_{645} = m_{16} + m_{45} + p + q + r.$

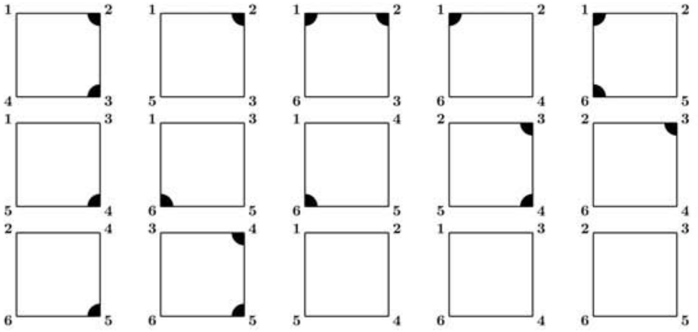
6 Q structure for I17SSS Q 113

Fixed quadrangles:

$(13, 24), (13, 25), (13, 26), (14, 26), (15, 26), (14, 35), (15, 36), (15, 46), (24, 35), (24, 36), (25, 46), (35, 46)$

Free quadrangles:

$(14, 25), (14, 36), (25, 36)$



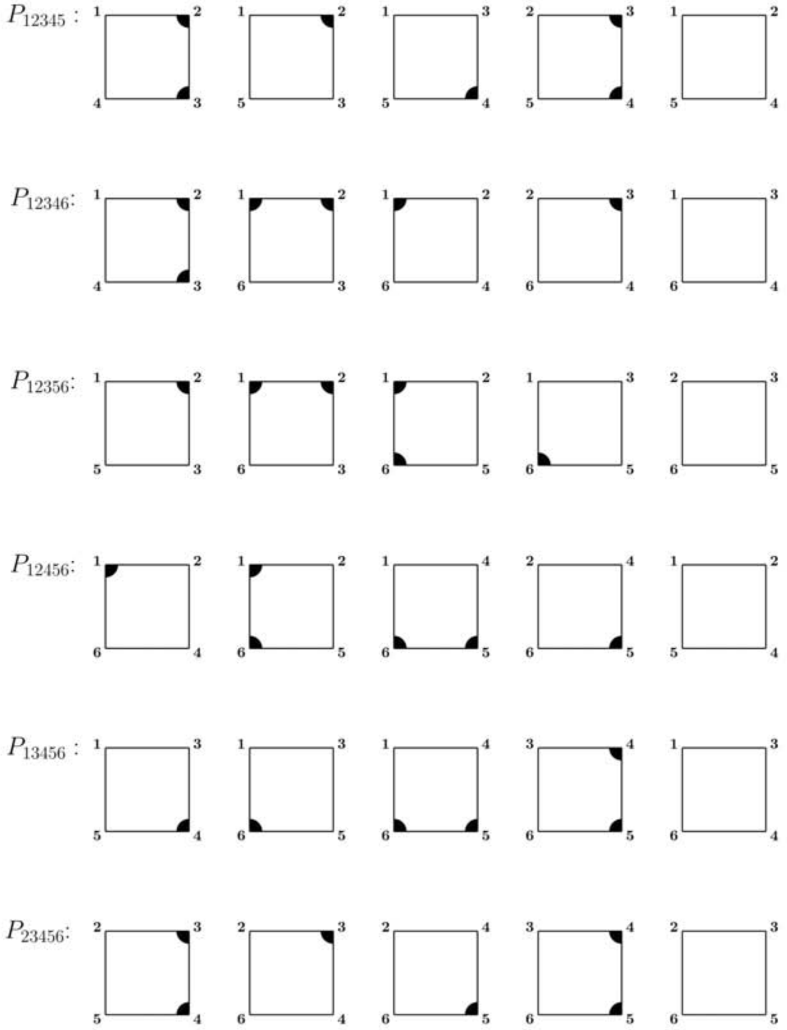


Figure 18: Quadrangle structure and Pentacles of Sturmfels-Yu metric 66 ($I_{17}SSS$).

Structures of 5-point subsets: As an example we work with type I17 SSS, Q type 113.

113	P_{12345}	$(13, 24)$	$(13, 25)$	$(14, 35)$	$(24, 35)$	$(14, 25)$
113	P_{12346}	$(13, 24)$	$(13, 26)$	$(14, 26)$	$(24, 36)$	$(13, 46)$
113	P_{12356}	$(13, 25)$	$(13, 26)$	$(15, 26)$	$(15, 36)$	$(25, 36)$
113	P_{12456}	$(14, 26)$	$(15, 26)$	$(15, 46)$	$(25, 46)$	$(14, 25)$
113	P_{13456}	$(14, 35)$	$(15, 36)$	$(15, 46)$	$(35, 46)$	$(13, 46)$
113	P_{23456}	$(24, 35)$	$(24, 36)$	$(25, 46)$	$(35, 46)$	$(25, 36)$

References

[1] Bilge, A.H., D. Çelik and Ş. Koçak (2015). Optimal Embeddings of Finite Metric Spaces into Graphs, Anadolu University Journal of Science and Technology B- Theoretical Science, 3 (2), 133 - 147.

[2] Bilge, A.H., D. Çelik and Ş. Koçak (2017). An equivalence class decomposition of finite metric spaces via Gromov products. Discrete Mathematics, Vol. 340 1928-1932.

[3] Incegul, M. and A.H. Bilge. Algorithms for the equivalence class decomposition classification of finite metric spaces via Gromov products".

[4] Bilge, A.H. and M. Incegul, A matrix representation for the Gromov product decomposition of finite metric spaces.

[5] Bilge, A.H. and M. Incegul, Gromov product decomposition of 7-point metric spaces.

[6] Jussi, V., (2004). "Gromov hyperbolic spaces" http://www.helsinki.fi/ jvaisala/grobok.pdf

[7] Kölbig, K. S. and Schwarz, F., (1979). A Program for Solving Systems of Homogeneous Linear Inequalities, Comput. Phys. Commun. 17, 375-382.

[8] Dines, L.L., (1927). On Positive Solutions of a System of Linear Equations, Annals of Mathematics, 28, 386-392.

[9] Dress, A., (1984). Trees, Tight Extensions of Metric Spaces and the Cohomological Dimension of Certain Groups: A Note on Combinatorial Properties of Metric Spaces Advances in Mathematics, 53, 321-402.

[10] Koolen, J., A. Lesser, V. Moulton (2009). Optimal Realizations of Generic Five-Point Metrics, European J. of Combin. 30, 1164-1171.

[11] Lesser, A. (2007). Optimal and Hereditarily Optimal Realizations of Metric Spaces, PhD thesis, Uppsala University.

[12] Imrich, W., J. Simoes-Pereira, C.M. Zamfirescu (1984). On Optimal Embeddings of Metrics in Graphs J. of Combin. Th. Ser.B 36, 1-15.

[13] Sturmfels, B., J. Yu (2004). Classification of Six-Point Metrics, Electron. J. of Combin. 11, R44.

Gromov Product Decomposition of 7-point Metric Spaces

Ayse Humeyra Bilge[*], **Metehan Incegul**[†]

May 9, 2021

Abstract

Let X be a finite metric space with elements P_i, $i = 1, \ldots, n$ and with the distance functions d_{ij}. The Gromov product of the triangle with vertices P_i, P_j and P_k at the vertex P_i is defined by $\Delta_{ijk} = \frac{1}{2}(d_{ij} + d_{ik} - d_{jk})$. A metric space is called Δ-generic, if the set of Gromov products at each P_i has a unique smallest element Δ_{ijk}. For a Δ-generic metric space, the map $P_i \rightarrow (P_j Pk)$, where $(P - jP_k)$ is the edge joining Pj to P_k is a well defined map called the "Gromov product structure" [Bilge, Celik and Kocak, "An equivalence class decomposition of finite metric spaces", *Discrete Mathmetics*, Vol 340, (2017) 1928-1932]. For $n = 5$, the 3 Δ-equivalence classes coincide with the classification of 5-point metrics. For $n = 6$, there are 26 Δ-equivalence classes refined by 339 metric classes. In this paper, we obtain the Δ-equivalence decomposition of 7-point metric spaces that consist of 431 equivalence classes.

Keywords: Finite metric spaces, Gromov product structure, weighted graphs

1 Introduction

An n-point metric space is described by the set of distances d_{ij} subject to triangle inequalities. The structure of a finite metric is described by the so-called "metric fan" and the metrics belonging to the interior of a metric fan are called "generic" [11]." In a series of papers we studied metric spaces in terms of the "Gromov products" as defined by Eqn.(2.1) [1],[2],[3],[5],[4]. In this approach, distance functions are expressed in terms of Gromov products and triangle inequalities are replaced by the compatibility conditions of the defining equations for the d_{ij}'s as given by (2.2). The solution of these equations turns out to be quite algorithmic and it is useful in obtaining split decomposition of the metric [3].

As it is well known [9], there are 1 and 3 inequivalent metrics on 4-point and 5-point spaces, respectively and the Δ-equivalence decomposition coincides with this classification. For $n = 6$, there are 339 inequivalent metrics as classified in [11]. We presented the Δ-equivalence classes of 6-point metrics in [2],[4] where we have shown that there are 26 equivalence classes.

In the present work, we study the Δ-equivalence decomposition of 7-point spaces by using the algorithm of [4] that consists of 4 steps. One first enumerates all Gromov product structures, then eliminates the ones that are not allowable, using Proposition 1 and Corollary 1. The third step that consists of the application of the full permutation group to allowable Gromov products is the most costly part of the algorithm. Finally at the fourth step, non-generic Gromov product structures are eliminated.

For 6-point spaces this algorithm runs on a standard computer in a reasonable time without the need for more sophisticated methods. For $n = 7$, the third step of the algorithm, consisting of applying the full

---

[*] Faculty of Engineering and Natural Sciences, Department of Industrial Engineering, Kadir Has University, Istanbul, Turkey(ayse.bilge@khas.edu.tr)

[†] Istanbul Bilgi University, Faculty of Engineering and Natural Sciences, Kazim Karabekir Cad. No: 2/13 Eyup 34060 Istanbul / Turkey

permutation group on 7 points to all allowable Gromov product structures turned out to be practically impossible to run on a standard computer. In the search of permutation invariants of Gromov product structures we ended up with a matrix representation, as presented in [**?**].

The matrix representation of a Gromov product structure is defined in Section 3 and the invariants of this matrix are used to split allowable Gromov product structures into smaller subsets on which the full permutation group is applied in a reasonable time.

As a result we obtain 431 Gromov product equivalence classes for 7-point metrics.

2 Preliminaries

Let X be a finite metric space with n elements P_i, $i = 1, \dots, n$ and let d_{ij} be the distance between P_i and P_j. We denote the triangle with vertices P_i, P_j and P_k by $(P_i P_j P_k)$. The quantity Δ_{ijk} defined by

$$\Delta_{ijk} = \Delta_{ikj} = \tfrac{1}{2}(d_{ij} + d_{ik} - d_{jk}) \tag{2.1}$$

is called the "Gromov product" of the triangle $(P_i P_j P_k)$ at the vertex P_i.

The triangle inequalities are equivalent to the non-negativity of the Δ_{ijk}'s. The solution set of the triangle inequalities is a convex polyhedral cone in $R^{n(n-1)/2}$ and the classification of metric spaces is given in terms of the combinatorial properties of this cone [11].

For any triangle $(P_i P_j P_k)$, the distances can be expressed in terms of the Gromov products as

$$d_{ij} = \Delta_{ijk} + \Delta_{jik}, \quad k = 1, \dots n, \quad k \neq i, j. \tag{2.2}$$

For an n-point metric space, there are $n(n-1)(n-2)/2$ Gromov products that has to satisfy $n-3$ equalities for each unordered pair (i,j). It follows that an n-point metric space is characterized by positive solutions of $n(n-1)(n-3)/2$ linear equations in $n(n-1)(n-2)/2$ unknowns. A finite metric space is called "Δ-generic", if the set of Gromov products at each P_i has a unique smallest element. We define Gromov product structures on Δ-generic spaces, as follows.

Definition 1.*Let X be a Δ-generic n-point metric space with elements P_a, let $E(X)$ be the set of edges of X and let Δ_{a,b_a,c_a} be the minimal Gromov Product at P_a. The Gromov product structure on X is the map $\mathcal{E}$ from X to the set $E(X)$ defined by $\mathcal{E}(P_a) = (P_{b_a} P_{c_a})$, where $(P_a P_b)$ is the edge joining P_a to P_b. Two Δ-generic finite metric spaces are "Δ-equivalent" if they have the same Gromov product structures, up to permutations of indices.*

The Gromov products at different P_i's not independent. Their relations are given by the Proposition 1 below, whose proof is straightforward.

Proposition 1.*Let Δ_{abc} be the Gromov product at node a and i be a node different from a, b and c. Then,*

$$\Delta_{abi} - \Delta_{abc} = \Delta_{cbi} - \Delta_{cai} = \Delta_{iac} - \Delta_{ibc} = \Delta_{bac} - \Delta_{bai}, \tag{2.3a}$$

$$\Delta_{aci} - \Delta_{abc} = \Delta_{bci} - \Delta_{bai} = \Delta_{iab} - \Delta_{ibc} = \Delta_{cab} - \Delta_{cai}. \tag{2.3b}$$

These equalities lead to a convenient algorithmic procedure for determining allowable Gromov product structures.

Corollary 1.*If Δ_{abc} is the minimal Gromov Product at P_a then, Δ_{cbi} and Δ_{cab} cannot be minimal at P_c, Δ_{bci} and Δ_{bac} cannot be minimal at P_b and Δ_{iac} and Δ_{iab} cannot be minimal at P_i.*

Thus, the minimality of Δ_{abc} leads to the exclusion of Gromov product structures containing Δ_{bci}, Δ_{cbi}, Δ_{iab}, Δ_{iac} from the set of allowable Gromov product structures. These elimination rules lead to an algorithmic method for obtaining Gromov Product equivalence classes. We start by all possible Gromov Product index sets, apply the elimination rules and collect together the ones that are mapped to each other under permutations. This process works well for $n = 5$ and $n = 6$ but it becomes practically impossible for $n = 7$. For $n \geq 7$ we will eliminate beforehand a large number of equivalent cases. This will be done by using "excess cycle lengths", to be defined below.

Definition 2. *If the Gromov products at nodes* $\{a_1, a_2, \ldots, a_k\}$ *are*

$$\{\Delta_{a_1 a_i a_2}, \Delta_{a_2 a_1 a_3}, \ldots, \Delta_{a_k, a_{k-1}, a_j}\},$$

then we say that the Gromov Product structure

$$\{(a_1 a_i a_2), (a_2 a_1 a_3), \ldots, (a_k, a_{k-1}, a_j)\}$$

contains a Δ*-chain of length* k*. If* $k = 1$ *then the chain has length* 1 *and* P_{a_1} *is called an 'isolated" point. For* $k > 1$ *The points* P_{a_1} *and* P_{a_k} *are "end points" of the chain while the points* P_{a_i}*,* $i \neq 1, k$ *are "interior points" of the chain. For* $k > 1$*, if* $i = k$ *and* $j = 1$*, then we say that there is a* Δ*-cycle of length* k*. A chain of length* 2 *has no interior points. A cycle has no end points.*

The cycle and chain structure allows to choose Gromov product structures in certain canonical forms and reduces considerably the burden of finding equivalence classes.

3 Matrix representation of Gromov product structures

An n-point metric space is represented by a weighted complete graph. If Δ_{aij} is the minimal Gromov product as P_a, then the edges joining P_a to the other P_b's can be sticked together by an amount Δ_{aij} and the edge joining P_i to P_j can be removed [1]. Repeating this procedure at each point we obtain the so-called "pendant-free" reduction of the complete graph.

In this section we define the matrix representation of a Gromov product structure as defined in [5], discuss its relation with the graph of pendant free reductions and we give the matrix representations for $n = 4$ and $n = 5$.

Definition 3. *Let* $S = \{\Delta_{1a_1b_1}, \Delta_{2a_2b_2}, \ldots, \Delta_{na_nb_n}\}$ *be a Gromov product structure for a finite metric space with* n *elements. The "matrix representation for* S*" is the matrix* G_S*, defined by* $G_S(i,j) = 1$*, if the* $j = a_i$ *or* $j = b_i$*, in* $\Delta_{ia_ib_i}$*, and* $G_S(i,j) = 0$ *otherwise.*

Note that G_S consists of zeros and ones only and it has exactly two 1's in each row. It follows that $\lambda = 2$ is always an eigenvalue with corresponding eigenvectors multiples of $X = (1, 1, \ldots, 1)^t$. We note that similarity of the matrix representations don't imply equivalence of Gromov product structures, because the change of basis matrix need not be a permutation matrix. Nevertheless the (integer valued) quantities traceG_S^k will be useful to discriminate between inequivalent Gromov product structures.

4-point spaces. For $n = 4$, there is unique Gromov product structure that can be given as

$$X^{(4)}: \quad \{\Delta_{124}, \Delta_{213}, \Delta_{324}, \Delta_{413}\}$$

and its matrix is

$$G = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{pmatrix}$$

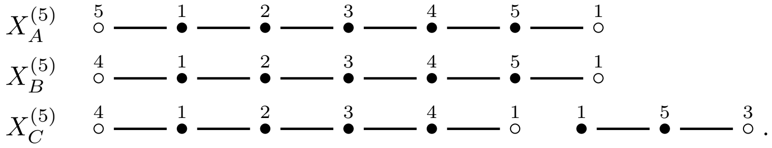
5-point spaces. For $n = 5$, there are 3 Gromov product structures corresponding to the metric classes

$$\begin{array}{ll} X_A^{(5)} & \{\Delta_{125}, \Delta_{213}, \Delta_{324}, \Delta_{435}, \Delta_{514}\}, \\ X_B^{(5)} & \{\Delta_{124}, \Delta_{213}, \Delta_{324}, \Delta_{435}, \Delta_{524}\}, \\ X_C^{(5)} & \{\Delta_{124}, \Delta_{213}, \Delta_{324}, \Delta_{413}, \Delta_{513}\}. \end{array}$$

and their matrix representations are

$$G_A = \begin{pmatrix} 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 \end{pmatrix}, \quad G_B = \begin{pmatrix} 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \end{pmatrix}, \qquad G_C = \begin{pmatrix} 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \end{pmatrix}.$$

Their Δ-chain and Δ-cycle structures are given by



We will now present certain results on the relation of matrix invariants to the properties of pendant free reductions.

An $n \times n$ matrix A is irreducible, if $\sum_{i=0}^{n} A^i$ has no zero element.

Definition 4. *A Gromov product structure is called "irreducible" if the corresponding matrix is irreducible. Otherwise it is called "reducible".*

If a metric space is reducible in the sense above, then, a proper subset is itself a metric spaces. Hence the parametrization of this subspace can be enlarged to the parametrization of the larger space.

Proposition 2. *Let G_S be the matrix representation of a Gromov product structure S, as defined in Definition 3. Then*

(i) The rank of G_S is the number of edges removed in the pendant free reduction.

(ii)If G_S is reducible, then the number of columns with no zero element in $\sum_{i=0}^{n} G_S^i$ gives the dimension of the largest irreducible subspace.

(iii) The symmetric part of the matrix G_S, expressed as $H_S = \text{floor} = \frac{1}{2}(G_S + G_S^t)$ gives information of the chain structure. The sum of the entries of the columns are 0, 1 or 2. P_i is respectively an "isolated point", an "end point" or an "interior point" according as the sum of the entries of the ith column of H_S is 0, 1 or 2.

4 Gromov Product Classification for $n = 6$

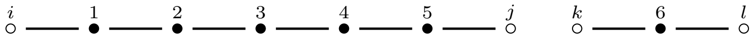
In this section we reiterate the Δ-equivalence decomposition of 6-point spaces in order to display the advantage of using chain structures. The structure of Δ-chains and and Δ-cycles (Definition 2) on an n-point space can be conveniently analysed by the "partition of integers" for n. The partition of integers for $n = 6$ is

$$\begin{array}{llllll} 6+0 & 5+1 & 4+1+1 & 3+1+1+1 & 2+1+1+1+1 & 1\times 6 \\ & 4+2 & 3+2+1 & 2+2+1+1 & & \\ & 3+3 & 2+2+2 & & & \end{array}$$

If there is a 6-cycle of 6-chain, without loss of generality, the chain structure can be chosen as

Here if $i = 6$ and $j = 1$ there is a 6-cycle. Otherwise there is a 6-chain. If there is a cycle all points are interior points. In the case of a chain, there are 2 end points and 4 interior points.

If there is a 5-cycle or a 5-chain, the chain structure is



Here if there is a 5-cycle, we have 5 interior points and 1 isolated point. If there is a 5-chain, there are 2 interior points, 2 end points and 1 isolated point.

For $n = 6$, not all partitions of the integer 6 leads to allowable types. The list of allowable types for $n = 6$ are below. The numbers in the last column gives the number of isolated points, end points and interior points respectively. Note that $4+2$ (Chain) and $3+3$ types are not distinguished by these invariants.

$6+0$(cycle)	○—●—●—●—●—●—●—○	(0,0,6)
$6+0$(chain)	○—●—●—●—●—●—●—○	(0,2,4)
$5+1$(cycle)	○—●—●—●—●—●—○ ○—●—○	(1,0,5)
$5+1$(chain)	○—●—●—●—●—●—○ ○—●—○	(1,2,3)
$4+2$(cycle)	○—●—●—●—●—○ ○—●—●—○	(0,2,4)
$4+2$(chain)	○—●—●—●—●—○ ○—●—●—○	(0,4,2)
$3+3$	○—●—●—●—○ ○—●—●—●—○	(0,4,2)
$4+1+1$(cycle)	○—●—●—●—●—○ ○—●—○ ○—●—○	(2,0,4)
$4+1+1$(chain)	○—●—●—●—●—○ ○—●—○ ○—●—○	(2,2,2)
$3+2+1$	○—●—●—●—●—○ ○—●—●—○ ○—●—○	(1,4,1)
$2+2+2$	○—●—●—○ ○—●—●—○ ○—●—●—○	(0,6,0)
$3+1\times 3$	○—●—●—●—○ ○—●—○ ○—●—○ ○—●—○	(3,1,2)
$2+2+1+1$	○—●—●—○ ○—●—●—○ ○—●—○ ○—●—○	(2,4,0)
$2+1\times 4$	○—●—●—○ ○—●—○ ○—●—○ ○—●—○ ○—●—○	(4,2,0)

Table 1: Cycle and chain configurations for 6-point spaces

The complete list of 6-point spaces grouped according to their cycle and chain structures is given below.

Type	R/I	Int.	End.	Disc.	Remv.	Gromov Product Structure	Type
6+0 (Cycle)	I	6	0	0	6	126,213,324,435,546,615	(I_{17})
6+0 (Chain)	I	4	2	0	4	124,213,324,435,546,635	I_7
	I				5	124,213,324,435,546,615	I_8
	I				5	124,213,324,435,546,625	I_{14}
	I				5	125,213,324,435,546,625	I_{15}
5+1 (Cycle)	R	5	0	1	5	125,213,324,435,514,613	R_8
5+1 (Chain)	I	3	2	1	5	125,213,324,435,546,613	I_{11}
	I				4	124,213,324,435,546,613	I_{13}
	R				3	124,213,324,435,524,624	R_2
	R				3	124,213,324,435,524,613	R_4
	R				4	124,213,324,435,524,615	R_6
4+2 (Cycle)	R	4	2	0	4	124,213,324,413,516,625	R_7
	R				4	124,213,324,413,516,635	R_9
4+2 (Chain)	I	2	4	0	4	124,213,324,435,526,635	I_5
	I				4	125,213,324,436,536,625	I_6
	I				4	124,213,324,435,516,635	I_9
	I				5	124,213,324,435,516,625	I_{10}
	I				5	124,213,324,435,526,615	I_{12}
4+1+1 (Cycle)	R	4	0	2	2	124,213,324,413,513,613	R_1
	R				2	124,213,324,413,513,624	R_3
	R				3	124,213,324,413,513,625	R_5
4+1+1 (Chain)	I	2	2	2	4	124,213,324,435,516,624	I_3
3+3	I	2	4	0	3	124,213,324,456,524,624	I_1
3+2+1	I	1	4	1	3	124,213,324,456,513,624	I_2
3+1+1+1	I	1	2	3	3	124,213,324,456,513,613	I_4
2+2+2	I	0	6	0	3	156,213,324,456,513,624	I_{16}

Table 2: Complete list of 6-point spaces

5 Gromov Product Chain Classification for $n = 7$

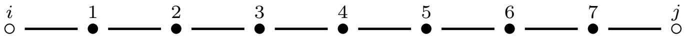
For $n = 7$, the partition of integers lead to the following chain configurations.

The partition of integers for $n = 7$ is

$7+0$	$6+1$	$5+1+1$	$4+1+1+1$	$3+1+1+1+1$	$2+1\times 5$	1×7
	$5+2$	$4+2+1$	$3+2+1+1$	$2+2+1+1+1$		
	$4+3$	$3+3+1$	$2+2+2+1$			

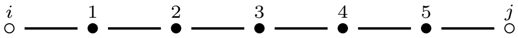
Without loss of generality, we choose the following canonical orderings. Algorithmically it is possible to terminate the chains. But even if we start with canonical forms with shorter chains, it is difficult to avoid the occurrence of longer chains. The grouping of the Gromov products by their matrix invariants help with this problem.

- There is a 7 cycle or a 7 chain:



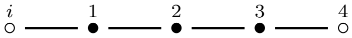
- There is a chain of length 6 which is not extendible to a longer chain :

- There is a chain of length 5 which is not extendible to a longer chain :

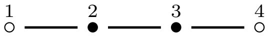


- There is a chain of length 4 which is not extendible to a longer chain :

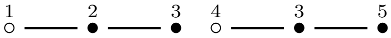
- There is a chain of length 3 which is not extendible to a longer chain :



- There is a chain of length 2 which is not extendible to a longer chain :



- The longest chain of length 2:



The Δ-equivalence decomposition of 7-point spaces is obtained by implementing the algorithms described below with MATLAB. Details of the algorithms are given in. We adapted the codes to $n = 7$. For each chain length, we start with canonical types generate all possible Gromov product structures. Then use Proposition 1 to select allowable ones. At this stage, the sets of allowable structures are too big for applying the full permutation group. Therefore, we use matrix invariants to obtain smaller groups to which permutation group acts in a reasonable time. This way the problem is reduced to a feasible size. Finally we eliminate the structures that are not generic. As a result we obtain 431 types listed in the Appendix. The number of types for each chain length is listed below.

Type	Isolated	End	Internal	Number
7 (Cycle)	0	0	7	1
7 (Chain)	0	2	5	8
6+1	1	2	4	8
5+2	0	4	3	43
4+3	0	4	3	
5+1+1	2	2	3	22
4+2+1	1	4	2	63
3+3+1	1	4	2	
3+2+2	0	6	1	27
4+1+141	3	2	2	27
3+2+1+1	2	4	1	42
2+2+2+1	1	6	0	14
3+1+1+1+1	4	2	1	12
2+2+1+1+1	3	4	0	13
2+1+1+1+1+1	5	2	0	6
1× 7	7	9	0	1

Table 3: Irreducible 7 point spaces

Type	Examples	Number
Contains $X^{(4)}$	124,213,324,413,513,613,713	24
Contains $X^{(5)}$	125,213,324,435,514,613,713	27
Contains $X^{(6)}$	126,213,324,435,546,615,713	93

Table 4: Reducible 7 point spaces

References

[1] Bilge, A. H., D. Celik and S. Kocak, Optimal Embeddings of Finite Metric Spaces into Graphs, Anadolu University, Journal of Science and Technology B-Theoretical Science, 3(2)(2015) 133-147.

[2] Bilge, A. H., D. Celik and S. Kocak, An equivalence class decomposition of finite metric spaces via Gromov products, Discrete Mathematics, Vol. 340, (2017), 1928-1932.

[3] Bilge, A. H., D. Celik and S. Kocak, The quadrangle structure and its applicaiton too the split decomposition of finite metric spaces. Submitted to :Discrete Mathematics.

[4] Algorithm paper

[5] Malta makalesi

[6] Blumenthal, L.M., Theory and Applications of Distance Geometry,Chelsea Pub.Comp., New York, 1970.

[7] Dress, A., Trees, Tight Extensions of Metric Spaces and the Cohomological Dimension of Certain Groups: A Note on Combinatorial Properties of Metric Spaces, Advances in Mathematics 53, 321-402, 1984.

[8] Imrich, W., J. Simoes-Pereira, C.M.Zamfirescu, On Optimal Embeddings of Metrics in Graphs, J. of Combin. Th. Ser.B 36, 1-15, 1984.

[9] Koolen, J., A. Lesser, V. Moulton, Optimal realizations of generic five-point metrics, European J. of Combin., 30, 1164-1171, 2009.

[10] Matousek, J., Thirty-three miniatures (mathematical and algorithmic applications of linear algebra), AMS, 2010.

[11] Sturmfels, B., J. Yu, Classification of six-point metrics, Electron. J. of Combin., 11, 2004, R44.

[12] Varone, S.C., A constructive algorithm for realizing a distance matrix, European J. of Oper. Research, 174, 102-111, 2006.

127	213	324	435	546	657	716
124	213	324	435	546	657	716
125	213	324	435	546	657	716
124	213	324	435	546	657	736
124	213	324	435	546	657	726
126	213	324	435	546	657	726
125	213	324	435	546	657	726
125	213	324	435	546	657	736
124	213	324	435	546	657	746

Table 5: 7-Cycle and 7-Chains

124	213	324	435	546	657	713
124	213	324	435	546	657	724
125	213	324	435	546	657	713
125	213	324	435	546	657	724
125	213	324	435	546	657	714
126	213	324	435	546	657	713
126	213	324	435	546	657	714
126	213	324	435	546	657	724

Table 6: 6+1 Decomposition

124	213	324	435	546	627	716
124	213	324	435	546	637	716
124	213	324	435	546	637	726
124	213	324	435	547	627	716
124	213	324	435	547	637	716
124	213	324	435	547	637	726
124	213	324	435	547	647	716
124	213	324	435	547	647	726
124	213	324	435	547	647	736
125	213	324	435	546	627	716
125	213	324	435	546	637	716
125	213	324	435	546	637	726
125	213	324	435	547	627	716
125	213	324	435	547	637	716
125	213	324	435	547	637	726
125	213	324	435	547	647	716
125	213	324	435	547	647	726
125	213	324	435	547	647	736
126	213	324	435	546	637	726
126	213	324	435	547	637	726
126	213	324	435	547	647	726

Table 7: 5+2 Decomposition

124	213	324	435	567	615	715
124	213	324	435	567	625	715
124	213	324	435	567	625	725
124	213	324	435	567	635	715
124	213	324	435	567	635	725
124	213	324	435	567	635	735
124	213	324	436	567	615	715
124	213	324	436	567	625	715
124	213	324	436	567	625	725
124	213	324	437	567	625	715
124	213	324	437	567	635	715
124	213	324	437	567	635	725
124	213	324	437	567	645	715
124	213	324	437	567	645	725
125	213	324	435	567	625	725
125	213	324	435	567	635	725
125	213	324	436	536	657	716
125	213	324	436	567	625	725
125	213	324	437	536	657	716
125	213	324	437	536	657	726
125	213	324	437	546	657	716
125	213	324	437	567	635	725

Table 8: 4+3 Decomposition

124	213	324	435	546	627	713
124	213	324	435	546	627	715
124	213	324	435	546	637	715
124	213	324	435	546	637	724
124	213	324	435	546	637	725
124	213	324	435	547	624	716
124	213	324	435	547	625	716
124	213	324	435	547	635	716
124	213	324	435	547	635	726
125	213	324	435	546	627	713
125	213	324	435	546	637	714
125	213	324	435	546	637	724
125	213	324	435	546	637	725
125	213	324	435	547	624	716
125	213	324	435	547	625	716
125	213	324	435	547	635	716
125	213	324	435	547	635	726
126	213	324	435	546	637	714
126	213	324	435	546	637	715
126	213	324	435	546	637	724
126	213	324	435	547	635	713
126	213	324	435	547	647	713

Table 9: 5+1+1 Decomposition

124	213	324	435	567	615	713
124	213	324	435	567	624	715
124	213	324	435	567	625	713
124	213	324	435	567	625	724
124	213	324	435	567	635	713
124	213	324	435	567	635	724
124	213	324	436	527	627	715
124	213	324	436	537	625	715
124	213	324	436	537	627	715
124	213	324	436	547	625	715
124	213	324	436	547	627	715
124	213	324	436	567	615	713
124	213	324	436	567	625	713
124	213	324	436	567	625	724
124	213	324	437	526	615	715
124	213	324	437	526	635	715
124	213	324	437	526	645	715
124	213	324	437	536	615	715
124	213	324	437	536	625	715
124	213	324	437	536	625	725
124	213	324	437	536	645	715
124	213	324	437	536	645	725
124	213	324	437	546	615	715
124	213	324	437	546	625	715
124	213	324	437	546	625	725
124	213	324	437	546	635	715
124	213	324	437	546	635	725
124	213	324	437	567	624	715
125	213	324	435	567	625	713
125	213	324	435	567	625	714
125	213	324	435	567	635	713
125	213	324	436	536	627	716
125	213	324	436	536	657	713
125	213	324	436	537	627	716
125	213	324	436	547	627	716
125	213	324	436	567	625	713
125	213	324	436	567	625	714
125	213	324	437	536	625	716
125	213	324	437	536	625	725
125	213	324	437	536	627	716
125	213	324	437	537	627	716
125	213	324	437	537	637	716
125	213	324	437	546	625	716
125	213	324	437	546	627	716

Table 10: 4+2+1 Decomposition

124	213	324	456	524	647	715
124	213	324	456	547	624	713
124	213	324	456	547	637	715
124	213	324	456	547	647	713
124	213	324	457	546	615	713
124	213	324	457	546	625	713
124	213	324	467	526	645	715
124	213	324	467	567	625	715
124	213	324	467	567	635	715
124	213	324	467	567	645	713
125	213	324	456	537	647	716
125	213	324	456	547	614	713
125	213	324	456	547	624	713
125	213	324	457	546	637	714
125	213	324	467	536	624	714
125	213	324	467	536	645	713
125	213	324	467	537	624	714
125	213	324	467	537	645	714
125	213	324	467	537	645	724

Table 11: 3+3+1 Decomposition

124	213	324	456	527	624	715
125	213	324	417	536	625	724
125	213	324	467	536	625	724
124	213	324	467	527	624	715
124	213	324	456	537	624	715
125	213	324	467	567	625	714
125	213	324	457	536	625	724
125	213	324	457	536	657	724
124	213	324	467	567	624	715
124	213	324	467	527	645	715
124	213	324	467	537	645	715
124	213	324	456	527	647	715
124	213	324	467	567	625	724
125	213	324	467	567	625	724
125	213	324	457	536	625	714
125	213	324	417	546	657	724
125	213	324	467	536	625	714
125	213	324	417	536	657	724
125	213	324	467	567	635	714
125	213	324	457	536	657	714
125	213	324	457	546	627	716
125	213	324	456	547	627	716
125	213	324	457	546	637	716
125	213	324	456	547	637	716
125	213	324	417	536	645	724
125	213	324	417	567	645	724
125	213	324	417	536	645	745

Table 12: 3+2+2 Decomposition

124	213	324	435	567	613	713
124	213	324	436	527	615	713
124	213	324	436	547	615	713
124	213	324	437	526	637	715
125	213	324	436	547	625	713
125	213	324	437	537	625	716
124	213	324	435	567	624	713
124	213	324	436	547	625	713
124	213	324	437	536	624	715
124	213	324	436	524	627	715
124	213	324	436	536	627	715
124	213	324	437	536	627	715
124	213	324	437	546	627	715
125	213	324	435	567	614	713
124	213	324	435	526	637	715
125	213	324	436	537	625	714
124	213	324	437	546	637	715
124	213	324	435	526	647	715
125	213	324	436	537	625	716
124	213	324	435	567	624	724
124	213	324	437	546	637	725
125	213	324	435	567	613	713
125	213	324	435	567	624	713
125	213	324	436	547	627	713
125	213	324	435	567	614	714
125	213	324	436	537	627	714
125	213	324	436	547	625	716

Table 13: 4+1+1+1 Decomposition

124	213	324	457	546	613	713
124	213	324	467	567	625	713
124	213	324	467	524	624	715
125	213	324	467	567	624	713
124	213	324	457	526	615	713
125	213	324	416	537	637	716
125	213	324	457	536	624	724
124	213	324	467	526	615	713
125	213	324	467	567	614	713
124	213	324	457	536	615	713
124	213	324	467	567	615	713
124	213	324	456	527	624	713
124	213	324	457	546	627	713
124	213	324	467	526	615	715
124	213	324	456	547	627	713
125	213	324	417	536	624	724
124	213	324	467	536	615	715
125	213	324	457	546	613	713
125	213	324	416	547	624	713
125	213	324	416	547	647	713
125	213	324	416	547	624	716
124	213	324	467	536	624	715
125	213	324	416	567	624	713
125	213	324	416	537	624	716
124	213	324	467	536	625	715
124	213	324	467	526	635	715
125	213	324	467	536	614	713
125	213	324	467	536	624	713
125	213	324	457	546	627	713
125	213	324	457	536	614	714
125	213	324	457	536	624	714
125	213	324	456	537	624	714
125	213	324	457	536	627	714
125	213	324	456	537	624	716
125	213	324	457	536	625	716
125	213	324	457	536	627	716
125	213	324	416	537	627	716
125	213	324	416	547	627	716
125	213	324	456	537	637	716
125	213	324	416	537	645	716
125	213	324	416	537	645	724
125	213	324	417	536	645	726

Table 14: 3+2+1+1 Decomposition

145	213	324	467	513	624	713
156	213	324	457	524	613	713
145	213	324	467	516	624	713
145	213	324	416	537	637	716
145	213	324	416	567	625	713
145	213	324	416	526	657	713
145	213	324	416	567	635	713
145	213	324	416	536	657	713
145	213	324	416	536	627	716
145	213	324	416	537	627	716
145	213	324	416	526	637	716
145	213	324	416	527	637	716
145	213	324	467	513	625	724
145	213	324	467	567	635	724

Table 15: 2+2+2+1 Decomposition

124	213	324	457	526	613	713
124	213	324	467	524	615	713
124	213	324	467	524	615	715
125	213	324	467	567	613	713
124	213	324	467	527	615	713
124	213	324	467	524	635	715
125	213	324	457	536	614	713
125	213	324	457	536	624	713
125	213	324	456	537	637	714
125	213	324	457	536	624	716
125	213	324	416	537	625	716
125	213	324	416	547	625	716

Table 16: 3+1+1+1+1 Decomposition

145	213	324	467	513	613	713
145	213	324	467	527	624	713
156	213	324	467	527	624	713
145	213	324	467	567	624	713
156	213	324	467	524	613	713
145	213	324	467	513	625	713
145	213	324	467	536	625	713
145	213	324	467	567	625	713
145	213	324	467	526	635	713
145	213	324	416	527	635	713
145	213	324	467	567	635	713
145	213	324	467	513	625	725
156	213	324	467	527	613	713

Table 17: 2+2+1+1+1 Decomposition

145	213	324	467	526	613	713
145	213	324	467	567	613	713
145	213	324	467	527	635	713
156	213	324	456	527	627	713
156	213	324	456	527	627	714
156	213	324	456	537	627	714

Table 18: 2+1+1+1+1+1 Decomposition

145	213	345	427	526	613	713

Table 19: 1+1+1+1+1+1+1 Decomposition

$R_1 \subset X_7$	124	213	324	413	513	613	713
	124	213	324	413	513	613	724
	124	213	324	413	513	613	725
	124	213	324	413	513	613	756
$R_3 \subset X_7$	124	213	324	413	513	624	716
	124	213	324	413	513	624	756
$R_5 \subset X_7$	124	213	324	413	513	625	716
	124	213	324	413	513	625	725
	124	213	324	413	513	625	745
	124	213	324	413	513	625	746
$R_7 \subset X_7$	124	213	324	413	516	625	713
	124	213	324	413	516	625	716
	124	213	324	413	516	625	735
	124	213	324	413	516	625	736
$R_9 \subset X_7$	124	213	324	413	516	635	713
	124	213	324	413	516	635	716
	124	213	324	413	516	635	724
	124	213	324	413	516	635	725
$X_4 \subset X_7$	124	213	324	413	536	627	715
	124	213	324	413	567	615	713
	124	213	324	413	567	615	715
	124	213	324	413	567	625	713
	124	213	324	413	567	625	715
	124	213	324	413	567	635	715
$I_3 \subset X_7$	124	213	324	435	516	624	713
	124	213	324	435	516	624	716
	124	213	324	435	516	624	724
	124	213	324	435	516	624	725
	124	213	324	435	516	624	735
	124	213	324	435	516	624	736
$I_{10} \subset X_7$	124	213	324	435	516	625	713
	124	213	324	435	516	625	716
	124	213	324	435	516	625	724
	124	213	324	435	516	625	725
	124	213	324	435	516	625	735
	124	213	324	435	516	625	736
$I_9 \subset X_7$	124	213	324	435	516	635	713
	124	213	324	435	516	635	716
	124	213	324	435	516	635	724
	124	213	324	435	516	635	725
	124	213	324	435	516	635	726
	124	213	324	435	516	635	735
	124	213	324	435	516	635	746
$R_4 \subset X_7$	124	213	324	435	524	613	713
	124	213	324	435	524	613	715
	124	213	324	435	524	613	726
	124	213	324	435	524	613	735
	124	213	324	435	524	613	746
	124	213	324	435	524	613	756

Table 20: Δ-Reducible 7-point spaces

$R_6 \subset X_7$	124	213	324	435	524	615	715
	124	213	324	435	524	615	726
	124	213	324	435	524	615	736
$R_2 \subset X_7$	124	213	324	435	524	624	713
	124	213	324	435	524	624	715
	124	213	324	435	524	624	716
	124	213	324	435	524	624	724
	124	213	324	435	524	624	736
$X_{5B} \subset X_7$	124	213	324	435	524	627	716
	124	213	324	435	524	637	716
	124	213	324	435	524	637	726
	124	213	324	435	524	647	716
	124	213	324	435	524	647	726
	124	213	324	435	524	657	716
$I_{12} \subset X_7$	124	213	324	435	526	615	713
	124	213	324	435	526	615	715
	124	213	324	435	526	615	724
	124	213	324	435	526	615	726
	124	213	324	435	526	615	735
	124	213	324	435	526	615	736
	124	213	324	435	526	615	746
$I_5 \subset X_7$	124	213	324	435	526	635	713
	124	213	324	435	526	635	715
	124	213	324	435	526	635	716
	124	213	324	435	526	635	724
	124	213	324	435	526	635	726
	124	213	324	435	526	635	735
	124	213	324	435	526	635	746
$I_{13} \subset X_7$	124	213	324	435	546	613	713
	124	213	324	435	546	613	715
	124	213	324	435	546	613	724
	124	213	324	435	546	613	725
	124	213	324	435	546	613	726
	124	213	324	435	546	613	735
	124	213	324	435	546	613	746
$I_8 \subset X_7$	124	213	324	435	546	615	713
	124	213	324	435	546	615	715
	124	213	324	435	546	615	724
	124	213	324	435	546	615	725
	124	213	324	435	546	615	726
	124	213	324	435	546	615	735
	124	213	324	435	546	615	736
	124	213	324	435	546	615	746

Table 21: Δ-Reducible 7-point spaces

$I_{14} \subset X_7$	124	213	324	435	546	625	713
	124	213	324	435	546	625	715
	124	213	324	435	546	625	716
	124	213	324	435	546	625	724
	124	213	324	435	546	625	725
	124	213	324	435	546	625	735
	124	213	324	435	546	625	736
	124	213	324	435	546	625	746
$I_7 \subset X_7$	124	213	324	435	546	635	713
	124	213	324	435	546	635	715
	124	213	324	435	546	635	716
	124	213	324	435	546	635	724
	124	213	324	435	546	635	725
$I_4 \subset X_7$	124	213	324	456	513	613	713
	124	213	324	456	513	613	724
	124	213	324	456	513	613	725
	124	213	324	456	513	613	756
$I_2 \subset X_7$	124	213	324	456	513	624	713
	124	213	324	456	513	624	716
	124	213	324	456	513	624	724
	124	213	324	456	513	624	725
	124	213	324	456	513	624	756
$I_1 \subset X_7$	124	213	324	456	524	624	713
	124	213	324	456	524	624	715
	124	213	324	456	524	624	724
$R_8 \subset X_7$	125	213	324	435	514	613	713
	125	213	324	435	514	613	714
	125	213	324	435	514	613	724
	125	213	324	435	514	613	726
	125	213	324	435	514	613	746
$X_{5A} \subset X_7$	125	213	324	435	514	627	716
	125	213	324	435	514	637	716
$I_{11} \subset X_7$	125	213	324	435	546	613	713
	125	213	324	435	546	613	714
	125	213	324	435	546	613	724
	125	213	324	435	546	613	725
	125	213	324	435	546	613	726
	125	213	324	435	546	613	735
	125	213	324	435	546	613	746
$I_{15} \subset X_7$	125	213	324	435	546	625	713
	125	213	324	435	546	625	714
	125	213	324	435	546	625	716
	125	213	324	435	546	625	724
	125	213	324	435	546	625	725
$I_6 \subset X_7$	125	213	324	436	536	625	713
	125	213	324	436	536	625	714
	125	213	324	436	536	625	716
	125	213	324	436	536	625	725
$I_{17} \subset X_7$	126	213	324	435	546	615	713
	126	213	324	435	546	615	714
$I_{16} \subset X_7$	156	213	324	456	513	624	713
	156	213	324	456	513	624	714

Table 22: Δ-Reducible 7-point spaces

Gromov Product Structures for Eight Point Metric Spaces

Ayşe H. Bilge$^{(1)}$, Burak Erkan Kaya$^{(1)}$
$^{(1)}$ Kadir Has University, Istanbul, Turkey

1 Introduction

Finite metric spaces are classified by the decomposition of their metric fan [8]. With respect to this classification there are respectively, 1, 3 and 339 classes of $n = 4$, 5 and 6 point spaces. The classification finite metric spaces with elements $n > 6$ by this method have not been attempted. In previous work [2], we defined a coarser equivalence called the Gromov product equivalence. The metric fan decomposition and the Gromov product equivalence coincide for $n = 4$ and 5. For $n = 6$ and $n = 7$ we found respectively 26 and 431 equivalence classes defined via Gromov product structures.

In the present work, we apply the same method to 8-point spaces and discuss the usefulness and the applications of such decompositions.

Let (X_n, d) be an n-point space with finite metric space with metric d. The Gromov product at a point P_i is defined as

$$\Delta_{ijk} = \frac{1}{2}(d_{ij} + d_{ik} - d_{jk}) \geq 0$$

A finite metric space is called Δ-generic, if the set of Gromov products at each point have a unique minimal element. In this case, the assignment

$$P_i \to min_{j,k}\{\Delta_{ijk}\}$$

defines the Gromov product structure. Sets of Gromov product structures gives an equivalence class decomposition.

In [2], we defined "rules" for eliminating those sets of minimal Gromov products that are not allowable. There is also an algorithm to decide whether an allowable set of minimal Gromov product set corresponds to a genetic metric [5].

In this report, we summarize the Gromov Product Decomposition of 8-point spaces. The basic approach consists of listing all possible sets of minimal Gromov products and eliminate those that are not allowed, based on a result given in [2]. Then, those structures that can be mapped to each other are identified, and a representative is selected. Finally, the ones corresponding to non-generic metrics are identified and removed. This is a straight forward algorithm, but as the number of points gets larger, it is practically impossible to implement. For this reason, we use techniques presented in [4] to split the problem into sub-problems that are easier to handle.

The algorithm presented here is programmed using Matlab, and it is run on the computers of at TUBITAK ULAKBIM, High Performance and Grid Computing Center (TRUBA resources).

2 Allowable collections of minimal Gromov products via chains and cycles

As noted above, it is possible to start with all possible sets of minimal Gromov products and eliminate those that are not allowed. But for $n = 8$, it is practically impossible to identify equivalent ones. Therefore, before finding allowable collections, we split the problem into subclasses determined by so-called chains and cycles.

The characterization of a Gromov product structure in terms of $\Delta - cycles$ and $\Delta - chains$ described in detail in [5]. We say that the Gromov product structure contains no chain of length larger than 1, if all minimal Gromov products are of the form Δ_{abc}, such that no $Delta_{bai}$ and no Δ_{caj} is minimal. This corresponds to the decomposition of the number 8 as a sum of 1's. If Δ_{abi} and Δ_{baj} are minimal, we say that P_a and P_b belong the a chain of Gromov products. If there are k such points, we say there is a Gromov product chain of length k. If for all minimal Gromov products Δ_{abc}, all triples P_a, P_b, P_c belong to the chain, we say that there is a Gromov product cycle.

These structures are closely related to the “partition of integers”. For $n = 8$, we have the following partitions:

$8+0$	$7+1$	$6+1+1$	$5+1+1+1$	$4+1+1\times 3$	$3+1+1\times 4$	$2+1\times 6$	1×8
	$6+2$	$5+2+1$	$4+2+1+1$	$3+2+1\times 3$	$2+2+1\times 4$		
	$5+3$	$4+3+1$	$4+2+1+1$	$3+3+1+1$			
	$4+4$						

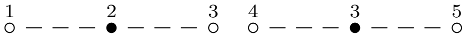
We will thus split the problem into sub-problems for those structures with k-cycles, these containing an 8-chain, those containing a chains of length at most k, for $k = 1, \ldots 8$. Here, we illustrate the procedure for $k = 1$ and $k = 2$.

2.1 Example 1: Gromov product structures with chain length 1

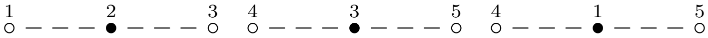
We illustrate the procedure first by describing Gromov product structures containing no chains of length greater than or equal to 2. Without loss of generality we start with the by assuming that minimal Gromov product value at P_2 is Δ_{213}. If there is no chain of length greater than or equal to 2, again without loss of generality we assume that Δ_{345} is minimal at P_3. Then the chosen P_2 the minimal Gromov product at P_1 assumed as Δ_{145}. There are 3 alternative subsets which includes Gromov products with chain lengths less than 2. By considering the previous studies, the minimality of Δ_{145} implies that Δ_{45i}, Δ_{54i}, Δ_{i14}, Δ_{i15} should be excluded. From the remaining 3 Gromov products, which are Δ_{167}, Δ_{168} and Δ_{167}, there is only one Gromov product which is Δ_{167}. If this rule is examined in detail that should be seen that, the edges of the chosen minimal products led to chose alternative Gromov products.

$S11$, $S12$ and $S13$ denote alternative subsets which are shown below.

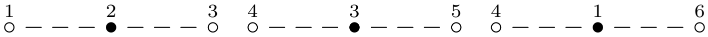
$$S1 = \{213, 345\} \tag{1}$$



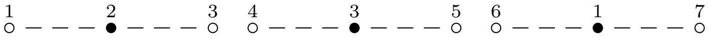
$$S11 = \{213, 345, 145\} \tag{2}$$



$$S12 = \{213, 345, 146\} \tag{3}$$

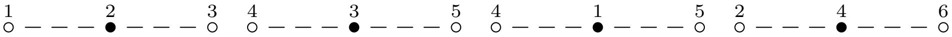


$$S13 = \{213, 345, 167\} \tag{4}$$



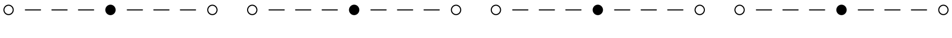
Regarding $S11$ there are 3 alternative Gromov products ($\Delta_{426}, \Delta_{427}$ and Δ_{428}) which may provide subsets with chain length less than 2. In the light of explanations given before one Gromov product must be selected from the set of $\{\Delta_{426}, \Delta_{427}, \Delta_{428}\}$. Without loss of generality we choose Δ_{426}, exclude remaining alternative Gromov products and obtain the set $S111$.

$$S111 = \{213, 345, 145, 426\} \tag{5}$$



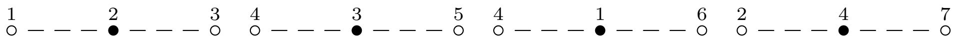
With same approach, by choosing Δ_{467} as the minimal Gromov product from the set $\{\Delta_{467}, \Delta_{468}, \Delta_{478}\}$, we can obtain another subset which is $S112$.

$$S112 = \{213, 345, 145, 467\} \tag{6}$$

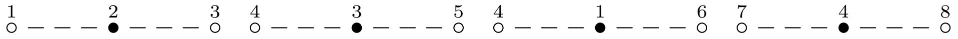


By using same approach in $S1$, we can obtain 2 alternative sets which provide subsets with chain length less than 2. Note that we choose the subset $\{213, 345, 146\}$ for $S2$. By choosing Δ_{146} as minimal Gromov product at P_1 there are 2 sets which are $\{\Delta_{427}, \Delta_{427}\}$ and $\{\Delta_{478}\}$. Since we must exclude one Gromov products in set $\{\Delta_{427}, \Delta_{428}\}$, we choose Δ_{427} as minimal Gromoov Product at P_4 and exclude Δ_{428}. Eventually we obtain 2 alternative subsets with chain length less than 2 by using $S12$.

$$S121 = \{213, 345, 146, 427\} \tag{7}$$

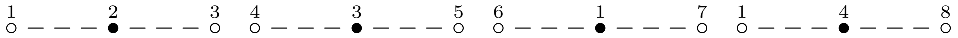


$$S121 = \{213, 345, 146, 478\} \tag{8}$$

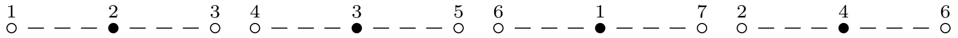


By applying same approach in the preceding steps we can obtain 5 alternative subsets with chain length less than 2 by using $S13$ which are shown below.

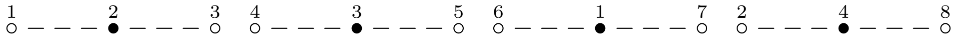
$$S131 = \{213, 345, 167, 418\} \tag{9}$$



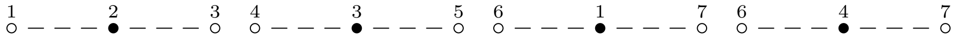
$$S132 = \{213, 345, 167, 426\} \tag{10}$$



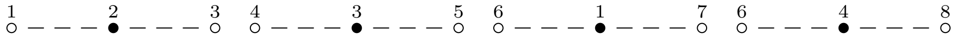
$$S133 = \{213, 345, 167, 428\} \tag{11}$$



$$S134 = \{213, 345, 167, 467\} \tag{12}$$



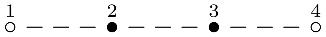
$$S135 = \{213, 345, 167, 468\} \tag{13}$$



2.2 Example 2: Gromov product structures with chain length ≤ 2

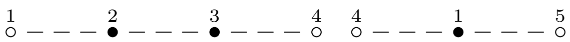
Without loss of generality we start with the by assuming that minimal Gromov product value at P_2 is Δ_{213} and Δ_{324} is minimal at P_3. Hence, we obtain a chain with length equal to 2.

$$\{213, 324\} \tag{14}$$



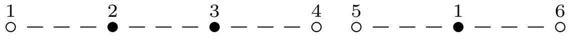
After these selections,there are 10 appropriate Gromov products may be assumed as minimal which can be shown in the set $\{\Delta_{145}, \Delta_{146}, \Delta_{147}, \Delta_{148}, \Delta_{156}, \Delta_{157}, \Delta_{158}, \Delta_{167}, \Delta_{168}, \Delta_{178}\}$. Since Gromov products in the set $\{\Delta_{145}, \Delta_{146}, \Delta_{147}, \Delta_{148}\}$ are equivalent, we assume Δ_{145} is the minimal product in the set $\{\Delta_{145}, \Delta_{146}, \Delta_{147}, \Delta_{148}\}$. Thus, we can represent a subset including one chain with length equal to 2 and one chain with length equal to 1 as $S21$.

$$S21 = \{213, 324, 145\} \tag{15}$$



Considering remaining Gromov products in the set $\{\Delta_{145}, \Delta_{146}, \Delta_{147}, \Delta_{148}, \Delta_{156}, \Delta_{157}, \Delta_{158}, \Delta_{167}, \Delta_{168}, \Delta_{178}\}$, we can choose Δ_{156} as the minimal Gromov product at P_1. It is straightforward that $\{\Delta_{156}, \Delta_{157}, \Delta_{158}\}$ are equivalent so $\{\Delta_{157}, \Delta_{158}\}$ must be excluded. After this selection $S22$ can be represented as follows:

$$S22 = \{213, 324, 156\} \tag{16}$$



All cases corresponding to the partitions of 8 are classified separately. Then the ones that overlap are identified and eliminated in steps.

3 Identification of equivalent structures

For each of the chain/cycle structures, we obtain sets of allowable collections of minimal Gromov products. The next step is to find the collections that can be mapped to each other by a permutation of indices. For this, we first eliminate the ones that cannot be mapped to each other. This is based on a matrix representation presented in [5]. If the matrices corresponding a pair of Gromov product structure are not iso-spectral, they cannot be equivalent. Using this criteria and other certain invariants, the collection of allowable collections of minimal Gromov products are further split into smaller pieces, and finally, the action of the permutation group is applied to select one representative form each equivalence class. The final step is the elimination of the collections of minimal Gromov products corresponding to non-generic metrics.

In Table 1, we present the number of allowable minimal Gromov product sets for each of the chain/cycle structures. The table contains a total of 26125 collections. This list still contains duplicate elements. The labels in the first column denote the chain/cycle type, while the labels in the first row denote the number of “end points” of the chains/cycles.

At the next step, subclasses with the same chain length are collected. The results are presented in Table 2. The labels in the first column denote the chain/cycle type, while the labels in the first row denote the number of ”end points” of the chains/cycles. The table contains a total of 15266 collections, and the list still contains duplicate elements.

In Table 1, we present the number of allowable minimal Gromov product sets for each of the chain/cycle structures. The table contains a total of 26125 collections. This list still contains duplicate elements. The labels in the first column denote the chain/cycle type, while the labels in the first row denote the number of “end points” of the chains/cycles.

At the next step, subclasses with the same chain length are collected. The results are presented in Table 2. The labels in the first column denote the chain/cycle type, while the labels in the first row denote the number of ”end points” of the chains/cycles. The table contains a total of 15266 collections, and the list still contains duplicate elements.

Finally, we grouped the subclasses by the number of endpoints and created new matrices by reordering the previous matrices in reverse. The numbers in the first column in Table 3 denote the number of end points, while the numbers in the second column denote the number of allowable minimal Gromov product sets regarding the number of end points of them. Table 3 contains 11470 collection and we eliminate all duplicate elements.

The numerical calculations reported in this paper were partially performed at TUBITAK ULAKBIM, High Performance and Grid Computing Center (TRUBA resources).

The codes used in this study and all list of allowable Gromov product sets are listed in `finitemetricspaces.khas.edu.tr`

Chain/cycle type	**00**	**02**	**04**	**06**	**08**	**10**	**12**	**14**	**16**
S111	-	-	-	-	-	-	-	-	27
S112	-	-	-	-	-	-	-	-	23
S121	-	-	-	-	-	-	-	-	28
S122	-	-	-	-	-	-	-	-	16
S131	-	-	-	-	-	-	-	-	20
S132	-	-	-	-	-	-	-	-	17
S133	-	-	-	-	-	-	-	-	13
S134	-	-	-	-	-	-	-	-	17
S135	-	-	-	-	-	-	-	-	3
S211	-	-	-	-	272	344	147	0	-
S212	-	-	-	-	813	1036	518	147	-
S221	-	-	-	-	813	1036	518	147	-
S222	-	-	-	-	482	458	185	52	-
S223	-	-	-	-	828	732	264	45	-
S224	-	-	-	-	241	216	76	14	-
S311	-	-	-	209	511	471	159	-	-
S321	-	-	-	236	801	812	240	-	-
S322	-	-	-	279	618	573	181	-	-
S323	-	-	-	312	725	494	122	-	-
S4cc	1	-	-	-	-	-	-	-	-
S4c1	-	0	8	48	54	-	-	-	-
S4c2	-	4	44	97	63	-	-	-	-
S4c3	-	4	39	46	24	-	-	-	-
S411	-	-	28	415	861	387	-	-	-
S421	-	-	28	415	861	387	-	-	-
S422	-	-	9	65	101	65	-	-	-
S423	-	-	28	287	521	237	-	-	-
S5c1	-	0	9	26	0	-	-	-	-
S5c2	-	3	17	20	0	-	-	-	-
S5c3	-	3	3	3	0	-	-	-	-
S51	-	-	141	501	324	-	-	-	-
S52	-	-	134	399	218	-	-	-	-
S53	-	-	225	785	572	-	-	-	-
S6c1	-	0	9	-	-	-	-	-	-
S6c2	-	0	5	-	-	-	-	-	-
S6c3	-	3	4	-	-	-	-	-	-
S61	-	-	107	222	-	-	-	-	-
S62	-	-	92	165	-	-	-	-	-
S63	-	-	83	127	-	-	-	-	-
S64	-	-	105	187	-	-	-	-	-
S7c1	-	1	-	-	-	-	-	-	-
S7c2	-	1	-	-	-	-	-	-	-
S71	-	-	52	-	-	-	-	-	-
S72	-	-	52	-	-	-	-	-	-
S73	-	-	40	-	-	-	-	-	-
S74	-	-	32	-	-	-	-	-	-
S75	-	-	23	-	-	-	-	-	-
S8c1	1	-	-	-	-	-	-	-	-
S81	-	13	-	-	-	-	-	-	-

Table 1: Numbers of allowable minimal Gromov product sets for each chain/cycle structure

Chain/cycle type	**00**	**02**	**04**	**06**	**08**	**10**	**12**	**14**	**16**
1-chain	-	-	-	-	-	-	-	-	31
2-chain	-	-	-	-	2043	1922	712	258	-
3-chain	-	-	-	671	1537	1595	473	-	-
4-cycle pair	1	-	-	-	-	-	-	-	-
4-cycle	-	4	45	99	71	-	-	-	-
4-chain	-	-	42	767	1483	689	-	-	-
5-cycle	-	3	17	28	0	-	-	-	-
5-chain	-	-	313	1028	666	-	-	-	-
6-cycle	-	3	12	-	-	-	-	-	-
6-chain	-	-	228	397	-	-	-	-	-
7-cycle	-	2	-	-	-	-	-	-	-
7-chain	-	-	112	-	-	-	-	-	-
8-cycle	1	-	-	-	-	-	-	-	-
8-chain	-	13	-	-	-	-	-	-	-

Table 2: Numbers of allowable minimal Gromov product sets for each chain/cycle length

Number of end points	**Number of Δ Class**
0	2
2	25
4	711
6	2438
8	3862
10	2960
12	1185
14	258
16	31
Total	**11470**

Table 3: Summary Table of remaining Δ classes of 8-point metric spaces after implementing the proposed algorithm 4 times

References

[1] Bandelt, H.-J., Dress, A. W. M. (1992). A Canonical Decomposition Theory for Metrics on a Finite Set, Advances in Mathematics 92, 47-105.

[2] Bilge, A.H., D. Çelik and Ş. Koçak (2017). An equivalence class decomposition of finite metric spaces via Gromov products. Discrete Mathematics, Vol. 340 1928-1932.

[3] Bilge, A.H. and M. Incegul, Matrix invariants of finite metric spaces, arXiv:2003.03335.

[4] Bilge, A.H. and M. Incegul, Gromov product decomposition of 7-point metric spaces, arXiv:1804.03051.

[5] Incegul, M., Algorithms for the equivalence class decomposition classification of finite metric spaces via Gromov products, Preprint. http://www.incegul.com/academic/algorithms.pdf

[6] Jussi, V., (2004). Gromov hyperbolic spaces, http://www.helsinki.fi/ jvaisala/grobok.pdf

[7] Koolen, J., A. Lesser, V. Moulton (2009). Optimal Realizations of Generic Five-Point Metrics, European J. of Combin. 30, 1164-1171.

[8] Sturmfels, B., J. Yu (2004). Classification of Six-Point Metrics, Electron. J. of Combin. 11, R44.

Quadrangle Structures of 6-Point Spaces

Ayşe H. Bilge(1), Derya Çelik(2), Burak Erkan Kaya(1)
Kadir Has University(1), Eskişehir Technical University(2)

Gromov product structures for 6-point spaces have been obtained in a previous work [2]. The list of Gromov product structures and the metrics in the Sturmfels-Yu classification for each type are given below in Tables 1 and 2, for “Irreducible” and “Reducible” spaces.

Table 1. List of Gromov product structures for irreducible spaces

	P_1	P_2	P_3	P_4	P_5	P_6	Sturmfels–Yu Type
I_1	Δ_{124}	Δ_{213}	Δ_{324}	Δ_{456}	Δ_{524}	Δ_{624}	68, 71, 95, 133, 137, 166, 200, 215, 222, 223, 227, 241, 260, 263, 275, 283, 286, 304, 306, 312, 317, 321, 331, 339
I_2	Δ_{124}	Δ_{213}	Δ_{324}	Δ_{456}	Δ_{513}	Δ_{624}	214, 249, 250, 259, 264, 285, 292, 295, 297, 298, 302, 303, 322, 325, 327, 329, 335
I_3	Δ_{124}	Δ_{213}	Δ_{324}	Δ_{435}	Δ_{516}	Δ_{624}	88, 190, 192, 251, 255, 272, 301
I_4	Δ_{124}	Δ_{213}	Δ_{324}	Δ_{456}	Δ_{513}	Δ_{613}	294, 300, 324, 334
I_5	Δ_{124}	Δ_{213}	Δ_{324}	Δ_{435}	Δ_{526}	Δ_{635}	9, 20, 40, 43, 67, 90, 101, 127, 132, 160, 204
I_6	Δ_{125}	Δ_{213}	Δ_{324}	Δ_{436}	Δ_{536}	Δ_{625}	4, 21, 42, 74, 87, 103, 120, 131, 159, 184, 185, 205, 258
I_7	Δ_{124}	Δ_{213}	Δ_{324}	Δ_{435}	Δ_{546}	Δ_{635}	11, 15, 24, 27, 45, 58, 73, 102, 115, 129, 135, 181, 221
I_8	Δ_{124}	Δ_{213}	Δ_{324}	Δ_{435}	Δ_{546}	Δ_{615}	8, 14, 18, 31, 36, 48, 78, 105, 123
I_9	Δ_{124}	Δ_{213}	Δ_{324}	Δ_{435}	Δ_{516}	Δ_{635}	142, 168, 188, 189, 220, 228, 239, 256, 265, 277, 290, 314
I_{10}	Δ_{124}	Δ_{213}	Δ_{324}	Δ_{435}	Δ_{516}	Δ_{625}	128, 161, 206
I_{11}	Δ_{125}	Δ_{213}	Δ_{324}	Δ_{435}	Δ_{546}	Δ_{613}	130, 158, 207, 208, 212, 257
I_{12}	Δ_{124}	Δ_{213}	Δ_{324}	Δ_{435}	Δ_{526}	Δ_{615}	1, 16, 28, 75
I_{13}	Δ_{124}	Δ_{213}	Δ_{324}	Δ_{435}	Δ_{546}	Δ_{613}	2, 17, 30, 60, 76, 149
I_{14}	Δ_{124}	Δ_{213}	Δ_{324}	Δ_{435}	Δ_{546}	Δ_{625}	34, 47, 49, 50, 64, 77, 92, 106, 107, 109, 121, 172
I_{15}	Δ_{125}	Δ_{213}	Δ_{324}	Δ_{435}	Δ_{546}	Δ_{625}	32, 33, 38, 55, 61, 63, 89, 114, 124, 136, 148, 162, 179, 209, 234, 273
I_{16}	Δ_{156}	Δ_{213}	Δ_{324}	Δ_{456}	Δ_{513}	Δ_{624}	247, 248, 262, 282, 293, 296, 316, 326, 332, 336
I_{17}	Δ_{126}	Δ_{213}	Δ_{324}	Δ_{435}	Δ_{546}	Δ_{615}	12, 13, 39, 65, 66

Table 2. List of Gromov product structures for reducible spaces

	P_1	P_2	P_3	P_4	P_5	P_6	Sturmfels–Yu Type
R_1	Δ_{124}	Δ_{213}	Δ_{324}	Δ_{413}	Δ_{513}	Δ_{613}	70, 81, 96, 100, 144, 151, 174, 176, 216, 218, 225, 226, 235, 240, 244, 261, 268, 280, 284, 287, 305, 308, 310, 311, 318, 320, 330, 338
R_2	Δ_{124}	Δ_{213}	Δ_{324}	Δ_{435}	Δ_{524}	Δ_{624}	69, 72, 80, 82, 84, 85, 94, 125, 126, 134, 138, 139, 143, 146, 153, 156, 157, 163, 167, 186, 187, 191, 193, 194, 195, 198, 199, 202, 211, 224, 231, 233, 252, 253, 266, 271, 278, 281, 315
R_3	Δ_{124}	Δ_{213}	Δ_{324}	Δ_{413}	Δ_{513}	Δ_{624}	7, 26, 59, 97, 118, 119, 152, 154, 175, 183, 217, 236, 237, 245, 246, 269, 274, 279, 288, 291, 299, 307, 309, 319, 323, 328, 333, 337
R_4	Δ_{124}	Δ_{213}	Δ_{324}	Δ_{435}	Δ_{524}	Δ_{613}	3, 25, 53, 54, 57, 83, 86, 98, 111, 113, 117, 140, 141, 145, 147, 164, 165, 171, 177, 182, 196, 197, 201, 203, 219, 229, 230, 232, 243, 254, 267, 270, 276, 289, 313
R_5	Δ_{124}	Δ_{213}	Δ_{324}	Δ_{413}	Δ_{513}	Δ_{625}	6, 23, 41, 51, 91, 112, 155, 170, 238
R_6	Δ_{124}	Δ_{213}	Δ_{324}	Δ_{435}	Δ_{524}	Δ_{615}	5, 22, 44, 110, 169, 210
R_7	Δ_{124}	Δ_{213}	Δ_{324}	Δ_{413}	Δ_{516}	Δ_{625}	10, 19, 29, 52, 108
R_8	Δ_{125}	Δ_{213}	Δ_{324}	Δ_{435}	Δ_{514}	Δ_{613}	35, 46, 62, 79, 93, 104, 122, 150, 173, 213
R_9	Δ_{124}	Δ_{213}	Δ_{324}	Δ_{413}	Δ_{516}	Δ_{635}	37, 56, 99, 116, 178, 180, 242

For 5-point spaces, Gromov product structures determine the quadrangle structure completely. For 6-point spaces, the Gromov product structure determines the structure of parts of quadrangles. The number of quadrangles whose structures are determined by the Gromov product structure are called "Fixed Quadrangles" , in the tables below. The remainf are called "Free Quadrangles". As each quadrangle may have 3 types, if there are *k* fixed quadrangles, the number possibilities corresponding to this Gromov type are 3^m, where m=15-k. These numbers are listed under the "Number of Classes" column. But not all of these possibilities are allowable. In order to determine the allowable ones, we use the linear programming tool of Matlab. Note that each configuration of quadrangles gives a set of inequalities amonf the distances, namely, the sum of diagonals should be larger than the sum of the sides. We introduce these inequalities as constraints in a lineer programming problem and see whether there is a feasible solution. The ones that lead to infeasible solutions are listed in the next column. Finally the column "Remaining Classes/Strumfels-Yu Types" gives the " gives the number of quadrangle classes for the given Gromov product type and the number of classes in the Sturmfels-Yu classification. The last column gives the list of quadrangle sturctures whose matrices are isospectral.

Table 1. The numbers of quadrangle types for each Gromov Product structure.

Type	Number of Fixed Quadrangles	Number of Free Quadrangles	Number of Classes	Number of Eliminated Classes	Remaining Classes/Sturmfels-Yu Types	Class Groups
I01	10	5	243	226	17/24	{1},{2,3,5,9},{4},{6,11,13},{7},{8},{10},{12},{14},{15},{16},{17}
I02	12	3	27	0	27/17	{1},{2,3,4,7,10,19},{5,25},{6,8,9,13,16,22},{11,21},{12,20},{14,27},{15,18},{17,24},{23,26}{6,8,9,13,16,22},{11,21},{12,20},{14,27},{15,18},{17,24},{23,26}
I03	13	2	9	3	6/7	{1},{2},{3},{4},{5},{6}
I04	13	2	9	6	3/4	{1},{2,3}
I05	13	2	9	3	6/11	{1},{2},{3},{4},{5},{6}
I06	12	3	27	23	4/13	{1},{2},{3},{4}
I07	11	4	81	62	19/13	{1},{2},{3},{4},{5},{6},{7},{8},{9},{10},{11},{12},{13},{14},{15},{16},{17},{18},{19}
I08	13	2	9	5	4/9	{1},{2},{3},{4}
I09	13	2	9	0	9/12	{1},{2},{3},{4},{5},{6},{7},{8,9}
I10	14	1	3	0	3/3	{1},{2},{3}
I11	13	2	9	3	6/6	{1},{2},{3},{4},{5},{6}
I12	12	3	27	25	2/4	{1},{2}
I13	13	2	9	5	4/6	{1},{2},{3},{4}
I14	13	2	9	3	6/12	{1},{2},{3},{4},{5},{6}
I15	12	3	27	13	14/16	{1},{2,4},{3,9},{5,7},{6,10},{8,12},{11},{14}
I16	11	4	81	27	54/10	{1,22},{2,3},{4,19},{5,21},{6,7,20},{8,9,44,45},{10,28,37,40,47,48,49,52},{11,12,13,16,30,34,38,39,41,51,53,54},{14,33,35},{15,17,18,36},{23,24},{25},{26},{27},{29,31,42,43,50},{32},{46},
I17	12	3	27	19	8/5	{1,3,4,7},{2,5,8},{6}

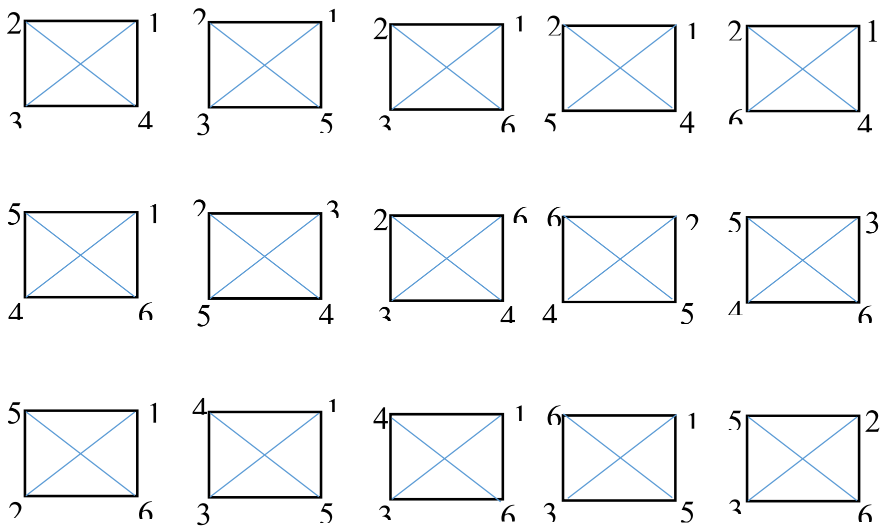
R01	10	5	243	226	17/28	{1},{2,3,5,9},{4,7,10,11},{6,13},{8,15},{12},{14},{16},{17}
R02	10	5	243	207	36/39	{1},{2,10,25},{3},{4,7},{5,28},{6},{8},{9},{11},{12},{13},{14},{15},{16,24},{17},{18,32},{19},{20},{21,36},{22,31},{23},{26},{27},{29},{30},{33},{34},{35}
R03	11	4	81	62	19/28	{1},{2,3,4,7,10,19,28},{5,8,12,13,16,20,22,25,29,30,31,46},{6,9,11,21,34,37},{14,15,26,27,33,36,39,40,48,51},{17,24,35,38,43},{18,23,32,47,49},{41,45,50,52},{42,53},{44},{54},{55,56,60},{57},{58,61},{59},{62}
R04	12	3	27	9	18/35	{1},{2},{3,7,13},{4,14},{5},{6},{8},{9},{10},{11},{12},{15},{16},{17},{18}
R05	13	2	9	5	4/9	{1},{2},{3},{4}
R06	13	2	9	5	4/6	{1},{2},{3},{4}
R07	14	1	3	1	2/5	{1},{2}
R08	12	3	27	14	13/10	{1,9},{2,3},{4,13},{5,10,11,12},{6,7},{8}
R09	14	1	3	1	2/7	{1},{2}

Since the Sturmfels-Yu classification is the finest, the number of quadrangle classes cannot be greater than the number of Sturmfel-Yu classes. But for example, for I17, there are 8 quadrangle classes and 5 Strumfles-Yu classes. In this case, it has been shown some of the quadrangle classes are mapped to each other by permutations, and the number of quadrangle classes is reduced to 4. These are similar reductions for othery Gromov product types. On the other hand, it would be possible to have less quadrangle classes. This is infact the case again for the Type I17. This type has been studied in detail in the manuscript posted to the web page, under the heading of “Finite metric spaces of dimensions n=5,6,7,8”. In this paper, quadrangle structures are used to obtain an explicit parametrization of 4 metrics in the class I17. It worths to mention that mode than one Strumfels-Yu type may have a common parametrization, distinguished by the ranges of a certain parameter.

In the tables below, the first row indicates the pars of points that lie on the diagonals. For example 1324 means that (P_1,P_3) and (P_2,P_4) are diagonals on the quadrangle with vertices P_1,P_2,P_3,P_4

We illustrate this on the first row of the next table:

1324 1325 1326 1524 1624 1456 2435 2436 2456 3456 1256 1345 1346 1356 2356



$I_1: \Delta_{124} \quad \Delta_{213} \quad \Delta_{324} \quad \Delta_{456} \quad \Delta_{524} \quad \Delta_{624}$

	Q(1,2, 3,4)	Q(1,2, 3,5)	Q(1,2, 3,6)	Q(1,2, 4,5)	Q(1,2, 4,6)	Q(1,4, 5,6)	Q(2,3, 4,5)	Q(2,3, 4,6)	Q(2,4, 5,6)	Q(3,4, 5,6)	Q(1,2, 5,6)	Q(1,3, 4,5)	Q(1,3, 4,6)	Q(1,3, 5,6)	Q(2,3, 5,6)
1	1324	1325	1326	1524	1624	1456	2435	2436	2456	3456	1256	1345	1346	1356	2356
2	1324	1325	1326	1524	1624	1456	2435	2436	2456	3456	1256	1345	1346	1356	2536
3	1324	1325	1326	1524	1624	1456	2435	2436	2456	3456	1256	1345	1436	1356	2356
4	1324	1325	1326	1524	1624	1456	2435	2436	2456	3456	1256	1345	1436	1356	2536
5	1324	1325	1326	1524	1624	1456	2435	2436	2456	3456	1256	1435	1346	1356	2356
6	1324	1325	1326	1524	1624	1456	2435	2436	2456	3456	1256	1435	1346	1356	2536
7	1324	1325	1326	1524	1624	1456	2435	2436	2456	3456	1256	1435	1436	1356	2356
8	1324	1325	1326	1524	1624	1456	2435	2436	2456	3456	1256	1435	1436	1356	2536
9	1324	1325	1326	1524	1624	1456	2435	2436	2456	3456	1526	1345	1346	1356	2356
10	1324	1325	1326	1524	1624	1456	2435	2436	2456	3456	1526	1345	1346	1356	2536
11	1324	1325	1326	1524	1624	1456	2435	2436	2456	3456	1526	1345	1436	1356	2356
12	1324	1325	1326	1524	1624	1456	2435	2436	2456	3456	1526	1345	1436	1356	2536
13	1324	1325	1326	1524	1624	1456	2435	2436	2456	3456	1526	1435	1346	1356	2356
14	1324	1325	1326	1524	1624	1456	2435	2436	2456	3456	1526	1435	1346	1356	2536
15	1324	1325	1326	1524	1624	1456	2435	2436	2456	3456	1526	1435	1436	1356	2356
16	1324	1325	1326	1524	1624	1456	2435	2436	2456	3456	1526	1435	1436	1356	2536
17	1324	1325	1326	1524	1624	1456	2435	2436	2456	3456	1526	1435	1436	1536	2536

$I_2: \Delta_{124} \quad \Delta_{213} \quad \Delta_{324} \, \Delta_{456} \, \Delta_{513} \, \Delta_{624}$

	Q(1,2, 3,4)	Q(1,2, 3,5)	Q(1,2, 3,6)	Q(1,2, 4,5)	Q(1,2, 4,6)	Q(1,3, 4,5)	Q(1,3, 5,6)	Q(1,4, 5,6)	Q(2,3, 4,5)	Q(2,3, 4,6)	Q(2,4, 5,6)	Q(3,4, 5,6)	Q(1,2, 5,6)	Q(1,3, 4,6)	Q(2,3, 5,6)
1	1324	1325	1326	1524	1624	1345	1356	1456	2435	2436	2456	3456	1256	1346	2356
2	1324	1325	1326	1524	1624	1345	1356	1456	2435	2436	2456	3456	1256	1346	2536
3	1324	1325	1326	1524	1624	1345	1356	1456	2435	2436	2456	3456	1256	1346	2635
4	1324	1325	1326	1524	1624	1345	1356	1456	2435	2436	2456	3456	1256	1436	2356
5	1324	1325	1326	1524	1624	1345	1356	1456	2435	2436	2456	3456	1256	1436	2536
6	1324	1325	1326	1524	1624	1345	1356	1456	2435	2436	2456	3456	1256	1436	2635
7	1324	1325	1326	1524	1624	1345	1356	1456	2435	2436	2456	3456	1256	1634	2356

8	1324	1325	1326	1524	1624	1345	1356	1456	2435	2436	2456	3456	1256	1634	2536
9	1324	1325	1326	1524	1624	1345	1356	1456	2435	2436	2456	3456	1256	1634	2635
10	1324	1325	1326	1524	1624	1345	1356	1456	2435	2436	2456	3456	1526	1346	2356
11	1324	1325	1326	1524	1624	1345	1356	1456	2435	2436	2456	3456	1526	1346	2536
12	1324	1325	1326	1524	1624	1345	1356	1456	2435	2436	2456	3456	1526	1346	2635
13	1324	1325	1326	1524	1624	1345	1356	1456	2435	2436	2456	3456	1526	1436	2356
14	1324	1325	1326	1524	1624	1345	1356	1456	2435	2436	2456	3456	1526	1436	2536
15	1324	1325	1326	1524	1624	1345	1356	1456	2435	2436	2456	3456	1526	1436	2635
16	1324	1325	1326	1524	1624	1345	1356	1456	2435	2436	2456	3456	1526	1634	2356
17	1324	1325	1326	1524	1624	1345	1356	1456	2435	2436	2456	3456	1526	1634	2536
18	1324	1325	1326	1524	1624	1345	1356	1456	2435	2436	2456	3456	1526	1634	2635
19	1324	1325	1326	1524	1624	1345	1356	1456	2435	2436	2456	3456	1625	1346	2356
20	1324	1325	1326	1524	1624	1345	1356	1456	2435	2436	2456	3456	1625	1346	2536
21	1324	1325	1326	1524	1624	1345	1356	1456	2435	2436	2456	3456	1625	1346	2635
22	1324	1325	1326	1524	1624	1345	1356	1456	2435	2436	2456	3456	1625	1436	2356
23	1324	1325	1326	1524	1624	1345	1356	1456	2435	2436	2456	3456	1625	1436	2536
24	1324	1325	1326	1524	1624	1345	1356	1456	2435	2436	2456	3456	1625	1436	2635
25	1324	1325	1326	1524	1624	1345	1356	1456	2435	2436	2456	3456	1625	1634	2356
26	1324	1325	1326	1524	1624	1345	1356	1456	2435	2436	2456	3456	1625	1634	2536
27	1324	1325	1326	1524	1624	1345	1356	1456	2435	2436	2456	3456	1625	1634	2635

$I_3: \Delta_{124}\ \ \Delta_{213}\ \ \Delta_{324}\ \ \Delta_{435}\ \ \Delta_{516}\ \Delta_{624}$

	Q(1,2, 3,4)	**Q(1,2, 3,5)**	**Q(1,2, 3,6)**	**Q(1,2, 4,5)**	**Q(1,2, 4,6)**	**Q(1,2, 5,6)**	**Q(1,3, 4,5)**	**Q(1,3, 5,6)**	**Q(1,4, 5,6)**	**Q(2,3, 4,5)**	**Q(2,3, 4,6)**	**Q(2,4, 5,6)**	**Q(3,4, 5,6)**	**Q(1,3, 4,6)**	**Q(2,3, 5,6)**
1	1324	1325	1326	1524	1624	1625	1435	1635	1645	2435	2436	2456	3546	1346	2536
2	1324	1325	1326	1524	1624	1625	1435	1635	1645	2435	2436	2456	3546	1346	2635
3	1324	1325	1326	1524	1624	1625	1435	1635	1645	2435	2436	2456	3546	1436	2536
4	1324	1325	1326	1524	1624	1625	1435	1635	1645	2435	2436	2456	3546	1436	2635

5	1324	1325	1326	1524	1624	1625	1435	1635	1645	2435	2436	2456	3546	1634	2536
6	1324	1325	1326	1524	1624	1625	1435	1635	1645	2435	2436	2456	3546	1634	2635

$I_4: \Delta_{124}\ \Delta_{213}\ \Delta_{324}\ \Delta_{456}\ \Delta_{513}\ \Delta_{613}$

	Q(1,2,3,4)	**Q(1,2,3,5)**	**Q(1,2,3,6)**	**Q(1,2,4,5)**	**Q(1,2,4,6)**	**Q(1,3,4,5)**	**Q(1,3,4,6)**	**Q(1,3,5,6)**	**Q(1,4,5,6)**	**Q(2,3,4,5)**	**Q(2,3,4,6)**	**Q(2,4,5,6)**	**Q(3,4,5,6)**	**Q(1,2,5,6)**	**Q(2,3,5,6)**
1	1324	1325	1326	1524	1624	1345	1346	1356	1456	2435	2436	2456	3456	1526	2356
2	1324	1325	1326	1524	1624	1345	1346	1356	1456	2435	2436	2456	3456	1256	2536
3	1324	1325	1326	1524	1624	1345	1346	1356	1456	2435	2436	2456	3456	1256	2635

$I_5: \Delta_{124}\ \Delta_{213}\ \Delta_{324}\ \Delta_{435}\ \Delta_{526}\ \Delta_{635}$

	Q(1,2,3,4)	**Q(1,2,3,5)**	**Q(1,2,3,6)**	**Q(1,2,4,5)**	**Q(1,2,4,6)**	**Q(1,2,5,6)**	**Q(1,3,4,5)**	**Q(1,3,5,6)**	**Q(2,3,4,5)**	**Q(2,3,4,6)**	**Q(2,3,5,6)**	**Q(2,4,5,6)**	**Q(3,4,5,6)**	**Q(1,3,4,6)**	**Q(1,4,5,6)**
1	1324	1325	1326	1524	1624	1526	1435	1635	2435	2436	2635	2645	3546	1346	1546
2	1324	1325	1326	1524	1624	1526	1435	1635	2435	2436	2635	2645	3546	1346	1645
3	1324	1325	1326	1524	1624	1526	1435	1635	2435	2436	2635	2645	3546	1436	1546
4	1324	1325	1326	1524	1624	1526	1435	1635	2435	2436	2635	2645	3546	1436	1645
5	1324	1325	1326	1524	1624	1526	1435	1635	2435	2436	2635	2645	3546	1634	1546
6	1324	1325	1326	1524	1624	1526	1435	1635	2435	2436	2635	2645	3546	1634	1645

$I_6: \Delta_{125}\ \Delta_{213}\ \Delta_{324}\ \Delta_{436}\ \Delta_{536}\ \Delta_{625}$

	Q(1,2,3,4)	**Q(1,2,3,5)**	**Q(1,2,3,6)**	**Q(1,2,4,5)**	**Q(1,2,5,6)**	**Q(1,3,4,6)**	**Q(1,3,5,6)**	**Q(2,3,4,5)**	**Q(2,3,4,6)**	**Q(2,3,5,6)**	**Q(2,4,5,6)**	**Q(3,4,5,6)**	**Q(1,2,4,6)**	**Q(1,3,4,5)**	**Q(1,4,5,6)**
1	1324	1325	1326	1425	1625	1436	1536	2435	2436	2536	2546	3546	1426	1435	1456
2	1324	1325	1326	1425	1625	1436	1536	2435	2436	2536	2546	3546	1426	1435	1546
3	1324	1325	1326	1425	1625	1436	1536	2435	2436	2536	2546	3546	1624	1435	1456
4	1324	1325	1326	1425	1625	1436	1536	2435	2436	2536	2546	3546	1624	1435	1546

$I_7: \Delta_{124}\ \Delta_{213}\ \Delta_{324}\ \Delta_{435}\ \Delta_{546}\ \Delta_{635}$

	Q(1,2,3,4)	**Q(12,3,5)**	**Q(1,2,3,6)**	**Q(1,2,4,5)**	**Q(1,2,4,6)**	**Q(1,3,4,5)**	**Q(1,4,5,6)**	**Q(2,3,4,5)**	**Q(2,3,4,6)**	**Q(2,4,5,6)**	**Q(3,4,5,6)**	**Q(1,2,5,6)**	**Q(1,3,4,6)**	**Q(1,3,5,6)**	**Q(2,3,5,6)**
1	1324	1325	1326	1524	1624	1435	1546	2435	2436	2546	3546	1526	1346	1356	2536
2	1324	1325	1326	1524	1624	1435	1546	2435	2436	2546	3546	1526	1346	1356	2635
3	1324	1325	1326	1524	1624	1435	1546	2435	2436	2546	3546	1526	1346	1536	2536

4	1324	1325	1326	1524	1624	1435	1546	2435	2436	2546	3546	1526	1346	1536	2635
5	1324	1325	1326	1524	1624	1435	1546	2435	2436	2546	3546	1526	1346	1635	2635
6	1324	1325	1326	1524	1624	1435	1546	2435	2436	2546	3546	1526	1436	1536	2536
7	1324	1325	1326	1524	1624	1435	1546	2435	2436	2546	3546	1526	1436	1536	2635
8	1324	1325	1326	1524	1624	1435	1546	2435	2436	2546	3546	1526	1436	1635	2635
9	1324	1325	1326	1524	1624	1435	1546	2435	2436	2546	3546	1526	1634	1635	2635
10	1324	1325	1326	1524	1624	1435	1546	2435	2436	2546	3546	1625	1346	1356	2536
11	1324	1325	1326	1524	1624	1435	1546	2435	2436	2546	3546	1625	1346	1356	2635
12	1324	1325	1326	1524	1624	1435	1546	2435	2436	2546	3546	1625	1346	1536	2536
13	1324	1325	1326	1524	1624	1435	1546	2435	2436	2546	3546	1625	1346	1635	2536
14	1324	1325	1326	1524	1624	1435	1546	2435	2436	2546	3546	1625	1346	1635	2635
15	1324	1325	1326	1524	1624	1435	1546	2435	2436	2546	3546	1625	1436	1536	2536
16	1324	1325	1326	1524	1624	1435	1546	2435	2436	2546	3546	1625	1436	1635	2536
17	1324	1325	1326	1524	1624	1435	1546	2435	2436	2546	3546	1625	1436	1635	2635
18	1324	1325	1326	1524	1624	1435	1546	2435	2436	2546	3546	1625	1634	1635	2536
19	1324	1325	1326	1524	1624	1435	1546	2435	2436	2546	3546	1625	1634	1635	2635

$I_8\colon \Delta_{124}\ \Delta_{213}\ \Delta_{324}\ \Delta_{435}\ \Delta_{546}\ \Delta_{615}$

	Q(1,2, 3,4)	**Q(1,2, 3,5)**	**Q(1,2, 3,6)**	**Q(1,2, 4,5)**	**Q(1,2, 4,6)**	**Q(1,2, 5,6)**	**Q(1,3, 4,5)**	**Q(1,3, 5,6)**	**Q(1,4, 5,6)**	**Q(2,3, 4,5)**	**Q(2,3, 4,6)**	**Q(2,4, 5,6)**	**Q(3,4, 5,6)**	**Q(1,3, 4,6)**	**Q(2,3, 5,6)**
1	1324	1325	1326	1524	1624	1526	1435	1536	1546	2435	2436	2546	3546	1346	2536
2	1324	1325	1326	1524	1624	1526	1435	1536	1546	2435	2436	2546	3546	1346	2635
3	1324	1325	1326	1524	1624	1526	1435	1536	1546	2435	2436	2546	3546	1436	2536
4	1324	1325	1326	1524	1624	1526	1435	1536	1546	2435	2436	2546	3546	1436	2635

$I_9\colon \Delta_{124}\ \Delta_{213}\ \Delta_{324}\ \Delta_{435}\ \Delta_{516}\ \Delta_{635}$

	Q(1,2, 3,4)	Q(1,2, 3,5)	Q(1,2, 3,6)	Q(1,2, 4,5)	Q(1,2, 4,6)	Q(1,2, 5,6)	Q(1,3, 4,5)	Q(1,3, 5,6)	Q(1,4, 5,6)	Q(2,3, 4,5)	Q(2,3, 4,6)	Q(2,3, 5,6)	Q(3,4, 5,6)	Q(1,3, 4,6)	Q(2,4, 5,6)
1	1324	1325	1326	1524	1624	1625	1435	1635	1645	2435	2436	2635	3546	1346	2456
2	1324	1325	1326	1524	1624	1625	1435	1635	1645	2435	2436	2635	3546	1346	2546
3	1324	1325	1326	1524	1624	1625	1435	1635	1645	2435	2436	2635	3546	1346	2645
4	1324	1325	1326	1524	1624	1625	1435	1635	1645	2435	2436	2635	3546	1436	2456
5	1324	1325	1326	1524	1624	1625	1435	1635	1645	2435	2436	2635	3546	1436	2546

6	1324	1325	1326	1524	1624	1625	1435	1635	1645	2435	2436	2635	3546	1436	2645
7	1324	1325	1326	1524	1624	1625	1435	1635	1645	2435	2436	2635	3546	1634	2456
8	1324	1325	1326	1524	1624	1625	1435	1635	1645	2435	2436	2635	3546	1634	2546
9	1324	1325	1326	1524	1624	1625	1435	1635	1645	2435	2436	2635	3546	1634	2645

$I_{10}: \Delta_{124}\ \Delta_{213}\ \Delta_{324}\ \Delta_{435}\ \Delta_{516}\ \Delta_{625}$

	Q(1,2,3,4)	**Q(1,2,3,5)**	**Q(1,2,3,6)**	**Q(1,2,4,5)**	**Q(1,2,4,6)**	**Q(1,2,5,6)**	**Q(1,3,4,5)**	**Q(1,3,5,6)**	**Q(1,4,5,6)**	**Q(2,3,4,5)**	**Q(2,3,4,6)**	**Q(2,3,5,6)**	**Q(2,4,5,6)**	**Q(3,4,5,6)**	**Q(1,3,4,6)**
1	1324	1325	1326	1524	1624	1625	1435	1635	1645	2435	2436	2536	2546	3546	1346
2	1324	1325	1326	1524	1624	1625	1435	1635	1645	2435	2436	2536	2546	3546	1436
3	1324	1325	1326	1524	1624	1625	1435	1635	1645	2435	2436	2536	2546	3546	1634

$I_{11}: \Delta_{124}\ \Delta_{213}\ \Delta_{324}\ \Delta_{435}\ \Delta_{546}\ \Delta_{613}$

	Q(1,2,3,4)	**Q(1,2,3,5)**	**Q(1,2,3,6)**	**Q(1,2,4,5)**	**Q(1,2,5,6)**	**Q(1,3,4,5)**	**Q(1,3,4,6)**	**Q(1,3,5,6)**	**Q(1,4,5,6)**	**Q(2,3,4,5)**	**Q(2,3,4,6)**	**Q(2,4,5,6)**	**Q(3,4,5,6)**	**Q(1,2,4,6)**	**Q(2,3,5,6)**
1	1324	1325	1326	1425	1625	1435	1346	1356	1546	2435	2436	2546	3546	1246	2536
2	1324	1325	1326	1425	1625	1435	1346	1356	1546	2435	2436	2546	3546	1246	2635
3	1324	1325	1326	1425	1625	1435	1346	1356	1546	2435	2436	2546	3546	1426	2536
4	1324	1325	1326	1425	1625	1435	1346	1356	1546	2435	2436	2546	3546	1426	2635
5	1324	1325	1326	1425	1625	1435	1346	1356	1546	2435	2436	2546	3546	1624	2536
6	1324	1325	1326	1425	1625	1435	1346	1356	1546	2435	2436	2546	3546	1624	2635

$I_{12}: \Delta_{124}\ \Delta_{213}\ \Delta_{324}\ \Delta_{435}\ \Delta_{526}\ \Delta_{615}$

	Q(1,2,3,4)	**Q(1,2,3,5)**	**Q(1,2,3,6)**	**Q(1,2,4,5)**	**Q(1,2,4,6)**	**Q(1,2,5,6)**	**Q(1,3,4,5)**	**Q(1,3,5,6)**	**Q(1,4,5,6)**	**Q(2,3,4,5)**	**Q(2,3,4,6)**	**Q(2,3,5,6)**	**Q(2,4,5,6)**	**Q(3,4,5,6)**	**Q(1,3,4,6)**
1	1324	1325	1326	1524	1624	1526	1435	1536	1546	2435	2436	2635	2645	3546	1346
2	1324	1325	1326	1524	1624	1526	1435	1536	1546	2435	2436	2635	2645	3546	1436

$I_{13}: \Delta_{124}\ \Delta_{213}\ \Delta_{324}\ \Delta_{435}\ \Delta_{546}\ \Delta_{613}$

	Q(1,2,3,4)	**Q(1,2,3,5)**	**Q(1,2,3,6)**	**Q(1,2,4,5)**	**Q(1,2,4,6)**	**Q(1,3,4,5)**	**Q(1,3,4,6)**	**Q(1,3,5,6)**	**Q(1,4,5,6)**	**Q(2,3,4,5)**	**Q(2,3,4,6)**	**Q(2,4,5,6)**	**Q(3,4,5,6)**	**Q(1,2,5,6)**	**Q(2,3,5,6)**
1	1324	1325	1326	1524	1624	1435	1346	1356	1546	2435	2436	2546	3546	1526	2536
2	1324	1325	1326	1524	1624	1435	1346	1356	1546	2435	2436	2546	3546	1526	2635
3	1324	1325	1326	1524	1624	1435	1346	1356	1546	2435	2436	2546	3546	1625	2536
4	1324	1325	1326	1524	1624	1435	1346	1356	1546	2435	2436	2546	3546	1625	2635

I_{14}: Δ_{124} Δ_{213} Δ_{324} Δ_{435} Δ_{546} Δ_{625}

	Q(1,2,3,4)	**Q(1,2,3,5)**	**Q(1,2,3,6)**	**Q(1,2,4,5)**	**Q(1,2,4,6)**	**Q(1,2,5,6)**	**Q(1,3,4,5)**	**Q(1,4,5,6)**	**Q(2,3,4,5)**	**Q(2,3,4,6)**	**Q(2,3,5,6)**	**Q(2,4,5,6)**	**Q(3,4,5,6)**	**Q(1,3,4,6)**	**Q(1,3,5,6)**
1	1324	1325	1326	1524	1624	1625	1435	1546	2435	2436	2536	2546	3546	1346	1356
2	1324	1325	1326	1524	1624	1625	1435	1546	2435	2436	2536	2546	3546	1346	1536
3	1324	1325	1326	1524	1624	1625	1435	1546	2435	2436	2536	2546	3546	1346	1635
4	1324	1325	1326	1524	1624	1625	1435	1546	2435	2436	2536	2546	3546	1436	1536
5	1324	1325	1326	1524	1624	1625	1435	1546	2435	2436	2536	2546	3546	1436	1635
6	1324	1325	1326	1524	1624	1625	1435	1546	2435	2436	2536	2546	3546	1634	1635

I_{15}: Δ_{125} Δ_{213} Δ_{324} Δ_{435} Δ_{546} Δ_{625}

	Q(1,2,3,4)	**Q(1,2,3,5)**	**Q(1,2,3,6)**	**Q(1,2,4,5)**	**Q(1,2,5,6)**	**Q(1,3,4,5)**	**Q(1,4,5,6)**	**Q(2,3,4,5)**	**Q(2,3,4,6)**	**Q(2,3,5,6)**	**Q(2,4,5,6)**	**Q(3,4,5,6)**	**Q(1,2,4,6)**	**Q(1,3,4,6)**	**Q(1,3,5,6)**
1	1324	1325	1326	1425	1625	1435	1546	2435	2436	2536	2546	3546	1246	1346	1356
2	1324	1325	1326	1425	1625	1435	1546	2435	2436	2536	2546	3546	1246	1346	1536
3	1324	1325	1326	1425	1625	1435	1546	2435	2436	2536	2546	3546	1246	1346	1635
4	1324	1325	1326	1425	1625	1435	1546	2435	2436	2536	2546	3546	1426	1346	1356
5	1324	1325	1326	1425	1625	1435	1546	2435	2436	2536	2546	3546	1426	1346	1536
6	1324	1325	1326	1425	1625	1435	1546	2435	2436	2536	2546	3546	1426	1346	1635
7	1324	1325	1326	1425	1625	1435	1546	2435	2436	2536	2546	3546	1426	1436	1536
8	1324	1325	1326	1425	1625	1435	1546	2435	2436	2536	2546	3546	1426	1436	1635
9	1324	1325	1326	1425	1625	1435	1546	2435	2436	2536	2546	3546	1624	1346	1356
10	1324	1325	1326	1425	1625	1435	1546	2435	2436	2536	2546	3546	1624	1346	1536
11	1324	1325	1326	1425	1625	1435	1546	2435	2436	2536	2546	3546	1624	1346	1635
12	1324	1325	1326	1425	1625	1435	1546	2435	2436	2536	2546	3546	1624	1436	1536
13	1324	1325	1326	1425	1625	1435	1546	2435	2436	2536	2546	3546	1624	1436	1635
14	1324	1325	1326	1425	1625	1435	1546	2435	2436	2536	2546	3546	1624	1634	1635

I_{16}: Δ_{156} Δ_{213} Δ_{324} Δ_{456} Δ_{513} Δ_{624}

	Q(1,2,3,4)	**Q(1,2,3,5)**	**Q(1,2,3,6)**	**Q(1,2,5,6)**	**Q(1,3,4,5)**	**Q(1,3,5,6)**	**Q(1,4,5,6)**	**Q(2,3,4,5)**	**Q(2,3,4,6)**	**Q(2,4,5,6)**	**Q(3,4,5,6)**	**Q(1,2,4,5)**	**Q(1,2,4,6)**	**Q(1,3,4,6)**	**Q(2,3,5,6)**
1	1324	1325	1326	1256	1345	1356	1456	2435	2436	2456	3456	1245	1246	1346	2356
2	1324	1325	1326	1256	1345	1356	1456	2435	2436	2456	3456	1245	1246	1346	2536
3	1324	1325	1326	1256	1345	1356	1456	2435	2436	2456	3456	1245	1246	1346	2635
4	1324	1325	1326	1256	1345	1356	1456	2435	2436	2456	3456	1245	1426	1346	2356
5	1324	1325	1326	1256	1345	1356	1456	2435	2436	2456	3456	1245	1426	1346	2536
6	1324	1325	1326	1256	1345	1356	1456	2435	2436	2456	3456	1245	1426	1346	2635

7	1324	1325	1326	1256	1345	1356	1456	2435	2436	2456	3456	1245	1426	1436	2356
8	1324	1325	1326	1256	1345	1356	1456	2435	2436	2456	3456	1245	1426	1436	2536
9	1324	1325	1326	1256	1345	1356	1456	2435	2436	2456	3456	1245	1426	1436	2635
10	1324	1325	1326	1256	1345	1356	1456	2435	2436	2456	3456	1245	1624	1346	2356
11	1324	1325	1326	1256	1345	1356	1456	2435	2436	2456	3456	1245	1624	1346	2536
12	1324	1325	1326	1256	1345	1356	1456	2435	2436	2456	3456	1245	1624	1346	2635
13	1324	1325	1326	1256	1345	1356	1456	2435	2436	2456	3456	1245	1624	1436	2356
14	1324	1325	1326	1256	1345	1356	1456	2435	2436	2456	3456	1245	1624	1436	2536
15	1324	1325	1326	1256	1345	1356	1456	2435	2436	2456	3456	1245	1624	1436	2635
16	1324	1325	1326	1256	1345	1356	1456	2435	2436	2456	3456	1245	1624	1634	2356
17	1324	1325	1326	1256	1345	1356	1456	2435	2436	2456	3456	1245	1624	1634	2536
18	1324	1325	1326	1256	1345	1356	1456	2435	2436	2456	3456	1245	1624	1634	2635
19	1324	1325	1326	1256	1345	1356	1456	2435	2436	2456	3456	1425	1246	1346	2356
20	1324	1325	1326	1256	1345	1356	1456	2435	2436	2456	3456	1425	1246	1346	2536
21	1324	1325	1326	1256	1345	1356	1456	2435	2436	2456	3456	1425	1246	1346	2635
22	1324	1325	1326	1256	1345	1356	1456	2435	2436	2456	3456	1425	1426	1346	2356
23	1324	1325	1326	1256	1345	1356	1456	2435	2436	2456	3456	1425	1426	1346	2536
24	1324	1325	1326	1256	1345	1356	1456	2435	2436	2456	3456	1425	1426	1346	2635
25	1324	1325	1326	1256	1345	1356	1456	2435	2436	2456	3456	1425	1426	1436	2356
26	1324	1325	1326	1256	1345	1356	1456	2435	2436	2456	3456	1425	1426	1436	2536
27	1324	1325	1326	1256	1345	1356	1456	2435	2436	2456	3456	1425	1426	1436	2635
28	1324	1325	1326	1256	1345	1356	1456	2435	2436	2456	3456	1425	1624	1346	2356
29	1324	1325	1326	1256	1345	1356	1456	2435	2436	2456	3456	1425	1624	1346	2536
30	1324	1325	1326	1256	1345	1356	1456	2435	2436	2456	3456	1425	1624	1346	2635
31	1324	1325	1326	1256	1345	1356	1456	2435	2436	2456	3456	1425	1624	1436	2356
32	1324	1325	1326	1256	1345	1356	1456	2435	2436	2456	3456	1425	1624	1436	2536
33	1324	1325	1326	1256	1345	1356	1456	2435	2436	2456	3456	1425	1624	1436	2635
34	1324	1325	1326	1256	1345	1356	1456	2435	2436	2456	3456	1425	1624	1634	2356
35	1324	1325	1326	1256	1345	1356	1456	2435	2436	2456	3456	1425	1624	1634	2536
36	1324	1325	1326	1256	1345	1356	1456	2435	2436	2456	3456	1425	1624	1634	2635
37	1324	1325	1326	1256	1345	1356	1456	2435	2436	2456	3456	1524	1246	1346	2356
38	1324	1325	1326	1256	1345	1356	1456	2435	2436	2456	3456	1524	1246	1346	2536
39	1324	1325	1326	1256	1345	1356	1456	2435	2436	2456	3456	1524	1246	1346	2635

40	1324	1325	1326	1256	1345	1356	1456	2435	2436	2456	3456	1524	1426	1346	2356
41	1324	1325	1326	1256	1345	1356	1456	2435	2436	2456	3456	1524	1426	1346	2536
42	1324	1325	1326	1256	1345	1356	1456	2435	2436	2456	3456	1524	1426	1346	2635
43	1324	1325	1326	1256	1345	1356	1456	2435	2436	2456	3456	1524	1426	1436	2356
44	1324	1325	1326	1256	1345	1356	1456	2435	2436	2456	3456	1524	1426	1436	2536
45	1324	1325	1326	1256	1345	1356	1456	2435	2436	2456	3456	1524	1426	1436	2635
46	1324	1325	1326	1256	1345	1356	1456	2435	2436	2456	3456	1524	1624	1346	2356
47	1324	1325	1326	1256	1345	1356	1456	2435	2436	2456	3456	1524	1624	1346	2536
48	1324	1325	1326	1256	1345	1356	1456	2435	2436	2456	3456	1524	1624	1346	2635
49	1324	1325	1326	1256	1345	1356	1456	2435	2436	2456	3456	1524	1624	1436	2356
50	1324	1325	1326	1256	1345	1356	1456	2435	2436	2456	3456	1524	1624	1436	2536
51	1324	1325	1326	1256	1345	1356	1456	2435	2436	2456	3456	1524	1624	1436	2635
52	1324	1325	1326	1256	1345	1356	1456	2435	2436	2456	3456	1524	1624	1634	2356
53	1324	1325	1326	1256	1345	1356	1456	2435	2436	2456	3456	1524	1624	1634	2536
54	1324	1325	1326	1256	1345	1356	1456	2435	2436	2456	3456	1524	1624	1634	2635

$I_{17}: \Delta_{126}\ \Delta_{213}\ \ \Delta_{324}\ \Delta_{435}\ \Delta_{546}\ \Delta_{615}$

	Q(1,2, 3,4)	**Q(1,2, 3,5)**	**Q(1,2, 3,6)**	**Q(1,2, 4,6)**	**Q(1,2, 5,6)**	**Q(1,3, 4,5)**	**Q(1,3, 5,6)**	**Q(1,4, 5,6)**	**Q(2,3, 4,5)**	**Q(2,3, 4,6)**	**Q(2,4, 5,6)**	**Q(3,4, 5,6)**	**Q(1,2, 4,5)**	**Q(1,3, 4,6)**	**Q(2,3, 5,6)**
1	1324	1325	1326	1426	1526	1435	1536	1546	2435	2436	2546	3546	1425	1346	2536
2	1324	1325	1326	1426	1526	1435	1536	1546	2435	2436	2546	3546	1425	1346	2635
3	1324	1325	1326	1426	1526	1435	1536	1546	2435	2436	2546	3546	1425	1436	2536
4	1324	1325	1326	1426	1526	1435	1536	1546	2435	2436	2546	3546	1425	1436	2635
5	1324	1325	1326	1426	1526	1435	1536	1546	2435	2436	2546	3546	1524	1346	2536
6	1324	1325	1326	1426	1526	1435	1536	1546	2435	2436	2546	3546	1524	1346	2635
7	1324	1325	1326	1426	1526	1435	1536	1546	2435	2436	2546	3546	1524	1436	2536
8	1324	1325	1326	1426	1526	1435	1536	1546	2435	2436	2546	3546	1524	1436	2635

$R_{01}: \Delta_{124}\ \Delta_{213}\ \ \Delta_{324}\ \Delta_{413}\ \Delta_{513}\ \Delta_{613}$

	Q(1,2, 3,4)	**Q(1,2, 3,5)**	**Q(1,2, 3,6)**	**Q(1,3, 4,5)**	**Q(1,2, 4,6)**	**Q(1,2, 5,6)**	**Q(1,3, 4,5)**	**Q(1,3, 4,6)**	**Q(1,3, 5,6)**	**Q(1,4, 5,6)**	**Q(2,3, 4,5)**	**Q(2,3, 4,6)**	**Q(2,3, 5,6)**	**Q(2,4, 5,6)**	**Q(3,4, 5,6)**
1	1324	1325	1326	1524	1624	1345	1346	1356	2435	2436	1256	1456	2356	2456	3456
2	1324	1325	1326	1524	1624	1345	1346	1356	2435	2436	1256	1456	2356	2456	3546
3	1324	1325	1326	1524	1624	1345	1346	1356	2435	2436	1256	1456	2536	2456	3456
4	1324	1325	1326	1524	1624	1345	1346	1356	2435	2436	1256	1456	2536	2456	3546
5	1324	1325	1326	1524	1624	1345	1346	1356	2435	2436	1256	1546	2356	2456	3456
6	1324	1325	1326	1524	1624	1345	1346	1356	2435	2436	1256	1546	2356	2456	3546

7	1324	1325	1326	1524	1624	1345	1346	1356	2435	2436	1256	1546	2536	2456	3456
8	1324	1325	1326	1524	1624	1345	1346	1356	2435	2436	1256	1546	2536	2456	3546
9	1324	1325	1326	1524	1624	1345	1346	1356	2435	2436	1526	1456	2356	2456	3456
10	1324	1325	1326	1524	1624	1345	1346	1356	2435	2436	1526	1456	2356	2456	3546
11	1324	1325	1326	1524	1624	1345	1346	1356	2435	2436	1526	1456	2536	2456	3456
12	1324	1325	1326	1524	1624	1345	1346	1356	2435	2436	1526	1456	2536	2456	3546
13	1324	1325	1326	1524	1624	1345	1346	1356	2435	2436	1526	1546	2356	2456	3456
14	1324	1325	1326	1524	1624	1345	1346	1356	2435	2436	1526	1546	2356	2456	3546
15	1324	1325	1326	1524	1624	1345	1346	1356	2435	2436	1526	1546	2536	2456	3456
16	1324	1325	1326	1524	1624	1345	1346	1356	2435	2436	1526	1546	2536	2456	3546
17	1324	1325	1326	1524	1624	1345	1346	1356	2435	2436	1526	1546	2536	2546	3546

$R_{02}: \Delta_{124}\ \Delta_{213}\ \ \Delta_{324}\ \Delta_{435}\ \Delta_{524}\ \Delta_{624}$

	Q(1,2, 3,4)	**Q(1,2, 3,5)**	**Q(1,2, 3,6)**	**Q(1,2, 4,5)**	**Q(1,2, 4,6)**	**Q(1,2, 5,6)**	**Q(1,3, 4,5)**	**Q(1,3, 4,6)**	**Q(1,3, 5,6)**	**Q(1,4, 5,6)**	**Q(2,3, 4,5)**	**Q(2,3, 4,6)**	**Q(2,3, 5,6)**	**Q(2,4, 5,6)**	**Q(3,4, 5,6)**
1	1324	1325	1326	1524	1624	1435	2435	2436	2456	3546	1256	1346	1356	1456	2356
2	1324	1325	1326	1524	1624	1435	2435	2436	2456	3546	1256	1346	1356	1456	2536
3	1324	1325	1326	1524	1624	1435	2435	2436	2456	3546	1256	1346	1356	1456	2635
4	1324	1325	1326	1524	1624	1435	2435	2436	2456	3546	1256	1346	1356	1546	2356
5	1324	1325	1326	1524	1624	1435	2435	2436	2456	3546	1256	1346	1356	1546	2536
6	1324	1325	1326	1524	1624	1435	2435	2436	2456	3546	1256	1346	1356	1546	2635
7	1324	1325	1326	1524	1624	1435	2435	2436	2456	3546	1256	1436	1356	1456	2356
8	1324	1325	1326	1524	1624	1435	2435	2436	2456	3546	1256	1436	1356	1456	2536
9	1324	1325	1326	1524	1624	1435	2435	2436	2456	3546	1256	1436	1356	1456	2635
10	1324	1325	1326	1524	1624	1435	2435	2436	2456	3546	1526	1346	1356	1456	2356
11	1324	1325	1326	1524	1624	1435	2435	2436	2456	3546	1526	1346	1356	1456	2536
12	1324	1325	1326	1524	1624	1435	2435	2436	2456	3546	1526	1346	1356	1456	2635
13	1324	1325	1326	1524	1624	1435	2435	2436	2456	3546	1526	1346	1356	1546	2356
14	1324	1325	1326	1524	1624	1435	2435	2436	2456	3546	1526	1346	1356	1546	2536
15	1324	1325	1326	1524	1624	1435	2435	2436	2456	3546	1526	1346	1356	1546	2635
16	1324	1325	1326	1524	1624	1435	2435	2436	2456	3546	1526	1346	1536	1546	2536
17	1324	1325	1326	1524	1624	1435	2435	2436	2456	3546	1526	1346	1536	1546	2635
18	1324	1325	1326	1524	1624	1435	2435	2436	2456	3546	1526	1436	1356	1456	2356
19	1324	1325	1326	1524	1624	1435	2435	2436	2456	3546	1526	1436	1356	1456	2536
20	1324	1325	1326	1524	1624	1435	2435	2436	2456	3546	1526	1436	1356	1456	2635
21	1324	1325	1326	1524	1624	1435	2435	2436	2456	3546	1526	1436	1536	1456	2536
22	1324	1325	1326	1524	1624	1435	2435	2436	2456	3546	1526	1436	1536	1456	2635
23	1324	1325	1326	1524	1624	1435	2435	2436	2456	3546	1526	1436	1536	1546	2536
24	1324	1325	1326	1524	1624	1435	2435	2436	2456	3546	1526	1436	1536	1546	2635
25	1324	1325	1326	1524	1624	1435	2435	2436	2456	3546	1625	1346	1356	1456	2356
26	1324	1325	1326	1524	1624	1435	2435	2436	2456	3546	1625	1346	1356	1456	2536

27	1324	1325	1326	1524	1624	1435	2435	2436	2456	3546	1625	1346	1356	1456	2635
28	1324	1325	1326	1524	1624	1435	2435	2436	2456	3546	1625	1346	1356	1546	2356
29	1324	1325	1326	1524	1624	1435	2435	2436	2456	3546	1625	1346	1356	1546	2536
30	1324	1325	1326	1524	1624	1435	2435	2436	2456	3546	1625	1346	1356	1546	2635
31	1324	1325	1326	1524	1624	1435	2435	2436	2456	3546	1625	1346	1536	1546	2536
32	1324	1325	1326	1524	1624	1435	2435	2436	2456	3546	1625	1436	1356	1456	2356
33	1324	1325	1326	1524	1624	1435	2435	2436	2456	3546	1625	1436	1356	1456	2536
34	1324	1325	1326	1524	1624	1435	2435	2436	2456	3546	1625	1436	1356	1456	2635
35	1324	1325	1326	1524	1624	1435	2435	2436	2456	3546	1625	1436	1536	1456	2536
36	1324	1325	1326	1524	1624	1435	2435	2436	2456	3546	1625	1436	1536	1546	2536

R_{03}: Δ_{124} Δ_{213} Δ_{324} Δ_{413} Δ_{513} Δ_{624}

	Q(1,2, 3,4)	**Q(1,2, 3,5)**	**Q(1,2, 3,6)**	**Q(1,2, 4,5)**	**Q(1,2, 4,6)**	**Q(1,2, 5,6)**	**Q(1,3, 4,5)**	**Q(1,3, 4,6)**	**Q(1,3, 5,6)**	**Q(1,4, 5,6)**	**Q(2,3, 4,5)**	**Q(2,3, 4,6)**	**Q(2,3, 5,6)**	**Q(2,4, 5,6)**	**Q(3,4, 5,6)**
1	1324	1325	1326	1524	1624	1345	1346	1356	2435	2436	2456	1256	1456	2356	3456
2	1324	1325	1326	1524	1624	1345	1346	1356	2435	2436	2456	1256	1456	2356	3546
3	1324	1325	1326	1524	1624	1345	1346	1356	2435	2436	2456	1256	1456	2356	3645
4	1324	1325	1326	1524	1624	1345	1346	1356	2435	2436	2456	1256	1456	2536	3456
5	1324	1325	1326	1524	1624	1345	1346	1356	2435	2436	2456	1256	1456	2536	3546
6	1324	1325	1326	1524	1624	1345	1346	1356	2435	2436	2456	1256	1456	2536	3645
7	1324	1325	1326	1524	1624	1345	1346	1356	2435	2436	2456	1256	1456	2635	3456
8	1324	1325	1326	1524	1624	1345	1346	1356	2435	2436	2456	1256	1456	2536	3546
9	1324	1325	1326	1524	1624	1345	1346	1356	2435	2436	2456	1256	1456	2536	3645
10	1324	1325	1326	1524	1624	1345	1346	1356	2435	2436	2456	1256	1546	2356	3456
11	1324	1325	1326	1524	1624	1345	1346	1356	2435	2436	2456	1256	1546	2356	3546
12	1324	1325	1326	1524	1624	1345	1346	1356	2435	2436	2456	1256	1546	2356	3645
13	1324	1325	1326	1524	1624	1345	1346	1356	2435	2436	2456	1256	1546	2536	3456
14	1324	1325	1326	1524	1624	1345	1346	1356	2435	2436	2456	1256	1546	2536	3546
15	1324	1325	1326	1524	1624	1345	1346	1356	2435	2436	2456	1256	1546	2536	3645
16	1324	1325	1326	1524	1624	1345	1346	1356	2435	2436	2456	1256	1546	2635	3456
17	1324	1325	1326	1524	1624	1345	1346	1356	2435	2436	2456	1256	1546	2635	3546
18	1324	1325	1326	1524	1624	1345	1346	1356	2435	2436	2456	1256	1546	2635	3645
19	1324	1325	1326	1524	1624	1345	1346	1356	2435	2436	2456	1256	1645	2356	3456
20	1324	1325	1326	1524	1624	1345	1346	1356	2435	2436	2456	1256	1645	2356	3546
21	1324	1325	1326	1524	1624	1345	1346	1356	2435	2436	2456	1256	1645	2356	3645
22	1324	1325	1326	1524	1624	1345	1346	1356	2435	2436	2456	1256	1645	2536	3456
23	1324	1325	1326	1524	1624	1345	1346	1356	2435	2436	2456	1256	1645	2536	3546
24	1324	1325	1326	1524	1624	1345	1346	1356	2435	2436	2456	1256	1645	2536	3645
25	1324	1325	1326	1524	1624	1345	1346	1356	2435	2436	2456	1256	1645	2635	3456
26	1324	1325	1326	1524	1624	1345	1346	1356	2435	2436	2456	1256	1645	2635	3546

27	1324	1325	1326	1524	1624	1345	1346	1356	2435	2436	2456	1256	1645	2635	3645
28	1324	1325	1326	1524	1624	1345	1346	1356	2435	2436	2456	1526	1456	2356	3456
29	1324	1325	1326	1524	1624	1345	1346	1356	2435	2436	2456	1526	1456	2356	3546
30	1324	1325	1326	1524	1624	1345	1346	1356	2435	2436	2456	1526	1456	2356	3645
31	1324	1325	1326	1524	1624	1345	1346	1356	2435	2436	2456	1526	1456	2536	3456
32	1324	1325	1326	1524	1624	1345	1346	1356	2435	2436	2456	1526	1456	2536	3546
33	1324	1325	1326	1524	1624	1345	1346	1356	2435	2436	2456	1526	1456	2536	3645
34	1324	1325	1326	1524	1624	1345	1346	1356	2435	2436	2456	1526	1456	2635	3456
35	1324	1325	1326	1524	1624	1345	1346	1356	2435	2436	2456	1526	1456	2635	3546
36	1324	1325	1326	1524	1624	1345	1346	1356	2435	2436	2456	1526	1456	2635	3645
37	1324	1325	1326	1524	1624	1345	1346	1356	2435	2436	2456	1526	1546	2356	3456
38	1324	1325	1326	1524	1624	1345	1346	1356	2435	2436	2456	1526	1546	2356	3546
39	1324	1325	1326	1524	1624	1345	1346	1356	2435	2436	2456	1526	1546	2356	3645
40	1324	1325	1326	1524	1624	1345	1346	1356	2435	2436	2456	1526	1546	2536	3456
41	1324	1325	1326	1524	1624	1345	1346	1356	2435	2436	2456	1526	1546	2536	3546
42	1324	1325	1326	1524	1624	1345	1346	1356	2435	2436	2456	1526	1546	2536	3645
43	1324	1325	1326	1524	1624	1345	1346	1356	2435	2436	2456	1526	1546	2635	3456
44	1324	1325	1326	1524	1624	1345	1346	1356	2435	2436	2456	1526	1546	2635	3546
45	1324	1325	1326	1524	1624	1345	1346	1356	2435	2436	2456	1526	1546	2635	3645
46	1324	1325	1326	1524	1624	1345	1346	1356	2435	2436	2456	1526	1645	2356	3456
47	1324	1325	1326	1524	1624	1345	1346	1356	2435	2436	2456	1526	1645	2356	3546
48	1324	1325	1326	1524	1624	1345	1346	1356	2435	2436	2456	1526	1645	2356	3645
49	1324	1325	1326	1524	1624	1345	1346	1356	2435	2436	2456	1526	1645	2536	3456
50	1324	1325	1326	1524	1624	1345	1346	1356	2435	2436	2456	1526	1645	2536	3645
51	1324	1325	1326	1524	1624	1345	1346	1356	2435	2436	2456	1526	1645	2635	3456
52	1324	1325	1326	1524	1624	1345	1346	1356	2435	2436	2456	1526	1645	2635	3546
53	1324	1325	1326	1524	1624	1345	1346	1356	2435	2436	2456	1526	1645	2635	3645
54	1324	1325	1326	1524	1624	1345	1346	1356	2435	2436	2456	1625	1625	2356	3456
55	1324	1325	1326	1524	1624	1345	1346	1356	2435	2436	2456	1625	1625	2356	3546
56	1324	1325	1326	1524	1624	1345	1346	1356	2435	2436	2456	1625	1625	2356	3645
57	1324	1325	1326	1524	1624	1345	1346	1356	2435	2436	2456	1625	1625	2536	3456
58	1324	1325	1326	1524	1624	1345	1346	1356	2435	2436	2456	1625	1625	2536	3546
59	1324	1325	1326	1524	1624	1345	1346	1356	2435	2436	2456	1625	1625	2536	3645
60	1324	1325	1326	1524	1624	1345	1346	1356	2435	2436	2456	1625	1625	2635	3456
61	1324	1325	1326	1524	1624	1345	1346	1356	2435	2436	2456	1625	1625	2635	3546
62	1324	1325	1326	1524	1624	1345	1346	1356	2435	2436	2456	1625	1625	2635	3645

R_{04}: Δ_{124} Δ_{213} Δ_{324} Δ_{435} Δ_{524} Δ_{613}

	Q(1,2, 3,4)	Q(1,2, 3,5)	Q(1,2, 3,6)	Q(1,2, 4,5)	Q(1,2, 4,6)	Q(1,2, 5,6)	Q(1,3, 4,5)	Q(1,3, 4,6)	Q(1,3, 5,6)	Q(1,4, 5,6)	Q(2,3, 4,5)	Q(2,3, 4,6)	Q(2,3, 5,6)	Q(2,4, 5,6)	Q(3,4, 5,6)
1	1324	1325	1326	1524	1624	1435	1346	1356	2435	2436	2456	3546	1256	2356	1456
2	1324	1325	1326	1524	1624	1435	1346	1356	2435	2436	2456	3546	1256	2356	1546
3	1324	1325	1326	1524	1624	1435	1346	1356	2435	2436	2456	3546	1256	2536	1456
4	1324	1325	1326	1524	1624	1435	1346	1356	2435	2436	2456	3546	1256	2536	1546
5	1324	1325	1326	1524	1624	1435	1346	1356	2435	2436	2456	3546	1256	2635	1456
6	1324	1325	1326	1524	1624	1435	1346	1356	2435	2436	2456	3546	1256	2635	1546
7	1324	1325	1326	1524	1624	1435	1346	1356	2435	2436	2456	3546	1526	2356	1456
8	1324	1325	1326	1524	1624	1435	1346	1356	2435	2436	2456	3546	1526	2356	1546
9	1324	1325	1326	1524	1624	1435	1346	1356	2435	2436	2456	3546	1526	2536	1456
10	1324	1325	1326	1524	1624	1435	1346	1356	2435	2436	2456	3546	1526	2536	1546
11	1324	1325	1326	1524	1624	1435	1346	1356	2435	2436	2456	3546	1526	2635	1456
12	1324	1325	1326	1524	1624	1435	1346	1356	2435	2436	2456	3546	1526	2635	1546
13	1324	1325	1326	1524	1624	1435	1346	1356	2435	2436	2456	3546	1625	2356	1456
14	1324	1325	1326	1524	1624	1435	1346	1356	2435	2436	2456	3546	1625	2356	1546
15	1324	1325	1326	1524	1624	1435	1346	1356	2435	2436	2456	3546	1625	2536	1456
16	1324	1325	1326	1524	1624	1435	1346	1356	2435	2436	2456	3546	1625	2536	1546
17	1324	1325	1326	1524	1624	1435	1346	1356	2435	2436	2456	3546	1625	2635	1456
18	1324	1325	1326	1524	1624	1435	1346	1356	2435	2436	2456	3546	1625	2635	1546

R_{05}: Δ_{124} Δ_{213} Δ_{324} Δ_{413} Δ_{513} Δ_{625}

	Q(1,2, 3,4)	Q(1,2, 3,5)	Q(1,2, 3,6)	Q(1,2, 4,5)	Q(1,2, 4,6)	Q(1,2, 5,6)	Q(1,3, 4,5)	Q(1,3, 4,6)	Q(1,3, 5,6)	Q(1,4, 5,6)	Q(2,3, 4,5)	Q(2,3, 4,6)	Q(2,3, 5,6)	Q(2,4, 5,6)	Q(3,4, 5,6)
1	1324	1325	1236	1524	1624	1625	1345	1346	1356	2435	2436	2536	2546	1546	3546
2	1324	1325	1236	1524	1624	1625	1345	1346	1356	2435	2436	2536	2546	1546	3645
3	1324	1325	1236	1524	1624	1625	1345	1346	1356	2435	2436	2536	2546	1645	3546
4	1324	1325	1236	1524	1624	1625	1345	1346	1356	2435	2436	2536	2546	1645	3645

R_{06}: Δ_{124} Δ_{213} Δ_{324} Δ_{435} Δ_{524} Δ_{615}

	Q(1,2, 3,4)	Q(1,2, 3,5)	Q(1,2, 3,6)	Q(1,2, 4,5)	Q(1,2, 4,6)	Q(1,2, 5,6)	Q(1,3, 4,5)	Q(1,3, 4,6)	Q(1,3, 5,6)	Q(1,4, 5,6)	Q(2,3, 4,5)	Q(2,3, 4,6)	Q(2,3, 5,6)	Q(2,4, 5,6)	Q(3,4, 5,6)
1	1324	1325	1326	1524	1624	1526	1435	1536	1546	2435	2436	2456	3546	1346	2536
2	1324	1325	1326	1524	1624	1526	1435	1536	1546	2435	2436	2456	3546	1346	2635
3	1324	1325	1326	1524	1624	1526	1435	1536	1546	2435	2436	2456	3546	1436	2356
4	1324	1325	1326	1524	1624	1526	1435	1536	1546	2435	2436	2456	3546	2436	2635

$\boldsymbol{R_{07}}$: $\boldsymbol{\Delta_{124}\ \Delta_{213}\ \ \Delta_{324}\ \Delta_{413}\ \Delta_{516}\ \Delta_{625}}$

	Q(1,2, 3,4)	**Q(1,2, 3,5)**	**Q(1,2, 3,6)**	**Q(1,2, 4,5)**	**Q(1,2, 4,6)**	**Q(1,2, 5,6)**	**Q(1,3, 4,5)**	**Q(1,3, 4,6)**	**Q(1,3, 5,6)**	**Q(1,4, 5,6)**	**Q(2,3, 4,5)**	**Q(2,3, 4,6)**	**Q(2,3, 5,6)**	**Q(2,4, 5,6)**	**Q(3,4, 5,6)**
1	1324	1325	1326	1524	1624	1625	1345	1346	1635	1645	2435	2436	2536	2546	3546
2	1324	1325	1326	1524	1624	1625	1345	1346	1635	1645	2435	2436	2536	2546	3645

$\boldsymbol{R_{08}}$: $\boldsymbol{\Delta_{125}\ \Delta_{213}\ \ \Delta_{324}\ \Delta_{435}\ \Delta_{514}\ \Delta_{613}}$

	Q(1,2, 3,4)	**Q(1,2, 3,5)**	**Q(1,2, 3,6)**	**Q(1,2, 4,5)**	**Q(1,2, 4,6)**	**Q(1,2, 5,6)**	**Q(1,3, 4,5)**	**Q(1,3, 4,6)**	**Q(1,3, 5,6)**	**Q(1,4, 5,6)**	**Q(2,3, 4,5)**	**Q(2,3, 4,6)**	**Q(2,3, 5,6)**	**Q(2,4, 5,6)**	**Q(3,4, 5,6)**
1	1324	1325	1326	1425	1625	1435	1346	1356	1456	2435	2436	3546	1246	2536	2546
2	1324	1325	1326	1425	1625	1435	1346	1356	1456	2435	2436	3546	1246	2635	2546
3	1324	1325	1326	1425	1625	1435	1346	1356	1456	2435	2436	3546	1426	2356	2456
4	1324	1325	1326	1425	1625	1435	1346	1356	1456	2435	2436	3546	1426	2536	2456
5	1324	1325	1326	1425	1625	1435	1346	1356	1456	2435	2436	3546	1426	2536	2546
6	1324	1325	1326	1425	1625	1435	1346	1356	1456	2435	2436	3546	1426	2635	2456
7	1324	1325	1326	1425	1625	1435	1346	1356	1456	2435	2436	3546	1426	2635	2546
8	1324	1325	1326	1425	1625	1435	1346	1356	1456	2435	2436	3546	1426	2635	2645
9	1324	1325	1326	1425	1625	1435	1346	1356	1456	2435	2436	3546	1624	2356	2456
10	1324	1325	1326	1425	1625	1435	1346	1356	1456	2435	2436	3546	1624	2536	2456
11	1324	1325	1326	1425	1625	1435	1346	1356	1456	2435	2436	3546	1624	2536	2546
12	1324	1325	1326	1425	1625	1435	1346	1356	1456	2435	2436	3546	1624	2635	2456
13	1324	1325	1326	1425	1625	1435	1346	1356	1456	2435	2436	3546	1624	2635	2546

$\boldsymbol{R_{09}}$: $\boldsymbol{\Delta_{124}\ \Delta_{213}\ \ \Delta_{324}\ \Delta_{413}\ \Delta_{516}\ \Delta_{625}}$

	Q(1,2, 3,4)	**Q(1,2, 3,5)**	**Q(1,2, 3,6)**	**Q(1,2, 4,5)**	**Q(1,2, 4,6)**	**Q(1,2, 5,6)**	**Q(1,3, 4,5)**	**Q(1,3, 4,6)**	**Q(1,3, 5,6)**	**Q(1,4, 5,6)**	**Q(2,3, 4,5)**	**Q(2,3, 4,6)**	**Q(2,3, 5,6)**	**Q(2,4, 5,6)**	**Q(3,4, 5,6)**
1	1324	1325	1326	1524	1624	1625	1345	1346	1635	1645	2435	2436	2536	2546	3546
2	1324	1325	1326	1524	1624	1625	1345	1346	1635	1645	2435	2436	2536	2546	3645

References

[1] Bilge, D. Çelik, Ş. Koçak, Optimal embeddings of finite metric spaces into graphs, Anadolu Univ. J. Sci. Technol. B- Theor. Sci.. 3 (2) (2015) 133–147.

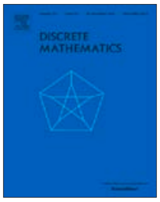
[2] Bilge, A.H., D. Çelik and S¸. Koçak (2017). An equivalence class decomposition of finite metric spaces via Gromov products. Discrete Mathematics, Vol. 340 1928-1932.

Discrete Mathematics 344 (2021) 112358

Contents lists available at ScienceDirect

Discrete Mathematics

journal homepage: www.elsevier.com/locate/disc



Gromov product structures, quadrangle structures and split metric decompositions for finite metric spaces

Ayşe Hümeyra Bilge [a], Derya Çelik [b,*], Şahin Koçak [c], Arash M. Rezaeinazhad [d]

[a] *Faculty of Engineering and Natural Sciences, Department of Industrial Engineering, Kadir Has University, Istanbul, Turkey*
[b] *Eskişehir Technical University, Department of Mathematics, 26470, Eskişehir, Turkey*
[c] *Anadolu University, Department of Mathematics, 26470, Eskişehir, Turkey*
[d] *Faculty of Management, Management Information Systems Department, Kadir Has University, Istanbul, Turkey*

ARTICLE INFO

Article history:
Received 25 August 2019
Received in revised form 8 February 2021
Accepted 9 February 2021
Available online 8 March 2021

Keywords:
Finite metric spaces
Split metric decompositions
Gromov products
Quadrangle structures

ABSTRACT

Let (X, d) be a finite metric space with elements P_i, $i = 1, \ldots, n$ and with distances $d_{ij} := d(P_i, P_j)$ for $i, j = 1, \ldots, n$. The “Gromov product” Δ_{ijk}, is defined as $\Delta_{ijk} = \frac{1}{2}(d_{ij} + d_{ik} - d_{jk})$. (X, d) is called Δ-generic, if, for each fixed i, the set of Gromov products Δ_{ijk} has a unique smallest element, $\Delta_{ij_ik_i}$. The Gromov product structure on a Δ-generic finite metric space (X, d) is the map that assigns the edge $E_{j_ik_i}$ to P_i. A finite metric space is called “quadrangle generic", if for all 4-point subsets $\{P_i, P_j, P_k, P_l\}$, the set $\{d_{ij}+d_{kl}, d_{ik}+d_{jl}, d_{il}+d_{jk}\}$ has a unique maximal element. The “quadrangle structure" on a quadrangle generic finite metric space (X, d) is defined as a map that assigns to each 4-point subset of X the pair of edges corresponding to the maximal element of the sums of distances. Two metric spaces (X, d) and (X, d') are said to be Δ-equivalent (Q-equivalent), if the corresponding Gromov product (quadrangle) structures are the same up to a permutation of X. We show that Gromov product classification is coarser than the metric fan classification. Furthermore it is proved that: (i) The isolation index of the 1-split metric δ_i is equal to the minimal Gromov product at the vertex P_i. (ii) For a quadrangle generic (X, d), the isolation index of the 2-split metric δ_{ij} is nonzero if and only if the edge E_{ij} is a side in every quadrangle whose set of vertices includes P_i and P_j. (iii) For a quadrangle generic (X, d), the isolation index of an m-split metric $\delta_{i_1\ldots i_m}$ is nonzero if and only if any edge $E_{i_ki_l}$ is a side in every quadrangle whose vertex set contains P_{i_k} and P_{i_l}. These results are applied to construct a totally split decomposable metric for $n = 6$.

1. Introduction

The understanding of finite metric spaces is an interesting issue by several respects (the most important being probably the phylogenetic analysis) and the main device for classifying them is the so-called metric fan, which we will recall below in Section 2. As there are too many sub-cones of the metric cone constituting the metric fan (for example 194160 sub-cones coming in 339 symmetry classes for a 6-point space), coarser classifications seem to be desirable. In a previous paper [2] we presented an equivalence class decomposition of finite metric spaces using the set of minimal Gromov products at each point of that space. We recall the definitions and some results concerning Gromov product structures.

---

* Corresponding author.
E-mail address: deryacelik@eskisehir.edu.tr (D. Çelik).

https://doi.org/10.1016/j.disc.2021.112358

Let (X, d) be a finite metric space with n elements P_i, $i = 1, \ldots, n$ $(n \geq 3)$ and let d_{ij} be the distance between P_i and P_j. Since a finite metric space can be considered as a weighted complete graph, the elements of X are also referred to as "vertices" or "nodes". In this line, E_{ij} and T_{ijk} denote respectively an edge and a triangle with corresponding vertices.

Definition 1. The quantity Δ_{ijk}, defined as

$$\Delta_{ijk} = \Delta_{ikj} = \frac{1}{2}(d_{ij} + d_{ik} - d_{jk}), \tag{1}$$

is called the Gromov product of the triangle T_{ijk} at the vertex P_i. We call a metric space Δ-generic, if for each P_i the set of Gromov products Δ_{ijk} has a unique minimal element.

The quantity Δ_{ijk} is attributed to M. L. Gromov since it originates from his work on δ-hyperbolic metric spaces [5, Page 27].

By the triangle inequality, the Gromov products Δ_{ijk} are non-negative numbers. The distances d_{ij} can be expressed in terms of the Gromov products as

$$d_{ij} = \Delta_{ijk_1} + \Delta_{jik_1} = \Delta_{ijk_2} + \Delta_{jik_2} = \cdots = \Delta_{ijk_{n-2}} + \Delta_{jik_{n-2}}, \tag{2}$$

where the indices k_l run from 1 to n, excluding i and j, leading to a total of $n-2$ equalities for each d_{ij}. In [2], we proved that a metric space can be defined using the Gromov products Δ_{ijk} as the primary ingredients in the sense that given a collection of Δ_{ijk} satisfying certain properties, then a metric space can be defined using Eq. (2), which produces the given collection of Δ_{ijk}.

The Gromov product structure for a Δ-generic finite metric space is defined as follows [2]:

Definition 2. Let (X, d) be a Δ-generic finite metric space. Let $P_i \in X$, and let $\Delta_{ij_ik_i}$ be the minimal Gromov product at P_i, $(i = 1, \ldots, n)$. The function that assigns the edge $E_{j_ik_i}$ to the vertex P_i is called the Gromov product structure on X. Two Δ-generic metric spaces (X, d) and (X, d') are called Δ-equivalent, if the corresponding Gromov product structures are the same up to a permutation of X.

Δ-equivalence classes for 5-point metrics coincide with the hypersimplex decomposition [7,8], obtained from the metric fan [2]. The Δ-equivalence classes and hypersimplex decomposition of 6-point spaces were given, respectively, in [2] and [8]. The algorithm for Δ-equivalence class decompositions is based on the following Proposition 1 and Corollary 1 of [2].

Proposition 1. *Let (X, d) be a finite metric space with n elements P_i, $i = 1, \ldots, n$. Then the following equations hold*

$$\Delta_{ijl} - \Delta_{ijk} = \Delta_{kjl} - \Delta_{kil} = \Delta_{lik} - \Delta_{ljk} = \Delta_{jik} - \Delta_{jil},$$

where $i, j, k, l = 1, 2, \ldots, n$.

Corollary 1. *Let (X, d) be a Δ-generic finite metric space and let Δ_{ijk} be the minimal Gromov product at node P_i. Then,*

(a) Δ_{jkl} cannot be minimal at node P_j, where $l \neq j, k$
(b) Δ_{kjl} cannot be minimal at node P_k, where $l \neq j, k$
(c) Δ_{lij} and Δ_{lik} cannot be minimal at node P_l, where $l \neq i, j, k$

In order to obtain a decomposition of finite metric spaces into Δ-equivalence classes, we start by the Cartesian product of the sets of Gromov products at each P_i, then use Corollary 1 in order to eliminate the ones that are not allowable. Then, the permutation group on n elements is acted on the list of allowable Gromov products to form the orbits under this group action. A representative from each Δ-equivalence class is selected to form a list. We note that Corollary 1 gives necessary conditions in the sense that the list obtained may contain structures that may not be realizable as generic Δ-equivalence classes. These can be eliminated by using an algorithm based on combinatorial arguments, as presented in [6].

Gromov product structures on a (generic) n-point metric space have a convenient representation by an $n \times n$ matrix M_Δ defined by $M_\Delta(i, j) = 1$ and $M_\Delta(i, k) = 1$ if Δ_{ijk} is the minimal Gromov product at P_i and 0 otherwise [3]. Thus, a Δ-equivalence class corresponds to the orbit of the matrix representation M_Δ of a Gromov Product structure, under the action $P^{-1}M_\Delta P$ where P is the permutation matrix on the vertices of the space.

In Section 2, the hypersimplex classification is recalled and its relation to the Gromov product structures is clarified. In Section 3, a new classification tool in terms of 4-point subsets of a finite metric space called "Quadrangle Structure" is defined. In Section 4, metrics of type 12, 13, 39, 65 and 66 of the hypersimplex classification [8] are studied considering their Gromov product and quadrangle structures. We will present elsewhere that for $n = 5$, all three classification concepts give the same classes as in [7] and from $n = 6$ on, all three classification concepts begin to differ. In Section 5, split metric decompositions are reviewed and necessary and sufficient conditions for the isolation index of an m-split to be nonzero, in terms of quadrangle structures, are derived. In Section 6, the tools developed throughout the paper are used to construct a totally split decomposable metric on a 6-point space.

2. The metric Fan and Gromov product structures

In this section, we recall the definitions concerning the metric fan classification of finite metric spaces and relate the Gromov product structures to them.

The set C_n of all pseudo-metrics $d = (d_{ij}) \in \mathbb{R}^{\binom{n}{2}}$ on a given n-point set X, is called the metric cone (Since for a pseudo-metric, $d_{ij} = d_{ji}$, C_n can be thought as a subset of $\mathbb{R}^{\binom{n}{2}}$.) A decomposition of C_n into some sub-cones can be defined as follows [8].

Consider the $\binom{n}{2} \times n$ matrix $\mathcal{A}$ where the rows are labeled by

$$(1, 2), (1, 3), \dots, (1, n), (2, 3), (2, 4), \dots, (2, n), \dots, (n-1, n)$$

and the (i, j)-row $(i < j)$ is given by $e_i + e_j = (0, \dots, 1, \dots, 1, \dots, 0) \in \mathbb{R}^n$.

Let $\mathcal{B}$ be an invertible $n \times n$ submatrix of $\mathcal{A}$ and denote the $\left[\binom{n}{2} - n\right] \times n$ matrix obtained by deleting $\mathcal{B}$ from $\mathcal{A}$ by $\mathcal{B}'$. Likewise, define $d_{\mathcal{B}} \in \mathbb{R}^n$ by choosing the components of $d \in \mathbb{R}^{\binom{n}{2}}$ corresponding to $\mathcal{B}$ and $d_{\mathcal{B}'} \in \mathbb{R}^{\binom{n}{2}-n}$ corresponding to $\mathcal{B}'$. Now consider the following system of equations and inequalities for $x \in \mathbb{R}^n$:

$$\mathcal{B}x = d_{\mathcal{B}} \quad \text{and} \quad \mathcal{B}'x > d_{\mathcal{B}'}.$$

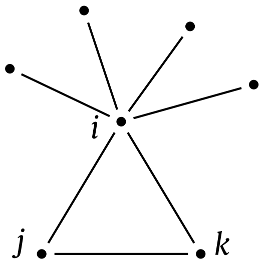
If this system has a solution we say that the matrix $\mathcal{B}$ is a "cell" or a "thrackle" for the metric d. We denote the collection of cells of a metric d by $Cell(d)$. This terminology stems from the fact that the row vectors of $\mathcal{B}$ can be viewed as the vertices of an $(n-1)$-simplex in $\mathbb{R}^n$. Still another interpretation is that, a row vector $e_i + e_j$ can be viewed as an edge of the complete graph K_n with n nodes so that a cell $\mathcal{B}$ can be viewed as a sub-graph of K_n.

Now we define two metrics d and d' on an n-point set X to be equivalent if they have the same collection of cells, i.e. $Cell(d) = Cell(d')$ (or what amounts to the same, the same collection of sub-graphs). The equivalence class of a metric d is a sub-cone of the metric cone and these sub-cones constitute altogether the metric fan.

We will now give a characterization of Gromov product structure in terms of sub-graphs in the metric fan picture.

Proposition 2. *Let $(X = \{P_1, P_2, \dots, P_n\}, d)$ be an n-point metric space. Then, (X, d) is Δ-generic with the Gromov product structure $P_i \mapsto E_{jk}, (i = 1, \dots, n)$ if and only if the following sub-graphs $G_{i,jk}$ of (the complete graph) K_n belong to the cell-collection $Cell(d)$ of the metric d.*

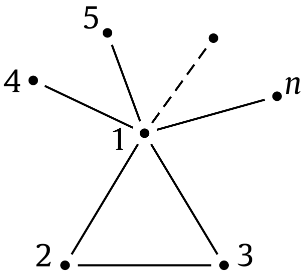
$G_{i,jk}$:



The unlabeled vertices have the labels from $\{1, 2, \dots, n\} \setminus \{i, j, k\}$.

Proof. For the ease of notation let $i = 1,\ j = 2,\ k = 3$. Then, the graph $G_{1,23}$ corresponds to the matrix $\mathcal{B}$ below:

$G_{1,23}$:



$$\longleftrightarrow \qquad \mathcal{B} = \begin{bmatrix} 1 & 1 & 0 & \dots & 0 \\ 1 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & \dots & 1 \\ 0 & 1 & 1 & \dots & 0 \end{bmatrix}$$

The system $\mathcal{B}x = d_{\mathcal{B}}$ and $\mathcal{B}'x > d_{\mathcal{B}'}$ reads as follows:

$$\begin{array}{rclrcl}
x_1 + x_2 & = & d_{12} & x_2 + x_4 & > & d_{24} \\
x_1 + x_3 & = & d_{13} & & \vdots & \\
x_1 + x_4 & = & d_{14} & x_2 + x_n & > & d_{2n} \\
& \vdots & & & \vdots & \\
x_1 + x_n & = & d_{1n} & x_r + x_s & > & d_{rs} \ \ (r \geq 2,\ s \geq 4,\ r < s) \\
x_2 + x_3 & = & d_{23} & & \vdots & \\
& & & x_{n-1} + x_n & > & d_{n-1\ n}.
\end{array}$$

If the system has a solution in terms of x_i's, then Δ_{123} is uniquely minimal among Δ_{1rs} for $1 \neq r \neq s \neq 1$ and vice versa. To show this, first consider the first 2 equations and the last one from left column to see that $x_1 = \Delta_{123}$, $x_2 = \Delta_{213}$ and $x_3 = \Delta_{312}$. From other equations it follows that $x_k = d_{1k} - \Delta_{123}$ for $4 \leq k \leq n$. By applying algebraic manipulations, it can be seen that the inequalities $x_2 + x_k > d_{2k}$ $(4 \leq k \leq n)$ are equivalent to $\Delta_{123} < \Delta_{12k}$, the inequalities $x_3 + x_k > d_{3k}$ $(4 \leq k \leq n)$ are equivalent to $\Delta_{123} < \Delta_{13k}$ and finally the inequalities $x_r + x_s > d_{rs}$ $(4 \leq r, s$ and $r < s)$ are equivalent to $\Delta_{123} < \Delta_{1rs}$. Applying this to every node of X, the proposition follows. ■

This proposition shows that Gromov product equivalence is weaker than the metric fan equivalence:

Proposition 3. *Let d and d' be two Δ-generic metrics on an n-point set $X = \{P_1, P_2, \ldots, P_n\}$. If d and d' are equivalent in the metric fan sense, i.e. Cell(d) = Cell(d'), then they have the same Gromov product structure (and so, a fortiori, they are equivalent in the Gromov product sense).*

Proof. If the Δ-generic metric d has the Gromov product structure $P_i \to E_{jk}$, for $i = 1, 2, \ldots, n$, then by Proposition 2 the sub-graphs $G_{i,jk}$ of (the complete graph) K_n belong to the cell-collection *Cell*(d) of the metric d. Since by assumption *Cell*(d) = *Cell*(d'), the sub-graphs $G_{i,jk}$ belong also to *Cell*(d'). This means, again by Proposition 2, that the metric d' has the same Gromov product structure. ■

3. Quadrangle structures

In this section we define a new kind of structure on a finite metric space.

Definition 3. An n-point finite metric space X is called "quadrangle generic", or Q-generic, if for every 4-point subset $\{P_i, P_j, P_k, P_l\} \subseteq X$, the set of distances

$$\{d_{ij} + d_{kl},\ d_{ik} + d_{jl}, d_{il} + d_{jk}\}$$

has a unique maximal element.

We remark that for a 4-point metric space the notions of Δ-genericness and Q-genericness coincide. Let $X = \{P_a, P_b, P_c, P_d\}$ be a 4-point metric space. If X is Δ-generic and if we assume without loss of generality that the minimal Gromov product at P_a is Δ_{abd}, then we have the relations

$$\Delta_{abc} - \Delta_{abd} = \tfrac{1}{2}(d_{ab} + d_{ac} - d_{bc}) - \tfrac{1}{2}(d_{ab} + d_{ad} - d_{bd}) = \tfrac{1}{2}(d_{ac} + d_{bd} - d_{ad} - d_{bc}) > 0$$
$$\Delta_{acd} - \Delta_{abd} = \tfrac{1}{2}(d_{ac} + d_{ad} - d_{cd}) - \tfrac{1}{2}(d_{ab} + d_{ad} - d_{bd}) = \tfrac{1}{2}(d_{ac} + d_{bd} - d_{ab} - d_{cd}) > 0,$$

which show that $d_{ac} + d_{bd}$ is the unique maximal element of the set $\{d_{ab} + d_{cd}, d_{ac} + d_{bd}, d_{ad} + d_{bc}\}$. Hence X is Q-generic. Conversely if $X = \{P_a, P_b, P_c, P_d\}$ is Q-generic with $d_{ac} + d_{bd}$ the maximal element of the set $\{d_{ab} + d_{cd}, d_{ac} + d_{bd}, d_{ad} + d_{bc}\}$, then X is Δ-generic with Δ_{abd} the minimal Gromov product at P_a. Likewise, Δ_{bac} is the minimal Gromov product at P_b, Δ_{cbd} at P_c and Δ_{dac} at P_d. Hence X is Δ-generic.

We now define the notion of a quadrangle structure:

Definition 4. A quadrangle structure on a Q-generic finite metric space (X, d) is a map which assigns to any 4-point subset $\{P_a, P_b, P_c, P_d\}$ of X the pair of edges corresponding to the maximal element of the set $\{d_{ab} + d_{cd}, d_{ac} + d_{bd}, d_{ad} + d_{bc}\}$.

If, for example, $d_{ac} + d_{bd}$ is the maximal element of the set $\{d_{ab} + d_{cd}, d_{ac} + d_{bd}, d_{ad} + d_{bc}\}$, then the pair $\{E_{ac}, E_{bd}\}$ is assigned to the 4-point subset $\{P_a, P_b, P_c, P_d\}$. We will say that the edges E_{ac} and E_{bd} are "diagonals" and the edges E_{ab}, E_{ad}, E_{bc}, E_{cd} are "sides".

We denote the 4-point subset $\{P_a, P_b, P_c, P_d\}$ without any restriction on the sides by $Q(a, b, c, d)$. In this notation the ordering of the indices is irrelevant. On the other hand, if $d_{ac} + d_{bd}$ is maximal, then the vertices should be ordered as (P_a, P_b, P_c, P_d) and we denote this structured quadrangle by $Q(abcd)$. Clearly, cyclic permutation and reversal of the order of the indices give equivalent quadrangles.

Definition 5. Two Q-generic finite metric spaces (X, d) and (X, d') are called Q-equivalent, if the corresponding quadrangle structures are same up to a permutation of X.

Although, quadrangle structures are defined independently of Gromov product structures, in what follows we prefer to work with quadrangle structures that are partially determined by a given Gromov product structure: If Δ_{ijk} is minimal at P_i, then for all 4-point subsets $\{P_i, P_j, P_k, P_l\}$, $d_{il} + d_{jk}$ is maximal. Thus, the Gromov product structure determines the structure of part of the quadrangles. Further discussions will be given in Section 4. It is worth to mention that for $n = 5$, the Gromov product structure determines the quadrangle structure completely.

The matrix representation for a quadrangle structure is defined as below.

Definition 6. The matrix M_Q of a quadrangle structure Q on an n-point metric space, is an $n_d \times n_d$ matrix $(n_d = n(n-1)/2)$ such that $M_Q(ab, cd) = 1$ if the edges E_{ab} and E_{cd} are diagonals in $\{P_a, P_b, P_c, P_d\}$, and $M_Q(ab, cd) = 0$ otherwise.

Note that here again the rows and columns are labeled by $(1, 2), (1, 3), \ldots, (1, n), (2, 3), (2, 4), \ldots, (2, n), \ldots, (n-1, n)$ and the rows and columns of this matrix related to the edges that are never diagonals in any of the quadrangles they occur, consist of zeros.

A Q-equivalence class thus corresponds to the orbit of the matrix representation M_Q of the quadrangle structure under the action $\tilde{P}^{-1} M_Q \tilde{P}$ where $\tilde{P}$ is the permutation matrix on the edges induced by the permutation of the vertices of the space.

4. The relation between Gromov product structures, quadrangle structures and the hypersimplex classification

In this section we illustrate the relations between Gromov product structures, quadrangle structures and the hypersimplex classification by an example.

Metrics that belong to the Δ-equivalence class I_{17}, introduced in [2] are characterized by the minimality of the Gromov products

$$\{\Delta_{126}, \Delta_{213}, \Delta_{324}, \Delta_{435}, \Delta_{546}, \Delta_{615}\}. \tag{3}$$

In [2], it was shown that the metrics numbered as 12, 13, 39, 65 and 66 in the hypersimplex classification [8] belong to this Δ-equivalence class, after relabeling of the vertices as

$$\begin{aligned} &(1, 2, 3, 4, 5, 6) \to (1, 2, 6, 4, 3, 5), \quad \text{for types } 12, 13 \\ &(1, 2, 3, 4, 5, 6) \to (1, 2, 6, 3, 5, 4), \quad \text{for types } 39, 65, 66. \end{aligned} \tag{4}$$

In order to determine the quadrangle structures of the metrics above one needs to find the maximal element of the sets $\{d_{ab} + d_{cd}, d_{ac} + d_{bd}, d_{ad} + d_{bc}\}$. For this aim, we use the Gromov product structure of class I_{17} to determine the quadrangle structure partially, which means to determine the structure of a part of the quadrangles using Gromov products.

For $n = 6$ there are 15 quadrangles. It can be seen that the structures of 12 of these quadrangles belonging to the metric class I_{17}, is determined by the minimality of Gromov products of this Δ-equivalence class. To make this clear, consider the quadrangle $Q(1, 2, 3, 4)$ for instance. For the class I_{17}, Δ_{213} is the unique minimal Gromov product at vertex 2, so $\Delta_{213} < \Delta_{214}$, or equivalently $\frac{1}{2}(d_{12} + d_{23} - d_{13}) < \frac{1}{2}(d_{12} + d_{24} - d_{14})$ or equivalently $d_{14} + d_{23} < d_{13} + d_{24}$. Similarly from $\Delta_{213} < \Delta_{234}$ one obtains that $d_{12} + d_{34} < d_{13} + d_{24}$. This means that for the quadrangle $Q(1, 2, 3, 4)$, E_{13} and E_{24} appear as diagonals and the structure of the quadrangle is determined as $Q(1234)$. This is in fact the way that we can determine the structure of a part of quadrangles by Gromov product structure. The fact that the structure of $Q(1234)$ could be determined by the minimality of Δ_{213} is depicted by an arc on the vertex 2 of $Q(1234)$ as shown in Fig. 1. The additional arc on vertex 3 of $Q(1234)$ means that this result could also be obtained by the minimality of Δ_{324}. For the class I_{17} these same calculations are done at each vertex and it happens that the structures of 12 quadrangles out of 15 are determined as shown in Fig. 1.

A priori, each of the quadrangles $Q(1, 2, 4, 5)$, $Q(1, 3, 4, 6)$ and $Q(2, 3, 5, 6)$ can have 3 structures; for example, for $Q(1, 2, 4, 5)$ diagonal pairs can be $\{E_{12}, E_{45}\}$, $\{E_{14}, E_{25}\}$ or $\{E_{15}, E_{24}\}$. But, by comparing inequalities among Gromov products, it is possible to see that the minimum of $\{\Delta_{124}, \Delta_{125}, \Delta_{145}\}$ is either Δ_{124} or Δ_{125}, but it cannot be Δ_{145}. To see this assume that Fig. 1 (without arcs) is given. The structure of $Q(1, 2, 3, 5)$ is given as $Q(1235)$ which implies $\Delta_{125} < \Delta_{135}$. Also the structure of $Q(1, 3, 4, 5)$ is given as $Q(1345)$ which implies $\Delta_{135} < \Delta_{145}$. Combining these two, results in $\Delta_{125} < \Delta_{145}$, which is to say that Δ_{145} cannot be the minimum hence $\{E_{12}, E_{45}\}$ is not a diagonal pair. It follows that the structure of the quadrangle $Q(1, 2, 4, 5)$ can be either $Q(1245)$ or $Q(1254)$. In the former case the ordering of the vertices is in agreement with the ordering of the vertices for the remaining quadrangles, hence we call $Q(1245)$ of type S, to sand for “straight” and in the latter case we call $Q(1254)$ of type T, to stand for “twisted”.

By similar arguments it can be seen that the three quadrangles $Q(1, 2, 4, 5)$, $Q(1, 3, 4, 6)$ and $Q(2, 3, 5, 6)$ can be of types S or T, which amounts to a total of 8 choices. But it can be shown that the quadrangle structures SST, STS and TSS can be mapped to each other by a permutation of indices and the same holds for types STT, STS and TTS. On the other hand, computations on the matrix of the quadrangle structures shows that the types SSS, SST, STT and TTT are inequivalent. It

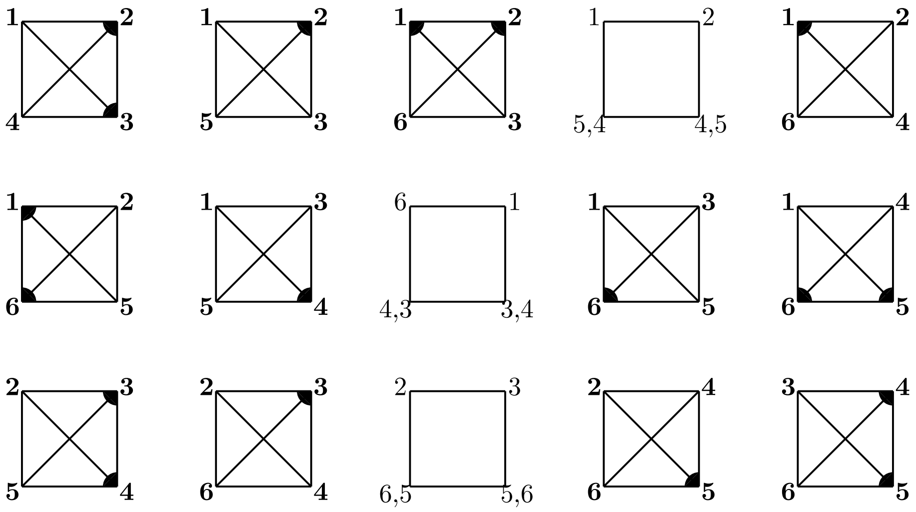


Fig. 1. The structure of the quadrangles of 6-point metric class I_{17} determined by its Gromov product structure.

follows that there are 4 distinct quadrangle types corresponding to a single Gromov product type.

$$\begin{aligned}
\text{Type}(SSS) &: Q(1245), Q(2356), Q(3461),\\
\text{Type}(SST) &: Q(1245), Q(2356), Q(3416),\\
\text{Type}(STT) &: Q(1245), Q(2365), Q(3416),\\
\text{Type}(TTT) &: Q(1254), Q(2365), Q(3416).
\end{aligned}$$

This example illustrates how the Gromov product structure partially determines the quadrangle structure of the given metric. It is worthwhile to mention that the reverse process is also important; selecting a diagonal for a quadrangle is to imply 8 relations between 12 Gromov products related to vertices of the underlying quadrangle. This means that the quadrangle structure can be used to obtain partial order relations among the Gromov products at each point. We should also note that, the metrics 12 and 13 of [8] both fall inside the structure which has all its 'free' quadrangles as twisted or TTT.

5. Split metric decompositions

A "split" $S = \{A, B\}$ of a finite set X is a partition of X into two non-empty subsets A and B. For simplicity we often identify the set of points of A with its index set. For each $P_a \in X$, we denote by $S(a)$ the subset A or B that contains P_a. Corresponding to each split S we define the pseudo-metric δ_S by

$$\delta_S(a, a') = \begin{cases} 1 & \text{if } S(a) \neq S(a'), \\ 0 & \text{if } S(a) = S(a'). \end{cases}$$

This pseudo-metric on X is called a split-metric or cut-metric on X [4]. As the split $S = \{A, B\}$ of the set X is already determined by A, this split metric is also denoted by δ_A. If $A \subset X$ has k elements, the split $\{A, B\}$ is called a k-split (or equivalently, an $(n-k)$-split). When $A = \{P_a\}$ or $\{P_a, P_b\}$ the corresponding 1-split and 2-split are simply denoted by δ_a and δ_{ab} respectively.

A metric on X is called totally split decomposable if it can be expressed as a linear combination (with non-negative coefficients) of the split metrics [1].

The isolation index of a split $S = \{A, B\}$ is defined as

$$\alpha_{A,B} = \frac{1}{2} min_{a,a' \in A, b,b' \in B}\{max\{d_{ab} + d_{a'b'}, d_{ab'} + d_{a'b}, d_{aa'} + d_{bb'}\} - (d_{aa'} + d_{bb'})\}.$$

Proposition 4. *Let (X, d) be a finite metric space with n elements P_i $(i = 1, \dots, n)$ and let $S = \{A, B\}$ be a split decomposition for X. Then,*

i. The isolation index for the 1-split with $A = \{P_a\}$ is the minimal Gromov product at P_a,
ii. If (X, d) is Q-generic, then the isolation index for the k-split with $A = \{P_{i_1}, \dots, P_{i_k}\}$ is non-zero if and only if for no pair of indices $a, a' \in A$, $E_{aa'}$ is a diagonal of the quadrangles $Q(a, a', b, b')$ where $b, b' \in B$.

Proof. **i.** If $A = \{P_a\}$ and $B = \{P_{b_1}, \dots, P_{b_{n-1}}\}$, then since $a = a'$, $d_{aa'} = 0$, the expression of the isolation index is

$$\begin{aligned}\alpha_{\{P_a\},B} &= \frac{1}{2} min_{b,b' \in B} \{max\{d_{ab} + d_{ab'}, d_{ab'} + d_{ab}, d_{bb'}\} - d_{bb'}\} \\ &= \frac{1}{2} min_{b,b' \in B} \{max\{d_{ab} + d_{ab'}, d_{bb'}\} - d_{bb'}\}.\end{aligned}$$

By triangle inequality, $d_{ab} + d_{ab'} \geq d_{bb'}$, thus

$$\alpha_{\{P_a\},B} = \frac{1}{2} min_{b,b' \in B} \{d_{ab} + d_{ab'} - d_{bb'}\}. \tag{5}$$

This means that the isolation index for the case $A = \{P_a\}$ is the minimal Gromov product at P_a.

ii. Let us now compute the isolation index for the case where A is a 2-point set $\{P_a, P_{a'}\}$, or by our abuse of notation $\{a, a'\}$. By definition

$$\alpha_{\{a,a'\},B} = \frac{1}{2} min_{b,b' \in B} \{max\{d_{ab} + d_{a'b'}, d_{ab'} + d_{a'b}, d_{aa'} + d_{bb'}\} - (d_{aa'} + d_{bb'})\}$$

If $E_{aa'}$ is a diagonal in at least one quadrangle $Q(a, a', b, b')$, then $d_{aa'} + d_{bb'}$ is maximal among $\{d_{ab} + d_{a'b'}, d_{ab'} + d_{a'b}, d_{aa'} + d_{bb'}\}$ and thus the isolation index $\alpha_{\{a,a'\},B}$ vanishes. If, on the other hand, $E_{aa'}$ is a side in every quadrangle $Q(a, a', b, b')$ for $b, b' \in B$, then $d_{aa'} + d_{bb'}$ is strictly less than $max\{d_{ab} + d_{a'b'}, d_{ab'} + d_{a'b}, d_{aa'} + d_{bb'}\}$ for all $b, b' \in B$ and consequently the isolation index $\alpha_{\{a,a'\},B}$ is strictly positive.

If A has more than 2 elements, the proof is essentially the same. ■

6. Totally split decomposable metrics for $n = 6$

In this section we will make use of the concepts developed so far, specially quadrangle structure, to construct a totally split decomposable metric for $n = 6$. Let us consider 6-point metrics that are Δ and Q-generic. We recall that there are always 6 1-splits whose isolation indices are the minimal Gromov products. We will look for 6-point metrics that have 15 splits or equivalently a total of 9 2-splits or 3-splits. According to Corollary 5 of [1] any such metric is totally split decomposable. We first show that the number of 2-splits is at most 6 in any split decomposition. For this we need to define 2-split chains and cycles.

Definition 7. A 2-split chain of length k is a collection of edges

$$\{E_{a_1a_2}, E_{a_2a_3}, \dots, E_{a_{j-1}a_j}, E_{a_ja_{j+1}}, \dots, E_{a_ka_{k+1}}\}$$

such that the isolation indices of the corresponding 2-splits are nonzero. If $a_{k+1} = a_1$, then we say that there is a 2-split cycle of length k.

Lemma 1. *In a finite Δ and Q-generic metric space (X, d) with n elements, the number of 2-splits is at most n.*

Proof. To show this consider the subgraph X_{2s} of the complete graph X consisting of the edges E_{ij} which represent the 2-splits with the belonging vertices included. We claim that the vertices of X_{2s} have degree of at most 2. To show this, assume that the vertex i has a degree of more than 2, hence we can select the edges E_{ij}, E_{ik} and E_{il} from X_{2s}. Since these edges represent the 2-splits, by Proposition 3, these edges are sides in every quadrangle that they appear in, specially $Q(i, j, k, l)$. But this cannot happen because if any two of them are sides in the quadrangle $Q(i, j, k, l)$ the third cannot be anymore. Thus the subgraph X_{2s}, is a disjoint union of some paths and cycles (which we call the maximal split chains and cycles). Now, any vertex of X belongs to at most one of these components and the number of edges of the subgraph X_{2s}, that is the number of 2-splits, is no more than the number of vertices of X_{2s}, and thus at most n. ■

By Lemma 1, for $n = 6$ there can be at most 6 2-splits and thus to use Corollary 5 of [1] we must have some 3-splits. To build up a totally split decomposable metric, first we start by selecting the 3-splits. For $n = 6$ there is a total of 10 3-splits. Assume that (A, A^c) is a 3-split, where A is a 3 point set. For any other 3-split (B, B^c), if A and B have empty intersection, then B is the complement of A. Furthermore, if A and B have 1 element in common, then A and the complement of B will have 2 elements in common, therefore we may assume that, for $n = 6$ any system of 3-splits contains

$$\delta_A = (\{P_1, P_2, P_3\}, \{P_4, P_5, P_6\}), \qquad \delta_B = (\{P_1, P_2, P_6\}, \{P_3, P_4, P_5\}).$$

(For simplicity we will refer to a split $\delta_A = (A, A^c)$ as A.)

Recall that if δ_A, where $A = \{P_a, P_b, P_c\}$, is a 3-split then by Proposition 3, E_{ab} is a side in every quadrangle that does not contain P_c as a vertex, likewise for E_{ac} and E_{bc} and the structure of the quadrangles that contain all three points P_a, P_b and P_c is not determined. Using this and going through all 15 quadrangles, it can be seen that the structure of the quadrangles $Q(1, 3, 4, 6)$, $Q(1, 3, 5, 6)$, $Q(2, 3, 4, 6)$, $Q(2, 3, 5, 6)$ are completely determined as shown in Fig. 2 as $Q(1346)$, $Q(1356)$, $Q(2346)$, $Q(2356)$.

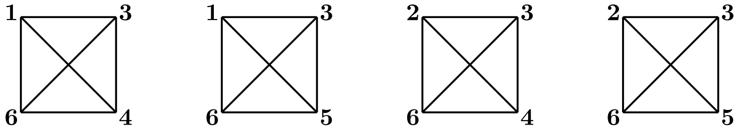


Fig. 2. The structures of the quadrangles $Q(1346)$, $Q(1356)$, $Q(2346)$, $Q(2356)$.

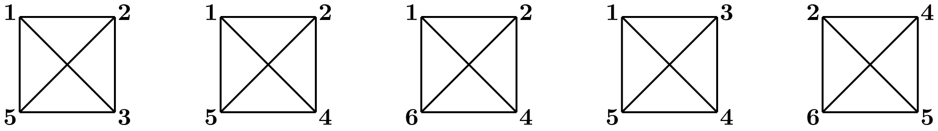


Fig. 3. The structures of the quadrangles $Q(1235)$, $Q(1245)$, $Q(1246)$, $Q(1345)$, $Q(2456)$.

To add a third 3-split to this list, first we note that in the quadrangle $Q(1346)$, E_{14} and E_{36} are diagonals, hence the 3-split $\{P_1, P_2, P_4\}$ cannot be added, otherwise it would violate Proposition 3. Similarly, in the quadrangle $Q(1356)$, E_{15} and E_{36} are diagonals, hence the splits $\{P_1, P_2, P_5\}$ and $\{P_1, P_4, P_5\}$ are eliminated. Likewise, in the quadrangle $Q(2346)$, E_{24} and E_{36} are diagonals, hence $\{P_1, P_3, P_6\}$ is eliminated. The quadrangle $Q(2356)$ is not helpful in eliminating more 3-splits from the list of 10 3-splits for a 6-point space.

It follows that the list of remaining allowable 3-splits is reduced to

$$\{P_1, P_3, P_4\},\ \{P_1, P_3, P_5\},\ \{P_1, P_4, P_6\},\ \{P_1, P_5, P_6\}.$$

Since the 3-splits above are invariant under the interchange of (P_4, P_5), as the third 3-split of the system, one can choose either $\{P_1, P_5, P_6\}$ or $\{P_1, P_3, P_5\}$. Both of these choices lead to a system of 3-splits of the form $\{P_1, P_2, P_3\}$, $\{P_1, P_2, P_6\}$, $\{P_1, P_5, P_6\}$, or $\{P_1, P_2, P_3\}$, $\{P_1, P_2, P_6\}$, $\{P_1, P_3, P_5\}$, respectively. Therefore one can choose the system of 3-splits

$$A = \{P_1, P_2, P_3\}, \qquad B = \{P_1, P_2, P_6\}, \qquad C = \{P_1, P_5, P_6\}.$$

Among the quadrangles, this third split C, when Proposition 3 is considered again, determines the structure of 5 more quadrangles as depicted in Fig. 3.

Now that the structure of 9 quadrangles are fully determined and for the quadrangles $Q(1, 2, 3, 4)$, $Q(1, 2, 3, 6)$, $Q(1, 2, 5, 6)$, $Q(1, 4, 5, 6)$, $Q(2, 3, 4, 5)$, $Q(3, 4, 5, 6)$ the edges E_{34}, E_{16}, E_{12}, E_{45}, E_{23}, E_{56} are sides, we can follow two paths: first is to add the fourth 3-split $\{P_1, P_3, P_5\}$, which in this case one can see that the structure of all of the quadrangles will be determined and addition of any 2-split will violate Proposition 3. In fact this case (sum of 4 3-splits) is the metric 339 of [8]. Second, instead of adding a fourth 3-split, one can see that without violating Proposition 3, 6 2-splits

$$\{P_1, P_2\},\ \{P_1, P_6\},\ \{P_2, P_3\},\ \{P_3, P_4\},\ \{P_4, P_5\},\ \{P_5, P_6\}.$$

can be added to the 3-splits discussed above which in turn determine the structure of all 15 quadrangles. This second case (sum of 6 2-splits plus sum of 3 3-splits) according to Corollary 5 of [1], is a totally split decomposable metric which is in fact the 6-point metric type 66 of [8]. According to [8] this is the only totally split decomposable metric for $n = 6$.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This work has been supported by the Scientific and Technological Research Council of Turkey (TUBITAK) under the project 118F412 titled “Analysis of Finite Metric Spaces via Gromov Products and their Applications to Phylogenetics”.

References

[1] H.-J. Bandelt, A.W.M. Dress, A canonical decomposition theory for metrics on a finite set, Adv. Math. 92 (1992) 47–105.
[2] A.H. Bilge, D. Çelik, Ş. Koçak, An equivalence class decomposition of finite metric spaces via Gromov products, Discrete Math. 340 (2017) 1928–1932.
[3] A.H. Bilge, M. Incegul, Matrix invariants of finite metric spaces, arXiv:2003.03335.
[4] M.M. Deza, M. Laurent, Geometry of Cuts and Metrics, in: Algorithms and Combinatorics, Springer, 2009.
[5] É. Ghys, P. de la Harpe, Sur les groupes hyperboliques d'apres Mikhael Gromov, in: Progress in Mathematics, Springer, 1990.

[6] M. Incegul, Algorithms for the equivalence class decomposition classification of finite metric spaces via Gromov products, Preprint, http://www.incegul.com/academic/algorithms.pdf.
[7] J. Koolen, A. Lesser, V. Moulton, Optimal realizations of generic five-point metrics, Eur. J. Comb. 30 (2009) 1164–1171.
[8] B. Sturmfels, J. Yu, Classification of six-point metrics, Electron. J. Comb. 11 (2004) R44.

Optimal Realizations, h-optimal Realizations and Tight Spans of Metric Spaces

Ayşe H. Bilge, Şahin Koçak, Derya Çelik, Arash M. Rezaeinazhad

Introduction

Every finite metric space (X, d) has at least one realization as a graph: let the vertices V be precisely the elements X, and construct the complete graph on these vertices with all edges x, y having weight $d(x, y)$, but we want to have a realization which is more efficient in capturing the structure of the metric space. This is generally done by deleting the paths that are realized by shortest paths.

Optimal Realizations, h-optimal Realizations and Tight Spans

Definition 1. The weighted graph $G = (V, E, w)$ is called a realization of the finite metric space (X, d) if there is a labeling function $\phi\colon X \to V$ such that for all $x,\ y \in X$ the weight of any path between $\phi(x)$ and $\phi(y)$ is equal to $d(x, y)$. Any such realization is called optimal if $||G||$ is minimal among all realizations of the metric space (X, d) [8].

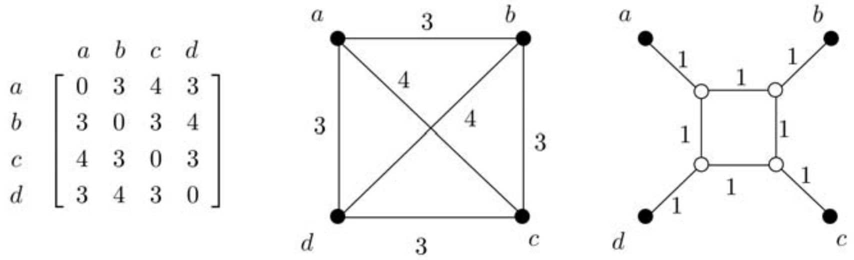


Figure 1: A 4-point metric and its optimal realization.

The problem of finding an optimal realization of a given metric was first investigated by Hakimi and Yau in 1964 [6]. It is known that in general there are several possible realizations for a metric space, thus motivating the definition of an optimal realization. The immediate question that comes in mind is that if there is a unique optimal realization or not and how one can compute it. These questions are studied by Dress [4] and Althöfer [1]. It is known that for every metric d, an optimal realization exists but is not necessarily unique [4, 7]. Althöfer [1] also showed that the realization problem is NP-hard by reducing it from the problem of finding minimal transversals in certain graphs. The following example for a 5-point space is given in [1].

$$\begin{array}{c} \\ a \\ b \\ c \\ d \\ e \end{array} \begin{array}{c} \begin{array}{ccccc} a & b & c & d & e \end{array} \\ \left[\begin{array}{ccccc} 0 & 2 & 4 & 2 & 2 \\ 2 & 0 & 2 & 4 & 2 \\ 4 & 2 & 0 & 2 & 2 \\ 2 & 4 & 2 & 0 & 2 \\ 2 & 2 & 2 & 2 & 0 \end{array}\right] \end{array}$$

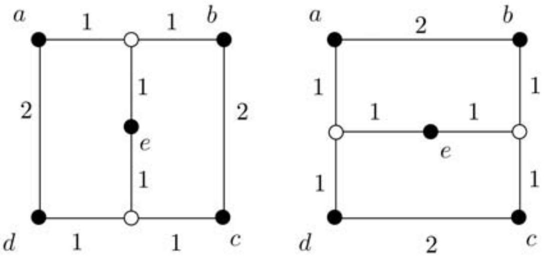


Figure 2: A 5-point metric and its two different optimal realization.

We define the concept of a hereditarily optimal or an h-optimal realization as following: For a metric space (X, d), by induction on $|X|$, the number of elements of X, if $|X| = 2$ then any optimal realization of d is defined to be h-optimal; so the h-optimal realization of a metric space with two elements a and b is simply the graph with vertices a and b joined by an edge of weight $d(a, b)$. For larger sets we require that the property of h-optimality is preserved on all subsets: if $|X| = k$ with $k \geq 3$ and if h-optimal realizations have been defined previously for all metric spaces Y with $|Y| < k$, then a realization $G = (V, E, w)$ of (X, d) is defined to be h-optimal if for any proper subset Y of X there is some subgraph $G' = (V', E', w|_{E'})$ of G such that G' is a hereditarily optimal realization of $(Y, d|_Y)$ and $\sum_{e \in E} w(e)$ s minimal among all such graphs.

To be able to show why are these definitions useful we also need to define the concept of tight span of a metric and clarify how optimal realizations and h-optimal realizations relate to each other and tight spans. In [4] it is shown that the h-optimal realization of a metric is always a subgraph of the 1-skeleton of its tight span and also the h-optimal realizations are unique and can be explicitly calculated. Beside their theoretical uses, it should also be noted that in phylogenetic analysis, h-optimal realizations can be useful tools for visualizing relationships present in biological data.

Define the unbounded polyhedron $P(X, d)$ by

$$P(X, d) = \{f \in \mathbb{R}^X | \text{ for all } x, y \in X : f(x) + f(y) \geq d(x, y)\}$$

It is clear from the definition that for all $f \in P(X, d)$, and all $x \in X$. Now define the tight span $T(X, d)$ as the set of functions in $P(X, d)$ that are minimal with respect to the pointwise partial ordering of $\mathbb{R}^X$ given by $f \leq g$ if and only if $f(x) \leq g(x)$ for all $x \in X$. For any $f \in P(X, d)$, one can associate its *tight-equality graph*, which will be denote by $K(f)$, with vertex set X and edge set

$$E = \{\{x, y\} | f(x) + f(y) = d(x, y)\}.$$

We have $f \in T(X, d)$ if and only if no vertex in $K(f)$ has degree 0.

To examine the relation between the tight span and h-optimal realizations, define a weighted graph $\Gamma_d = (V_d, E_d, w_d)$ for which

$$\begin{aligned} V_d &= \{f \in P(X, d) | \text{ K(f) is connected and not bipartite }\}, \\ Ed &= \{\{f, g\} \in \binom{V_d}{2} | \text{ K((f + g)/2) is connected and bipartite }\}, \\ w_d : E_d \to \mathbb{R}^+ \quad & w_d(\{f, g\}) = \max_{x \in X} |f(x) - g(x)|. \end{aligned}$$

Dress [4] shows that this graph Γ_d is an h-optimal realization of (X, d), and moreover, any other h-optimal realization of (X, d) is essentially isomorphic to Γ_d in the sense that it becomes isomorphic after removal of all vertices of degree two.

Now consider the 1-skeleton of the tight span, i.e. the weighted graph $T(X, d)^{(1)} = (F_0, F_1, w_d)$ that has vertex set F_0 consisting of the vertices of $P(X, d)$, edge set F_1 consisting of those $\{f, g\} \in \binom{F_0}{2}$ for which f and g are the vertices of a 1-dimensional face in $P(X, d)$, and weighting wd as defined for Γ_d above, then we have that the graph Γ_d is a subgraph of the graph $T(X, d)^{(1)}$ for any metric space (X, d).

In the study of tight span of a metric space, Dress states two conjectures: First that the set of edge-weightings of a given graph that make it an optimal realization forms a convex polytope, this polytope always consists precisely of one point, which implies that any given metric has a finite number of optimal realizations. Second any optimal realization of a given metric can be obtained from its h-optimal realization by deleting some (possibly empty) subset of edges, and possibly thereafter ignoring any vertices of degree 2.

Althöfer has provided a counterexample for both of these conjectures [1]: There exists a metric space on 6-points such that has uncountably many optimal realizations. This is given in the Figure 3 below:

$$\begin{array}{c|cccccc} & a & b & c & d & e & f \\ \hline a & 0 & 1 & 2 & 3 & 3 & 2 \\ b & 1 & 0 & 1 & 3 & 4 & 3 \\ c & 2 & 1 & 0 & 2 & 3 & 3 \\ d & 3 & 3 & 2 & 0 & 1 & 2 \\ e & 3 & 4 & 3 & 1 & 0 & 1 \\ f & 2 & 3 & 3 & 2 & 1 & 0 \end{array}$$

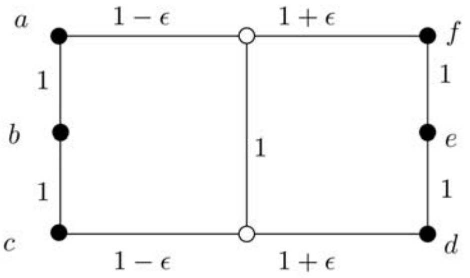


Figure 3: A 6-point metric and its optimal realizations for each $\epsilon \in [\frac{-1}{2}, \frac{1}{2}]$.

For the example above, there are two "extremal" optimal realizations. After Althöfer's counterexample, a natural question arises that "Given a graph G and a set W of edge-weightings that make G an optimal realization of some metric, can the extremals of W be derived from the h-optimal realization of the metric?"

It is easy to see that for all metrics on four or fewer elements, the optimal realization is unique, and equal to the h-optimal realization and to the 1-skeleton of the tight span. In [8], it is shown that there exist metrics on five points having more than one optimal realization, but that the maximum possible number of optimal realizations for a fixed metric on five points is three and so the answer for the question above for metrics with at most five elements is yes. The answer for metrics with more than five elements stays open. By the way the optimal realization of a metric and 1-skeleton of its tight span coincide precisely when the metric is totally split-decomposable [5].

In the Figure below, we give a family of optimal realizations of metric with Gromov type $I_{17}SSS$ [2] for all $1 \leq \gamma, \delta, \epsilon \leq 2$.

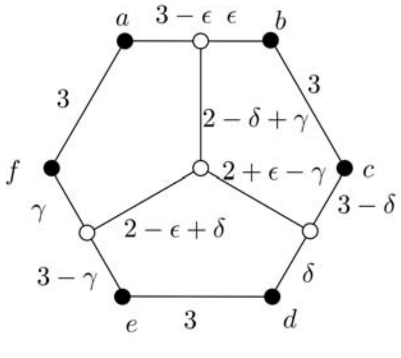


Figure 4: A family of optimal realizations for 6-point metric type $I_{17}SSS$ for all $1 \leq \gamma, \delta, \epsilon \leq 2$.

To obtain this realization for $I_{17}SSS$ in general with metric parameters, we need to first look at the graph of type I_{17} for which the distances that are realized by other paths are removed, given in Figure 5 below:

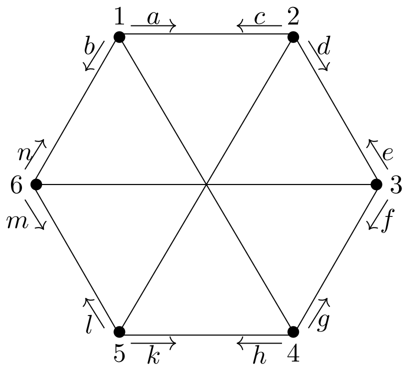


Figure 5: Graph of 6-point metric Gromov type I_{17}.

As shown in the graph above in Figure 5, we can move the sides inside the 6-gon in the directions pointed out by arrows by an amount of the parameters assigned to each one. Note that we have relabeled the points of spaces as numbers from 1 to 6. The amount of this deplacements are bound to the values given in the Table 1 below (Parameters for 4 types are given in the Appendix):

Table 1:

	SSS	SST	STT	TTT
$a \leq \Delta_{124}$	$m_{16}+p$	$m_{16}+p+r$	$m_{16}+p+q+r$	$m_{16}+q+r$
$b \leq \Delta_{146}$	$m_{12}+q$	$m_{12}+q+r$	$m_{12}+r$	$m_{12}+r$
$c \leq \Delta_{215}$	$m_{23}+p$	$m_{23}+p$	$m_{23}+p+q$	$m_{23}+q$
$d \leq \Delta_{235}$	$m_{12}+r$	m_{12}	m_{12}	m_{12}
$e \leq \Delta_{326}$	$m_{34}+r$	m_{34}	m_{34}	m_{34}
$f \leq \Delta_{346}$	$m_{23}+q$	$m_{23}+q$	m_{23}	m_{23}
$g \leq \Delta_{413}$	$m_{45}+q$	$m_{45}+q+r$	$m_{45}+r$	$m_{45}+r$
$h \leq \Delta_{415}$	$m_{34}+p$	$m_{34}+p+r$	$m_{34}+p+q+r$	$m_{34}+q+r$
$k \leq \Delta_{524}$	$m_{56}+p$	$m_{56}+p$	$m_{56}+p+q$	$m_{56}+q$
$l \leq \Delta_{526}$	$m_{45}+r$	m_{45}	m_{45}	m_{45}
$m \leq \Delta_{635}$	$m_{16}+r$	m_{16}	m_{16}	m_{16}
$n \leq \Delta_{613}$	$m_{56}+q$	$m_{56}+q$	m_{56}	m_{56}

Now if we merge arrows pointing to each other Figure 4 is obtained and we have the following parameters:

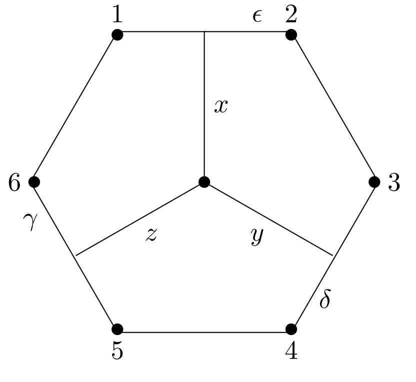


Figure 6: Optimal reduction of 6-point metric Gromov type $I_{17}SSS$.

According to Figure 6, we have:

$$\begin{aligned} d_{14} &= (d_{12}-\epsilon)+x+y+\delta, \\ d_{25} &= \epsilon+x+z+(d_{56}-\gamma), \\ d_{36} &= (d_{34}-\delta)+y+z+\gamma. \end{aligned}$$

And thus

$$x = -\delta + \gamma + \frac{1}{2}(-d_{12} + d_{14} + d_{25} + d_{34} - d_{36} - d_{56}),$$
$$y = \epsilon - \gamma + \frac{1}{2}(-d_{12} + d_{14} - d_{25} - d_{34} + d_{36} + d_{56}),$$
$$z = \delta - \epsilon + \frac{1}{2}(d_{12} - d_{14} + d_{25} - d_{34} + d_{36} - d_{56}).$$

We need the parameters to satisfy the following inequalities:
$\epsilon \leq \Delta_{215}, \quad \delta \leq \Delta_{413}, \quad \gamma \leq \Delta_{635},$
$d_{12} - \epsilon \leq \Delta_{124}, \quad d_{34} - \delta \leq \Delta_{346}, \quad d_{56} - \gamma \leq \Delta_{526}.$
So
$d_{12} - \Delta_{214} \leq \epsilon \leq \Delta_{215}, \quad d_{34} - \Delta_{346} \leq \delta \leq \Delta_{413}, \quad d_{56} - \Delta_{526} \leq \gamma \leq \Delta_{635}.$

Notice that $d_{12} - \Delta_{214} = \frac{1}{2}(d_{12} - d_{14} + d_{24}) = \frac{1}{2}(d_{12} - d_{14} + d_{23} + d_{34})$ since $d_{24} = d_{23} + d_{34}$, and $\Delta_{215} = \frac{1}{2}(d_{12} + d_{25} - d_{15}) = \frac{1}{2}(d_{12} + d_{25} - d_{16} - d_{56})$ since $d_{15} = d_{16} + d_{56}$. Thus

$$\frac{1}{2}(d_{12} + d_{23} + d_{34} - d_{14}) \leq \epsilon \leq \frac{1}{2}(d_{12} + d_{25} - d_{16} - d_{56}).$$

Likewise we have

$$\frac{1}{2}(d_{34} + d_{45} + d_{56} - d_{36}) \leq \delta \leq \frac{1}{2}(d_{14} + d_{34} - d_{12} - d_{23}),$$

and

$$\frac{1}{2}(d_{56} + d_{12} + d_{16} - d_{25}) \leq \gamma \leq \frac{1}{2}(d_{36} + d_{56} - d_{34} - d_{45}),$$

We will ge the parameterization of the metrics in the appendix. According to this parameters, for type SSS we have:

$$m_{23} \leq \epsilon \leq m_{23} + p, \quad m_{45} \leq \delta \leq m_{45} + q, \quad m_{16} \leq \gamma \leq m_{16} + r.$$

and

$$x = -\delta + \gamma + m_{12} - m_{16} + m_{45} + q,$$
$$y = \epsilon - \gamma + m_{16} - m_{23} + m_{34} + r,$$
$$z = \delta - \epsilon + m_{13} - m_{45} + m_{56} + p.$$

The realization given in Figure 6 is optimal realization of type $I_{17}SSS$ for γ, δ and ϵ satisfying the above inequalities. For this realization, if we let $\epsilon = m_{23} + u$, $\delta = m_{45} + v$ and $\gamma = m_{16} + w$ for $u \leq p$, $v \leq q$ and $w \leq r$, we will have:

$$x = m_{12} + w + (q - v), \quad y = m_{34} + u + (r - w), \quad z = m_{56} + v + (p - u).$$

It is also possible to obtain suboptimal realizations of metric type I_{17}. For this, one needs to shift the edges inside of the hexagon given in Figure 5 toward the arrows assigned to its two ends. This shifts of course are not arbitrary and bound to the restrictions given in Table 1. The results are given in the Figures below:

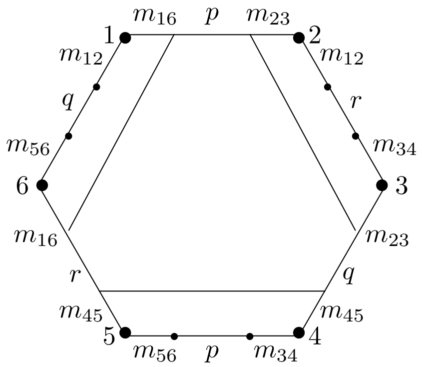


Figure 7: Suboptimal realization of metric type $I_{17}SSS$.

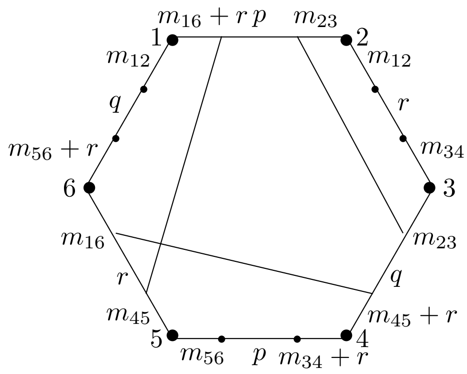


Figure 8: Suboptimal realization of metric type $I_{17}SST$.

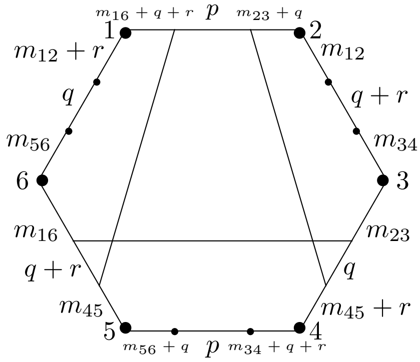


Figure 9: Suboptimal realization of metric type $I_{17}STT$.

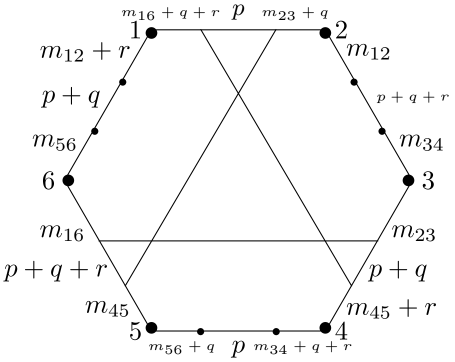


Figure 10: Suboptimal realization of metric type $I_{17}TTT$.

Optimal Realizations of 5-Point Metric Spaces

In the following, we will start with the pendant free reductions and use certain "moves" as defined in [3] to reduce the total weight and reach the optimal realizations. This kind of operations are generally done by adjoining new vertices to the original graph, which in this case the added vertices are called secondary vertices and the original vertices as primary, discarding some edges or adding new edges between the enlarged set of vertices and assigning weights to the new edges in a way that the distance between primary nodes are unchanged but the weight of the graph, namely $||G||$, is reduced.

The first move, which is called *joining edges*, is done in the following way: Consider a vertex u and all (or some) of the other nodes $v_1, v_2, \ldots, v_k$ of G, which are neighbors of u. Calculate the Gromov products of all triangles $T_{uv_iv_j}$ with $1 \leq i, j \leq k$ at vertex u and call the minimum m_u. Now delete all the edges between u and v_i's, introduce a new vertex v and connect v_i's to v by edges of weight $w_{uv_i} - m_u$ for $1 \leq i \leq k$ and also u to v by an edge of weight m_u; hence the nodes v_i become connected to u by two edges through v and the total weight of the graph is reduced by an amount of $(k-1)m_u$.

The second move, which is called *edge removing*, is done by deleting the edge between two nodes u and v if it can be avoided by a shortest path. This move reduces $||G||$ by an amount of the weight of the deleted edge.

The "$\Delta - Y$" transform is a consequence of the above moves and can be applied to any triangle with 1-connected vertices in G. It is called a $\Delta - Y$ transform, because a triangle shape (Δ) turns to a Y shape after the operation.

We should also note that what we mean by *underlying graph of a metric*, is the complete graph with the same set of vertices as the metric space and all the edges with weight d_{ij} removed for which there is a point in space p_k such that $d_{ij} = d_{ik} + d_{kj}$.

For Type A with the Gromov product structure as $\{\Delta_{125}, \Delta_{213}, \Delta_{324}, \Delta_{435}, \Delta_{514}\}$, when edge removing operations are applied and passed to pendant-free reduction, a 5-cycle given as below is obtained. The optimal realization given in [8] is a 5-cycle with edges connected to each of its nodes (Type (a) of [8]).

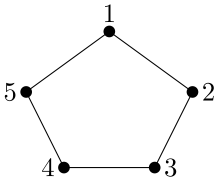


Figure 11: Optimal reduction of metric Type A

For Type B with the Gromov product structure as $\{\Delta_{125}, \Delta_{213}, \Delta_{325}, \Delta_{425}, \Delta_{514}\}$, the underlying graph is given below:

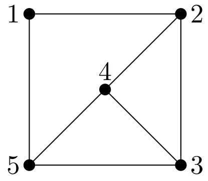


Figure 12: Underlying graph of metric Type B

By applying a $\Delta - Y$ transform to T_{345} we have the following:

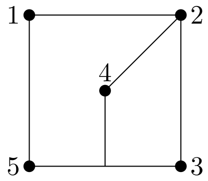


Figure 13: Graph with $\Delta - Y$ transformed.

In this step, one can follow two different approaches which reduce the metric to Type (b) or (c) of [8]. To observe the process closely we need to point out that the parameterization of Type B is given as follows:

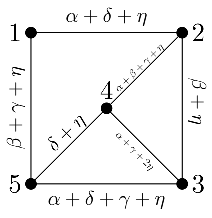


Figure 14: Underlying graph of metric Type B with distances parameterized

Here we have $\Delta_{345} = \alpha+\gamma+\eta$, $\Delta_{435} = \eta$ and $\Delta_{534} = \delta$, and applying a $\Delta - Y$ transform to T_{345} will result in the following:

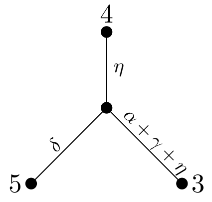


Figure 15: T_{345} of Type B after $\Delta - Y$ transform

So the Type B with parameters are as following:

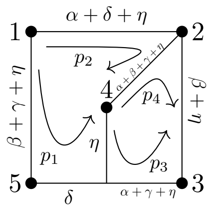


Figure 16: Metric Type B with the parameters

Now according to the graph above, we have d_{14} equal to $\beta+\delta+\gamma+2\eta$ (path p_1) or equal to $2\alpha+\beta+\gamma+\delta+2\eta$ (path p_2). Path p_2 is longer than path p_1 by an amount of 2α. Likewise d_{34} is equal to $\alpha+\gamma+2\eta$ (path p_3) or equal to $\alpha+2\beta+\gamma+2\eta$ (path p_4). Here path p_4 is longer than path p_3 by a difference of 2β. It should be noted that $\alpha = \Delta_{214}$ and $\beta = \Delta_{234}$ and two scenarios are possible: either $\alpha > \beta$ or $\beta > \alpha$. If $\alpha > \beta$, in order to decrease the total weight of the graph, we will introduce a new node called v on the edge joining 1 to 2 as shown below:

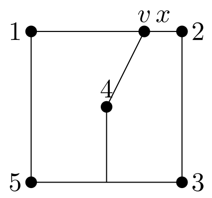


Figure 17: Reduction of Type B to (b)

This will reduce the total weight as $x = \Delta_{214}$ and that results the Type B to reduced into (b) of [8] and the metric will be as following:

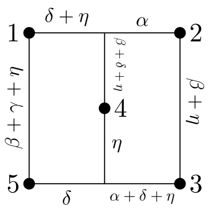


Figure 18: Reduction of Type B to (b) when $\alpha > \beta$ (parameters given)

In the other case, when $\beta > \alpha$, if we do the same operation as before, but this time for the edge joining 2 to 3 we will have the following reduction:

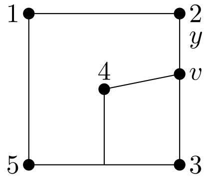


Figure 19: Reduction of Type B to (c)

This reduces the weight of graph as $y = \Delta_{234}$ and turns it into Type (c) given as below:

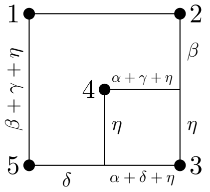


Figure 20: Reduction of Type B to (c) when $\beta > \alpha$ (parameters given)

For Type C which the underlying graph with the parameters given is depicted below, the following can be done:

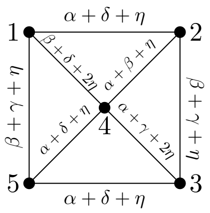


Figure 21: Underlying graph of metric Type C with the metric parameterized

Since $\Delta_{124} = \delta + \eta$, $\Delta_{214} = \alpha$ and $\Delta_{412} = \beta + \eta$, applying a $\Delta - Y$ transform to T_{124} will result in the following:

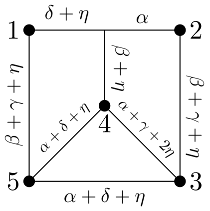


Figure 22: Type C with a $\Delta - Y$ transform applied to T_{124}

Finally considering that $\Delta_{345} = \alpha + \eta$, $\Delta_{435} = \gamma + \eta$ and $\Delta_{534} = \delta$, applying another $\Delta - Y$ transform to T_{345} will result in the following:

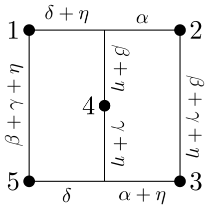


Figure 23: Type C with a second $\Delta - Y$ transform applied to T_{345} and reduced to (b)

Appendix: Type I_{17} Metric Parameters

Type SSS:

$$\begin{aligned}
d_{12} &= m_{16} + m_{23} + p,\\
d_{23} &= m_{12} + m_{34} + r,\\
d_{34} &= m_{23} + m_{45} + q,\\
d_{45} &= m_{34} + m_{56} + p,\\
d_{56} &= m_{16} + m_{45} + r,\\
d_{16} &= m_{12} + m_{56} + q,\\
d_{14} &= m_{12} + m_{16} + m_{34} + m_{45} + p + q + r,\\
d_{25} &= m_{12} + m_{23} + m_{45} + m_{56} + p + q + r,\\
d_{36} &= m_{16} + m_{23} + m_{34} + m_{56} + p + q + r.
\end{aligned}$$

$$\begin{aligned}
\Delta_{123} &= m_{16} + m_{23} + p,\\
\Delta_{124} &= m_{16} + p,\\
\Delta_{125} &= m_{16},\\
\Delta_{126} &= 0,\\
\Delta_{134} &= m_{12} + m_{16} + m_{34} + p + r,\\
\Delta_{135} &= m_{12} + m_{16} + r,\\
\Delta_{136} &= m_{12},\\
\Delta_{145} &= m_{12} + m_{16} + m_{45} + q + r,\\
\Delta_{146} &= m_{12} + q,\\
\Delta_{156} &= m_{12} + m_{56} + q.
\end{aligned}$$

$$\begin{aligned}
\Delta_{213} &= 0,\\
\Delta_{214} &= m_{23},\\
\Delta_{215} &= m_{23} + p,\\
\Delta_{216} &= m_{16} + m_{23} + p,\\
\Delta_{234} &= m_{12} + m_{34} + r,\\
\Delta_{235} &= m_{12} + r,\\
\Delta_{236} &= m_{12},\\
\Delta_{245} &= m_{12} + m_{23} + m_{45} + q + r,\\
\Delta_{246} &= m_{12} + m_{23} + q,\\
\Delta_{256} &= m_{12} + m_{23} + m_{56} + p + q.
\end{aligned}$$

$$\begin{aligned}
\Delta_{312} &= m_{12} + m_{34} + r,\\
\Delta_{314} &= m_{23},\\
\Delta_{315} &= m_{23} + m_{34} + p,\\
\Delta_{316} &= m_{16} + m_{23} + m_{34} + p + r,\\
\Delta_{324} &= 0,\\
\Delta_{325} &= m_{34},\\
\Delta_{326} &= m_{34} + r,\\
\Delta_{345} &= m_{23} + m_{45} + q,\\
\Delta_{346} &= m_{23} + q,\\
\Delta_{356} &= m_{23} + m_{34} + m_{56} + p + q.
\end{aligned}$$

$$\begin{aligned}
\Delta_{412} &= m_{12} + m_{34} + m_{45} + q + r,\\
\Delta_{413} &= m_{45} + q,\\
\Delta_{415} &= m_{34} + p,\\
\Delta_{416} &= m_{16} + m_{34} + m_{45} + p + r,\\
\Delta_{423} &= m_{23} + m_{45} + q,\\
\Delta_{425} &= m_{34},\\
\Delta_{426} &= m_{34} + m_{45} + r,\\
\Delta_{435} &= 0,\\
\Delta_{436} &= m_{45},\\
\Delta_{456} &= m_{34} + m_{56} + p.
\end{aligned}$$

$$\begin{aligned}
\Delta_{512} &= m_{12} + m_{45} + m_{56} + q + r, \\
\Delta_{513} &= m_{45} + m_{56} + q, \\
\Delta_{514} &= m_{56}, \\
\Delta_{516} &= m_{16} + m_{45} + r, \\
\Delta_{523} &= m_{23} + m_{45} + m_{56} + p + q, \\
\Delta_{524} &= m_{56} + p, \\
\Delta_{526} &= m_{45} + r, \\
\Delta_{534} &= m_{34} + m_{56} + p, \\
\Delta_{536} &= m_{45}, \\
\Delta_{546} &= 0.
\end{aligned}$$

$$\begin{aligned}
\Delta_{612} &= m_{12} + m_{56} + q, \\
\Delta_{613} &= m_{56} + q, \\
\Delta_{614} &= m_{56}, \\
\Delta_{615} &= 0, \\
\Delta_{623} &= m_{16} + m_{23} + m_{56} + p + q, \\
\Delta_{624} &= m_{16} + m_{56} + p, \\
\Delta_{625} &= m_{16}, \\
\Delta_{634} &= m_{16} + m_{34} + m_{56} + p + r, \\
\Delta_{635} &= m_{16} + r, \\
\Delta_{645} &= m_{16} + m_{45} + r.
\end{aligned}$$

Type SST:

$$\begin{aligned}
d_{12} &= m_{16} + m_{23} + p + r, \\
d_{23} &= m_{12} + m_{34} + r, \\
d_{34} &= m_{23} + m_{45} + q + r, \\
d_{45} &= m_{34} + m_{56} + p + r, \\
d_{56} &= m_{16} + m_{45} + r, \\
d_{16} &= m_{12} + m_{56} + q + r, \\
d_{14} &= m_{12} + m_{16} + m_{34} + m_{45} + p + q + 3r, \\
d_{25} &= m_{12} + m_{23} + m_{45} + m_{56} + p + q + r, \\
d_{36} &= m_{16} + m_{23} + m_{34} + m_{56} + p + q + r.
\end{aligned}$$

$$\begin{aligned}
\Delta_{123} &= m_{16} + m_{23} + p + r, \\
\Delta_{124} &= m_{16} + p + r, \\
\Delta_{125} &= m_{16} + r, \\
\Delta_{126} &= 0, \\
\Delta_{134} &= m_{12} + m_{16} + m_{34} + p + 2r, \\
\Delta_{135} &= m_{12} + m_{16} + r, \\
\Delta_{136} &= m_{12} + r, \\
\Delta_{145} &= m_{12} + m_{16} + m_{45} + q + 2r, \\
\Delta_{146} &= m_{12} + q, \\
\Delta_{156} &= m_{12} + m_{56} + q + r.
\end{aligned}$$

$$\begin{aligned}
\Delta_{213} &= 0, \\
\Delta_{214} &= m_{23}, \\
\Delta_{215} &= m_{23} + p, \\
\Delta_{216} &= m_{16} + m_{23} + p + r, \\
\Delta_{234} &= m_{12} + m_{34} + r, \\
\Delta_{235} &= m_{12}, \\
\Delta_{236} &= m_{12} + r, \\
\Delta_{245} &= m_{12} + m_{23} + m_{45} + q + r, \\
\Delta_{246} &= m_{12} + m_{23} + q + r, \\
\Delta_{256} &= m_{12} + m_{23} + m_{56} + p + q + r.
\end{aligned}$$

$$\begin{aligned}
\Delta_{312} &= m_{12} + m_{34} + r, \\
\Delta_{314} &= m_{23}, \\
\Delta_{315} &= m_{23} + m_{34} + p + r, \\
\Delta_{316} &= m_{16} + m_{23} + m_{34} + p + r, \\
\Delta_{324} &= 0, \\
\Delta_{325} &= m_{34} + r, \\
\Delta_{326} &= m_{34}, \\
\Delta_{345} &= m_{23} + m_{45} + q + r, \\
\Delta_{346} &= m_{23} + q, \\
\Delta_{356} &= m_{23} + m_{34} + m_{56} + p + q + r.
\end{aligned}$$

$$\begin{aligned}
\Delta_{412} &= m_{12} + m_{34} + m_{45} + q + 2r, \\
\Delta_{413} &= m_{45} + q + r, \\
\Delta_{415} &= m_{34} + p + r, \\
\Delta_{416} &= m_{16} + m_{34} + m_{45} + p + 2r, \\
\Delta_{423} &= m_{23} + m_{45} + q + r, \\
\Delta_{425} &= m_{34} + r, \\
\Delta_{426} &= m_{34} + m_{45} + r, \\
\Delta_{435} &= 0, \\
\Delta_{436} &= m_{45} + r, \\
\Delta_{456} &= m_{34} + m_{56} + p + r.
\end{aligned}$$

$$\begin{aligned}
\Delta_{512} &= m_{12} + m_{45} + m_{56} + q + r, \\
\Delta_{513} &= m_{45} + m_{56} + q + r, \\
\Delta_{514} &= m_{56}, \\
\Delta_{516} &= m_{16} + m_{45} + r, \\
\Delta_{523} &= m_{23} + m_{45} + m_{56} + p + q + r, \\
\Delta_{524} &= m_{56} + p, \\
\Delta_{526} &= m_{45}, \\
\Delta_{534} &= m_{34} + m_{56} + p + r, \\
\Delta_{536} &= m_{45} + r, \\
\Delta_{546} &= 0.
\end{aligned}$$

$$\begin{aligned}
\Delta_{612} &= m_{12} + m_{56} + q + r, \\
\Delta_{613} &= m_{56} + q, \\
\Delta_{614} &= m_{56}, \\
\Delta_{615} &= 0, \\
\Delta_{623} &= m_{16} + m_{23} + m_{56} + p + q + r, \\
\Delta_{624} &= m_{16} + m_{56} + p + r, \\
\Delta_{625} &= m_{16} + r, \\
\Delta_{634} &= m_{16} + m_{34} + m_{56} + p + r, \\
\Delta_{635} &= m_{16}, \\
\Delta_{645} &= m_{16} + m_{45} + r.
\end{aligned}$$

Type STT:

$$\begin{aligned}
d_{12} &= m_{16} + m_{23} + p + 2q + r, \\
d_{23} &= m_{12} + m_{34} + q + r, \\
d_{34} &= m_{23} + m_{45} + q + r, \\
d_{45} &= m_{34} + m_{56} + p + 2q + r, \\
d_{56} &= m_{16} + m_{45} + q + r, \\
d_{16} &= m_{12} + m_{56} + q + r, \\
d_{14} &= m_{12} + m_{16} + m_{34} + m_{45} + p + 2q + 3r, \\
d_{25} &= m_{12} + m_{23} + m_{45} + m_{56} + p + 2q + r, \\
d_{36} &= m_{16} + m_{23} + m_{34} + m_{56} + p + 2q + r.
\end{aligned}$$

$$\begin{aligned}
\Delta_{123} &= m_{16} + m_{23} + p + 2q + r, \\
\Delta_{124} &= m_{16} + p + q + r, \\
\Delta_{125} &= m_{16} + q + r, \\
\Delta_{126} &= 0, \\
\Delta_{134} &= m_{12} + m_{16} + m_{34} + p + 2q + 2r, \\
\Delta_{135} &= m_{12} + m_{16} + q + r, \\
\Delta_{136} &= m_{12} + q + r, \\
\Delta_{145} &= m_{12} + m_{16} + m_{45} + q + 2r, \\
\Delta_{146} &= m_{12} + r, \\
\Delta_{156} &= m_{12} + m_{56} + q + r.
\end{aligned}$$

$$\begin{aligned}
\Delta_{213} &= 0, \\
\Delta_{214} &= m_{23} + q, \\
\Delta_{215} &= m_{23} + p + q, \\
\Delta_{216} &= m_{16} + m_{23} + p + 2q + r, \\
\Delta_{234} &= m_{12} + m_{34} + q + r, \\
\Delta_{235} &= m_{12}, \\
\Delta_{236} &= m_{12} + q + r, \\
\Delta_{245} &= m_{12} + m_{23} + m_{45} + q + r, \\
\Delta_{246} &= m_{12} + m_{23} + q + r, \\
\Delta_{256} &= m_{12} + m_{23} + m_{56} + p + 2q + r.
\end{aligned}$$

$$\begin{aligned}
\Delta_{312} &= m_{12} + m_{34} + q + r, \\
\Delta_{314} &= m_{23} + q, \\
\Delta_{315} &= m_{23} + m_{34} + p + 2q + r, \\
\Delta_{316} &= m_{16} + m_{23} + m_{34} + p + 2q + r, \\
\Delta_{324} &= 0, \\
\Delta_{325} &= m_{34} + q + r, \\
\Delta_{326} &= m_{34}, \\
\Delta_{345} &= m_{23} + m_{45} + q + r, \\
\Delta_{346} &= m_{23}, \\
\Delta_{356} &= m_{23} + m_{34} + m_{56} + p + 2q + r.
\end{aligned}$$

$$\begin{aligned}
\Delta_{412} &= m_{12} + m_{34} + m_{45} + q + 2r, \\
\Delta_{413} &= m_{45} + r, \\
\Delta_{415} &= m_{34} + p + q + r, \\
\Delta_{416} &= m_{16} + m_{34} + m_{45} + p + 2q + 2r, \\
\Delta_{423} &= m_{23} + m_{45} + q + r, \\
\Delta_{425} &= m_{34} + q + r, \\
\Delta_{426} &= m_{34} + m_{45} + q + r, \\
\Delta_{435} &= 0, \\
\Delta_{436} &= m_{45} + q + r, \\
\Delta_{456} &= m_{34} + m_{56} + p + 2q + r.
\end{aligned}$$

$$\begin{aligned}
\Delta_{512} &= m_{12} + m_{45} + m_{56} + q + r,\\
\Delta_{513} &= m_{45} + m_{56} + q + r,\\
\Delta_{514} &= m_{56} + q,\\
\Delta_{516} &= m_{16} + m_{45} + q + r,\\
\Delta_{523} &= m_{23} + m_{45} + m_{56} + p + 2q + r,\\
\Delta_{524} &= m_{56} + p + q,\\
\Delta_{526} &= m_{45},\\
\Delta_{534} &= m_{34} + m_{56} + p + 2q + r,\\
\Delta_{536} &= m_{45} + q + r,\\
\Delta_{546} &= 0.
\end{aligned}$$

$$\begin{aligned}
\Delta_{612} &= m_{12} + m_{56} + q + r,\\
\Delta_{613} &= m_{56},\\
\Delta_{614} &= m_{56} + q,\\
\Delta_{615} &= 0,\\
\Delta_{623} &= m_{16} + m_{23} + m_{56} + p + 2q + r,\\
\Delta_{624} &= m_{16} + m_{56} + p + 2q + r,\\
\Delta_{625} &= m_{16} + q + r,\\
\Delta_{634} &= m_{16} + m_{34} + m_{56} + p + 2q + r,\\
\Delta_{635} &= m_{16},\\
\Delta_{645} &= m_{16} + m_{45} + q + r.
\end{aligned}$$

Type TTT:

$$\begin{aligned}
d_{12} &= m_{16} + m_{23} + p + 2q + r,\\
d_{23} &= m_{12} + m_{34} + p + q + r,\\
d_{34} &= m_{23} + m_{45} + p + q + r,\\
d_{45} &= m_{34} + m_{56} + p + 2q + r,\\
d_{56} &= m_{16} + m_{45} + p + q + r,\\
d_{16} &= m_{12} + m_{56} + p + q + r,\\
d_{14} &= m_{12} + m_{16} + m_{34} + m_{45} + p + 2q + 3r,\\
d_{25} &= m_{12} + m_{23} + m_{45} + m_{56} + p + 2q + r,\\
d_{36} &= m_{16} + m_{23} + m_{34} + m_{56} + p + 2q + r.
\end{aligned}$$

$$\begin{aligned}
\Delta_{123} &= m_{16} + m_{23} + p + 2q + r,\\
\Delta_{124} &= m_{16} + q + r,\\
\Delta_{125} &= m_{16} + p + q + r,\\
\Delta_{126} &= 0,\\
\Delta_{134} &= m_{12} + m_{16} + m_{34} + p + 2q + 2r,\\
\Delta_{135} &= m_{12} + m_{16} + p + q + r,\\
\Delta_{136} &= m_{12} + p + q + r,\\
\Delta_{145} &= m_{12} + m_{16} + m_{45} + p + q + 2r,\\
\Delta_{146} &= m_{12} + r,\\
\Delta_{156} &= m_{12} + m_{56} + p + q + r.
\end{aligned}$$

$$\begin{aligned}
\Delta_{213} &= 0,\\
\Delta_{214} &= m_{23} + p + q,\\
\Delta_{215} &= m_{23} + q,\\
\Delta_{216} &= m_{16} + m_{23} + p + 2q + r,\\
\Delta_{234} &= m_{12} + m_{34} + p + q + r,\\
\Delta_{235} &= m_{12},\\
\Delta_{236} &= m_{12} + p + q + r,\\
\Delta_{245} &= m_{12} + m_{23} + m_{45} + p + q + r,\\
\Delta_{246} &= m_{12} + m_{23} + p + q + r,\\
\Delta_{256} &= m_{12} + m_{23} + m_{56} + p + 2q + r.
\end{aligned}$$

$$\begin{aligned}
\Delta_{312} &= m_{12} + m_{34} + p + q + r,\\
\Delta_{314} &= m_{23} + p + q,\\
\Delta_{315} &= m_{23} + m_{34} + p + 2q + r,\\
\Delta_{316} &= m_{16} + m_{23} + m_{34} + p + 2q + r,\\
\Delta_{324} &= 0,\\
\Delta_{325} &= m_{34} + p + q + r,\\
\Delta_{326} &= m_{34},\\
\Delta_{345} &= m_{23} + m_{45} + p + q + r,\\
\Delta_{346} &= m_{23},\\
\Delta_{356} &= m_{23} + m_{34} + m_{56} + p + 2q + r.
\end{aligned}$$

$$\begin{aligned}
\Delta_{412} &= m_{12} + m_{34} + m_{45} + p + q + 2r,\\
\Delta_{413} &= m_{45} + r,\\
\Delta_{415} &= m_{34} + q + r,\\
\Delta_{416} &= m_{16} + m_{34} + m_{45} + p + 2q + 2r,\\
\Delta_{423} &= m_{23} + m_{45} + p + q + r,\\
\Delta_{425} &= m_{34} + p + q + r,\\
\Delta_{426} &= m_{34} + m_{45} + p + q + r,\\
\Delta_{435} &= 0,\\
\Delta_{436} &= m_{45} + p + q + r,\\
\Delta_{456} &= m_{34} + m_{56} + p + 2q + r.
\end{aligned}$$

$$\begin{aligned}
\Delta_{512} &= m_{12} + m_{45} + m_{56} + p + q + r,\\
\Delta_{513} &= m_{45} + m_{56} + p + q + r,\\
\Delta_{514} &= m_{56} + p + q,\\
\Delta_{516} &= m_{16} + m_{45} + p + q + r,\\
\Delta_{523} &= m_{23} + m_{45} + m_{56} + p + 2q + r,\\
\Delta_{524} &= m_{56} + q,\\
\Delta_{526} &= m_{45},\\
\Delta_{534} &= m_{34} + m_{56} + p + 2q + r,\\
\Delta_{536} &= m_{45} + p + q + r,\\
\Delta_{546} &= 0.
\end{aligned}$$

$$\begin{aligned}
\Delta_{612} &= m_{12} + m_{56} + p + q + r,\\
\Delta_{613} &= m_{56},\\
\Delta_{614} &= m_{56} + p + q,\\
\Delta_{615} &= 0,\\
\Delta_{623} &= m_{16} + m_{23} + m_{56} + p + 2q + r,\\
\Delta_{624} &= m_{16} + m_{56} + p + 2q + r,\\
\Delta_{625} &= m_{16} + p + q + r,\\
\Delta_{634} &= m_{16} + m_{34} + m_{56} + p + 2q + r,\\
\Delta_{635} &= m_{16},\\
\Delta_{645} &= m_{16} + m_{45} + p + q + r.
\end{aligned}$$

References

[1] Althöfer, I. 1988. On optimal realizations of finite metric spaces by graphs, Discrete & Computational Geometry, 3, 103-122.

[2] Bilge, A. H., Çelik, D., Koçak, Ş. and Rezaeinazhad, A. M., 2021. Gromov product structures, quadrangle structures and split metric decompositions for finite metric spaces. Discrete Mathematics, 344(6), p.112358.

[3] Çelik, D., Bilge, A.H. and Koçak, Ş., 2015. Optimal embeddings of finite metric spaces into graphs. Anadolu Üniversitesi Bilim Ve Teknoloji Dergisi-B Teorik Bilimler, 3(2), pp.133-147.

[4] Dress, A. W. M. 1984. Trees, tight extensions of metric spaces, and the cohomological dimension of certain groups: a note on combinatorial properties of metric spaces, Adv. in Math., 53, 321 – 402.

[5] Dress, A., Huber, K.T., Lesser, A. and Moulton, V., 2006. Hereditarily optimal realizations of consistent metrics. Annals of Combinatorics, 10(1), pp.63-76.

[6] Hakimi, S. L. and Yau, S. S. 1964. Distance matrix of a graph and its realizability, Quart. AppL Math., 22, 305-317.

[7] lmrich, W., Simres-Pereira, J. M. S. and Zamfirescu, C. M. 1984. On optimal embeddings of metrics in graphs, J. Combin. Theory Ser. B, 36, 1-15.

[8] Koolen, J., A. Lesser, V. Moulton 2009. Optimal Realizations of Generic Five-Point Metrics, European J. of Combin. 30, 1164-1171.

Tight Reductions of Finite Metric Spaces

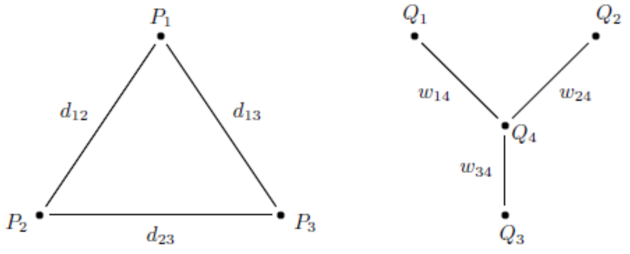
Derya Çelik
Eskişehir Teknik Üniversitesi

Optimal reduction of a finite metric space is its isometric embedding in a weighted graph with minimal total weight. The optimal embedding problem is NP hard and only a few results are known.

In the present work we discuss “sub-optimal” of “local optimal” of “tight” realizations o finite metric spaces. These are embeddings of a metric space in a graph in such a way that the total weight is no more reducible by “moves” (Bilge vd, 2015). These moves are illustrated here for 4 and 5 point spaces. An alternative approach for obtaining tight reductions, is to use triangulations of the pendant free reduction and applying Delta-Y transform to disjoint triangles.

The Delta-Y transform:

The building block of this approach is the so-called Delta-Y transform that consists of the representation of a 3-point metric space space on a graph with 4 nodes, as shown below.



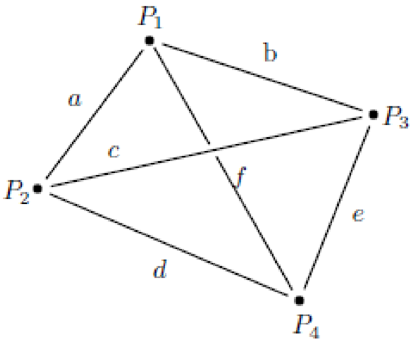
It can be seen that the weights w_{i4} are the (minimal) Gromov products at the nodes Q_i i=1,2,3 and $w_{i4}+w_{j4}=d_{ij}$..

Delta-moves.

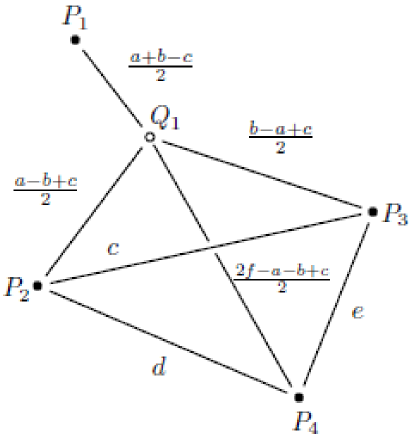
For metric spaces with 4 or more elements, the edges connected to a node P_i can be joined together by an amount equal to the value of the minimal Gromov product at that point,, say $\Delta(P_iP_jP_k)$ up to a new point Q_i. After that step, if for example $\Delta(P_iP_jP_k)$ is minima, the edge Q_jQ_k can be removes because the distance is realized by an indirect path. After that step not all edges connected to Q_i can be joined. Nevertheless, a subset of these edges can be chosen and they can be joined by an amount dictated by the amount of the minimal Gromov products involving these edges and a corresponding edge can be removed. These operationa are called Δ-moves. Successive application of Δ-moves and Δ-Y transformations leads, in the generic case, to a graph in which all nodes have degree 1 or 3. Such a graph representation is not necessarily optimal. But as its weight cannot be decreased bu local operations we call it “local optimal”, or “sub-optimal” or “tight”.

Optimal reductions of 4-point spaces:

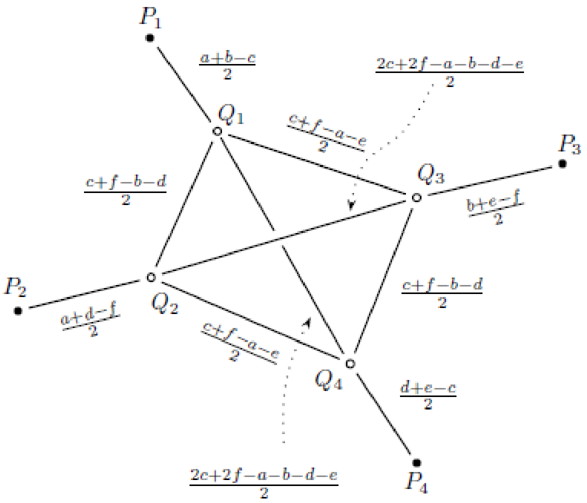
Let $X_4=\{P_1,P_2,P_3,P_4\}$ be a generic 4-point metric space. Assume $\Delta(P_1,P_2,P_3)$ is minimal at P_1. It can be shown that this assumption determines the structure of the space. The complete graph that represents X_4 is given below.



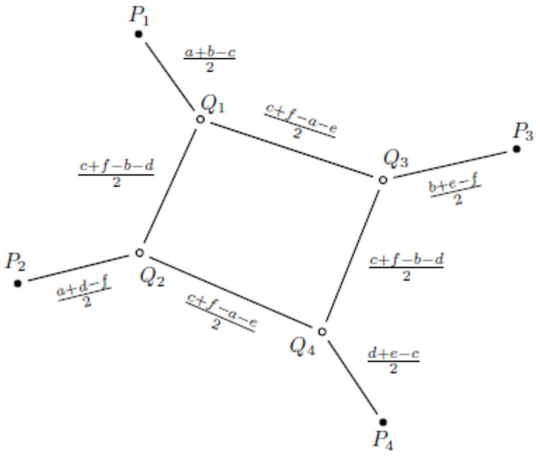
We will illustrate the reduction procedure on this example. If Δ(x,y,z) is minimal, the point x can be joined to a new point x’ by a path of length Δ(x,y,z) so that the path (yx’z) is equal to the edge (yz), hence the edge (yz) is removed. and the graph below is obtained.



Subsequently, the same operations are applied to all points and the graph below is obtained.

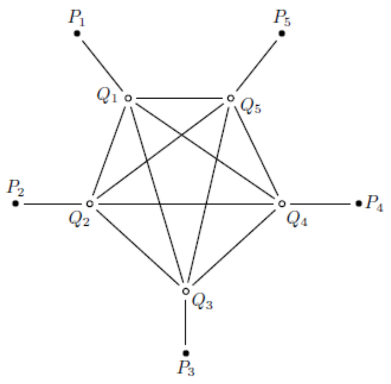


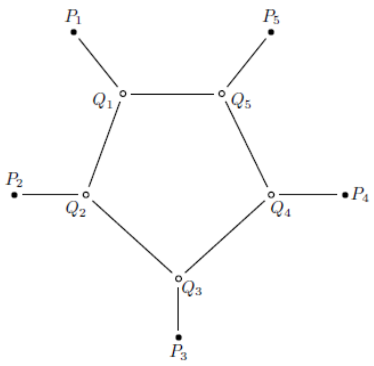
The last move is the removal of unnecessary edges. From the previous graph it can be seen the length of the diagonals are equal to the sum of the sides forming triangles, hence they can be removed to give the graph below, which is known to be the optimal reduction.



Optimal reductions of 5-point spaces:

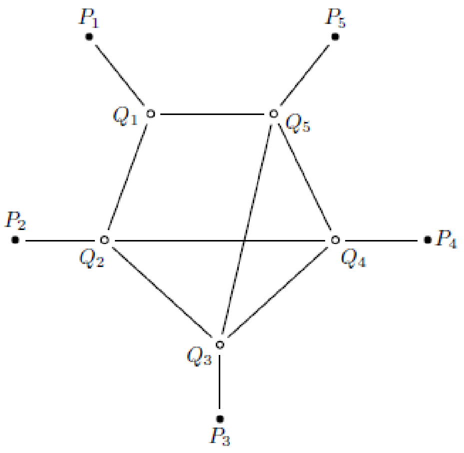
Optimal reductions of 5-point spaces are obtained by (Koolen, vd, 2009). We repeat these derivations for completeness. The moves for Types A, B, C are summarized below. The pendant free reduction of Type A given below is already optimal.



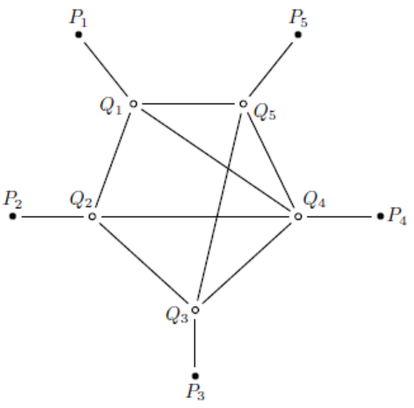


For types B and C we obtain the following pendant free reductions.

B:

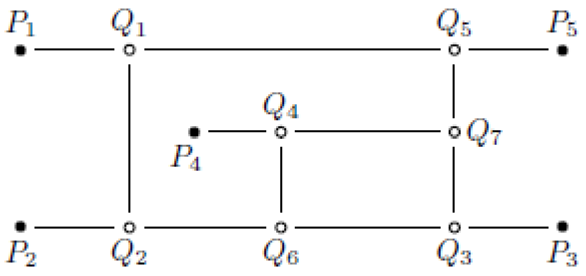


C:

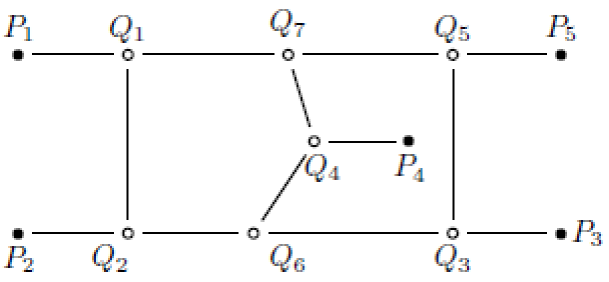


Their optimal reductions are shown below.

A2:



B1:



Triangulations of pendant-free reductions of 6-point spaces:

The Gromov product classification of 6-point spaces have been obtained in (Bilge vd, 2015). These equivalence classes are grouped in reducible and irreducible ones. The lists of irreducible and reducible classes are given below.

Table 1. Gromov product structures for irreducible 6 point spaces

	P_1	P_2	P_3	P_4	P_5	P_6	Sturmfels–Yu Type
I_1	Δ_{124}	Δ_{213}	Δ_{324}	Δ_{456}	Δ_{524}	Δ_{624}	68, 71, 95, 133, 137, 166, 200, 215, 222, 223, 227, 241, 260, 263, 275, 283, 286, 304, 306, 312, 317, 321, 331, 339
I_2	Δ_{124}	Δ_{213}	Δ_{324}	Δ_{456}	Δ_{513}	Δ_{624}	214, 249, 250, 259, 264, 285, 292, 295, 297, 298, 302, 303, 322, 325, 327, 329, 335
I_3	Δ_{124}	Δ_{213}	Δ_{324}	Δ_{435}	Δ_{516}	Δ_{624}	88, 190, 192, 251, 255, 272, 301
I_4	Δ_{124}	Δ_{213}	Δ_{324}	Δ_{456}	Δ_{513}	Δ_{613}	294, 300, 324, 334
I_5	Δ_{124}	Δ_{213}	Δ_{324}	Δ_{435}	Δ_{526}	Δ_{635}	9, 20, 40, 43, 67, 90, 101, 127, 132, 160, 204
I_6	Δ_{125}	Δ_{213}	Δ_{324}	Δ_{436}	Δ_{536}	Δ_{625}	4, 21, 42, 74, 87, 103, 120, 131, 159, 184, 185, 205, 258
I_7	Δ_{124}	Δ_{213}	Δ_{324}	Δ_{435}	Δ_{546}	Δ_{635}	11, 15, 24, 27, 45, 58, 73, 102, 115, 129, 135, 181, 221
I_8	Δ_{124}	Δ_{213}	Δ_{324}	Δ_{435}	Δ_{546}	Δ_{615}	8, 14, 18, 31, 36, 48, 78, 105, 123
I_9	Δ_{124}	Δ_{213}	Δ_{324}	Δ_{435}	Δ_{516}	Δ_{635}	142, 168, 188, 189, 220, 228, 239, 256, 265, 277, 290, 314
I_{10}	Δ_{124}	Δ_{213}	Δ_{324}	Δ_{435}	Δ_{516}	Δ_{625}	128, 161, 206
I_{11}	Δ_{125}	Δ_{213}	Δ_{324}	Δ_{435}	Δ_{546}	Δ_{613}	130, 158, 207, 208, 212, 257
I_{12}	Δ_{124}	Δ_{213}	Δ_{324}	Δ_{435}	Δ_{526}	Δ_{615}	1, 16, 28, 75
I_{13}	Δ_{124}	Δ_{213}	Δ_{324}	Δ_{435}	Δ_{546}	Δ_{613}	2, 17, 30, 60, 76, 149
I_{14}	Δ_{124}	Δ_{213}	Δ_{324}	Δ_{435}	Δ_{546}	Δ_{625}	34, 47, 49, 50, 64, 77, 92, 106, 107, 109, 121, 172
I_{15}	Δ_{125}	Δ_{213}	Δ_{324}	Δ_{435}	Δ_{546}	Δ_{625}	32, 33, 38, 55, 61, 63, 89, 114, 124, 136, 148, 162, 179, 209, 234, 273
I_{16}	Δ_{156}	Δ_{213}	Δ_{324}	Δ_{456}	Δ_{513}	Δ_{624}	247, 248, 262, 282, 293, 296, 316, 326, 332, 336
I_{17}	Δ_{126}	Δ_{213}	Δ_{324}	Δ_{435}	Δ_{546}	Δ_{615}	12, 13, 39, 65, 66

Table 2. Gromov product structures for reducible 6 point spaces

	P_1	P_2	P_3	P_4	P_5	P_6	Sturmfels–Yu Type
R_1	Δ_{124}	Δ_{213}	Δ_{324}	Δ_{413}	Δ_{513}	Δ_{613}	70, 81, 96, 100, 144, 151, 174, 176, 216, 218, 225, 226, 235, 240, 244, 261, 268, 280, 284, 287, 305, 308, 310, 311, 318, 320, 330, 338
R_2	Δ_{124}	Δ_{213}	Δ_{324}	Δ_{435}	Δ_{524}	Δ_{624}	69, 72, 80, 82, 84, 85, 94, 125, 126, 134, 138, 139, 143, 146, 153, 156, 157, 163, 167, 186, 187, 191, 193, 194, 195, 198, 199, 202, 211, 224, 231, 233, 252, 253, 266, 271, 278, 281, 315
R_3	Δ_{124}	Δ_{213}	Δ_{324}	Δ_{413}	Δ_{513}	Δ_{624}	7, 26, 59, 97, 118, 119, 152, 154, 175, 183, 217, 236, 237, 245, 246, 269, 274, 279, 288, 291, 299, 307, 309, 319, 323, 328, 333, 337
R_4	Δ_{124}	Δ_{213}	Δ_{324}	Δ_{435}	Δ_{524}	Δ_{613}	3, 25, 53, 54, 57, 83, 86, 98, 111, 113, 117, 140, 141, 145, 147, 164, 165, 171, 177, 182, 196, 197, 201, 203, 219, 229, 230, 232, 243, 254, 267, 270, 276, 289, 313
R_5	Δ_{124}	Δ_{213}	Δ_{324}	Δ_{413}	Δ_{513}	Δ_{625}	6, 23, 41, 51, 91, 112, 155, 170, 238
R_6	Δ_{124}	Δ_{213}	Δ_{324}	Δ_{435}	Δ_{524}	Δ_{615}	5, 22, 44, 110, 169, 210
R_7	Δ_{124}	Δ_{213}	Δ_{324}	Δ_{413}	Δ_{516}	Δ_{625}	10, 19, 29, 52, 108
R_8	Δ_{125}	Δ_{213}	Δ_{324}	Δ_{435}	Δ_{514}	Δ_{613}	35, 46, 62, 79, 93, 104, 122, 150, 173, 213
R_9	Δ_{124}	Δ_{213}	Δ_{324}	Δ_{413}	Δ_{516}	Δ_{635}	37, 56, 99, 116, 178, 180, 242

The pendant free reduction of a metric space is the weighted graph that is obtained by joining edged at each vertex by an amount of the minimal Gromov product at that vertex and remowing the edges for which the distances are realized by an indirect path. If Δ_{ijk} is minimal This is obtained from the complete graph by removing the edges (P_j,P_k)

Decompositon of graphs into disjoint triangles is discuused in (DeLoera vd, 2010, Reid and Rosa 2009). We aim to partition the edges in the pendant free reductions into disjoint triangles and then apply Delta-Y transform to reduce total weight.

We can encounter the following situations.

(i) The pendant free reduction can be partioned into disjoint triangles. These disjoint triangles are the triangles that are shaded in the diagrams.
(ii) We need to add an edge that has been removed in order to obtain partitioning into disjoint triangles. In these cases, the edge that has been added is shown in red.
(iii) After applying the steps above, there is an edge that still remains. This is shown as a single edge.

For example, in Type I1, the edges (P1,P3), (P2,P4) and (P5,P6) are removed in the pendant free reduction. The remaining edges are

(P1,P2), (P1,P4), (P1,P5), (P1,P6), (P2,P3), (P2,P5), (P2,P6), (P3,P4), (P3,P5), (P3,P6), (P4,P5), (P4,P6).

These are grouped into disjoint triangles as

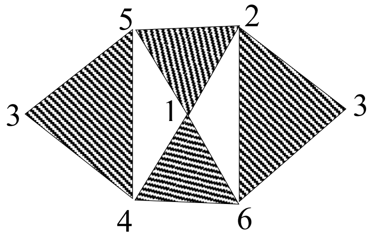
(P1,P2), (P1,P5), (P2,P5),

(P1,P4), (P1,P6), (P4,P6).

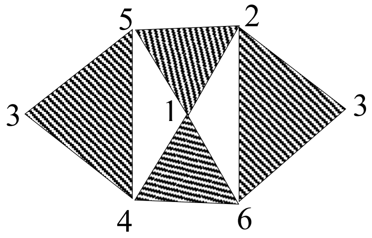
(P2,P3), (P2,P6), (P3,P6),

(P3,P4), (P3,P5), (P4,P5),

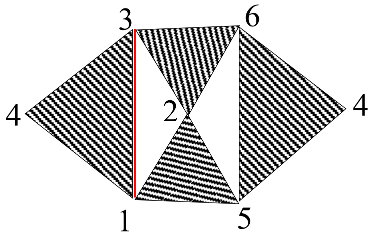
I_1: Δ_{124} Δ_{213} Δ_{324} Δ_{456} Δ_{524} Δ_{624}



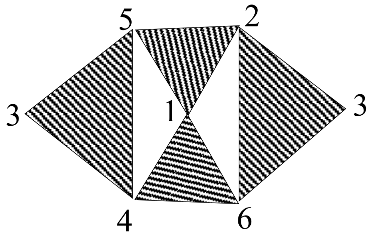
I_2: Δ_{124} Δ_{213} Δ_{324} Δ_{456} Δ_{513} Δ_{624}



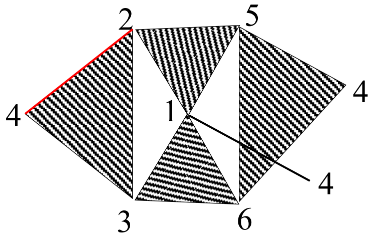
I_3: Δ_{124} Δ_{213} Δ_{324} Δ_{435} Δ_{516} Δ_{624}



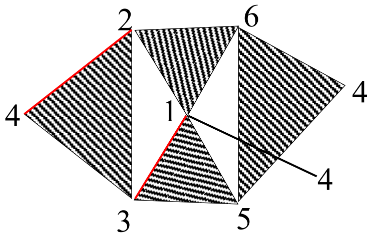
I_4: Δ_{124} Δ_{213} Δ_{324} Δ_{456} Δ_{513} Δ_{613}



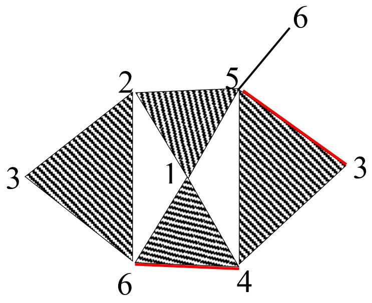
I_5: Δ_{124} Δ_{213} Δ_{324} Δ_{435} Δ_{526} Δ_{635}



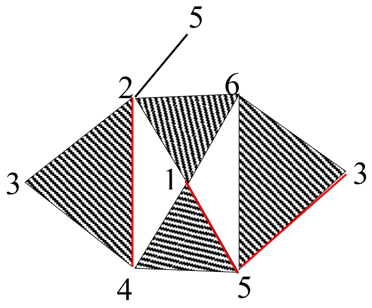
I_6: Δ_{125} Δ_{213} Δ_{324} Δ_{436} Δ_{536} Δ_{625}



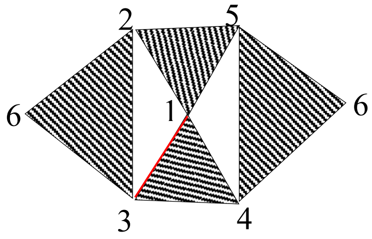
I_7: Δ_{124} Δ_{213} Δ_{324} Δ_{435} Δ_{546} Δ_{635}



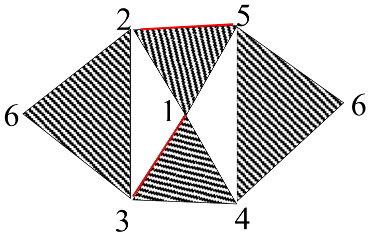
I_8: Δ_{124} Δ_{213} Δ_{324} Δ_{435} Δ_{546} Δ_{615}



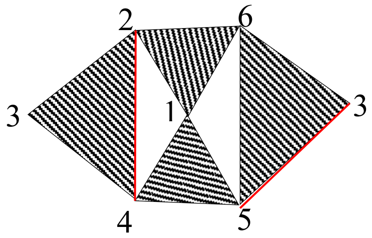
I_9: Δ_{124} Δ_{213} Δ_{324} Δ_{435} Δ_{516} Δ_{635}



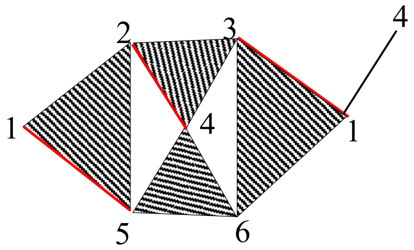
I_{10}: Δ_{124} Δ_{213} Δ_{324} Δ_{435} Δ_{516} Δ_{625}



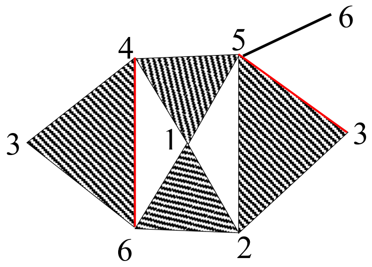
I_{11}: Δ_{125} Δ_{213} Δ_{324} Δ_{435} Δ_{546} Δ_{613}



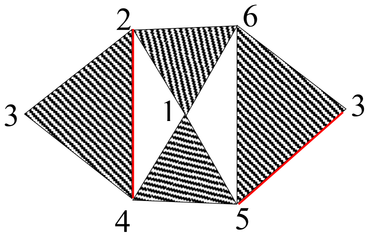
I_{12}: Δ_{124} Δ_{213} Δ_{324} Δ_{435} Δ_{526} Δ_{615}



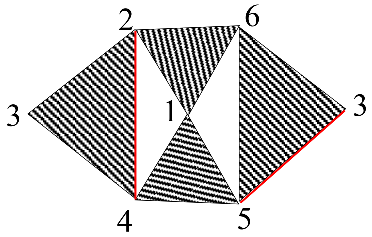
I_{13}: Δ_{124} Δ_{213} Δ_{324} Δ_{435} Δ_{546} Δ_{613}



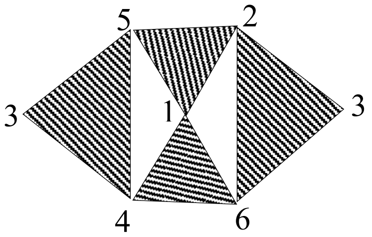
I_{14}: Δ_{124} Δ_{213} Δ_{324} Δ_{435} Δ_{546} Δ_{625}



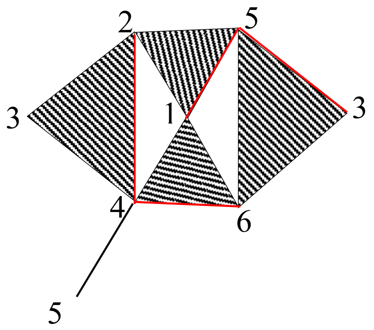
I_{15}: Δ_{125} Δ_{213} Δ_{324} Δ_{435} Δ_{546} Δ_{625}



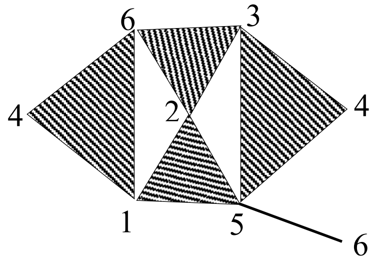
I_{16}: Δ_{156} Δ_{213} Δ_{324} Δ_{456} Δ_{513} Δ_{624}



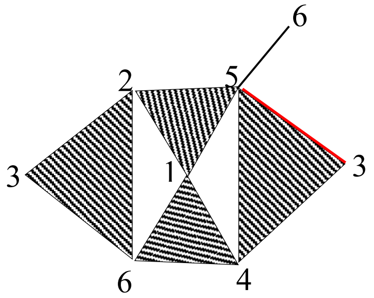
I_{17}: Δ_{126} Δ_{213} Δ_{324} Δ_{435} Δ_{546} Δ_{615}



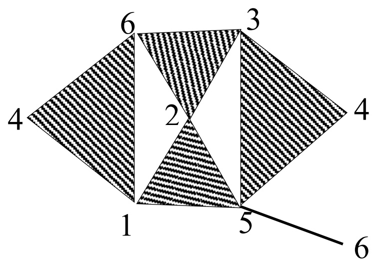
R_1: Δ_{124} Δ_{213} Δ_{324} Δ_{413} Δ_{513} Δ_{613}



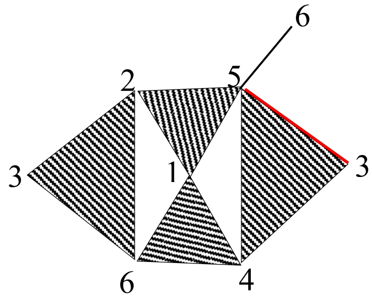
R_2: Δ_{124} Δ_{213} Δ_{324} Δ_{435} Δ_{524} Δ_{624}



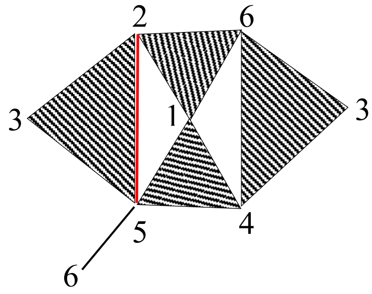
R_3: Δ_{124} Δ_{213} Δ_{324} Δ_{413} Δ_{513} Δ_{624}



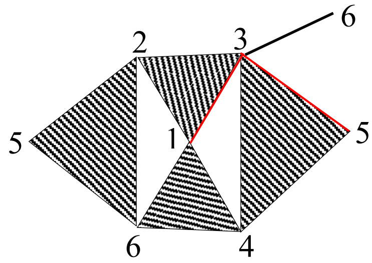
R_4: Δ_{124} Δ_{213} Δ_{324} Δ_{435} Δ_{524} Δ_{613}



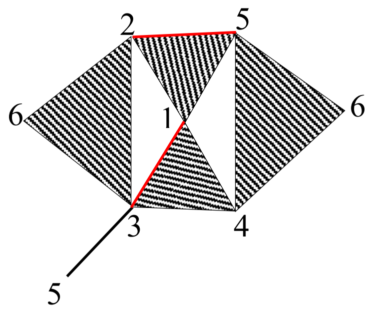
R_5: Δ_{124} Δ_{213} Δ_{324} Δ_{413} Δ_{513} Δ_{625}



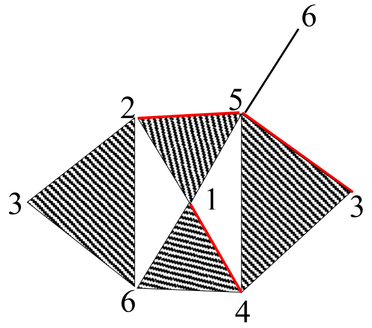
R_6: Δ_{124} Δ_{213} Δ_{324} Δ_{435} Δ_{524} Δ_{615}



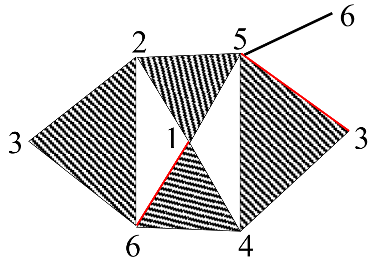
R_7: Δ_{124} Δ_{213} Δ_{324} Δ_{413} Δ_{516} Δ_{625}



R_8: Δ_{125} Δ_{213} Δ_{324} Δ_{435} Δ_{514} Δ_{613}



R_9: Δ_{124} Δ_{213} Δ_{324} Δ_{413} Δ_{516} Δ_{635}



As a result of these triangle decompositions we obtain 8 types as given below in Tables 3 and 4. In these tables the first column displays the Gromov product structure type, the second column is the list of edges that have been removed in the pendant free reduction. The edges that are added are shown in bold. The graph of the removed set actually be used to obtain courses subdivisions of the metric cone , which would be useful in understanding metric spaces with elements n>6, since the number of Gromov product spaces increases very rapidly. The last column it the triangle type. For example, ın Type A, the edges d13, d24, d56 are removed. The remaining edges are decomposed into 4 disjoint triangles. The total weight of the pendant free reduced by ½. For Type B, the edges **d13**, d16, d24, d35 are removed. The edge **d13** is added to obtain decomposition into triangles

Table 3 Triangulation types of pendant free reductions of irreducible 6-point spaces

GPS Type	Removed Edges	Graph of Removed Edges	T-Type
I1	d13, d24, d56	1---3 2---4 5---6	A
I2	d13, d24, d56	1---3 2---4 5---6	A
I4	d13, d24, d56	1---3 2---4 5---6	A
I16	d13, d24, d56	1---3 2---4 5---6	A
I3	**d13**, d16, d24, d35	6---1---3---5 2---4	B
I9	**d13**, d16, d24, d35	6---1---3---5 2---4	B
I5	d13, **d24**, d26, d35	5---3---1 4---2---6	C
I6	**d13, d24**, d25, d36	1---3---6 4---2---5	C
I7	d13, d24, **d35, d46**	5---3---1 2---4---6	C
I13	d13, d24, **d35, d46**	5---3---1 2---4---6	C
I8	d13, **d15**, **d24, d35**, d46	1---3---5---1 2---4---6	D
I12	**d13, d15, d24,** d26, d35	1---3---5---1 4---2---6	D
I10	**d13**, d16, d24, **d25**, d35	6---1---3---5---2---4	E
I11	d13, **d24,** d25, **d35,** d46	6---4---2---5---3---1	E
I14	d13, **d24**, d25, **d35,** d46	6---4---2---5---3---1	E
I15	d13, **d24,** d25, **d35,** d46	6---4---2---5---3---1	E

I17	d13, **d15, d24,** d26, **d35, d46**	1---3---5---1 2---4---6---2	F

Table 3 Triangulation types of pendant free reductions of reducible 6-point spaces

R1	d13, d24	1---3 2---4	G
R3	d13, d24	1---3 2---4	G
R2	d13, d24, **d35**	1---3---5 2---4	H
R4	d13, d24, **d35**	1---3---5 2---4	H
R5	d13, d24, **d25**	4---2---5 1---3	H
R6	**d13**, d15, d24, **d35**	1---3---5---1 2---4	B
R9	d13, **d16**, d24, **d35**	5---3---1---6 2---4	B
R7	**d13**, d16, d24, **d25**	3---1---6 4---2---5	C
R8	d13, **d14**, d24, **d25**, **d35**	3---1---4---2---5---3	E

References

Bilge, D. Çelik, Ş. Koçak, Optimal embeddings of finite metric spaces into graphs, Anadolu Univ. J. Sci. Technol. B- Theor. Sci.. 3 (2) (2015) 133–147.

Bilge, A.H., D. Çelik and S¸. Koçak (2017). An equivalence class decomposition of finite metric spaces via Gromov products. Discrete Mathematics, Vol. 340 1928-1932.

De Loera, J.A., J. Rambau, F. Santos, (2010), Triangulations Structures for Algorithms and Applications, Springer, 2010.

Koolen, J., Lesser, A., Moulton, V. 2009. “Optimal realization of generic 5-point metrics”, European J. Combin., 30 (5), 1164–1171.

Reid, C., A. Rosa, (2010), Steiner systems S(2,4,v)- a survey, Electronic Journal of Combinatorics, Article Number DS18,

Examples of Distance Matrices

Burak Erkan Kaya

Kadir Has Univerisity

In this report, certain datasets, chosen mainly from phyogenetics are used to evaluate distance functions. These dissimilarity measures may fail to satisfy triangle inequalities. The percentage of triangle inqualities that are violated are computed and Gromov product structures are obtained.

Datasets:

During the search of suitable databases, sample databases in SplitsTree and MEGA programs, which are widely used in phylogenetic applications, were examined. In these programs, databases used in the current studies in the literature are available for researchers to review. These programs allow an existing database to be handled with different distance calculation methods. Thus; different distance matrices can be obtained by using different calculation methods for each distance dataset. Totally thirty-six distance matrices have been obtained with different calculation methods and databases. The distance matrices obtained from each dataset and the findings obtained by examining them are presented in the following sections. A summary table of the findings obtained as the results of examining all databases is added at the end of this report.

Methods for the computation of distance matrices:

Many different methods are used to determine the distance between nucleotides in molecular phylogenetic analysis studies. While all methods are based on Markov chains, they differ in the use of parameters that determine the rates at which one nucleotide will change with another. We describe some of these methods here.

F81 Method (Felsenstein 1981): This model was created on the basis of the JC69 method, where the base frequencies are taken equally. The difference from the JC69 method is that the calculation is done without taking the base frequencies equally.

HKY85 Method (Hasegawa, Kishino ve Yano 1985): It is a method based on created with the F81 and the K80 method. While the K80 method utilises the base frequencies equally as in the JC69 method; F81 method accepts the conversion rate for the purine-pyrimidine transitions as one, and the purity-purine or pyrimidine-pyrimidine exchange rate is based on a K parameter. In the HKY85 method, calculations are made by using different change and conversion rates as in K80 method and different base frequencies as in F81 method.

GTR Method (Tavare 1986): In this method is the calculations are made using a base frequency vector and different parameters for each change.

F84 Method (Felsenstein 1984): This model was obtained by synthesising F81 and K80 methods. Base frequencies are not taken equally, and different parameters are used for exchange and conversion rates.

The Bee, Algea and Primates datasets are analyzed in detail in Appendices 3.2, 3.4 and 3.5, respectively.

Bee Dateset (6-Element): As a result of the analysis of the bee dataset with different distance calculation methods, the results presented below were obtained. Bee dataset consists of 'A.andrenof', 'A.mellifer', 'A.dorsata', 'A.cerana', 'A.florea', 'A.koschev' taxa.

a) Distance Matrix Calculated by the BASEFREQ Method:

0	0.0562	0.074	0.0444	0.0089	0.0473
0.0562	0	0.0237	0.0148	0.0592	0.0089
0.074	0.0237	0	0.0296	0.0769	0.0266
0.0444	0.0148	0.0296	0	0.0473	0.0059
0.0089	0.0592	0.0769	0.0473	0	0.0503
0.0473	0.0089	0.0266	0.0059	0.0503	0

b) Distance Matrix Calculated by the F81 Method:

0	0.096	0.1113	0.1027	0.0044	0.0794
0.096	0	0.0994	0.096	0.0994	0.1078
0.1113	0.0994	0	0.1268	0.1147	0.1113
0.1027	0.096	0.1268	0	0.1061	0.1061
0.0044	0.0994	0.1147	0.1061	0	0.0827
0.0794	0.1078	0.1113	0.1061	0.0827	0

c) Distance Matrix Calculated by the F84 Method:

0	0.096	0.1114	0.1028	0.0044	0.0795
0.096	0	0.0994	0.0961	0.0994	0.1079
0.1114	0.0994	0	0.1269	0.1148	0.1113
0.1028	0.0961	0.1269	0	0.1062	0.1061
0.0044	0.0994	0.1148	0.1062	0	0.0827
0.0795	0.1079	0.1113	0.1061	0.0827	0

d) Distance Matrix Calculated by the GTR Method:

0	0.3401	0.3563	0.3403	0.2457	0.3211
0.3401	0	0.3348	0.3358	0.3488	0.3545
0.3563	0.3348	0	0.3565	0.3652	0.3537
0.3403	0.3358	0.3565	0	0.3489	0.3376
0.2457	0.3488	0.3652	0.3489	0	0.3297
0.3211	0.3545	0.3537	0.3376	0.3297	0

e) Distance Matrix Calculated by the HAMMING Method:

0	0.0901	0.1034	0.096	0.0044	0.0753
0.0901	0	0.0931	0.0901	0.0931	0.1004
0.1034	0.0931	0	0.1167	0.1064	0.1034
0.096	0.0901	0.1167	0	0.099	0.099
0.0044	0.0931	0.1064	0.099	0	0.0783
0.0753	0.1004	0.1034	0.099	0.0783	0

f) Distance Matrix Calculated by the HKY85 Method:

0	0.0983	0.114	0.1054	0.0045	0.0808
0.0983	0	0.1022	0.0982	0.1019	0.1107
0.114	0.1022	0	0.1313	0.1178	0.1143
0.1054	0.0982	0.1313	0	0.1091	0.1093
0.0045	0.1019	0.1178	0.1091	0	0.0843
0.0808	0.1107	0.1143	0.1093	0.0843	0

g) Distance Matrix Calculated by the K3ST Method:

0	0.0968	0.1122	0.1035	0.0044	0.0798
0.0968	0	0.1	0.0967	0.1001	0.1086
0.1122	0.1	0	0.1284	0.1156	0.112
0.1035	0.0967	0.1284	0	0.1069	0.107
0.0044	0.1001	0.1156	0.1069	0	0.0831
0.0798	0.1086	0.112	0.107	0.0831	0

h) Distance Matrix Calculated by the LOGDET Method:

0	0.0884	0.1093	0.0967	0.0062	0.0767
0.0884	0	0.087	0.0891	0.0919	0.1025
0.1093	0.087	0	0.1141	0.1128	0.1084
0.0967	0.0891	0.1141	0	0.1	0.0955
0.0062	0.0919	0.1128	0.1	0	0.08
0.0767	0.1025	0.1084	0.0955	0.08	0

The number and the proport,on of triangle inequalities that are not satisfied is given below.

Dataset	Method	Number of Taxa	Number of Tri. Ineq.	Tri. Inq. satisfied	Tri. Inq. not satisfied	%
Bee	BASEFREQ	6	20	14	6	30.00
Bee	F81	6	20	20	0	0.00
Bee	F84	6	20	20	0	0.00
Bee	GTR	6	20	20	0	0.00
Bee	HAMMING	6	20	20	0	0.00
Bee	HKY85	6	20	20	0	0.00
Bee	K3ST	6	20	20	0	0.00
Bee	LOGDET	6	20	20	0	0.00

Algea Dataset (8 elements):

a)Distance Matrix Calculated by the BASEFREQ Method:

0	0.0143	0.0295	0.0718	0.0668	0.166	0.0295	0.1503
0.0143	0	0.0152	0.0575	0.0525	0.1517	0.0152	0.136
0.0295	0.0152	0	0.0423	0.0373	0.1364	0.0065	0.1208
0.0718	0.0575	0.0423	0	0.0133	0.0942	0.0423	0.0785
0.0668	0.0525	0.0373	0.0133	0	0.0991	0.0373	0.0835
0.166	0.1517	0.1364	0.0942	0.0991	0	0.1364	0.0185
0.0295	0.0152	0.0065	0.0423	0.0373	0.1364	0	0.1208
0.1503	0.136	0.1208	0.0785	0.0835	0.0185	0.1208	0

Gromov Product Structures

P_1	P_2	P_3	Minimal Δ
1	2	8	0.0143
2	1	3	0
2	1	4	0
2	1	5	0
2	1	6	0
2	1	7	0
3	1	6	-0.0001
3	2	6	-0.0001
4	3	8	0
4	7	8	0
5	1	6	-0.0001
5	2	6	-0.0001
6	3	8	0.017
6	7	8	0.017
7	1	6	-0.0001
7	2	6	-0.0001
8	1	6	0.0014

b) Distance Matrix Calculated by the F81Method:

0	0.0266	0.03	0.1218	0.0828	0.1504	0.1343	0.1556
0.0266	0	0.0425	0.132	0.0938	0.1596	0.1446	0.1608
0.03	0.0425	0	0.1064	0.0668	0.1343	0.1315	0.1461
0.1218	0.132	0.1064	0	0.1068	0.1575	0.1585	0.175
0.0828	0.0938	0.0668	0.1068	0	0.126	0.1284	0.1259
0.1504	0.1596	0.1343	0.1575	0.126	0	0.1789	0.149
0.1343	0.1446	0.1315	0.1585	0.1284	0.1789	0	0.1501
0.1556	0.1608	0.1461	0.175	0.1259	0.149	0.1501	0

Gromov Product Structures

P_1	P_2	P_3	Minimal Δ
1	2	3	0.007
2	1	8	0.0159
3	1	6	0.0069
4	1	6	0.0644
5	3	8	0.0233
6	4	8	0.0658
7	1	8	0.0644
8	6	7	0.0601

c) Distance Matrix Calculated by the F84 Method:

0	0.0266	0.0301	0.1228	0.0835	0.1538	0.1354	0.1601
0.0266	0	0.0426	0.1328	0.0945	0.1626	0.1454	0.1647
0.0301	0.0426	0	0.1069	0.0671	0.1368	0.1324	0.1492
0.1228	0.1328	0.1069	0	0.1075	0.1597	0.1597	0.1782
0.0835	0.0945	0.0671	0.1075	0	0.1284	0.1294	0.1281
0.1538	0.1626	0.1368	0.1597	0.1284	0	0.1816	0.1513
0.1354	0.1454	0.1324	0.1597	0.1294	0.1816	0	0.1517
0.1601	0.1647	0.1492	0.1782	0.1281	0.1513	0.1517	0

Gromov Product Structures

P_1	P_2	P_3	Minimal Δ
1	2	3	0.0071
2	1	8	0.0156
3	1	6	0.0066
4	1	6	0.0644
5	3	8	0.023
6	4	8	0.0664
7	1	8	0.0635
8	6	7	0.0607

d) Distance Matrix Calculated by the HAMMING Method:

0	0.0261	0.0294	0.1124	0.0784	0.1363	0.123	0.1405
0.0261	0	0.0413	0.121	0.0882	0.1438	0.1315	0.1447
0.0294	0.0413	0	0.0992	0.0639	0.1229	0.1207	0.1328
0.1124	0.121	0.0992	0	0.0996	0.1421	0.1429	0.1561
0.0784	0.0882	0.0639	0.0996	0	0.116	0.118	0.1159
0.1363	0.1438	0.1229	0.1421	0.116	0	0.1592	0.1352
0.123	0.1315	0.1207	0.1429	0.118	0.1592	0	0.136
0.1405	0.1447	0.1328	0.1561	0.1159	0.1352	0.136	0

Gromov Product Structures

P_1	P_2	P_3	Minimal Δ
1	2	3	0.0071
2	1	8	0.0151
3	1	5	0.0074
4	1	6	0.0591
5	3	8	0.0235
6	4	8	0.0606
7	1	8	0.0592
8	6	7	0.056

e) Distance Matrix Calculated by the GTR Method:

0	0.0492	0.0521	0.1473	0.1066	0.1757	0.1635	0.1813
0.0492	0	0.065	0.1563	0.1177	0.1843	0.1747	0.1873
0.0521	0.065	0	0.1265	0.0868	0.1568	0.1582	0.1688
0.1473	0.1563	0.1265	0	0.1275	0.1782	0.1819	0.195
0.1066	0.1177	0.0868	0.1275	0	0.1443	0.1542	0.1435
0.1757	0.1843	0.1568	0.1782	0.1443	0	0.1999	0.1677
0.1635	0.1747	0.1582	0.1819	0.1542	0.1999	0	0.1698
0.1813	0.1873	0.1688	0.195	0.1435	0.1677	0.1698	0

Gromov Product Structures

P_1	P_2	P_3	Minimal Δ
1	2	3	0.0181
2	1	8	0.0276
3	1	4	0.0157
4	3	6	0.0739
5	3	8	0.0307
6	4	8	0.0754
7	1	8	0.076
8	6	7	0.0688

f) Distance Matrix Calculated by the HKY85 Method:

0	0.0267	0.03	0.1235	0.0831	0.1508	0.1365	0.1554
0.0267	0	0.0427	0.1346	0.0948	0.161	0.1478	0.1613
0.03	0.0427	0	0.1082	0.0673	0.1348	0.1339	0.1467
0.1235	0.1346	0.1082	0	0.1082	0.1597	0.162	0.1771
0.0831	0.0948	0.0673	0.1082	0	0.1263	0.1302	0.1264
0.1508	0.161	0.1348	0.1597	0.1263	0	0.1819	0.1505
0.1365	0.1478	0.1339	0.162	0.1302	0.1819	0	0.1525
0.1554	0.1613	0.1467	0.1771	0.1264	0.1505	0.1525	0

Gromov Product Structures

P_1	P_2	P_3	Minimal Δ
1	2	3	0.007
2	1	8	0.0163
3	1	6	0.007
4	1	6	0.0662
5	3	8	0.0235
6	4	8	0.0666
7	1	8	0.0668
8	6	7	0.0605

g) Distance Matrix Calculated by the LOGDET Method:

0	0.0279	0.032	0.1275	0.0857	0.1499	0.1418	0.1584
0.0279	0	0.045	0.1369	0.0971	0.1594	0.1531	0.1653
0.032	0.045	0	0.1095	0.0682	0.1341	0.1383	0.1494
0.1275	0.1369	0.1095	0	0.1117	0.1598	0.1641	0.1793
0.0857	0.0971	0.0682	0.1117	0	0.1261	0.1352	0.1274
0.1499	0.1594	0.1341	0.1598	0.1261	0	0.1763	0.1511
0.1418	0.1531	0.1383	0.1641	0.1352	0.1763	0	0.1507
0.1584	0.1653	0.1494	0.1793	0.1274	0.1511	0.1507	0

Gromov Product Structures

P_1	P_2	P_3	Minimal Δ
1	2	3	0.0075
2	1	8	0.0174
3	1	4	0.007
4	3	6	0.0676
5	3	8	0.0231
6	4	8	0.0658
7	1	8	0.067
8	6	7	0.0628

h) Distance Matrix Calculated by the K3ST Method:

0	0.0266	0.0301	0.1229	0.0836	0.1539	0.1354	0.1601
0.0266	0	0.0426	0.1328	0.0945	0.1628	0.1455	0.1647
0.0301	0.0426	0	0.1069	0.0672	0.1368	0.1324	0.1492
0.1229	0.1328	0.1069	0	0.1076	0.1597	0.1597	0.1783
0.0836	0.0945	0.0672	0.1076	0	0.1284	0.1294	0.1281
0.1539	0.1628	0.1368	0.1597	0.1284	0	0.1816	0.1513
0.1354	0.1455	0.1324	0.1597	0.1294	0.1816	0	0.1517
0.1601	0.1647	0.1492	0.1783	0.1281	0.1513	0.1517	0

Gromov Product Structures

P_1	P_2	P_3	Minimal Δ
1	2	3	0.0071
2	1	8	0.0156
3	1	6	0.0065
4	1	6	0.0644
5	3	8	0.023
6	4	8	0.0664
7	1	8	0.0635
8	6	7	0.0607

List of triangle inequalites that are not satisfied:

Dataset	Method	Number of Taxa	Number of Tri. Ineq.	Tri. Inq. satisfied	Tri. Inq. not satisfied	%
Algea	BASEFREQ	8	56	22	34	60.71
Algea	F81	8	56	56	0	0.00
Algea	F84	8	56	56	0	0.00
Algea	GTR	8	56	56	0	0.00
Algea	HAMMING	8	56	56	0	0.00
Algea	HKY85	8	56	56	0	0.00
Algea	K3ST	8	56	56	0	0.00
Algea	LOGDET	8	56	56	0	0.00
Algea	F84	8	56	56	0	0.00

Primates-mtDNA Dataset (12 Elements): Primates-mtDNA dataset consists of 'Lemur_catta', 'Homo_sapiens', 'Pan', 'Gorilla', 'Pongo', 'Hylobates', 'Macaca_fuscata', 'M._mulatta', 'M._fascicularis', 'M._sylvanus', 'Saimiri_sciureus', 'Tarsius_syrichta' taxa.

a) Distance Matrix Calculated by the BASEFREQ Method:

0	0.13772	0.11314	0.12431	0.15884	0.12878	0.08409	0.07292	0.08409	0.04632	0.03408	0.03356
0.13772	0	0.02458	0.01341	0.04617	0.03799	0.05363	0.0648	0.05363	0.09385	0.1323	0.15586
0.11314	0.02458	0	0.01341	0.05534	0.04022	0.03352	0.04246	0.03799	0.06927	0.11889	0.14245
0.12431	0.01341	0.01341	0	0.04617	0.02905	0.04022	0.0514	0.04022	0.08045	0.12112	0.14468
0.15884	0.04617	0.05534	0.04617	0	0.05217	0.07475	0.08592	0.07898	0.11497	0.15989	0.18345
0.12878	0.03799	0.04022	0.02905	0.05217	0	0.04469	0.05587	0.04469	0.08492	0.10772	0.14668
0.08409	0.05363	0.03352	0.04022	0.07475	0.04469	0	0.01117	0.01788	0.04022	0.0876	0.11116
0.07292	0.0648	0.04246	0.0514	0.08592	0.05587	0.01117	0	0.01341	0.02905	0.07643	0.09999
0.08409	0.05363	0.03799	0.04022	0.07898	0.04469	0.01788	0.01341	0	0.04022	0.0809	0.10446
0.04632	0.09385	0.06927	0.08045	0.11497	0.08492	0.04022	0.02905	0.04022	0	0.05632	0.07988
0.03408	0.1323	0.11889	0.12112	0.15989	0.10772	0.0876	0.07643	0.0809	0.05632	0	0.04198
0.03356	0.15586	0.14245	0.14468	0.18345	0.14668	0.11116	0.09999	0.10446	0.07988	0.04198	0

b) Distance Matrix Calculated by the F81 Method:

0	0.3965	0.3984	0.3722	0.3709	0.3759	0.354	0.3649	0.3796	0.3612	0.3592	0.307
0.3965	0	0.0951	0.1117	0.1811	0.2084	0.2778	0.2794	0.3024	0.3125	0.3398	0.421
0.3984	0.0951	0	0.1143	0.194	0.2173	0.2941	0.3058	0.3314	0.3024	0.3576	0.419
0.3722	0.1117	0.1143	0	0.1882	0.2173	0.2843	0.2826	0.3227	0.2958	0.3363	0.4074
0.3709	0.1811	0.194	0.1882	0	0.216	0.2962	0.2995	0.3215	0.2912	0.3563	0.3889
0.3759	0.2084	0.2173	0.2173	0.216	0	0.2991	0.2875	0.3142	0.2925	0.3328	0.3978
0.354	0.2778	0.2941	0.2843	0.2962	0.2991	0	0.0366	0.0888	0.1354	0.3649	0.4074
0.3649	0.2794	0.3058	0.2826	0.2995	0.2875	0.0366	0	0.0989	0.1301	0.3722	0.4112
0.3796	0.3024	0.3314	0.3227	0.3215	0.3142	0.0888	0.0989	0	0.1341	0.3612	0.4016
0.3612	0.3125	0.3024	0.2958	0.2912	0.2925	0.1354	0.1301	0.1341	0	0.3612	0.4151
0.3592	0.3398	0.3576	0.3363	0.3563	0.3328	0.3649	0.3722	0.3612	0.3612	0	0.417
0.307	0.421	0.419	0.4074	0.3889	0.3978	0.4074	0.4112	0.4016	0.4151	0.417	0

c) Distance Matrix Calculated by the F84 Method:

0	0.4009	0.4032	0.3755	0.3754	0.3796	0.3581	0.3697	0.3861	0.3655	0.362	0.3086
0.4009	0	0.0978	0.1152	0.1865	0.2146	0.2855	0.2871	0.3134	0.3249	0.3489	0.4304
0.4032	0.0978	0	0.1179	0.2009	0.2247	0.3042	0.3172	0.3473	0.3136	0.3689	0.4279
0.3755	0.1152	0.1179	0	0.1947	0.2245	0.2933	0.2912	0.3375	0.3063	0.3455	0.4163
0.3754	0.1865	0.2009	0.1947	0	0.2217	0.3059	0.3094	0.335	0.2998	0.3635	0.3954
0.3796	0.2146	0.2247	0.2245	0.2217	0	0.3087	0.2955	0.326	0.3013	0.3396	0.4053
0.3581	0.2855	0.3042	0.2933	0.3059	0.3087	0	0.037	0.0912	0.141	0.3763	0.4151
0.3697	0.2871	0.3172	0.2912	0.3094	0.2955	0.037	0	0.1021	0.1353	0.3844	0.4192
0.3861	0.3134	0.3473	0.3375	0.335	0.326	0.0912	0.1021	0	0.1398	0.3721	0.4093
0.3655	0.3249	0.3136	0.3063	0.2998	0.3013	0.141	0.1353	0.1398	0	0.3721	0.4232
0.362	0.3489	0.3689	0.3455	0.3635	0.3396	0.3763	0.3844	0.3721	0.3721	0	0.4223
0.3086	0.4304	0.4279	0.4163	0.3954	0.4053	0.4151	0.4192	0.4093	0.4232	0.4223	0

d) Distance Matrix Calculated by the GTR Method:

0	0.4882	0.4937	0.46	0.4668	0.4638	0.453	0.463	0.4786	0.4577	0.446	0.3997
0.4882	0	0.1947	0.21	0.2937	0.3107	0.3843	0.387	0.4076	0.4242	0.4373	0.5169
0.4937	0.1947	0	0.214	0.3084	0.3203	0.4097	0.4239	0.4509	0.4171	0.4618	0.5168
0.46	0.21	0.214	0	0.3015	0.3166	0.3894	0.3865	0.4322	0.4031	0.4377	0.5052

0.4668	0.2937	0.3084	0.3015	0	0.3183	0.4126	0.415	0.4411	0.405	0.4569	0.489
0.4638	0.3107	0.3203	0.3166	0.3183	0	0.4079	0.3903	0.4202	0.4002	0.4247	0.4932
0.453	0.3843	0.4097	0.3894	0.4126	0.4079	0	0.1317	0.1846	0.2468	0.4737	0.5118
0.463	0.387	0.4239	0.3865	0.415	0.3903	0.1317	0	0.1979	0.2396	0.4816	0.5163
0.4786	0.4076	0.4509	0.4322	0.4411	0.4202	0.1846	0.1979	0	0.2428	0.4619	0.5035
0.4577	0.4242	0.4171	0.4031	0.405	0.4002	0.2468	0.2396	0.2428	0	0.4689	0.5239
0.446	0.4373	0.4618	0.4377	0.4569	0.4247	0.4737	0.4816	0.4619	0.4689	0	0.5129
0.3997	0.5169	0.5168	0.5052	0.489	0.4932	0.5118	0.5163	0.5035	0.5239	0.5129	0

e) Distance Matrix Calculated by the HAMMING Method:

0	0.308	0.3091	0.2934	0.2926	0.2956	0.2822	0.2889	0.2979	0.2867	0.2854	0.252
0.308	0	0.0893	0.1038	0.1609	0.1819	0.2321	0.2333	0.2489	0.2556	0.2732	0.3221
0.3091	0.0893	0	0.106	0.1709	0.1886	0.2433	0.2511	0.2679	0.2489	0.2844	0.321
0.2934	0.1038	0.106	0	0.1665	0.1886	0.2366	0.2355	0.2623	0.2444	0.271	0.3143
0.2926	0.1609	0.1709	0.1665	0	0.1877	0.2447	0.2469	0.2615	0.2413	0.2836	0.3035
0.2956	0.1819	0.1886	0.1886	0.1877	0	0.2467	0.2388	0.2567	0.2422	0.2688	0.3087
0.2822	0.2321	0.2433	0.2366	0.2447	0.2467	0	0.0357	0.0837	0.1239	0.2889	0.3143
0.2889	0.2333	0.2511	0.2355	0.2469	0.2388	0.0357	0	0.0926	0.1194	0.2934	0.3166
0.2979	0.2489	0.2679	0.2623	0.2615	0.2567	0.0837	0.0926	0	0.1228	0.2867	0.311
0.2867	0.2556	0.2489	0.2444	0.2413	0.2422	0.1239	0.1194	0.1228	0	0.2867	0.3188
0.2854	0.2732	0.2844	0.271	0.2836	0.2688	0.2889	0.2934	0.2867	0.2867	0	0.3199
0.252	0.3221	0.321	0.3143	0.3035	0.3087	0.3143	0.3166	0.311	0.3188	0.3199	0

f) Distance Matrix Calculated by the HKY85 Method:

0	0.4271	0.4285	0.4004	0.3961	0.4037	0.3768	0.3882	0.4027	0.385	0.3859	0.3277
0.4271	0	0.0942	0.1107	0.1814	0.2096	0.2838	0.2857	0.3078	0.3176	0.3521	0.4473
0.4285	0.0942	0	0.1133	0.1938	0.2181	0.2995	0.3111	0.3354	0.3076	0.37	0.4459
0.4004	0.1107	0.1133	0	0.1881	0.2183	0.2896	0.2883	0.3268	0.3008	0.348	0.4318
0.3961	0.1814	0.1938	0.1881	0	0.2184	0.3021	0.3056	0.3266	0.2978	0.3741	0.4138
0.4037	0.2096	0.2181	0.2183	0.2184	0	0.3055	0.2945	0.3202	0.299	0.3473	0.4227
0.3768	0.2838	0.2995	0.2896	0.3021	0.3055	0	0.0364	0.088	0.1337	0.3782	0.4337
0.3882	0.2857	0.3111	0.2883	0.3056	0.2945	0.0364	0	0.0978	0.1284	0.3858	0.4381
0.4027	0.3078	0.3354	0.3268	0.3266	0.3202	0.088	0.0978	0	0.1322	0.3747	0.427
0.385	0.3176	0.3076	0.3008	0.2978	0.299	0.1337	0.1284	0.1322	0	0.3747	0.4425
0.3859	0.3521	0.37	0.348	0.3741	0.3473	0.3782	0.3858	0.3747	0.3747	0	0.4501
0.3277	0.4473	0.4459	0.4318	0.4138	0.4227	0.4337	0.4381	0.427	0.4425	0.4501	0

g) Distance Matrix Calculated by the K3ST Method:

0	0.40187	0.40437	0.37647	0.37599	0.38014	0.35816	0.36996	0.3862	0.36559	0.36204	0.30856
0.40187	0	0.09778	0.11526	0.18651	0.2147	0.28568	0.28739	0.31372	0.32569	0.34898	0.43145
0.40437	0.09778	0	0.11791	0.20092	0.22473	0.30437	0.31756	0.34763	0.31405	0.36901	0.42899
0.37647	0.11526	0.11791	0	0.19478	0.22451	0.29335	0.29128	0.33796	0.30676	0.34554	0.41714
0.37599	0.18651	0.20092	0.19478	0	0.22179	0.30634	0.30988	0.33605	0.30046	0.36364	0.39769
0.38014	0.2147	0.22473	0.22451	0.22179	0	0.30906	0.29582	0.32657	0.30189	0.33962	0.40629
0.35816	0.28568	0.30437	0.29335	0.30634	0.30906	0	0.03701	0.09118	0.14096	0.3763	0.41546
0.36996	0.28739	0.31756	0.29128	0.30988	0.29582	0.03701	0	0.10208	0.13529	0.38438	0.42022
0.3862	0.31372	0.34763	0.33796	0.33605	0.32657	0.09118	0.10208	0	0.13983	0.37228	0.40958
0.36559	0.32569	0.31405	0.30676	0.30046	0.30189	0.14096	0.13529	0.13983	0	0.37219	0.42374
0.36204	0.34898	0.36901	0.34554	0.36364	0.33962	0.3763	0.38438	0.37228	0.37219	0	0.4225
0.30856	0.43145	0.42899	0.41714	0.39769	0.40629	0.41546	0.42022	0.40958	0.42374	0.4225	0

h) Distance Matrix Calculated by the LOGDET Method:

0	0.37817	0.38003	0.3508	0.35067	0.36098	0.34267	0.35277	0.37086	0.34446	0.34735	0.2936
0.37817	0	0.10136	0.11919	0.19437	0.2217	0.28383	0.2866	0.30958	0.31771	0.3438	0.40966
0.38003	0.10136	0	0.11929	0.20446	0.22705	0.30461	0.31853	0.34736	0.30746	0.36373	0.40649
0.3508	0.11919	0.11929	0	0.20128	0.22682	0.28784	0.2852	0.33211	0.29653	0.34341	0.39853
0.35067	0.19437	0.20446	0.20128	0	0.2234	0.30467	0.30687	0.33466	0.29251	0.35401	0.3762
0.36098	0.2217	0.22705	0.22682	0.2234	0	0.31136	0.2949	0.32671	0.29975	0.3367	0.39289
0.34267	0.28383	0.30461	0.28784	0.30467	0.31136	0	0.0393	0.0929	0.14569	0.37609	0.40333
0.35277	0.2866	0.31853	0.2852	0.30687	0.2949	0.0393	0	0.10593	0.13923	0.38446	0.40852
0.37086	0.30958	0.34736	0.33211	0.33466	0.32671	0.0929	0.10593	0	0.14508	0.3688	0.39951
0.34446	0.31771	0.30746	0.29653	0.29251	0.29975	0.14569	0.13923	0.14508	0	0.36837	0.41224
0.34735	0.3438	0.36373	0.34341	0.35401	0.3367	0.37609	0.38446	0.3688	0.36837	0	0.41656
0.2936	0.40966	0.40649	0.39853	0.3762	0.39289	0.40333	0.40852	0.39951	0.41224	0.41656	0

The list of triangle inequalities that are not satisfied is given below.

Dataset	Method	Number	Number of	Tri. Inq.	Tri. Inq. not	%
Primates-mtDNA	BASEFREQ	12	220	183	37	16.82
Primates-mtDNA	F81	12	220	220	0	0.00
Primates-mtDNA	F84	12	220	220	0	0.00
Primates-mtDNA	GTR	12	220	220	0	0.00
Primates-mtDNA	HAMMING	12	220	220	0	0.00
Primates-mtDNA	HKY85	12	220	220	0	0.00
Primates-mtDNA	K3ST	12	220	220	0	0.00
Primates-mtDNA	LOGDET	12	220	220	0	0.00

Crab-rRNA Dataset (13 Elements): The crab-rRNA dataset consists of a total of thirteen taxa consisting of sequences for the gene encoding the mitochondrial large subunit ribosomal RNA. The distance matrix and the results are as shown below.

Distance Matrix:

0	0.43988	0.4308	0.34872	0.34832	0.37547	0.39137	0.36516	0.36516	0.39951	0.39313	0.36575	0.36181
0.43988	0	0.27713	0.36457	0.35041	0.37858	0.35923	0.37767	0.38311	0.32078	0.33668	0.33976	0.33976
0.4308	0.27713	0	0.35716	0.34478	0.33806	0.33041	0.37094	0.35545	0.37	0.35765	0.36905	0.37301
0.34872	0.36457	0.35716	0	0.01733	0.07763	0.0696	0.08024	0.07732	0.1395	0.12267	0.11977	0.11681
0.34832	0.35041	0.34478	0.01733	0	0.07735	0.06647	0.07995	0.07704	0.13257	0.11894	0.1102	0.11313
0.37547	0.37858	0.33806	0.07763	0.07735	0	0.05729	0.02782	0.02521	0.14952	0.13236	0.12912	0.12613
0.39137	0.35923	0.33041	0.0696	0.06647	0.05729	0	0.06846	0.06	0.14467	0.13082	0.13089	0.12786
0.36516	0.37767	0.37094	0.08024	0.07995	0.02782	0.06846	0	0.00738	0.15274	0.1355	0.12563	0.12266
0.36516	0.38311	0.35545	0.07732	0.07704	0.02521	0.06	0.00738	0	0.15274	0.1355	0.12892	0.12594
0.39951	0.32078	0.37	0.1395	0.13257	0.14952	0.14467	0.15274	0.15274	0	0.01995	0.05785	0.05785
0.39313	0.33668	0.35765	0.12267	0.11894	0.13236	0.13082	0.1355	0.1355	0.01995	0	0.04932	0.04668
0.36575	0.33976	0.36905	0.11977	0.1102	0.12912	0.13089	0.12563	0.12892	0.05785	0.04932	0	0.00244
0.36181	0.33976	0.37301	0.11681	0.11313	0.12613	0.12786	0.12266	0.12594	0.05785	0.04668	0.00244	0

The list of triangle inequalities that are not satisfied:

Dataset	Method	Number of Taxa	Number of Tri. Ineq.	Tri. Inq. satisfied	Tri. Inq. not satisfied	%
Crab_rRNA	F86	13	286	283	3	1.05

Hum_Dist Dataset (15 Elements): The distance matrix from the human communities belonging to 15 nations and the examination are as shown in the tables below.

Distance Matrix:

0	0.028	0.0342	0.071	0.1172	0.1164	0.1188	0.1203	0.1289	0.1267	0.1335	0.1523	0.1569	0.1371	0.1408
0.028	0	0.0244	0.0686	0.1189	0.1168	0.1193	0.117	0.1334	0.1284	0.1388	0.1579	0.1551	0.1416	0.1406
0.0342	0.0244	0	0.0355	0.0962	0.0957	0.0972	0.0953	0.1052	0.1033	0.1128	0.1414	0.1402	0.1144	0.1193
0.071	0.0686	0.0355	0	0.0863	0.087	0.0859	0.087	0.0945	0.0922	0.1025	0.1274	0.1315	0.0987	0.1103
0.1172	0.1189	0.0962	0.0863	0	0.0041	0.0045	0.007	0.048	0.0479	0.0598	0.0961	0.0899	0.0587	0.0645
0.1164	0.1168	0.0957	0.087	0.0041	0	0.0022	0.005	0.0501	0.0509	0.0601	0.1075	0.0955	0.0659	0.0707
0.1188	0.1193	0.0972	0.0859	0.0045	0.0022	0	0.0067	0.0547	0.0543	0.0648	0.1091	0.1006	0.0674	0.0704
0.1203	0.117	0.0953	0.087	0.007	0.005	0.0067	0	0.0492	0.0493	0.0581	0.1045	0.094	0.0616	0.0663
0.1289	0.1334	0.1052	0.0945	0.048	0.0501	0.0547	0.0492	0	0.0053	0.0203	0.0559	0.05	0.0373	0.0488
0.1267	0.1284	0.1033	0.0922	0.0479	0.0509	0.0543	0.0493	0.0053	0	0.0188	0.0497	0.0517	0.0362	0.0484
0.1335	0.1388	0.1128	0.1025	0.0598	0.0601	0.0648	0.0581	0.0203	0.0188	0	0.0676	0.0595	0.0553	0.0635
0.1523	0.1579	0.1414	0.1274	0.0961	0.1075	0.1091	0.1045	0.0559	0.0497	0.0676	0	0.0396	0.0712	0.0793
0.1569	0.1551	0.1402	0.1315	0.0899	0.0955	0.1006	0.094	0.05	0.0517	0.0595	0.0396	0	0.0875	0.095
0.1371	0.1416	0.1144	0.0987	0.0587	0.0659	0.0674	0.0616	0.0373	0.0362	0.0553	0.0712	0.0875	0	0.0313
0.1408	0.1406	0.1193	0.1103	0.0645	0.0707	0.0704	0.0663	0.0488	0.0484	0.0635	0.0793	0.095	0.0313	0

The list of triangle inequalities that are not satisfied

Dataset	Method	Number of Taxa	Number of Tri. Ineq.	Tri. Inq. satisfied	Tri. Inq. not satisfied	%
Hum_Dist	F86	15	455	423	32	7.03

Mammals Dataset (30 Elements): Mammals consists of a total of 30 taxa: 'Homo_sap' , 'Pan_trog' , 'Pan_panisc' , 'Gorilla' , 'Pongo_pygB' , 'Pongo_pygS' , 'Hylobat_la' , 'harbor_sel' , 'grey_seal' , 'Felis_cat' , 'Equus_cab' , 'Rhinoceros' , 'Bos_ta_cow' , 'fin_whale' , 'blue_whale' , 'Rattus_norv' , 'Mus_mouse' , 'guinea_pig' , 'rabbit' , 'hedgehog' , 'wallaroo' , 'opossum' , 'platypus' , 'chicken' , 'Xenopus_la' , 'carp' , 'loach' , 'trout' , 'Atlant_cod' ve '_bichir'. Since the distance matrix of the mammals dataset is not suitable to be shown in the report, only the results obtained with the distance matrices calculated by BASEFREQ, HAMMING and LOGDET methods are shown below. Distance matrices are given in the website.

The list of triangle inequalities that are not satisfied

Dataset	Method	Number of Taxa	Number of Tri. Ineq.	Tri. Inq. satisfied	Tri. Inq. not satisfied	%
Mammals	BASEFREQ	30	4060	4060	0	0.00
Mammals	HAMMING	30	4060	4060	0	0.00
Mammals	LOGDET	30	4060	4060	0	0.00

Chainletters Dataset (33 Elements): Chainletters dateset consist of a total of 33 taxa. Since the distance matrix of the Chainletters dataset is not suitable to be shown in the report, only the result is given below. The distance matrix is given in the website.

The list of triangle inequalities that are not satisfied

Dataset	Method	Number of Taxa	Number of Tri. Ineq.	Tri. Inq. satisfied	Tri. Inq. not satisfied	%
Chainletters	F86	33	5456	5456	0	0.00

D_Loop_Vigillant Dataset (70 Elements): Consisting of a total of 70 taxa which are from D-loop human mtDNA data oluşan is examined, and the result is shown below. The distance matrix related to this dataset is given in the website.

The list of triangle inequalities that are not satisfied

Dataset	Method	Number of	Number of Tri.	Tri. Inq.	Tri. Inq. not	%
D_Loop_Vigilant	F86	70	52394	50521	1873	3.57

South Dataset (7 Elements):

The South dataset consists of a total of seven taxa: "Christchurch", "Wellington", "Dunedin", "Greymouth", "Invercargill", "Picton" and "Queenstown". The distance matrix and the results are as shown below.

Distance Matrix

0	403	361	249	575	322	535
403	0	764	424	968	139	938
361	764	0	600	214	654	310
249	424	600	0	717	343	555
575	968	214	717	0	897	185
322	139	654	343	897	0	857
535	938	310	555	185	857	0

The list of triangle inequalities that are not satisfied

Dataset	Method	Number of Taxa	Number of Tri. Ineq.	Tri. Inq. satisfied	Tri. Inq. not satisfied	%
South	F86	7	35	29	6	17.14

Gromov Product Structures

P_1	P_2	P_3	Minimal Δ
1	2	3	0
1	2	7	0
1	5	6	0
1	6	7	0
2	5	6	105
3	5	6	-14.5
4	2	7	20.5
4	6	7	20.5

5	3	7	44.5
6	2	3	14.5
7	4	5	11.5

mtCDNA Dataset (7 elements): The distance matrix obtained from the 7 species in total including tulips such as homo sapiens, chimpanzees and orangutan and its examination is as shown below.

Distance Matrix:

0	0.356	0.345	0.498	0.962	0.934	1.343
0.356	0	0.114	0.464	0.947	0.943	1.349
0.345	0.114	0	0.446	0.94	0.946	1.318
0.498	0.464	0.446	0	0.994	0.988	1.378
0.962	0.947	0.94	0.994	0	0.202	1.432
0.934	0.943	0.946	0.988	0.202	0	1.394
1.343	1.349	1.318	1.378	1.432	1.394	0

The list of triangle inequalities that are not satisfied

Dataset	Method	Number of Taxa	Number of Tri. Ineq.	Tri. Inq. satisfied	Tri. Inq. not satisfied	%
mtCDNA	F86	7	35	35	0	0.00

Gromov Product Structures

P_1	P_2	P_3	Minimal Δ
1	3	6	0.1665
2	3	6	0.0555
3	2	7	0.0415
4	3	6	0.244
5	3	6	0.098
6	5	7	0.082
7	3	6	0.883

Chloroplast_Martin Dataset(10 Elements):Chloroplast_Martin dataset consists of ten taxa, including 'Synechocys', 'Odontella', 'Porphyra', 'Cyanophora', 'Euglema', 'Marchantia', 'Pinus', 'Nicotiana', 'Zea' and 'Oryza'. The distance matrix and results are as shown below.

Distance Matrisi :

0	0.42	0.33	0.33	0.51	0.44	0.46	0.47	0.49	0.49
0.42	0	0.36	0.4	0.5	0.47	0.5	0.51	0.52	0.52
0.33	0.36	0	0.32	0.48	0.41	0.44	0.45	0.47	0.47
0.33	0.4	0.32	0	0.48	0.4	0.44	0.45	0.47	0.47
0.51	0.5	0.48	0.48	0	0.47	0.5	0.5	0.51	0.51
0.44	0.47	0.41	0.4	0.47	0	0.22	0.24	0.28	0.28
0.46	0.5	0.44	0.44	0.5	0.22	0	0.22	0.25	0.25
0.47	0.51	0.45	0.45	0.5	0.24	0.22	0	0.15	0.15
0.49	0.52	0.47	0.47	0.51	0.28	0.25	0.15	0	0.03
0.49	0.52	0.47	0.47	0.51	0.28	0.25	0.15	0.03	0

The list of triangle inequalities that are not satisfied

Dataset	Method	Number of	Number of Tri. Ineq.	Tri. Inq. satisfied	Tri. Inq. not satisfied	%
Chloroplast Martin	F86	10	120	120	0	0.00

Gromov Product Structures

P_1	P_2	P_3	Minimal Δ
1	3	4	0.17
2	3	5	0.19

3	1	2	0.135
4	1	6	0.145
5	2	8	0.245
5	2	9	0.245
5	2	10	0.245
6	4	7	0.09
7	6	9	0.095
7	6	10	0.095
8	6	9	0.055
8	6	10	0.055
9	8	10	0.015
10	8	9	0.015

Drosophila_Adh Dataset (11 Elements): The distance matrix of this dataset, created with 11 adh sequences, is given below, and the results are as shown in the table.

Distance Matrix:

0	0.154	0.2356	0.3075	0.2924	0.2563	0.3066	0.3049	0.2878	0.3147	0.3126
0.154	0	0.2339	0.2984	0.2988	0.233	0.288	0.2827	0.2762	0.2996	0.2767
0.2356	0.2339	0	0.3238	0.3264	0.252	0.3031	0.3064	0.2939	0.3036	0.2858
0.3075	0.2984	0.3238	0	0.0988	0.2536	0.2091	0.2041	0.1922	0.2067	0.2006
0.2924	0.2988	0.3264	0.0988	0	0.2558	0.2116	0.2244	0.2151	0.2219	0.2121
0.2563	0.233	0.252	0.2536	0.2558	0	0.2053	0.205	0.2053	0.2082	0.2077
0.3066	0.288	0.3031	0.2091	0.2116	0.2053	0	0.0377	0.0634	0.0633	0.097
0.3049	0.2827	0.3064	0.2041	0.2244	0.205	0.0377	0	0.0547	0.0617	0.088
0.2878	0.2762	0.2939	0.1922	0.2151	0.2053	0.0634	0.0547	0	0.0591	0.0792
0.3147	0.2996	0.3036	0.2067	0.2219	0.2082	0.0633	0.0617	0.0591	0	0.0972
0.3126	0.2767	0.2858	0.2006	0.2121	0.2077	0.097	0.088	0.0792	0.0972	0

The list of triangle inequalities that are not satisfied

Dataset	Method	Number	Number of	Tri. Inq.	Tri. Inq. not	%
Drosophila_Adh	F86	11	120	120	0	0.00

Violation of triangle inequalities for all datasets

No	Dataset	Method	Number	Number of	Tri. Inq.	Tri. Inq. not	%
1	Algea	BASEFREQ	8	56	22	34	60.71
2	Algea	F81	8	56	56	0	0.00
3	Algea	F84	8	56	56	0	0.00
4	Algea	GTR	8	56	56	0	0.00
5	Algea	HAMMING	8	56	56	0	0.00
6	Algea	HKY85	8	56	56	0	0.00
7	Algea	K3ST	8	56	56	0	0.00
8	Algea	LOGDET	8	56	56	0	0.00
9	Algea	F84	8	56	56	0	0.00
10	Bee	BASEFREQ	6	20	14	6	30.00
11	Bee	F81	6	20	20	0	0.00
12	Bee	F84	6	20	20	0	0.00
13	Bee	GTR	6	20	20	0	0.00
14	Bee	HAMMING	6	20	20	0	0.00
15	Bee	HKY85	6	20	20	0	0.00
16	Bee	K3ST	6	20	20	0	0.00
17	Bee	LOGDET	6	20	20	0	0.00
18	Chainletters	F86	33	5456	5456	0	0.00
19	south	F86	7	35	29	6	17.14
20	Distance Between	Euclidean	81	82160	51923	30237	36.80
21	Mammals	BASEFREQ	30	4060	4060	0	0.00
22	Mammals	HAMMING	30	4060	4060	0	0.00
23	Mammals	LOGDET	30	4060	4060	0	0.00
24	Primates-mtDNA	BASEFREQ	12	220	183	37	16.82
25	Primates-mtDNA	F81	12	220	220	0	0.00
26	Primates-mtDNA	F84	12	220	220	0	0.00
27	Primates-mtDNA	GTR	12	220	220	0	0.00
28	Primates-mtDNA	HAMMING	12	220	220	0	0.00
29	Primates-mtDNA	HKY85	12	220	220	0	0.00
30	Primates-mtDNA	K3ST	12	220	220	0	0.00
31	Primates-mtDNA	LOGDET	12	220	220	0	0.00
32	Chloroplast Martin	F86	10	120	120	0	0.00
33	Crab_rRNA	F86	13	286	283	3	1.05
34	D_Loop_Vigilant	F86	70	52394	50521	1873	3.57
35	Drosophila_Adh	F86	10	120	120	0	0.00
36	Hum_Dist	F86	15	455	423	32	7.03
37	mtCDNA	F86	7	35	35	0	0.00

Non-generic metrics in Phylogenetics: The Bee family

Ayşe Hümeyra Bilge
Kadir Has University

In phylognetics, discrepencies between species are measured by certain metrics defined via the dissimilarites in their DNA of Protein Coding sequences. In many cases, these dissimilarity measures do not satisfy triangle inequalities, hence they fail to be metrics in a mathematical sense. Researchers in phylogenetics aim to generate evolutionary trees starting from these dissimilarity measures, even if the metric they start with is not a tree metric. We recall that a metric is a tree metric, if and only if for every quadruple of points (i,j,k,l) the the largest two of the numbers

$$d_{ij}+d_{kl},\ d_{ik}+d_{jl},\ d_{il}+d_{jk}$$

are equal. This is a quite strong condition and it holds only in elementary textbook examples. Thus realistic distance functions that one is likely to encounter in phylogenetics are not expected to have tree structures. Nevertheless, the examples we studied also show that they may also fail to be "generic" in the senses we defined for the classifications via Gromov products or quadrangles.

In this report we discuss a specific example which constituted the starting point of our approach to the study of the examples from phylogenetics.

This is the example given on the webpage
http://page.math.tu-berlin.de/~joswig/tightspans/index.html
to illustrate the construction of the tight span of the metric below, for the distances between the members of a Bee family:

Table 1: Distance Functions

A.andrenof	A.mellifer	A.dorsata	A.cerana	A.florea	A.koschev
0.0	0.090103395	0.10339734	0.09601182	0.0044313148	0.07533235
0.090103395	0.0	0.09305761	0.090103395	0.09305761	0.10044313
0.10339734	0.09305761	0.0	0.116691284	0.106351554	0.10339734
0.09601182	0.090103395	0.116691284	0.0	0.098966025	0.098966025
0.0044313148	0.09305761	0.106351554	0.098966025	0.0	0.07828656
0.07533235	0.10044313	0.10339734	0.098966025	0.07828656	0.0

When we compute their Gromov products we see that the metric is generic:

Table 2. Minimal Gromov Products

P_a	P_b	P_c		Δ_{abc}
1.0000	2.0000	5.0000		0.0007
2.0000	3.0000	4.0000		0.0332
3.0000	2.0000	6.0000		0.0480
4.0000	2.0000	6.0000		0.0443
5.0000	1.0000	4.0000		0.0037
6.0000	3.0000	5.0000		0.0377

From this table, we see that although the metric is generic, the minimal Gromov product at node 1 is small compared with the others.

We add random perturbations of order 10^{-6} to the metric and we compute the Gromov products for each perturbation. In most cases, although the perturbed metric remains generic, its Gromov product type changes. As an example we present the full list of Gromov products for one such perturbation. The values that are very close to each other are marked by (*).

Table 3. Gromov products of perturbed metrics for the bee family

$\Delta_{123} = 0.050221562500000$	$\Delta_{213} = 0.039881832500000$	$\Delta_{312} = 0.053175777500000$
$\Delta_{124} = 0.048005910000000$	$\Delta_{214} = 0.042097485000000$	$\Delta_{314} = 0.062038402000000$
$\Delta_{125} = 0.000738549900000$ *	$\Delta_{215} = 0.089364845100000$	$\Delta_{315} = 0.102658789600000$
$\Delta_{126} = 0.032496307500000$	$\Delta_{216} = 0.057607087500000$	$\Delta_{316} = 0.065731165000000$
$\Delta_{134} = 0.041358938000000$	$\Delta_{234} = 0.033234860500000$ *	$\Delta_{324} = 0.059822749500000$
$\Delta_{135} = 0.000738550400000$ *	$\Delta_{235} = 0.039881833000000$	$\Delta_{325} = 0.053175777000000$
$\Delta_{136} = 0.037666175000000$	$\Delta_{236} = 0.045051700000000$	$\Delta_{326} = 0.048005910000000$ *
$\Delta_{145} = 0.000738554900000$ *	$\Delta_{245} = 0.042097490000000$	$\Delta_{345} = 0.062038406500000$
$\Delta_{146} = 0.036189072500000$	$\Delta_{246} = 0.045790250000000$	$\Delta_{346} = 0.060561299500000$
$\Delta_{156} = 0.000738552400000$ *	$\Delta_{256} = 0.057607090000000$	$\Delta_{356} = 0.065731167000000$
$\Delta_{412} = 0.048005910000000$	$\Delta_{512} = 0.003692764900000$ *	$\Delta_{612} = 0.042836042500000$
$\Delta_{413} = 0.054652882000000$	$\Delta_{513} = 0.003692764400000$ *	$\Delta_{613} = 0.037666175000000$ *
$\Delta_{415} = 0.095273265100000$	$\Delta_{514} = 0.003692759900000$ *	$\Delta_{614} = 0.039143277500000$
$\Delta_{416} = 0.059822747500000$	$\Delta_{516} = 0.003692762400000$ *	$\Delta_{615} = 0.074593797600000$
$\Delta_{423} = 0.056868534500000$	$\Delta_{523} = 0.053175777000000$	$\Delta_{623} = 0.055391430000000$
$\Delta_{425} = 0.048005905000000$	$\Delta_{524} = 0.050960120000000$	$\Delta_{624} = 0.054652880000000$
$\Delta_{426} = 0.044313145000000$ *	$\Delta_{526} = 0.035450520000000$	$\Delta_{625} = 0.042836040000000$
$\Delta_{435} = 0.054652877500000$	$\Delta_{534} = 0.044313147500000$	$\Delta_{634} = 0.042836040500000$
$\Delta_{436} = 0.056129984500000$	$\Delta_{536} = 0.040620387000000$	$\Delta_{635} = 0.037666173000000$ *
$\Delta_{456} = 0.059822745000000$	$\Delta_{546} = 0.039143280000000$	$\Delta_{645} = 0.039143280000000$

As the next step we compute the Hamming distance, directly from their sequence codes, as

Table 4. Hamming distance metric for the Bee family

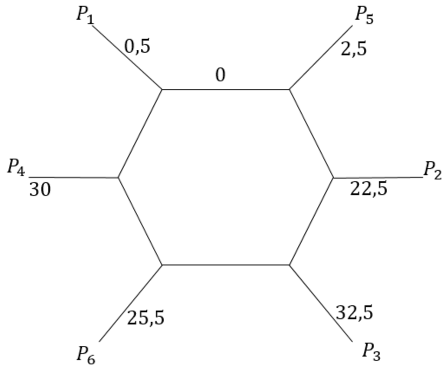
	P_1	P_2	P_3	P_4	P_5	P_6
P_1	0	61	70	65	3	51
P_2	61	0	63	61	63	68
P_3	70	63	0	79	72	70
P_4	65	61	79	0	67	67
P_5	3	63	72	67	0	53
P_6	51	68	70	67	53	0

Gromov products are computed and it is seen that the metric is not generic.

Table 5. The non-generic Gromov product structure for the Bee family

P_a	P_b	P_c	Δ_{abc}
1	2	5	0.5
1	3	5	0.5
1	4	5	0.5
1	5	6	0.5
2	3	4	22.5
3	2	6	32.5
4	2	6	30.0
5	1	2	2.5
5	1	3	2.5
5	1	4	2.5
5	1	6	2.5
6	1	3	25.5
6	3	5	25.5

Note that the distance between P_1 and P_5 is exactly 3=0.5+2.5, that is the 6-point graph, the distance between auxiliary points is zero. This means, the metric effectively a 5-point metric, with a branching in one of the auxiliary nodes, as shown below. It can be seen that as a 5-point it is of Type B. This example shows that, dissimilarity measures used in phylogenetics should be treated with care, in order to determine stable properties of species to be analysed.



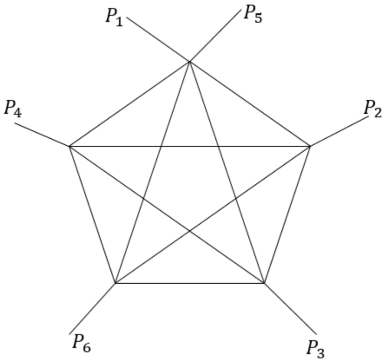


Table 6. Hamming distance metric for the reduced Bee family

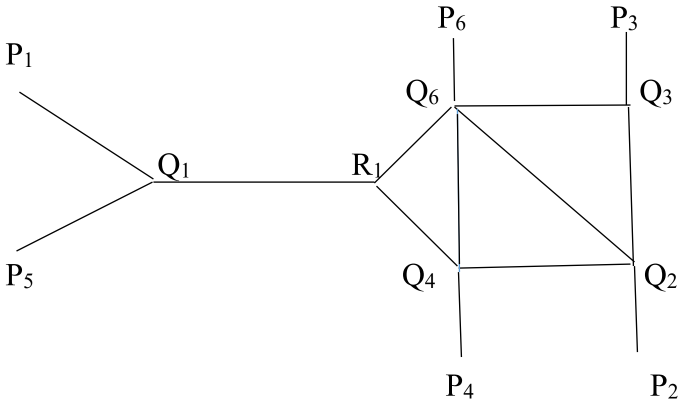
	Q_1	Q_2	Q_3	Q_4	Q_6
Q_1	0	38.0	37.0	34.5	25.0
Q_2	38.0	0	8.0	8.5	20.0
Q_3	37.0	8.0	0	16.5	12.0
Q_4	34.5	8.5	16.5	0	11.5
Q_6	25.0	20.0	12.0	11.5	0

Gromov products are computed and it is seen that the metric is now generic.

Table 7. Gromov product structure for the reduced Bee family

Qa	Qb	Qc	Δ_{abc}
1	2	6	21.5
2	3	4	0
3	2	6	0
4	2	6	0
6	1	3	0

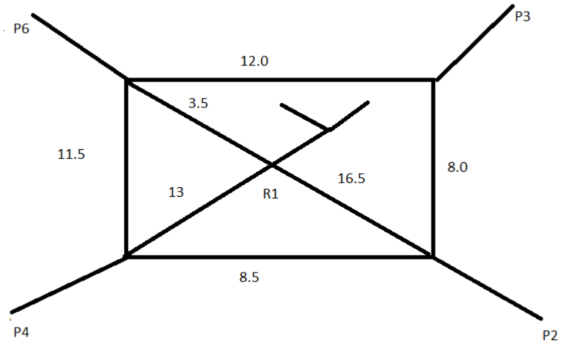
The reduced metric has Type B and the graph representation is as below.



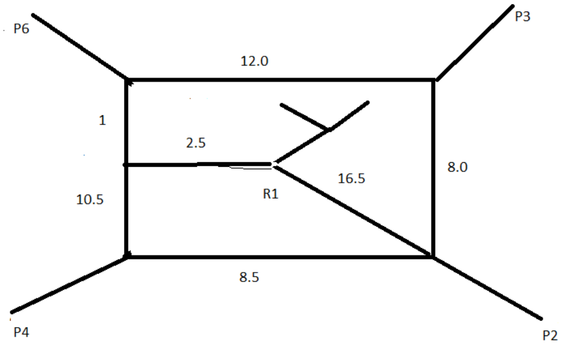
The distances in the graph above are

$d(P_1,Q_1)=0.5$, $d(P_5,Q_1)=2.5$, $d(P_6,Q_6)=25.5$, $d(P_3,Q_3)=32.5$,
$d(P_2,Q_2)=22.5$, $d(P_4,Q_4)=30.0$, $d(Q_1,R_1)=21.5$, $d(R_1,Q_6)=3.5$,
$d(R_1,Q_4)=13.0$, $d(Q_6,Q_3)=12.0$, $d(Q_3,Q_2)=8.0$, $d(Q_2,Q_4)=8.5$,
$d(R_1,Q_2)=16.5$, $d(Q_6,Q_4)=11.5$.

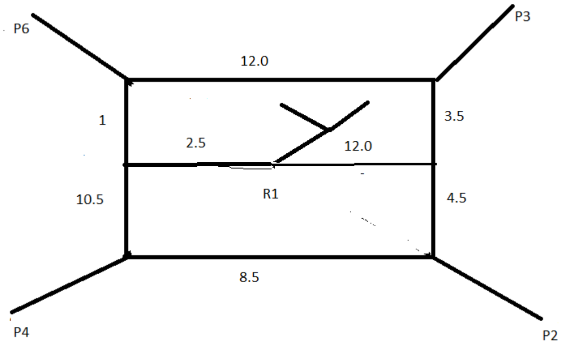
We redraw the graph to get a form more suitable to the optimal reduction of Type B metric.



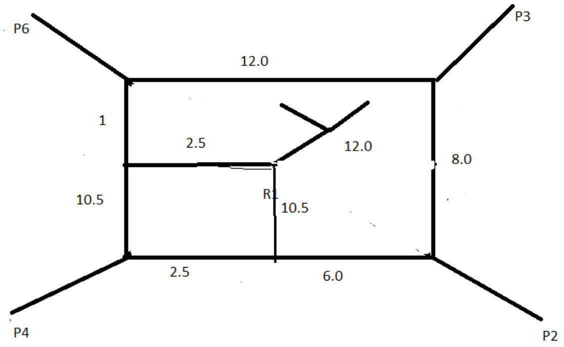
In this graph we apply Delta-Y transform to the triangle (R_1,Q_6,Q_4), to get the figure below.



At this step we can move the line segment (R_1,Q_2) to right or to left:



If we move to left we get the figure above, in which to total length is decreased by 4.5. If we move to right we get the next graph.



In the second graph, the total length is decreased by 6. Thus the second graph is the optimal realization. But if we consider these reduction as alternatives for ancestor reconstruction, both graphs seem plausible.

A New Method for Constructing Phylogenetic Trees

Ayşe Hümeyra Bilge,[1*] Demet Akten Akdogan,[2]
Arash M. Rezaeinazhad,[3] Burak Erkan KAYA[1]

[1]Department of Industrial Engineering, Kadir Has University,
[2]Department of Molecular Biology and Genetics, Kadir Has University
[3]Department of Management Information Systems, Kadir Has University
[*]Ayşe Hümeyra Bilge; E-mail: ayse.bilge@khas.edu.tr

We use finite metric space methods to decide whether a given distance function can be approximated by a tree.

1 Introduction

In phylognetics, discrepancies between species are measured by certain metrics, such as Hamming, JC69 (*1*), K80 (*2*), F81 (*3*), K81 (*4*), HKY85 (*5*), defined via the dissimilarities in their DNA or Protein Coding sequences. In many cases, these dissimilarity measures do not satisfy triangle inequalities, hence they fail to be metrics in a mathematical sense. Researchers in phylogenetics aim to generate evolutionary trees starting from these dissimilarity measures, even if the metric they start with is not a tree metric. In this work we obtain criteria for deciding whether a given metric can be approximated by a tree metric.

In the mathematical sense, a set of distances d_{ij} among objects is a "metric" if these distances satisfy the triangle inequality,

$$d_{ij} + d_{jk} \geq d_{ik}, \tag{1}$$

for each i, j, k, i.e, for every triple of points $\{P_i, P_j, P_k\}$. Thus, when we find a candidate for an approximate tree metric, we have to ensure that triangle inequalities are satisfied.

Our criterion for deciding on the existence of a tree metric will be given in terms of the metric structure of quadruples of points $\{P_i, P_j, P_k, P_l\}$. The collection of these points (without any order relation implied, will be denoted as a a quadrangle $Q(i, j, k, l)$. The order relations among the quantities

$$a = d_{ij} + d_{kl}, \quad b = d_{ik} + d_{jl}, \quad c = d_{il} + d_{jk} \tag{2}$$

determine the shape of the graphs that represent the distances among these points. We may assume that $a \geq b \geq c$. If $a = b = c$ then the graph is star shaped, if $a = b > c$ then it is tree-like, and if $a > b$ (irrelevant of the order relation between b and c, it has a loop. It is known that such a metric is representable by a tree, if and only if for every quadruple of points (i, j, k, l) the largest two of the numbers are equal.

This is a quite strong condition that holds exactly only for very special, simple examples. Thus realistic distance functions that one is likely to encounter in phylogenetics do not satisfy this criterion *exactly*.

In order to determine the degree of approximation we performed a number of simulations by giving random perturbations to the sequences. We compute distance matrices for every taxa by using Hamming distance. Then with 200 trials and by altering the 10% of DNA of Protein Coding sequence members the stability of the metric is studied. The process of altering the sequence for each trial is done by random perturbations in the sequence itself, meaning that 10% of characters in the total number of taxon of each family sequences are randomly selected and randomly replaced by either one of the four characters 'a', 'g', 't' or 'c'. We implemented our algorithm by using a program that is coded in Python and MATLAB. In this study, we discussed some specific families as examples; the Bee, Algae, Bat, Zebra Fish families. We have seen that this was in fact leading to about 10% change in the metrics.

In order to decide whether the metric can be represented by a tree or not, we proceed as follows. For each quadrangle $Q(i, j, k, l)$, we compute the quantities a, b, c defined in (2) and order them as $a \geq b \geq c$. We compute the differences $a - b$ and $a - c$. We then compute the ratios r_{max} and r_{min},

$$r_{max} = max\{(a-b)/a, (a-c)/a\} \quad r_{min} = min\{(a-b)/a, (a-c)/a\}$$

for all quadrangles and pick a threshold level α. If

$$max(r_{min}) < \alpha, \tag{3}$$

then the tree metric $\tilde{d}$ approximating the given metric d should have $\tilde{a} = \tilde{b}$. Note that

$$a(1-\alpha) < b < a.$$

If $|\tilde{d_{ij}} - d_{ij}| < \epsilon$, then it can be seen that (3) will hold, provided that

$$\epsilon < \frac{b-(1-\alpha)a}{2(2-\alpha)}. \tag{4}$$

The critical question is to be able to choose ϵ is small enough to ensure that the perturbed distance functions $\tilde{d}$ satisfy triangle inequalities.

2 Bee Family

The Bee family is an example studied by [ref]. It consists of 6 species: A.andrenof; A.mellifer; A.dorsata; A.cerana; A.florea; A.koschev. DNA of Protein Coding sqeuence of this family is obtained from a well-known phylogenetic program of Huson and Bryant (*6*) namely SplitsTree4. The original distance matrix (D_1) and the minimal Gromov products that appear after these alterations are given below:

$$D_1 = \begin{bmatrix} 0 & 61 & 70 & 65 & 3 & 51 \\ 61 & 0 & 63 & 61 & 63 & 68 \\ 70 & 63 & 0 & 79 & 72 & 70 \\ 65 & 61 & 79 & 0 & 67 & 67 \\ 3 & 63 & 72 & 67 & 0 & 53 \\ 51 & 68 & 70 & 67 & 53 & 0 \end{bmatrix} \tag{5}$$

The quadrangle excess's are listed in the tables below. For all generic quadrangles $r_{min} < 0.12$, hence the metric can be approximated by a tree metric at level $\alpha = 0.12$.

$Q(i,j,k,l)$	$d_{ij}+d_{kl}$	$d_{ik}+d_{jl}$	$d_{il}+d_{jk}$	r_{min}	r_{max}
$Q(1,2,3,4)$	140	131	128	0.0643	0.0857
$Q(1,2,3,5)$	133	133	66	0.5038	0.5038
$Q(1,2,3,6)$	131	138	114	0.0507	0.1739
$Q(1,2,4,5)$	128	128	64	0.5000	0.5000
$Q(1,2,4,6)$	128	133	112	0.0376	0.1579
$Q(1,2,5,6)$	114	71	114	0.3772	0.3772
$Q(1,3,4,5)$	137	137	82	0.4015	0.4015
$Q(1,3,4,6)$	137	135	130	0.0146	0.0511
$Q(1,3,5,6)$	123	73	123	0.4065	0.4065
$Q(1,4,5,6)$	118	70	118	0.4068	0.4068
$Q(2,3,4,5)$	130	133	142	0.0634	0.0845
$Q(2,3,4,6)$	130	131	147	0.1088	0.1156
$Q(2,3,5,6)$	116	133	140	0.0500	0.1714
$Q(2,4,5,6)$	114	130	135	0.0370	0.1556
$Q(3,4,5,6)$	132	139	137	0.0144	0.0504

3 Algea Family

In this paragraph we will investigate the Algae family, this family has 8 members consisting of Tobacco, Rice, Marchantia, Chlamydomonas, Chlorella, Euglena, Anacystis Nidulans and Olithodiscus and the length of each DNA sequence is 920. (*6*) The original distance matrix (D_2) of the Algea family and the minimal Gromov products that appear after these alterations are given below:

$$D_2 = \begin{bmatrix} 0 & 25 & 28 & 107 & 85 & 134 & 114 & 131 \\ 25 & 0 & 38 & 114 & 93 & 140 & 121 & 134 \\ 28 & 38 & 0 & 94 & 71 & 121 & 111 & 123 \\ 107 & 114 & 94 & 0 & 106 & 141 & 134 & 147 \\ 85 & 93 & 71 & 106 & 0 & 113 & 120 & 119 \\ 134 & 140 & 121 & 141 & 113 & 0 & 154 & 133 \\ 114 & 121 & 111 & 134 & 120 & 154 & 0 & 126 \\ 131 & 134 & 123 & 147 & 119 & 133 & 126 & 0 \end{bmatrix} \tag{6}$$

Here the metric is tree reducible at level $\alpha = 0.11$.

$Q(i,j,k,l)$	$d_{ij} + d_{kl}$	$d_{ik} + d_{jl}$	$d_{il} + d_{jk}$	r_{min}	r_{max}
$Q(1,2,3,4)$	119	142	145	0.0207	0.1793
$Q(1,2,3,5)$	96	121	123	0.0163	0.2195
$Q(1,2,3,6)$	146	168	172	0.0233	0.1512
$Q(1,2,3,7)$	136	149	152	0.0197	0.1053
$Q(1,2,3,8)$	148	162	169	0.0414	0.1243
$Q(1,2,4,5)$	131	200	199	0.0050	0.3450
$Q(1,2,4,6)$	166	247	248	0.0040	0.3306
$Q(1,2,4,7)$	159	228	228	0.3026	0.3026
$Q(1,2,4,8)$	172	241	245	0.0163	0.2980
$Q(1,2,5,6)$	138	225	227	0.0088	0.3921
$Q(1,2,5,7)$	145	206	207	0.0048	0.2995
$Q(1,2,5,8)$	144	219	224	0.0223	0.3571
$Q(1,2,6,7)$	179	255	254	0.0039	0.2980
$Q(1,2,6,8)$	158	268	271	0.0111	0.4170
$Q(1,2,7,8)$	151	248	252	0.0159	0.4008
$Q(1,3,4,5)$	134	178	179	0.0056	0.2514
$Q(1,3,4,6)$	169	228	228	0.2588	0.2588
$Q(1,3,4,7)$	162	218	208	0.0459	0.2569
$Q(1,3,4,8)$	175	230	225	0.0217	0.2391
$Q(1,3,5,6)$	141	206	205	0.0049	0.3155
$Q(1,3,5,7)$	148	196	185	0.0561	0.2449
$Q(1,3,5,8)$	147	208	202	0.0288	0.2933
$Q(1,3,6,7)$	182	245	235	0.0408	0.2571
$Q(1,3,6,8)$	161	257	252	0.0195	0.3735
$Q(1,3,7,8)$	154	237	242	0.0207	0.3636
$Q(1,4,5,6)$	220	226	240	0.0583	0.0833
$Q(1,4,5,7)$	227	219	220	0.0308	0.0352
$Q(1,4,5,8)$	226	232	237	0.0211	0.0464
$Q(1,4,6,7)$	261	268	255	0.0261	0.0485
$Q(1,4,6,8)$	240	281	272	0.0320	0.1459
$Q(1,4,7,8)$	233	261	265	0.0151	0.1208
$Q(1,5,6,7)$	239	254	227	0.0591	0.1063
$Q(1,5,6,8)$	218	253	244	0.0356	0.1383
$Q(1,5,7,8)$	211	233	251	0.0717	0.1594
$Q(1,6,7,8)$	260	247	285	0.0877	0.1333

$Q(i,j,k,l)$	$d_{ij}+d_{kl}$	$d_{ik}+d_{jl}$	$d_{il}+d_{jk}$	r_{min}	r_{max}
$Q(2,3,4,5)$	144	185	187	0.0107	0.2299
$Q(2,3,4,6)$	179	235	234	0.0043	0.2383
$Q(2,3,4,7)$	172	225	215	0.0444	0.2356
$Q(2,3,4,8)$	185	237	228	0.0380	0.2194
$Q(2,3,5,6)$	151	214	211	0.0140	0.2944
$Q(2,3,5,7)$	158	204	192	0.0588	0.2255
$Q(2,3,5,8)$	157	216	205	0.0509	0.2731
$Q(2,3,6,7)$	192	251	242	0.0359	0.2351
$Q(2,3,6,8)$	171	263	255	0.0304	0.3498
$Q(2,3,7,8)$	164	244	245	0.0041	0.3306
$Q(2,4,5,6)$	227	234	246	0.0488	0.0772
$Q(2,4,5,7)$	234	227	227	0.0299	0.0299
$Q(2,4,5,8)$	233	240	240	0.0292	0.0292
$Q(2,4,6,7)$	268	274	262	0.0219	0.0438
$Q(2,4,6,8)$	247	287	275	0.0418	0.1394
$Q(2,4,7,8)$	240	268	268	0.1045	0.1045
$Q(2,5,6,7)$	247	260	234	0.0500	0.1000
$Q(2,5,6,8)$	226	259	247	0.0463	0.1274
$Q(2,5,7,8)$	219	240	254	0.0551	0.1378
$Q(2,6,7,8)$	266	254	288	0.0764	0.1181
$Q(3,4,5,6)$	207	212	227	0.0661	0.0881
$Q(3,4,5,7)$	214	205	217	0.0138	0.0553
$Q(3,4,5,8)$	213	218	229	0.0480	0.0699
$Q(3,4,6,7)$	248	255	252	0.0118	0.0275
$Q(3,4,6,8)$	227	268	264	0.0149	0.1530
$Q(3,4,7,8)$	220	258	257	0.0039	0.1473
$Q(3,5,6,7)$	225	241	224	0.0664	0.0705
$Q(3,5,6,8)$	204	240	236	0.0167	0.1500
$Q(3,5,7,8)$	197	230	243	0.0535	0.1893
$Q(3,6,7,8)$	247	244	277	0.1083	0.1191
$Q(4,5,6,7)$	260	261	247	0.0038	0.0536
$Q(4,5,6,8)$	239	260	260	0.0808	0.0808
$Q(4,5,7,8)$	232	253	267	0.0524	0.1311
$Q(4,6,7,8)$	267	267	301	0.1130	0.1130
$Q(5,6,7,8)$	239	253	273	0.0733	0.1245

4 Lasiurus (Red Bats) Family

The last family we investigated is Lasiurus family which consists of 18 red bats species. As seen in Table 1, although some of red bats' species or locations are the same they were defined via different tissue vouchers. We evaluated DBY squences of Lasiurus family whose length is 1362. The original distance matrix (D) is given below:

$$D = \begin{bmatrix} 0 & 856 & 856 & 942 & 847 & 808 & 847 & 917 & 925 & 917 & 922 & 922 & 931 & 931 & 944 & 857 & 946 & 938 \\ & 0 & 1 & 857 & 834 & 866 & 834 & 963 & 962 & 963 & 932 & 932 & 937 & 937 & 852 & 1 & 912 & 936 \\ & & 0 & 858 & 834 & 866 & 834 & 963 & 963 & 963 & 932 & 932 & 938 & 938 & 853 & 2 & 913 & 937 \\ & & & 0 & 916 & 922 & 916 & 885 & 864 & 885 & 824 & 824 & 707 & 707 & 149 & 856 & 217 & 864 \\ & & & & 0 & 529 & 0 & 918 & 939 & 918 & 927 & 927 & 918 & 918 & 921 & 833 & 919 & 919 \\ & & & & & 0 & 529 & 909 & 925 & 909 & 923 & 923 & 925 & 925 & 914 & 866 & 902 & 945 \\ & & & & & & 0 & 918 & 939 & 918 & 927 & 927 & 918 & 918 & 921 & 833 & 919 & 919 \\ & & & & & & & 0 & 363 & 0 & 790 & 790 & 870 & 870 & 875 & 963 & 877 & 836 \\ & & & & & & & & 0 & 363 & 830 & 830 & 868 & 868 & 867 & 962 & 863 & 809 \\ & & & & & & & & & 0 & 790 & 790 & 870 & 870 & 875 & 963 & 877 & 836 \\ & & & & & & & & & & 0 & 0 & 824 & 824 & 823 & 932 & 826 & 816 \\ & & & & & & & & & & & 0 & 824 & 824 & 823 & 932 & 826 & 816 \\ & & & & & & & & & & & & 0 & 0 & 701 & 936 & 698 & 862 \\ & & & & & & & & & & & & & 0 & 701 & 936 & 698 & 862 \\ & & & & & & & & & & & & & & 0 & 851 & 83 & 872 \\ & & & & & & & & & & & & & & & 0 & 911 & 935 \\ & & & & & & & & & & & & & & & & 0 & 872 \\ & & & & & & & & & & & & & & & & & 0 \end{bmatrix} \tag{7}$$

The number of quadrangles is 3060. Among these 598 are exactly degenerate. The remaining 2462 the maximum of r_{min} is 0.0773, thus the distance metric for the family is tree reducible at the level 0.0773.

Table 1: Lasirius (Red Bats) Family.

P_i	Genus	Species	Tissue voucher	Locality
1	Lasiurus	atratus	F54400	Guyana
2	Lasiurus	blossevillii	F44348	Brazil
3	Lasiurus	blossevillii	F44354	Brazil
4	Lasiurus	borealis	AK21073	Kansas
5	Lasiurus	cinereus	AK11014	Queretaro
6	Lasiurus	cinereus	NK11502	Bolivia
7	Lasiurus	cinereus	NK8096	Baja Cal
8	Lasiurus	ega	AK1635	Tamaulipas
9	Lasiurus	ega	NK15304	Bolivia
10	Lasiurus	ega	SP12622	Guatemala
11	Lasiurus	egregius	F54554	Guyana
12	Lasiurus	egregius	F54845	Suriname
13	Lasiurus	pfeifferi	TK32016	Cuba
14	Lasiurus	pfeifferi	TK32029	Cuba
15	Lasiurus	seminolus	AK21348	Texas
16	Lasiurus	frantzii	F34142	Guatemala
17	Lasiurus	varius	AK16070	Argentina
18	Lasiurus	xanthinus	NK11103	New Mexico

5 Fish family

This family has 20 members. Hamming distances are given below. This family is not tree-reducible. Quadrangle conditions are computed. Those quadrangles that are not tree reducible are listed. The largest α value is 0.3333.

$$D = \left[\begin{array}{cccccccccccccccccccc}
0 & 1 & 2 & 6 & 5 & 6 & 7 & 6 & 4 & 4 & 5 & 15 & 28 & 26 & 24 & 42 & 37 & 34 & 38 & 39 \\
1 & 0 & 1 & 7 & 6 & 5 & 8 & 7 & 5 & 5 & 6 & 16 & 27 & 27 & 23 & 42 & 36 & 33 & 39 & 40 \\
2 & 1 & 0 & 6 & 5 & 4 & 7 & 6 & 6 & 6 & 7 & 15 & 26 & 26 & 23 & 42 & 35 & 34 & 39 & 40 \\
6 & 7 & 6 & 0 & 1 & 2 & 9 & 8 & 4 & 4 & 5 & 17 & 26 & 26 & 25 & 45 & 39 & 38 & 39 & 40 \\
5 & 6 & 5 & 1 & 0 & 1 & 8 & 7 & 3 & 3 & 4 & 16 & 25 & 25 & 24 & 44 & 38 & 37 & 38 & 39 \\
6 & 5 & 4 & 2 & 1 & 0 & 9 & 8 & 4 & 4 & 5 & 17 & 24 & 26 & 23 & 44 & 37 & 36 & 39 & 40 \\
7 & 8 & 7 & 9 & 8 & 9 & 0 & 1 & 9 & 9 & 10 & 18 & 28 & 25 & 24 & 40 & 36 & 35 & 38 & 39 \\
6 & 7 & 6 & 8 & 7 & 8 & 1 & 0 & 8 & 8 & 9 & 17 & 27 & 24 & 23 & 39 & 37 & 34 & 37 & 38 \\
4 & 5 & 6 & 4 & 3 & 4 & 9 & 8 & 0 & 2 & 3 & 15 & 26 & 24 & 26 & 44 & 39 & 36 & 40 & 41 \\
4 & 5 & 6 & 4 & 3 & 4 & 9 & 8 & 2 & 0 & 1 & 17 & 26 & 26 & 26 & 44 & 39 & 36 & 40 & 41 \\
5 & 6 & 7 & 5 & 4 & 5 & 10 & 9 & 3 & 1 & 0 & 18 & 27 & 25 & 25 & 45 & 40 & 37 & 39 & 42 \\
15 & 16 & 15 & 17 & 16 & 17 & 18 & 17 & 15 & 17 & 18 & 0 & 29 & 29 & 28 & 48 & 41 & 40 & 42 & 41 \\
28 & 27 & 26 & 26 & 25 & 24 & 28 & 27 & 26 & 26 & 27 & 29 & 0 & 33 & 30 & 50 & 42 & 40 & 46 & 47 \\
26 & 27 & 26 & 26 & 25 & 26 & 25 & 24 & 24 & 26 & 25 & 29 & 33 & 0 & 23 & 43 & 35 & 32 & 31 & 34 \\
24 & 23 & 23 & 25 & 24 & 23 & 24 & 23 & 26 & 26 & 25 & 28 & 30 & 23 & 0 & 39 & 35 & 27 & 26 & 29 \\
42 & 42 & 42 & 45 & 44 & 44 & 40 & 39 & 44 & 44 & 45 & 48 & 50 & 43 & 39 & 0 & 48 & 46 & 48 & 47 \\
37 & 36 & 35 & 39 & 38 & 37 & 36 & 37 & 39 & 39 & 40 & 41 & 42 & 35 & 35 & 48 & 0 & 21 & 41 & 44 \\
34 & 33 & 34 & 38 & 37 & 36 & 35 & 34 & 36 & 36 & 37 & 40 & 40 & 32 & 27 & 46 & 21 & 0 & 35 & 38 \\
38 & 39 & 39 & 39 & 38 & 39 & 38 & 37 & 40 & 40 & 39 & 42 & 46 & 31 & 26 & 48 & 41 & 35 & 0 & 5 \\
39 & 40 & 40 & 40 & 39 & 40 & 39 & 38 & 41 & 41 & 42 & 41 & 47 & 34 & 29 & 47 & 44 & 38 & 5 & 0
\end{array}\right] \tag{8}$$

$Q(i,j,k,l)$	$d_{ij}+d_{kl}$	$d_{ik}+d_{jl}$	$d_{il}+d_{jk}$	r_{min}	r_{max}
$Q(1,3,6,9)$	6	12	8	0.3333	0.5000
$Q(1,3,6,10)$	6	12	8	0.3333	0.5000
$Q(1,3,6,11)$	7	13	9	0.3077	0.4615
$Q(1,2,3,5)$	6	8	6	0.2500	0.2500
$Q(1,2,3,4)$	7	9	7	0.2222	0.2222
$Q(1,2,3,8)$	7	9	7	0.2222	0.2222
$Q(3,5,6,9)$	9	7	7	0.2222	0.2222
$Q(3,5,6,10)$	9	7	7	0.2222	0.2222
$Q(1,2,3,7)$	8	10	8	0.2000	0.2000
$Q(2,5,6,9)$	10	8	6	0.2000	0.4000
$Q(2,5,6,10)$	10	8	6	0.2000	0.4000
$Q(3,4,6,9)$	10	8	8	0.2000	0.2000
$Q(3,4,6,10)$	10	8	8	0.2000	0.2000
$Q(3,5,6,11)$	10	8	8	0.2000	0.2000
$Q(1,2,6,9)$	5	11	9	0.1818	0.5455
$Q(1,2,6,10)$	5	11	9	0.1818	0.5455
$Q(1,3,5,6)$	3	9	11	0.1818	0.7273
$Q(1,3,5,9)$	5	11	9	0.1818	0.5455
$Q(1,3,5,10)$	5	11	9	0.1818	0.5455
$Q(2,3,6,9)$	5	11	9	0.1818	0.5455
$Q(2,3,6,10)$	5	11	9	0.1818	0.5455
$Q(2,4,6,9)$	11	9	7	0.1818	0.3636
$Q(2,4,6,10)$	11	9	7	0.1818	0.3636
$Q(2,5,6,11)$	11	9	7	0.1818	0.3636
$Q(3,4,6,11)$	11	9	9	0.1818	0.1818

$Q(i,j,k,l)$	$d_{ij}+d_{kl}$	$d_{ik}+d_{jl}$	$d_{il}+d_{jk}$	r_{min}	r_{max}
$Q(3,6,9,12)$	19	23	19	0.1739	0.1739
$Q(1,2,5,6)$	2	10	12	0.1667	0.8333
$Q(1,2,6,11)$	6	12	10	0.1667	0.5000
$Q(1,3,4,6)$	4	10	12	0.1667	0.6667
$Q(1,3,4,9)$	6	12	10	0.1667	0.5000
$Q(1,3,4,10)$	6	12	10	0.1667	0.5000
$Q(1,3,5,11)$	6	12	10	0.1667	0.5000
$Q(1,3,6,8)$	10	12	10	0.1667	0.1667
$Q(1,3,8,9)$	10	12	10	0.1667	0.1667
$Q(1,3,8,10)$	10	12	10	0.1667	0.1667
$Q(2,3,5,9)$	4	12	10	0.1667	0.6667
$Q(2,3,5,10)$	4	12	10	0.1667	0.6667
$Q(2,3,6,11)$	6	12	10	0.1667	0.5000
$Q(2,4,6,11)$	12	10	8	0.1667	0.3333
$Q(1,2,4,6)$	3	11	13	0.1538	0.7692
$Q(1,2,6,8)$	9	13	11	0.1538	0.3077
$Q(1,3,4,11)$	7	13	11	0.1538	0.4615
$Q(1,3,6,7)$	11	13	11	0.1538	0.1538
$Q(1,3,7,9)$	11	13	11	0.1538	0.1538
$Q(1,3,7,10)$	11	13	11	0.1538	0.1538
$Q(1,3,8,11)$	11	13	11	0.1538	0.1538
$Q(1,5,8,9)$	13	9	11	0.1538	0.3077
$Q(1,5,8,10)$	13	9	11	0.1538	0.3077
$Q(2,3,4,9)$	5	13	11	0.1538	0.6154
$Q(2,3,4,10)$	5	13	11	0.1538	0.6154
$Q(2,3,5,11)$	5	13	11	0.1538	0.6154
$Q(2,3,8,9)$	9	13	11	0.1538	0.3077
$Q(2,3,8,10)$	9	13	11	0.1538	0.3077
$Q(3,5,6,8)$	13	11	7	0.1538	0.4615

$Q(i,j,k,l)$	$d_{ij}+d_{kl}$	$d_{ik}+d_{jl}$	$d_{il}+d_{jk}$	r_{min}	r_{max}
$Q(1,2,6,7)$	10	14	12	0.1429	0.2857
$Q(1,3,7,11)$	12	14	12	0.1429	0.1429
$Q(1,4,8,9)$	14	10	12	0.1429	0.2857
$Q(1,4,8,10)$	14	10	12	0.1429	0.2857
$Q(1,5,7,9)$	14	10	12	0.1429	0.2857
$Q(1,5,7,10)$	14	10	12	0.1429	0.2857
$Q(1,5,8,11)$	14	10	12	0.1429	0.2857
$Q(1,6,8,9)$	14	10	12	0.1429	0.2857
$Q(1,6,8,10)$	14	10	12	0.1429	0.2857
$Q(2,3,4,11)$	6	14	12	0.1429	0.5714
$Q(2,3,7,9)$	10	14	12	0.1429	0.2857
$Q(2,3,7,10)$	10	14	12	0.1429	0.2857
$Q(2,3,8,11)$	10	14	12	0.1429	0.2857
$Q(2,5,6,8)$	14	12	8	0.1429	0.4286
$Q(2,5,8,9)$	14	10	12	0.1429	0.2857
$Q(2,5,8,10)$	14	10	12	0.1429	0.2857
$Q(3,4,6,8)$	14	12	8	0.1429	0.4286
$Q(3,5,6,7)$	14	12	8	0.1429	0.4286
$Q(3,6,8,9)$	12	10	14	0.1429	0.2857
$Q(3,6,8,10)$	12	10	14	0.1429	0.2857
$Q(2,15,17,19)$	64	62	74	0.1351	0.1622
$Q(3,15,17,19)$	64	61	74	0.1351	0.1757
$Q(6,15,17,19)$	64	63	74	0.1351	0.1486
$Q(1,4,7,9)$	15	11	13	0.1333	0.2667
$Q(1,4,7,10)$	15	11	13	0.1333	0.2667
$Q(1,4,8,11)$	15	11	13	0.1333	0.2667
$Q(1,5,7,11)$	15	11	13	0.1333	0.2667
$Q(1,6,7,9)$	15	11	13	0.1333	0.2667
$Q(1,6,7,10)$	15	11	13	0.1333	0.2667
$Q(1,6,8,11)$	15	11	13	0.1333	0.2667
$Q(1,6,12,13)$	35	39	45	0.1333	0.2222
$Q(2,3,7,11)$	11	15	13	0.1333	0.2667
$Q(2,4,6,8)$	15	13	9	0.1333	0.4000
$Q(2,4,8,9)$	15	11	13	0.1333	0.2667
$Q(2,4,8,10)$	15	11	13	0.1333	0.2667
$Q(2,5,6,7)$	15	13	9	0.1333	0.4000
$Q(2,5,7,9)$	15	11	13	0.1333	0.2667
$Q(2,5,7,10)$	15	11	13	0.1333	0.2667
$Q(2,5,8,11)$	15	11	13	0.1333	0.2667
$Q(3,4,6,7)$	15	13	9	0.1333	0.4000
$Q(3,6,7,9)$	13	11	15	0.1333	0.2667
$Q(3,6,7,10)$	13	11	15	0.1333	0.2667
$Q(3,6,8,11)$	13	11	15	0.1333	0.2667

$Q(i,j,k,l)$	$d_{ij}+d_{kl}$	$d_{ik}+d_{jl}$	$d_{il}+d_{jk}$	r_{min}	r_{max}
$Q(2,14,15,19)$	53	54	62	0.1290	0.1452
$Q(3,14,15,19)$	52	54	62	0.1290	0.1613
$Q(6,14,15,19)$	52	54	62	0.1290	0.1613
$Q(1,4,7,11)$	16	12	14	0.1250	0.2500
$Q(1,6,7,11)$	16	12	14	0.1250	0.2500
$Q(1,6,9,15)$	32	27	28	0.1250	0.1563
$Q(1,6,10,15)$	32	27	28	0.1250	0.1563
$Q(2,4,6,7)$	16	14	10	0.1250	0.3750
$Q(2,4,7,9)$	16	12	14	0.1250	0.2500
$Q(2,4,7,10)$	16	12	14	0.1250	0.2500
$Q(2,4,8,11)$	16	12	14	0.1250	0.2500
$Q(2,5,7,11)$	16	12	14	0.1250	0.2500
$Q(3,6,7,11)$	14	12	16	0.1250	0.2500
$Q(13,15,17,19)$	71	68	81	0.1235	0.1605
$Q(1,3,9,13)$	28	30	34	0.1176	0.1765
$Q(1,3,10,13)$	28	30	34	0.1176	0.1765
$Q(1,8,9,14)$	30	28	34	0.1176	0.1765
$Q(2,4,7,11)$	17	13	15	0.1176	0.2353
$Q(3,8,9,14)$	30	30	34	0.1176	0.1176
$Q(6,9,12,15)$	32	43	38	0.1163	0.2558
$Q(8,9,12,15)$	36	43	38	0.1163	0.1628
$Q(13,14,15,19)$	59	61	69	0.1159	0.1449
$Q(1,3,11,13)$	29	31	35	0.1143	0.1714
$Q(1,7,9,14)$	31	29	35	0.1143	0.1714
$Q(1,8,11,14)$	31	29	35	0.1143	0.1714
$Q(2,8,9,14)$	31	29	35	0.1143	0.1714
$Q(3,7,9,14)$	31	31	35	0.1143	0.1143
$Q(3,8,11,14)$	31	31	35	0.1143	0.1143
$Q(7,9,12,15)$	37	44	39	0.1136	0.1591
$Q(2,9,14,15)$	28	53	47	0.1132	0.4717
$Q(6,8,13,16)$	58	63	71	0.1127	0.1831
$Q(1,2,3,12)$	16	18	16	0.1111	0.1111
$Q(1,7,11,14)$	32	30	36	0.1111	0.1667
$Q(1,8,9,13)$	32	31	36	0.1111	0.1389
$Q(1,8,10,13)$	32	31	36	0.1111	0.1389
$Q(2,7,9,14)$	32	30	36	0.1111	0.1667
$Q(2,8,11,14)$	32	30	36	0.1111	0.1667
$Q(3,7,11,14)$	32	32	36	0.1111	0.1111
$Q(6,7,13,16)$	59	64	72	0.1111	0.1806
$Q(8,15,17,19)$	64	63	72	0.1111	0.1250

6 Primates

This family has 12 members. The normalized Hamming distance matrix is given below. It is shown that the metric is tree-reducible at the level $\alpha = 0.069$.

$$\begin{matrix}
0 & 0.3080 & 0.3091 & 0.2934 & 0.2926 & 0.2956 & 0.2822 & 0.2889 & 0.2979 & 0.2867 & 0.2854 & 0.2520 \\
0.3080 & 0 & 0.0893 & 0.1038 & 0.1609 & 0.1819 & 0.2321 & 0.2333 & 0.2489 & 0.2556 & 0.2732 & 0.3221 \\
0.3091 & 0.0893 & 0 & 0.1060 & 0.1709 & 0.1886 & 0.2433 & 0.2511 & 0.2679 & 0.2489 & 0.2844 & 0.3210 \\
0.2934 & 0.1038 & 0.1060 & 0 & 0.1665 & 0.1886 & 0.2366 & 0.2355 & 0.2623 & 0.2444 & 0.2710 & 0.3143 \\
0.2926 & 0.1609 & 0.1709 & 0.1665 & 0 & 0.1877 & 0.2447 & 0.2469 & 0.2615 & 0.2413 & 0.2836 & 0.3035 \\
0.2956 & 0.1819 & 0.1886 & 0.1886 & 0.1877 & 0 & 0.2467 & 0.2388 & 0.2567 & 0.2422 & 0.2688 & 0.3087 \\
0.2822 & 0.2321 & 0.2433 & 0.2366 & 0.2447 & 0.2467 & 0 & 0.0357 & 0.0837 & 0.1239 & 0.2889 & 0.3143 \\
0.2889 & 0.2333 & 0.2511 & 0.2355 & 0.2469 & 0.2388 & 0.0357 & 0 & 0.0926 & 0.1194 & 0.2934 & 0.3166 \\
0.2979 & 0.2489 & 0.2679 & 0.2623 & 0.2615 & 0.2567 & 0.0837 & 0.0926 & 0 & 0.1228 & 0.2867 & 0.3110 \\
0.2867 & 0.2556 & 0.2489 & 0.2444 & 0.2413 & 0.2422 & 0.1239 & 0.1194 & 0.1228 & 0 & 0.2867 & 0.3188 \\
0.2854 & 0.2732 & 0.2844 & 0.2710 & 0.2836 & 0.2688 & 0.2889 & 0.2934 & 0.2867 & 0.2867 & 0 & 0.3199 \\
0.2520 & 0.3221 & 0.3210 & 0.3143 & 0.3035 & 0.3087 & 0.3143 & 0.3166 & 0.3110 & 0.3188 & 0.3199 & 0
\end{matrix} \tag{9}$$

References

1. T. H. JUKES, C. R. CANTOR, *Mammalian Protein Metabolism*, H. MUNRO, ed. (Academic Press, 1969), pp. 21 – 132.

2. M. Kimura, *Journal of molecular evolution* **16**, 111 (1980).
3. J. Felsenstein, *Journal of molecular evolution* **17**, 368 (1981).
4. M. Kimura, *Proceedings of the National Academy of Sciences - PNAS* **78**, 454 (1981).
5. M. Hasegawa, H. Kishino, T.-a. Yano, *Journal of Molecular Evolution* **22**, 160 (1985).
6. D. H. Huson, D. Bryant .

Phylogenetic Data for Biological Organisms

The most crucial step in phylogenetic constructions is the *multiple sequence alignment* of DNA or protein sequences of N distinct species (taxons) that will provide the degree of similarity or differences which is often expressed with N(N-1)/2 pairwise distances. The biological organisms which display conflicting bifurcating tree-based models due to reticulate events [1. 2] were specifically selected for this study.

Reticulate evolution is a direct result of horizontal gene transfer (HGT). hybridization or recombination. HGT is the direct transfer of genes between organisms and frequently encountered in bacteria and prokaryotes [3. 4]. Because of HGT. the phylogenetic analysis of different genes that belong to the same set of taxons can give rise to different phylogenetic trees. In hybridization which is ubiquitious in plants. two species hybridize and merge their chromosomes to produce a hybrid species which is rarely reproductive and rarely evolves as a distinct species [5. 6]. In recombination. similar DNA sequences are shuffled to produce new combinations of the genetic material. This may result in different phylogenetic constructions for different segments of an individual gene [7]. In all these cases mentioned so far. the problem is addressed using a phylogenetic network model that incorporates different gene variations.

A variety of organisms was studied in this project; bee, fish, algea, HIV virus. and muskrats. The distance matrices were either generated using the aligned DNA/RNA/protein sequences or directly adopted from previously published works.

4.1. Bee family

Already aligned short DNA sequences (677 nucleotides) of six species of bees were considered to generate the distance matrix given below. This data set originates from the example file bees.nex of the SpitsTree program by Huson and Bryant (Huson and Bryant, 2006)

	A.andrenof	A.mellifer	A.dorsata	A.cerana	A.florea	A.koschev
A.andreno	0.0					
A.mellifer	0.090103395	0.0				
A.dorsata	0.10339734	0.09305761	0.0			
A.cerana	0.09601182	0.090103395	0.116691284	0.0		
A.florea	0.004431314	0.09305761	0.106351554	0.098966025	0.0	
A.koschev	0.07533235	0.10044313	0.10339734	0.098966025	0.07828656	0.0

4.2 Barbus fish family:

The following distance values were reported by Khaefi and his coworkers (Khaefi *et al.* 2018) for 33 gene sequences of cytochrome oxidase subunit I of two genera. Luciobarbus and Barbus. Later the sequences were aligned and intergentic distances were calculated by Mega software and Kimura 2-parameter distance method.

	B. lacerta	*L. mursa*	*L. capito*	*L. pectoralis*	*L. barbulus*	*L. xanthopterus*	*L. kersin* *L. barbulus*	*L. kersin*	*L. esocinus*	*L. barbulus* *L. xanthopterus*
B. lacerta	0.0									
L. mursa	11.03	0.0								
L. capito	9.72	9.75	0.0							
L. pectoralis	8.49	9.59	2.48	0.0						
L. barbulus	9.44	9.97	3.07	1.79	0.0					
L. xanthopterus	8.85	9.62	2.51	1.44	1.01	0.0				
L. kersin *L. barbulus*	9.28	9.73	2.60	1.72	0.71	1.01	0.0			
L. kersin	10.37	10.15	4.35	3.63	3.28	3.08	3.37	0.0		
L. esocinus	8.85	9.51	2.60	1.35	0.91	0.09	1.10	2.98	0.0	
L. barbulus *L. xanthopterus*	8.8	9.74	2.22	1.35	1.83	0.27	0.73	2.98	0.36	0.0

4.3 Algea family: Selected RNA sequences were used to calculate eight different distance values using the aligned RNA sequences of 920 nucleotides following the distance tables. This data set originates from the example file bees.nex of the SpitsTree program by Huson and Bryant (Huson and Bryant, 2006)

Hamming Distance:

	Tobacco	Rice	Marchantia	Chlamydo-monas	Chlorella	Euglena	Anacystis_ nidulans	Olithodiscus
Tobacco	0.00							
Rice	0.026115343	0.00						
Marchantia	0.02937976	0.041304346	0.00					
Chlamydomonas	0.112445414	0.12104689	0.099236645	0.00				
Chlorella	0.07836644	0.088202864	0.06394708	0.09955752	0.00			
Euglena	0.13626374	0.14379802	0.12294182	0.14207049	0.1160221	0.00		
Anacystis_nidulans	0.12295974	0.13152175	0.12065218	0.14285715	0.11797133	0.15916575	0.00	
Olithodiscus	0.14052288	0.14472252	0.132753	0.15611354	0.11589404	0.13516484	0.13601741	0.00

BaseFreq Distance:

	Tobacco	Rice	Marchantia	Chlamydo-monas	Chlorella	Euglena	Anacystis_ nidulans	Olithodiscus
Tobacco	0.00							
Rice	0.014278568	0.00						
Marchantia	0.029512519	0.01523395	0.00					
Chlamydomonas	0.071786016	0.057507448	0.0422735	0.00				
Chlorella	0.06681672	0.052538153	0.0373042	0.013322151	0.00			
Euglena	0.1659604	0.15168183	0.13644788	0.094174385	0.099143684	0.00		
Anacystis_nidulans	0.029512519	0.01523395	0.0065288357	0.0422735	0.0373042	0.13644788	0.00	
Olithodiscus	0.1503268	0.13604823	0.12081428	0.07854078	0.08351008	0.018511336	0.12081428	0.00

Ledget Distance

	Tobacco	Rice	Marchantia	Chlamydomonas	Chlorella	Euglena	Anacystis_nidulans	Olithodiscus
Tobacco	0.00							
Rice	0.027883343	0.00						
Marchantia	0.032022476	0.04496163	0.00					
Chlamydomonas	0.12754661	0.13690494	0.10953057	0.00				
Chlorella	0.085726134	0.09710287	0.06816578	0.111699775	0.00			
Euglena	0.14987901	0.15935506	0.13410266	0.15984116	0.12607436	0.00		
Anacystis_nidulans	0.14177562	0.15307303	0.13834897	0.16407073	0.1351617	0.17630866	0.00	
Olithodiscus	0.15844822	0.16529733	0.14937684	0.17930831	0.12735935	0.15113328	0.15074529	0.00

F81 Distance:

	Tobacco	Rice	Marchantia	Chlamydomonas	Chlorella	Euglena	Anacystis_nidulans	Olithodiscus
Tobacco	0.00							
Rice	0.026580855	0.00						
Marchantia	0.02997069	0.042485278	0.00					
Chlamydomonas	0.12182498	0.13201238	0.10644535	0.00				
Chlorella	0.08277024	0.09383561	0.06683883	0.106815256	0.00			
Euglena	0.15038092	0.15964499	0.1342754	0.1575107	0.12604433	0.00		
Anacystis_nidulans	0.13429683	0.14460838	0.13154183	0.15848182	0.12835385	0.17890327	0.00	
Olithodiscus	0.15560384	0.16078967	0.14610295	0.17503878	0.12589285	0.14903925	0.15007998	0.00

F84 Distance:

	Tobacco	Rice	Marchantia	Chlamydomonas	Chlorella	Euglena	Anacystis_nidulans	Olithodiscus
Tobacco	0.00							
Rice	0.026627325	0.00						
Marchantia	0.030086145	0.042571124	0.00					
Chlamydomonas	0.122818515	0.13278906	0.106896415	0.00				
Chlorella	0.083544426	0.094461486	0.06713761	0.107504696	0.00			
Euglena	0.15384829	0.16262908	0.13676623	0.15965487	0.12844175	0.00		
Anacystis_nidulans	0.13536446	0.14544883	0.13243945	0.15970297	0.12943254	0.18158154	0.00	
Olithodiscus	0.1600877	0.16469441	0.14917389	0.17824154	0.12807469	0.15133785	0.15168762	0.00

GTR Distance:

	Tobacco	Rice	Marchantia	Chlamydomonas	Chlorella	Euglena	Anacystis_nidulans	Olithodiscus
Tobacco	0.00							
Rice	0.04921772	0.00						
Marchantia	0.052097257	0.06498793	0.00					
Chlamydomonas	0.14725205	0.1562832	0.12645796	0.00				
Chlorella	0.106617145	0.11774201	0.08677514	0.12747823	0.00			
Euglena	0.1757383	0.18427755	0.15677047	0.17820384	0.14431894	0.00		
Anacystis_nidulans	0.1634779	0.17473312	0.15815634	0.18185239	0.1541929	0.19986351	0.00	
Olithodiscus	0.18133584	0.18731333	0.16877677	0.19501169	0.14353554	0.16774793	0.16979857	0.00

HKY85 Distance:

	Tobacco	Rice	Marchantia	Chlamydo-monas	Chlorella	Euglena	Anacystis_ nidulans	Olithodiscus
Tobacco	0.00							
Rice	0.026657386	0.00						
Marchantia	0.030001977	0.04273202	0.00					
Chlamydomonas	0.12347773	0.13456242	0.10818272	0.00				
Chlorella	0.08313749	0.09475303	0.06732381	0.10818313	0.00			
Euglena	0.15078707	0.16101041	0.13483661	0.15969175	0.12634157	0.00		
Anacystis_nidulans	0.13651815	0.14784408	0.13386476	0.16202244	0.13022801	0.18186255	0.00	
Olithodiscus	0.15538442	0.16132386	0.14666674	0.17711866	0.12638576	0.15054101	0.15248564	0.00

K3ST Distance:

	Tobacco	Rice	Marchantia	Chlamydo-monas	Chlorella	Euglena	Anacystis_ nidulans	Olithodiscus
Tobacco	0.00							
Rice	0.026627647	0.00						
Marchantia	0.030089103	0.042572476	0.00					
Chlamydomonas	0.122862965	0.13284372	0.106946655	0.00				
Chlorella	0.08356431	0.0945029	0.06716182	0.10756703	0.00			
Euglena	0.1539113	0.16279978	0.13679734	0.15970647	0.12844944	0.00		
Anacystis_nidulans	0.13536446	0.14545302	0.1324399	0.15970346	0.1294492	0.1815821	0.00	
Olithodiscus	0.16011366	0.16470289	0.1491919	0.17826827	0.12807469	0.15133785	0.15173626	0.00

Aligned RNA sequences of eight algea species, each composed of 920 nucleotides:

```
Tobacco            AAGAACCUGCCCUUGGGAGCUGGAAACGGCUGC UAAUACCCCGUAGGGGCUCGCGUCUGAUUAGCU AGUUGGUGGCAAGCUUACCAAGGCGAUGACUGG
Rice               AAGAACCUGCCCUUGGGAACUGGAAACGGUUGC UAAUACCCCGUAGGGGCUCGCGUCUGAUUAGCU AGUUGGUGGCAAGCUUACCAAGGCGAUGACUGG
Marchantia         AAGAACCUGCCCUUGGGAGCUGGAAACGGUUGC UAAUACCCCAUAGGGGCUUGCGUCUGAUUAGCU AGUUGGUGGUAAGCUUACCAAGGCGACGACUGG
Chlamydomonas      AAGAACCUACCUAUCGGAUUGGGAAACUGUUGC UAAUACCCCAUAGGGGCUUGCGUCUGAUUAGCU AGUUGGUGGUAAGCCUCCCAAGGCCACGACUGG
Chlorella          AAGAACCUACCUUUAGGAACUGGAAACGGUUGC UAAUACCCCAUAUGGGCUUGCGUCUGAUUAGCU UGUUGGUGGUAAGCUUACCAAGGCAACGACUGG
Euglena            AAGAAUCUGCGCUUGGGAGAUGGAAACGUUUGC UAAUGCCUCAUAUGAGCUUGCAUCUGAUUAGCU UGUUGGUGGUAAGCUUACCAAGGCGACGACUGA
Anacystis_nidulans GAGAAUCUGCCUACAGGAGUUGGAAACGACUGC UAAUACCCGAUAUGAGCUCGCGUCUGAUUAGCU AGUUGGUGGUAAGCCUACCAAGGCGACGACUGG
Olithodiscus       GAGAAUCUGCCUUUAGGAUUUGGAAACGAAUGC UAAUACCUUAUAGGAGCUCGCGGCUGAUUAGCU AGUUGGUGGUAAGCCUACCAAGGCGAUCACUGA

Tobacco            UCCGAGAGGAUGAUCAGCCACACUGGGACUGAG ACACGGCCCAGACUCCUACGGGAGGCAGCAGUG GGGAAUUUUCCGCAAUGGGCGAAAGACGGAGCA
Rice               UCCGAGAGGAUGAUCAGCCACACUGGGACUGAG ACACGGCCCAGACUCCUACGGGAGGCAGCAGUG GGGAAUUUUCCGCAAUGGGCGAAAGACGGAGCA
Marchantia         UCUGAGAGGAUGAUCAGCCACACUGGGACUGAG ACACGGCCCAGACUCUUACGGGAGGCAGCAGUG GGGAAUUUUCCGCAAUGGGCGAAACACGGAGCA
Chlamydomonas      UCUGAGAGGAUGAUCAGCCACACUGGGACUGAG ACACGGCCCAGACUCCUACGGGAGGCAGCAGUG AGGAAUUUUUCGCAAUGGGCGCAAGACGGAGCA
Chlorella          UCUGAGAGGAUGAUCAGCCACACUGGGACUGAG ACACGGCCCAGACUUCUACGGAAGGCAGCAGUG AGGAAUUUUCCGCAAUGGGCGAAAGACGGAGCA
Euglena            UUUGAGAGGAUGAUCAGCCACACUGGGAUUGAG A-ACGGAACAGACUUCUACGGAAGGCAGCAGUG AGGAAUUUUCCGCAAUGGGCGCAAGACGGAGCA
Anacystis_nidulans UCUGAGAGGAUGAUCAGCCACACUGGGACUGAG ACACGGCCCAGACUCCUACGGGAGGCAGCAGUG GGGAAUUUUCCGCAAUGGGCGCAAGACGGAGCA
Olithodiscus       UUUGAGAGGAUGAUCAGUCACACUGGAACUGAG ACACGGUCCAGACUCCUACGGGAGGCAGCAGUG GGGAAUUUUCUGCAAUGGGCGAAAGACAGAGCA

Tobacco            AUGCCGCGUGGAGGUAGAAGGGGGUCGUGACUU UUCCCGGAGAAGAUGACGGUAUCUGGGGAAUAA GCAUCGGCUAACUCUGAAUCCCAGGGCUCAACC
Rice               AUGCCGCGUGGAGGUGGAAGGGGUCGUCAACUU UUCUCGGAGAAGAUGACGGUAUCUGAGGAAUAA GCAUCGGCUAACUCUGAAUCCCAGGGCUCAACC
Marchantia         AUGCCGCGUGGAGGUAGAAGGGGGUCGUAACUU UUCUCAGAGAAGAUGACGGUAUCUGAGGAAUAA GCAUCGGCUAACUCUGAAUCCCAGGGCUCAACC
Chlamydomonas      AUGCCGCGUGCAGGAAGAAGGGGGUCGUAACUU UUCUCAGAGAAGAUGACGGUAUCUGAGGAAUAA GCACCGGCUAACUCUGAAUACCAGGGCUCAACC
Chlorella          AUGCCGCGUGGAGGAUGACAGGGGUCGUAACUU UUCUCAGAGAAGAUGACGGUAUCUGAGGAAUAA GCAUCGGCUAACUCCGAAGAUCAGGGCUUAACC
Euglena            AUACCGCGUGAAGGAAGACGGGGGUUGAAACUU UUCUCAAAGAAGAUGACGGUAUUUGAGGAAUAA GCAUCGGCUAAUUCCGAAAGUCAAAGCUUAACU
Anacystis_nidulans ACGCCGCGUGGGGGAGGAAGGGGACUGUAACUU UUCUCAGGGAAGAUGACGGUACCUGAGGAAUAA GCCUCGGCUAAUUCCGAAGCGUGGGGCUCAACC
Olithodiscus       AUACCGCGUGAGGGAUGACAGGGAUUGUAACUU UUUUCAGGGAGGAUGACGUUACCUGAAGAAUAA GCAUCGGCUAACUCCGAAUCUUAAGGCUCAACC

Tobacco            CUGGACAGGCGGUGGACUGGAGUACGGUAGGGG CAGAGGGAAUUGUGGAGCGGUGAAAUGCGUAGA GAUCGGAAAGAACACCAACGGCGAAAGCACUCU
Rice               CUGGACAGGCGGUGGACUGGAGUACGGUAGGGG CAGAGGGAAUUGUGGAGCGGUGAAAUGCAUUGA GAUCGGAAAGAACACCAACGGCGAAAGCACUCU
Marchantia         CUGGACAGGCGGUGGACUGGAGUACGGUAGGGG CAGAGGGAAUUGUGGAGCGGUGAAAUGCGUAGA GAUCGGAAAGAACACCAAUGGCGAAAGCACUCU
Chlamydomonas      UUGGACCGGCAUUGGACUUGAGUACGGUAGGGG CAGAGGGAAUUGUGGAGCGGUGAAAUGCGUAGA GAUAUGGAGGAACACCAGUGGCGAAGGCGCUCU
Chlorella          CUGAGUCGGCAGUAGACUUGAGUACGGUAGGGG CAGAGGGAAUUGUGUAGCGGUGAAAUGCGUAGA GAUCGGGAAGAACACCAAUGGCGAAAG-ACUCU
Euglena            UUGGAAGGGCAUUAAACUUGAGUAUGGUAGGGG UGAAGGGAAUUGUGUAGCGGUGAAAUGCGUAGA GAUUGGAAAGAACACCAAUGGCGAAGGCACUUU
Anacystis_nidulans UCAUACAGGCAAUGGACUAGAGUAUGGUAGGGG UAGCGGGAAUUGUGUAGCGGUGAAAUGCGUAGA UAUCUGGAAGAACACCAGCGGCGAAAGCGCGCU
Olithodiscus       UUAAACCAGCAAUGGACUUGAGUGUGGUAGAGG UAGAGGGAAUUGUGGAGCGGUGAAAUGCGUAGA UAUUGGAAAGACCACCAAUGGCGAAGGCACUCU

Tobacco            GCUGGGCCGACACUGACACUGAGAGACGAAAGC UAGGGGAGCGAAUGGGAUUAGAUACCCCAGUAG UCCUAGCCGUAAACGAUGGAUACUAGGCGCUGC
Rice               GCUGGGCCGACACUGACACUGAGAGACGAAAGC UAGGGGAGCAAAUGGGAUUAGAGACCCCAGUAG UCCUAGCCGUAAACGAUGGAUACUAGGUGCUGC
Marchantia         UCUGGGCCGACACUGACACUGAGAGACGAAAGC UAGGGGAGCAAAUGGGAUUAGAUACCCCAGUAG UCCUAGCCGUAAACGAUGGAUACUAAGCGCUGC
Chlamydomonas      GCUGGGCCGAAACUGACACUGAGAGACGAAAGC UGGGGGAGCGAAUAGGAUUAGAUACCCUAGUAG UCCCAGCCGUAAACUAUGGAGACUAAGUGCUGC
Chlorella          GCUGGGCCGAAACUGACACUCAGAGACGAAAGC UAGGGGAGCGAAUGGGAUUAGAUACCCCAGUAG UCCUAGCCGUAAACGAUGGAUACUAAGUGCUGC
Euglena            UCUAGGCCAAUACUGACGCUGAGAAACGAAAGC UGAGGGAGCAAACAGGAUUAGAUACCCU-GUAG UCUUGGCCGUAAACUAUGGAUACUAAGUGGUGC
Anacystis_nidulans ACUGGGCCAUAACUGACGCUCAUGGACGAAAGC UAGGGGAGCGAAAGGGAUUAGAUACCCCUGUAG UCCUAGCCGUAAACGAUGAACACUAGGUGUUGC
Olithodiscus       ACUGGGCCAUUACUGACACUGAGAGACGAAAGC UAGGGGAGCAAAUGGGAUUAGAUACCCCAGUAG UCCUAGCUGUAAACGAUGGAUACUAGAUGUCGC

Tobacco            AGUGCUGUAGCUAACGCGUUAAGUAUCCCGCCU GGGGAGUACGUUCGCAAGAAUGAAACUCAAAGG AAUUGACGGGGGCCCGCACAAGCGGUGGAGCAU
Rice               AGUGCUGUAGCUAACGCGUUAAGUAUCCCGCCU GGGGAGUACGUUCGCAAGAAUGAAACUCAAAGG AAUUGACGGGGGCCCGCACAAGCGGUGGAGCAU
Marchantia         AGUGCUGUAGCUAACGCGUUAAGUAUCCCGCCU GGGGAGUACGUUCGCAAGAAUGAAACUCAAAGG AAUUGACGGGGGCCCGCACAAGCGGUGGAGCAU
Chlamydomonas      AGUGCUGUAGCUAACGCGUUAAGUCUCCCGCCU GGGGAGUAUGCUCGCAAGAGUGAAACUCAAAGG AAUUGACGGGA--CCGCACAAGCGGUGGAUUAU
Chlorella          AGUACUGUAGCUAACGCGUGAAGUAUCCCGCCU GGGGAGUAUGCUCGCAAGAGUGAAACUCAAAGG AAUUGACGGGGGCCCGCACAAGCGGUGGAGCAU
Euglena            ACUGCUGUAGUUAACACGUUAAGUAUCCCGCCU GGGGAGUACGCUUGCACAAGUGAAACUCAAAGG AAUUGACGGGGGCCCGCACAAGCGGUGGAGCAU
Anacystis_nidulans AGUGCCGUAGCCAACGCGUUAAGUGUUCCGCCU GGGGAGUACGCACGCAAGUUGGAAACUCAAAGG AAUUGACGGGGGCCCGCACAAGCGGUGGAGUAU
Olithodiscus       GGUAUUAUAGCUAACGCGUUAAGUAUCCCGCCU GGGAAGUAUGCUCGCAAGAGUGAAACUCAAAGG AAUUGACGGGGGCCCGCACAAGCGGUGGAGCAU

Tobacco            GUGGUUUAAUUCGAUGCAAAGCGAAGAACCUUA CCAGGGCUUGACAAUCCUCUUGAAAACAGGUGG UGCAUGGCUGUCGUCAGCGUGCCGUAAGUGUUG
Rice               GUGGUUUAAUUCGAUGCAAAGCGAAGAACCUUA CCAGGGCUUGACAAUCCUCUUGAAAACAGGUGG UGCAUGGCUGUCGUCAGCGUGCCGUAAGUGUUG
Marchantia         GUGGUUUAAUUCGAUGCAACGCGAAGAACCUUA CCAGGGCUUGACAAUCUUUUUGAAAACAGGUGG UGCAUGGCUGUCGUCAGCGUGCCGUAAGUGUUG
Chlamydomonas      GUGGAUUAAUUCGAUACAACGCGAAGAACCUUA CCAGGGUUUGACAACCUCUCAGAAAACAGGUGG UGCAUGGCUGUCGUCAGCGUGCUGUGAAUGUAU
Chlorella          GUGGUUUAAUUCGAUGCAACGCGAAGAACCUUA CCAGGGCUUGACA-----------ACAGGUGG UGCAUGGCUGUCGUCAGCGUGUCGUGAGUGUUG
Euglena            GUGGUUUAAUUCGAUGCAACACGAAGAACCUUA CCAGGAUUUGACAA-------GAAAACAGGUGG UGCAUGGCUGUCGUCAGCGUGUCGUGAGUGUUG
Anacystis_nidulans GUGGUUUAAUUCGAUGCAACGCGAAGAACCUUA CCAGGGUUUGACAAUCUCUUGGAAAACAGGUGG UGCAUGGCUGUCGUCAGCGUGUCGUGAGUGUUG
Olithodiscus       GUGGUUUAAUUCGAUGCAACGCGAAGAACCUUA CCAGGGUUUGACAAGUUUUUUGAAAACAGGUGG UGCAUGGCUGUCGUCAGCGUGUCGUGAGUGUUG

Tobacco            GGUUAAGUCCCGCAACGAGCGCAACCCUCGUGU UCGGAGGAAGGUGAGGAUGACGUCAAGUCAUCA UGCCCCUUAGGGCGACACACGUGCUACAAUGGC
Rice               GGUUAAGUCUCGCAACGAGCGCAACCCUCGUGU UCGGAGGAAGGAGAGGAUGAGGCCAAGUCAUCA UGCCCCUUAGGGCGACACACGUGCUACAAUGGG
Marchantia         GGUUAAGUCCCGCAACGAGCGCAACCCUCUUGU UCGGAGGAAGGUGAGGAUGACGUCAAGUCAGCA UGCCCCUUAGGGCGACACACGUGCUACAAUGGC
Chlamydomonas      AGUUAAGUCUCAUAACGAGCGCAACCCUCGUCU UCGGAGGAAGGUGAGGAUGACGUCAAGUCAGCA UGCCCCUUAGGGCUUCACACGUAAUACAAUGGU
Chlorella          GGUUAAGUCCCGCAACGAGCGCAACCCUUGUUU UCGGAGGAAGGUGAGGAUGACGUCAAGUCAGCA UGCCCCUUAGGGCGACACACGUGCUACAAUGGC
Euglena            GGUUAAGUCCCGCAACGAGCGCAACCCUUUUUU UCGGAGGAAGGUGAGGACGACGUCAAGUCAUCA UGCCCCUUAGGGCUACACACGUGCUACAAUGGU
Anacystis_nidulans GGUUAAGUCCCGCAACGAGCGCAACCCACGUUU UCGGAGGAAGGUGUGGACGACGUCAAGUCAUCA UGCCCCUUAGGGCUACACACGUACUACAAUGCU
Olithodiscus       GGUUAAGUCCCGCAACGAGCGUAACCCUUGUUU UCGGAGGAAGGUAAGGACGACGUCAAGUCAUCA UGCCCCUUAGGGCUACACACGUGCUACAUUGGG

Tobacco            CGGGACAAAGGGUCGCGAUCCCGCGAG-GUGAG CUAAAACCCGUCCUCAGUUCGGAUUGCAGGCUG CAACUCGCCUGCAUGAAGCCGGAAUCGCUAGUA
Rice               CGGGACAAAGGGUCGCGAUCUCGCGAGGGUGAG CUAAAACCCGUCCUCAGUUCGGAUUGCAGGCUG CAACUCGCCUGCAUGAAGCAGGAAUCGCUAGUA
Marchantia         CGGGACAAAGGGUCGCGACCUCGCGAGAGAAAG CUAAAACCCGGCCUCAGUUCGGAUUGCAGGCUG CAACUCGCCUGCAUGAAGCCGGAAUCGCUAGUA
Chlamydomonas      UGGGACAAUCAGAAGCGA-CUCGUGAGAGCUAG CGGCAACCCAACCUCAGUUCGGAUUGUAGGCUG CAACUCGCCUACAUGAAGCCGGAAUCGCUAGUA
Chlorella          CAGGACAAUGAGAUGCUACCUCGCGAGAGCAAG CUAAAACCUGGUCUCAGUUCGGAUUGCAGGCUG CAACUCGCCUGCAUGAAGUCGGAAUCGCUAGUA
Euglena            UAAGACAAUAAGUUGCAAUUUUGUGAAAAUGAG CUAAAACUUAGCCUAAGUUCGGAUUGUAGGCUG AAACUCGCCUACAUGAAGCCGGAAUCGCUAGUA
Anacystis_nidulans CCGGACAGCGAGACGCGAAGCCGCGAGGUGAAG CAAAAACCGGGGCUCAGUUCAGAUUGCAGGCUG CAACUCGCCUGCAUGAAGGCGGAAUCGCUAGUA
Olithodiscus       UAGAACAAUAAGUUGCUAAGUUGCGAAACCAAG CGAAAAUCUACUCUAAGUUCGGAUUGUAGGCUG CAACUCGCCUACAUAAAGAUGGAAUCGCUAGUA

Tobacco            AUCGCCGGUCAGAUACGGCGGUGAAUUCG
Rice               AUCGCCGGUCAGAUACGGCGGUGAAUCCG
Marchantia         AUCGCCGGUCAGAUACGGCGGUGAAUCCG
Chlamydomonas      AUCGCCAGUCAGAUAUGGCGGUGAAUACG
Chlorella          AUCGCUGGUCAGAUACAGCGGUGAAUACG
Euglena            AUCGCCGGUCAGAUACGGCGGUGAAUACG
Anacystis_nidulans AUCGCAGGUCAGAUACUGCGGUGAAUACG
Olithodiscus       AUCGCUGGUCAGACACAGCGGUGAAU-CG
```

4.4. Muskrat family:

Samples from nine local populations of muskrats of the La Houille River Basin were extracted and skull measurements were collected for 144 individuals (Le Boulange et al., 1996). Finally, the following distance matrix was determined.

Population zones	C	E	J	L	M	N	O	T	Z
C	0.0								
E	2.1380	0.0							
J	2.2713	2.9579	0.0						
L	1.7135	2.3927	1.7772	0.0					
M	1.5460	1.9818	2.4575	1.0125	0.0				
N	2.6979	3.3566	1.9900	1.8520	2.6954	0.0			
O	2.9985	3.6848	3.4484	2.4272	2.6816	2.3108	0.0		
T	2.3859	2.3169	2.4666	1.4545	1.7581	2.2105	2.5041	0.0	
Z	2.3107	2.3648	1.8086	1.6609	2.0516	2.2954	3.4301	2.0413	0.0

4.5. Human/Simian Immunodeficiency Virus (HIV/SIV) Relationship.

4.5.1 Data Extraction:

Three protein coding regions from sixteen different isolated strains of HIV and SIV were extracted from GenBank using their accession numbers. Three protein coding regions belong to gag protein (GAG). pol polyprotein (POL) and envelope polyprotein precursor (ENV). MATLAB® was used as the working environment.

```
>> data = {'HIV-1 (Zaire)'               'K03454'   [1 2 8]  ;
          'HIV1-NDK (Zaire)'             'M27323'   [1 2 8]  ;
          'HIV-2 (Senegal)'              'M15390'   [1 2 8]  ;
          'HIV2-MCN13'                   'AY509259' [1 2 8]  ;
          'HIV-2UC1 (IvoryCoast)'        'L07625'   [1 2 8]  ;
          'SIVMM251 Macaque'             'M19499'   [1 2 8]  ;
          'SIVAGM677A Green monkey'      'M58410'   [1 2 7]  ;
          'SIVlhoest L''Hoest monkeys'   'AF075269' [1 2 7]  ;
          'SIVcpz Chimpanzees Cameroon'  'AF115393' [1 2 8]  ;
          'SIVmnd5440 Mandrillus sphinx' 'AY159322' [1 2 8]  ;
          'SIVAGM3 Green monkeys'        'M30931'   [1 2 7]  ;
          'SIVMM239 Simian macaque'      'M33262'   [1 2 8]  ;
          'CIVcpzUS Chimpanzee'          'AF103818' [1 2 8]  ;
          'SIVmon Cercopithecus Monkeys' 'AY340701' [1 2 8]  ;
          'SIVcpzTAN1 Chimpanzee'        'AF447763' [1 2 8]  ;
          'SIVsmSL92b Sooty Mangabey'    'AF334679' [1 2 8]  ;
          };
>> numViruses = size(data.1)
```

Then. the **getgenbank** function was used to copy the data from GenBank into a structure in MATLAB®. The **SearchURL** field of the structure contains the address of the actual GenBank record. This record can be browsed using the **web** command.

```
>> acc_num = data{1.2};
>> lentivirus = getgenbank(acc_num);
>> web(lentivirus(1).SearchURL)
```

This example was given for the first nucleotide sequence with the accession id ‘K03454’. The information about the remaining 15 sequences was retrieved from NCBI GenBank in a similar fashion using the following loop construction.

```
>> for ind = 2:numViruses
    lentivirus(ind) = getgenbank(data{ind.2});
   end
>> load('lentivirus.mat')
```

`lentivirus.mat` holds all the information about the three protein coding region of 16 species. Now. we can extract the coding sequence information (CDS) of GAG. POL and ENV and save them separately.

```
>> for ind = 1:numViruses
    temp_seq = lentivirus(ind).Sequence;
    temp_seq = regexprep(temp_seq.'[nry]'.'a');
    CDSs = lentivirus(ind).CDS(data{ind.3});
    gag(ind).Sequence = temp_seq(CDSs(1).indices(1):CDSs(1).indices(2));
    pol(ind).Sequence = temp_seq(CDSs(2).indices(1):CDSs(2).indices(2));
    env(ind).Sequence = temp_seq(CDSs(3).indices(1):CDSs(3).indices(2));
   end
```

All three protein-coding sequences for each sixteen viral species was given in Appendix A.

4.5.2 Determination of Distance Matrices

Once we retrieve the sequence information. we are ready to calculate the distance matrices. The distance between 16 species was measured using the “Tajima-Nei” method (Tajima ve Nei. 1994) [11]. which takes into account unequal rates of substitution among different nucleotide pairs. The expected number of nucleotide substitutions per site between two sequences is given by

$$\delta = -b_1 \log_e(1 - \pi/b_1) \qquad (1)$$

Where $b_1 = 1 - \sum q_i^2$ and π is the proportion of sites with different nucleotides.

Here. q_i is the equilibrium frequency of the i^{th} nucleotide ($i = 1.\ 2.\ 3.\ 4$ corresponding to the nucleotides A. T. G. C). When the rate of nucleotide substitution is the same for all nucleotide pairs. $q_i = 1/4$. so that $b_1 = 3/4$ which makes the above expression to

$$\delta = -3/4 \log_e(1 - 4\pi/3) \qquad (2)$$

which corresponds to Jukes Cantor model [12].

However. as the “Tajima-Nei” method is only defined for nucleotides. the nucleotide sequences were used instead of translated amino acid sequences. The following **`seqpdist`** built-in function was used for that purpose:

```
>> gagd =
     seqpdist(gag.'method'.'TajimaNei'.'Alphabet'.'NT'.'indel'.'pair');
```

which yields the following table.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	0															
2	0.045	0														
3	0.343	0.346	0													
4	0.367	0.362	0.071	0												
5	0.326	0.348	0.194	0.199	0											
6	0.352	0.350	0.190	0.204	0.179	0										
7	0.363	0.373	0.341	0.345	0.288	0.318	0									
8	0.357	0.380	0.397	0.384	0.344	0.350	0.387	0								
9	0.302	0.304	0.343	0.363	0.328	0.351	0.365	0.380	0							
10	0.368	0.357	0.320	0.348	0.326	0.351	0.352	0.363	0.371	0						
11	0.356	0.352	0.337	0.341	0.322	0.312	0.282	0.355	0.386	0.323	0					
12	0.363	0.363	0.200	0.208	0.187	0.009	0.325	0.359	0.366	0.353	0.321	0				
13	0.276	0.264	0.354	0.386	0.359	0.359	0.363	0.370	0.202	0.377	0.354	0.357	0			
14	0.391	0.398	0.389	0.385	0.375	0.362	0.403	0.396	0.392	0.363	0.380	0.370	0.399	0		
15	0.289	0.293	0.336	0.360	0.310	0.347	0.359	0.335	0.306	0.332	0.373	0.356	0.304	0.355	0	
16	0.367	0.377	0.219	0.228	0.220	0.211	0.344	0.339	0.366	0.323	0.345	0.216	0.342	0.364	0.358	0

In addition to Tajima-Nei. there exist several other distance methods for calculating pairwise distances between nucleotides only and they are listed below:

1. **Kimura**: transitional and transversional nucleotide substitutions are treated separately [13]
2. **Tamura**: transitional. transversional nucleotide substitutions as well as GC content are considered separately [14].
3. **Hasegawa**: transitional. transversional nucleotide substitutions as well as background nucleotide frequencies are considered separately [15].
4. **Nei-Tamura**: transitional nucleotide substitution between purines. the transitional nucleotide substitution between pyrimidines. the transversional nucleotide substitution. and the background nucleotide frequencies are considered separately [14].

Also. two methods exist for calculating pairwise distances between nucleotides and amino acids and they are:

1. **p-distance**: Also called Hamming distance. This is the fraction of sites at which two sequences are different.

2. **Jukes-Cantor:** for nucleotides: $\delta = -3/4\log_e(1-4\pi/3)$

   for amino acids: $\delta = -19/20\log_e(1-20\pi/19)$ [12]

Translated DNA sequences:

In addition to DNA sequences. amino acid sequences obtained by translating the corresponding DNA sequences were also used to determine distance matrices. To illustrate. for POL polyprotein coding gene. the 'Jukes-Cantor' method was used to measure the distance between species. As 'Jukes-Cantor' method was defined for amino-acid sequence only. the nucleotide was translated into its corresponding amino acid sequence using **`nt2aa`** built-in function:

```
for ind = 1:numViruses
    aapol(ind).Sequence = nt2aa(pol(ind).Sequence);
end
```

Ex. Translated POL coding region 'HIV-1 (Zaire)' 'K03454'

```
'MNLPGKWKPKMIGGIGGFIKVRQYDQIPIEICGQKAIGTVLVGPTPVNIIGRNLLTQIGCTLNFPISPIETVPVKLKPGMDG
PKVKQWPLTEEKIKALTEICTDMEKEGKISRIGPENPYNTPIFAIKKKDSTKWRKLVDFRELNKRTQDFWEVQLGIPHPA
GLKKKKSVTVLDVGDAYFSVPLDEDFRKYTAFTISSINNETPGIRYQYNVLPQGWKGSPAIFQSSMTKILEPFRKQNPEM
VIYQYMDDLYVGSDLEIGQHRTKIEKLREHLLRWGFTRPDKKHQKEPPFLWMGYELHPDKWTVQSIKLPEKESWTVN
DIQNLVERLNWASQIYPGIKVRQLCKLLRGTKALTEVIPLTEEAELELAENREILKEPVHGVYYDPSKDLIAEIQKQGHG
QWTYQIYQEPFKNLKTGKYARMRGAHTNDVKQLAEAVQRISTESIVIWGRTPKFRLPIQKETWETWWAEYWQATWIP
EWEFVNTPPLVKLWYQLEKEPIIGAETFYVDGAANRETKLGKAGYVTDRGRQKVVPLTDTTNQKTELQAINLALQDSG
LEVNIVTDSQYALGIIQAQPDKSESELVNQIIEQLIKKEKVYLAWVPAHKGIGGNEQVDKLVSQGIRKVLFLDGIDKAQE
EHEKYHNNWRAMASDFNLPPVVAKEIVASCDKCQLKGEAMHGQVDCSPGIWQLDCTHLEGKVILVAVHVASGYIEAE
VIPAETGQETAYFLLKLAGRWPVKVVHTDNGSNFTSAAVKAACWWAGIKQEFGIPYNPQSQGVVESMNKELKKIIGQV
RDQAEHLKTAVQMAVFIHNFKRRRGIGGYSAGERIIDIIATDIQTKELQKQIIKIQNFRVYYRDSRDPIWKGPAKLLWKG
EGAVVIQDKSDIKVVPRRKVKIIRDYGKQMAGDDCVASRQDED*'
```

The following **`seqpdist`** built-in function was then used to determine the distance matrix:

```
pold = seqpdist(aapol.'method'.'Jukes-Cantor'.'indel'.'pair');
```

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	1
1	0															
2	0.03	0														
3	0.50	0.49	0													
4	0.51	0.55	0.06	0												
5	0.49	0.53	0.15	0.19	0											
6	0.52	0.55	0.15	0.20	0.22	0										
7	0.46	0.51	0.50	0.58	0.58	0.56	0									
8	0.54	0.57	0.51	0.58	0.57	0.56	0.58	0								
9	0.17	0.20	0.50	0.57	0.56	0.58	0.53	0.58	0							
1	0.37	0.40	0.46	0.51	0.49	0.48	0.46	0.55	0.42	0						
1	0.50	0.53	0.46	0.57	0.53	0.53	0.34	0.56	0.55	0.46	0					
1	0.52	0.54	0.15	0.19	0.22	0.00	0.55	0.56	0.57	0.48	0.54	0				
1	0.19	0.20	0.47	0.54	0.54	0.54	0.50	0.56	0.14	0.42	0.57	0.54	0			
1	0.61	0.64	0.55	0.62	0.61	0.61	0.60	0.67	0.62	0.61	0.62	0.62	0.62	0		
1	0.32	0.35	0.49	0.53	0.51	0.52	0.48	0.56	0.35	0.40	0.51	0.52	0.34	0.61	0	
1	0.50	0.53	0.25	0.30	0.29	0.26	0.57	0.58	0.56	0.48	0.55	0.26	0.52	0.58	0.52	0

4.5.3. GAG protein-coding genes for 16 species

1. 'HIV-1 (Zaire)' 'K03454'

```
'atgagagcgaggggggatagagagaaattgtcaaaactggtggaaatggggcatcatgctccttgggatattgatgacctgtagtgctgcagacaatctgtgggtcacagtttattatggggtgcctgtatggaaggaagcaaccacca
ctctattttgtgcatcagatgctaaatcatatgaaacagaggcacataatatctgggccacacatgcctgtgtacccacggaccccaacccacaagaaatagcactggaaaatgtgacagaaaactttaacatgtggaaaaataacatgg
tggaacagatgcatgaggatataatcagtttatgggatcaaagcctaaaaccatgtgtaaaattaaccccactctgtgtcactttaaactgtagtgatgaattgaggaacaatggcactatggggaacaatgtcactacagaggagaaag
gaatgaaaaactgctctttcaatgtaaccacagtactaaaagataagaagcagcaagtatatgcactttttttatagacttgatatagtaccaatagacaatgatagtagtaccaatagtaccaattataggttaataaattgtaatacctcagcc
attacacaggcttgtccaaaggtatcctttgagccaattcccatacattattgtgccccagctggttttgcgattctaaagtgtagagataagaagttcaatggaacaggcccatgcacaaatgtcagcacagtacaatgtacacatggaatt
aggccagtggtgtcaactcaactgctgttgaatggcagtctagcagaagaagaggtcataattagatccgaaaatctcacaaacaatgctaaaaacataatagcacatcttaatgaatctgtaaaaattacctgtgcaaggccctatcaaa
atacaagacaaagaacacctataggactagggcaatcactctatactacaagatcaagatcaataataggacaagcacattgtaatattagtagagcacaatggagtaaaactttacaacaagtagctagaaaattaggaacccttcttaa
caaaacaataataaagtttaaaccatcctcaggaggggacccagaaattacaacacacagttttaattgtggaggggaattcttctactgtaatacatcaggactgtttaatagtacatggaatattagtgcatggaataatattacagagtca
aataatagcacaaacacaaacatcacactccaatgcagaataaaacaaattataaagatggtggcaggcaggaaagcaatatatgcccctcctatcgaaagaaacattctatgttcatcaaatattacagggctactattgacaagagat
ggtggtataaataatagtactaacgagacctttagacctggaggaggagatatgagggacaattggagaagtgaattatataaatataaggtagtacaaattgaaccactaggagtagcacccaccagggcaaagagaagagtggtg
gaaagagaaaaaagagcaataggattaggagctatgttccttgggttcttgggagcagcaggaagcacgatgggcgcacggtcagtgacgctgacggtacaggccagacaattaatgtctggtatagtgcaacagcaaaacaatttg
ctgagggctatagaggcgcaacagcatctgttgcaactcacggtctggggcattaaacagctccaggcaagaatcctggctgtggaaagatacctaaaggatcaacagctcctaggaatttggggttgctctggaaaacacatttgca
ccactaatgtgccctggaactctagttggagtaatagatctctaaatgagatttggcagaacatgacctggatggagtgggaaagagaaattgacaattacacaggcttaatatatagcttaattgaggaatcgcagacccagcaagaaa
agaatgaaaaagaattgttggaattggacaagtgggcaagtttgtggaattggtttagcataacacaatggctgtggtatataaaaatattcataatgataataggaggcttgataggtttaagaatagtttttgctgtgctttctttagtaaata
gagttaggcagggatactcacctctgtcgtttcagaccctcctcccagccccgaggggacccgacaggcccgaaggaacagaagaagaaggtggagagcgaggcagagacagatccgtgagattgctgaacggattctcggcac
ttatctgggacgacctgcggagcctgtgcctcttcagctaccaccgcttgagagacttaatcttaattgcagtgaggattgtagaacttctgggacgcagggggtgggacatcctcaaatatctgtggaatctcctacagtattggagtca
ggaactgaggaacagtgctagtagcttgtttgatgccatagcaatagcagtagctgaggggacagatagagttatagaaataatacaaagagcttgcagagctgttcttaacatacccagaagaataagacagggcttagaaaggtcttt
actttaa'
```

2. 'HIV1-NDK (Zaire)' 'M27323'

```
'atgagagcgagggagaaagagaggaattgtcaaaacttgtggaaatggggcatcatgctccttgggatgttgatgacctgtagtgctgcagaagatttgtgggtcacagtttattatggggtgcctatatggaaggaagcaactaccac
tctattttgtgcatcagatgctaaagcatataaaaaagaggcacataatatctgggccacacatgcctgtgtacccacggaccccaacccacaagaaatagaattggaaaatgtgacagaaaactttaacatgtggaaaaataacatggt
ggaacaaatgcatgaggatataatcagtttatgggatcaaagcctaaaaccatgtgtaaaattaactccactctgtgtcactttaaactgcactgatgaattgaggaacagcaagggcaatgggaaggtagaagaggaggaaaaaagg
aaaaactgctctttcaatgtaagggataagagggagcaagtatatgcactttttttataaacttgatatagtgccaatagacaataataataggaccaatagtactaattataggttaataaattgtgatacctcaaccattacacaggcttgtcc
aaagatatcctttgaaccaattcccatacatttttgtgccccagctggttttgcaattctaaagtgtagagataagaagttcaatgggacaggcccatgctcaaatgtcagcacagtacaatgtacacatggaattagaccagtggtgtcaac
tcaactgctgttgaatggcagtctagcagaagaagagatcataattagatctgaaaatctcacaaacaatgttaaaaccataatagtacagcttaatgcatctatagtaattaattgtacaaggccctacaaatatacaaggcaaaggacatc
gataggactaaggcaatcactctatacaataacaggaaaaaagaagaaaacaggatacataggacaagcacattgtaaaattagcagagcagaatggaataaagctttacaacaggtagctacaaaactaggaaaccttcttaacaaa
acaacaataacttttaagccatcctcaggaggggacccagaaattacatcacacatgcttaattgtggaggggacttcttctactgtaatacatcaagactgtttaatagtacatggaatcagactaatagtacagggttcaataatggcaca
gtcacactcccatgcagaataaaacaaattgtaaacttgtggcagagagtaggaaaagcaatgtatgcccctcccatcgaaggactaattaaatgttcatcaaatattacagggctactattaacaagagatggtggtgcaaataatagttc
tcatgagaccatcagacctggaggaggagatatgagggacaattggagaagtgaattatataaatataaagtagtaaaaattgaaccaataggagtagcacccaccaaggcaaggagaagagtggtggaaagagaaaaaagagca
ataggactaggagctgtgttccttgggttcttgggagcagcaggaagcacgatgggcgcagcgtcagtgacgctgacggtacaggccagacagttaatgtctggtatagtgcaccagcaaaacaatttgctgagggctatagaggcg
caacaacatctgttgcaactcacggtctggggcattaaacagctccaggcaagagtcctggctgtggaaagatacctaagggatcaacagctcctaggaatttggggttgctctggaaggcacatttgcaccactaatgtaccctggaa
ctctagttggagtaatagatctctagatgagatttggcagaacatgacctggatggagtgggaaagagaaattgacaattacacaggcttaatatacagcttaattgaagaatcgcagatccagcaagaaaagaatgaaaaagaattatt
```

ggaattggacaaatgggcaagtttgtggaattggtttagcataacaaaatggctgtggtatataaaattattcataatgatagtaggaggcctgataggtttaagaatagtttttgctgtgctttctgtagtaaatagagttaggcagggatact
cacctctgtcatttcagaccctcctcccagtcccgaggggacccgacaggcccgaagaaatagaagaagaaggtggagagcgaggcagagacagatccattcgattggtgaacggattatttgcacttttctgggacgacctgagga
acctgtgcctcttcagctaccaccgcttgagagactcaatcttaattgcagcgaggattgtggaacttctgggacgcagggggtgggaagccctcaaatacctgtggaacctcctgcagtattggagtcaggaactgaggaatagtgct
agtagcttgcttgataccatagcaatagcagtagctgagaggacagatagggttatagaagtagtacaaagagcttgcagagctattcttaacgtacccagaagaataagacagggcttggaaaggcttttgctataa'

3. 'HIV-2 (Senegal)' 'M15390'

'atgatgaatcagctgcttattgccattttattagctagtgcttgcttagtatattgcacccaatatgtaactgttttctatggcgtacccacgtggaaaaatgcaaccattcccctcttttgtgcaaccagaaatagggatacttggggaaccata
cagtgcttgcctgacaatgatgattatcaggaaataactttgaatgtaacagaggcttttgatgcatggaataatacagtaacagaacaagcaatagaagatgtctggcatctattcgagacatcaataaaaccatgtgtcaaactaacacc
tttatgtgtagcaatgaaatgcagcagcacagagagcagcacagggaacaacacaacctcaaagagcacaagcacaaccacaaccacacccacagaccaggagcaagagataagtgaggatactccatgcgcacgcgcagaca
actgctcaggattgggagaggaagaaacgatcaattgccagttcaatatgacaggattagaaagagataagaaaaaacagtataatgaaacatggtactcaaaagatgtggtttgtgagacaaataatagcacaaatcagacccagtgt
tacatgaaccattgcaacacatcagtcatcacagaatcatgtgacaagcactattgggatgctataaggtttagatactgtgcaccaccgggttatgccctattaagatgtaatgataccaattattcaggctttgcacccaactgttctaaag
tagtagcttctacatgcaccaggatgatggaaacgcaaacttccacatggtttggctttaatggcactagagcagagaatagaacatatatctattggcatggcagagataatagaactatcatcagcttaaacaaatattataatctcagttt
gcattgtaagaggccagggaataagatagtgaaacaaataatgcttatgtcaggacatgtgtttcactcccactaccagccgatcaataaaagacccagacaagcatggtgctggttcaaaggcaaatggaaagacgccatgcagga
ggtgaaggaaacccttgcaaaacatcccaggtatagaggaaccaatgacacaaggaatattagctttgcagcgccaggaaaaggctcagacccagaagtagcatacatgtggactaactgcagaggagagtttctctactgcaacat
gacttggttcctcaattggatagagaataagacacaccgcaattatgcaccgtgccatataaagcaaataattaacacatggcataaggtagggagaaatgtatatttgcctcccagggaaggggagctgtcctgcaactcaacagtaac
cagcataattgctaacattgactggcaaaacaataatcagacaaacattacctttagtgcagaggtggcagaactatacagattggagttgggagattataaattggtagaaataacaccaattggcttcgcacctacaaaagaaaaaag
atactcctctgctcacgggagacatacaagaggtgtgttcgtgctagggttcttgggttttctcgcaacagcaggttctgcaatgggcgcggcgtccctgaccgtgtcggctcagtcccggactttactggccgggatagtgcagcaac
agcaacagctgttggacgtggtcaagagacaacaagaactgttgcgactgaccgtctggggaacgaaaaacctccaggcaagagtcactgctatagagaagtacctacaggaccaggcgcggctaaattcatggggatgtgcgttt
agacaagtctgccacactactgtaccatgggttaatgattccttagcacctgactgggacaatatgacgtggcaggaatgggaaaaacaagtccgctacctggaggcaaatatcagtaaaagtttagaacaggcacaaattcagcaag
agaaaaatatgtatgaactacaaaaattaaatagctgggatatttttggcaattggtttgacttaacctcctgggtcaagtatattcaatatggagtgcttataatagtagcagtaatagctttaagaatagtgatatatgtagtacaaatgttaag
taggcttagaaagggctataggcctgttttctcttcccccccccggttatatccaacagatccatatccacaaggaccggggacagccagccaacgaagaaacagaagaagacggtggaagcaacggtggagacagatactggccct
ggccgatagcatatatacatttcctgatccgccagctgattcgcctcttgaccagactatacagcatctgcagggacttactatccaggagcttcctgaccctccaactcatctaccagaatctcagagactggctgagacttagaacagc
cttcttgcaatatgggtgcgagtggatccaagaagcattccaggccgccgcgagggctacaagagagactcttgcgggcgcgtgcaggggcttgtggagggtattggaacgaatcgggaggggaatactcgcggttccaagaagg
atcagacagggagcagaaatcgccctcctgtga'

4. 'HIV2-MCN13' 'AY509259'

'atgatgggtggtagaaatcagctgcttgttgccattttgctaactagtacttgcttgatatattgcaccaattatgtgactgttttctatggcatacccgcgtggagaaatgcatccattcccctcttttgtgcaaccaagaatagggatacttgg
ggaaccatacagtgcttgccagacaatgatgattatcaggagataactttgaatgtgacagaggctttcgatgcatgggataatacagtaacagaacaagcaatagaagatgtctggaatctatttgagacatcaataaaaccatgtgtca
aattaacgcctttatgtgtagcaatgagatgtaacaacacagatgcaaggaacacaaccacacccacaacagcatccccgcgtacaataaaacccgtgacagagataagtgagaattcctcatgcatacgcgcaaacaactgctcag
gattgggagaagaagaggtggtcaattgtcaattcaatatgacaggattagagagagataagaaaaagcaatatagtgagacatggtactcgaaggatgtagtttgtgaaggaaatggcaccacagatacatgttacatgaaccattgc
aacacatcggtcatcacagagtcatgtgacaagcactattgggatgctatgaggtttagatactgtgcaccaccaggttttgccctactaagatgcaatgataccaattattcaggctttgcgcccaattgctctaaggtagtagctgctaca
tgcaccagaatgatggaaacgcaaacttctacatggtttggctttaatggcactagagcagaaaatagaacatttatctattggcatggtagggataacagaactatcatcagcttaaacaaatattataatctcactatacattgtaagaggc
caggaaataagacagtggtaccaataacacttatgtcagggttaaggtttcactcccagccggtcatcaataaaagacccagacaagcatggtgttggttcaaaggtgaatggaagggagccatgcaggaggtgaaggaaacccttg
caaaacatcccaggtataaaggaaccaatgaaacaaagaatattaactttacagcaccaggaaagggctcagacccagaggtggcatacatgtggactaactgcagaggagaatttctctactgcaacatgacttggttcctcaattgg
atagaaaataagacacaccgcaattatgtaccgtgccatataagacaaataattaacacctggcataaggtagggaaaaatgtatatttgcctcccagggaaggggagttgacctgcaactcaacagtaactagcataattgctaacatt
gatgcaaatggaaataatacaaatattacctttagtgcagaggtggcagaactataccgattagagttgggagattataaattggtagaaataacaccaattggcttcgcacctacagcagaaaaaagatactcctctactccaatgagga
acaagagaggtgtgttcgtgctagggttcttgggttttctcgcaacagcaggctctgcaatgggcgcggcgtccttaacgctgtcggctcagtctcggactttactggccgggatagtgcagcaacagcaacagctgttggacgtggtc
aagagacaacaggaaatgttgcgactgaccgtctggggaacaaaaaatctccaggcaagagtcactgctatcgagaagtacttaaaggaccaggcgcaactaaattcatggggatgtgcatttagacaagtctgccacactactgtac
catgggtaaatgataccttaacgcctgagtggaacaatatgacgtggcaagaatgggaaggcaaaatccgcgacctggaggcaaatatcagtcaacaattagaacaagcacaaattcagcaagagaagaatatgtatgaactacaaa
agttaaatagctgggatgtttttggtaactggtttgacttaacctcctggatcaagtatattcaatatggagtttatataataataggaatagtagttcttagaatagtaatatatatagtacagatgttaagtagacttagaaagggctataggcct
gttttctcttcccccccccggttacctccaacagatccatatccacaaggactgggaacagccagccagagaagaaacagaagaagacgttggaaacaacgttggagacagctcgtggccttggccgataagatatatacatttcctgat
ccaccagctgattcgcctcttggccggactatacaacatctgcaggaacttactatccaggatctccctgaccctccgaccagttttccagagtcttcagagggcactgacagcaatcagagactggctaagaactgacgcagcctactt
gcagtatgggtgcgagtggatccaaggagcgttccaggccttcgcaagggctacgagagagactcttgcgggcacgtggagagacttgtggggggcactgcagcggatcgggaggggaatacttgcagtcccaagaagaatca
ggcagggagcagagatcgccctcctatga'

5. 'HIV-2UC1 (IvoryCoast)' 'L07625'

'atggcacacactagcaatcacctgtttattttgctcctacttataagtgtctatgggtttctgggtcataagaaaaactatgtcactgtcttctatggcatacctgcatggaggaacgcaacggttcctctcttctgtgcaaccacaaacaggg
atacctggggaactgtacagtgcctcccagacaatggggattacaccgagatcagtgtaaatataacagaggcttttgatgcatggaataatacagtaacagaacaggcagtagatgatgtgtggagtctttttgaaacctccataaaac
catgtgttaagctgacgccgctgtgcgtggcaatgagatgtaataacacaggaaccaatactactactaaacctattactacacctattactactactaaaccatcagaaaacttactcaacgacacaagtccatgtattaaaaatgacacct
gcccgggaataggactagaaaacacagttgactgttactttaacatgacaggactaagaagagatgagaaaaaacaatataaggacacctggtacgaaaaagatttagagtgcaacggcaacagcaccagcaccatatgttacatga
gaacttgtaatacctcagtgatccaagagtcatgtgataaacattactgggacagcttaagatttaggtactgtgctcccccaggatatgctctgctaagatgtaatgacaccaactattcaggctttatgccaaaatgtagtaaagtagtagt
gtcctcttgcacaagaatgatggagacacagacctctacatggtttggcttcaatggtacgaggacagaaaataggacatatatgtattggcatagtaaagataataggaccatcataagcttgaataagtattataatctaacaatgcattg
cagaaggccagggaacaagacagtcataccaataacaatcatgtcaggactgaatttccattcgcagcctctcaataccagacccagacaggcttggtgctggtttaagggaaactggatagaagccataagggaggtaaaggaga
ccatcataaaacatcccaggtataaaggaacaaataacaccgagaggataagattggtagggccctcggcaggttcagacccagaggtgagacatatgtggactaactgtaggggagaattcttctattgtaatatgacctggtttttaa
actgggtggagaacagaacaggtacaactcagaagaattatgtgacgtgccacatcaaacagatagtcaacacgtggcacaaggttggaaaatacgtatatttgcctccaagggagggtacgctctcctgtaattccagtgtcaccagt
cttatagccaacattgacgtgtattatgatggcaatgataccaagaccaatattaccatgagtgcagaagtgggagaactgtacagattggaactgggggactacaaattagtagaaataacaccgattggctttgcacctacagagataa
aaagatattcctcaacgacaccgaggaataaaagaggtgtaatggtgctagggttcttgggacttcttgcaatggcaggttctgcaatgggcgcaacgtccttgacgctgtcagctcagtcccggactttactggctgggatagtgcagc
agcagcaacagctgctggacgtggtcaaaagacaacaggaactgttgcggctgaccgtctggggaacgaaaaacctccagacaagagtcactgccatcgagaaatacctaaaggaccaggcactactaaattcgtggggatgtgc
ttttaggcaggtttgtcacacgactgtaccatggccaaatgaaactctcacaccagactgggaaaacatgacatggcaacagtgggaaaagcgagtcaatttcttggatgcaaatataacagccctgttagaagaggctcagatacaac
aagaaaggaacatgtatgaattacagaaattaaatagctgggatgtttttggtaattggtttgacttcacctcctggatggcatacatcaggttaggactatatgtagtagcaggattaatagtgttaagaatagtaatatacattatgcagatg
ctagcaagacttaggaagggctataggccagtattctcctcccctccctcttatactcaacagatccctatccgcaaacaccggggacagccagccaacgaagaaacagaagacgaaggtggaaacgaaggggcttacagatcttg
gccctggcagatagaatacgctcactttctgattcgccaactgaggaacctcttgatttggctgtacaacggctgcaggaacttactgttgaagacctcccaaatcctccaaccagcactccaaccgctcaggctttcacttgcatacctcc
agtatgggatcagctggttccaagaagcaatccaagcagcaacgagggctgcgagagagactcttgcgaacacaggaagagccctatggaaagctctcagaaggacagcggaagcaatcatcgcgatccccaggaggatcaga
caaggacttgaactcgccctcttgtga'

6. 'SIVMM251 Macaque' 'M19499'

'atgggatgtcttgggaatcagctgcttatcgccatcttgcttttaagtgtctatgggatctattgtactcaatatgtcacagtcttttatggtgtaccagcttggaggaatgcgacaattcccctcttctgtgcaaccaagaatagggatacttgg
ggaacaactcagtgcctaccagataatggtgattattcagaattggcccttaatgttacagaaagctttgatgcttgggagaatacagtcacagaacaggcaatagaggacgtatggcaactctttgagacctcaataaagccttgtgtaa
aattatccccattatgcattactatgagatgcaataaaagtgagacagatagatggggattgacaaaatcatcaacaacaataacaacagcagcaccaacatcagcaccagtatcagaaaaaatagacatggtcaatgagactagttctt
gtatagctcagaataattgcacaggcttggaacaagagcaaatgataagctgtaaattcaccatgacagggttaaaaagagacaagacaaaggagtacaatgaaacttggtactctacagatttggtttgtgaacaagggaatagcact
gataatgaaagcagatgctacatgaatcactgtaacacttctgttatccaagaatcttgtgacaaacattattgggatactattagatttaggtattgtgcacctccaggttatgctttgcttagatgtaatgacacaaattattcaggctttatgcc
taaatgttctaaggtggtggtctcttcatgcacaaggatgatggagacacagacttctacttggtttggctttaatggaactagagcagaaaatagaacttatatttactggcatggtagggataataggactataattagtttaaataagtatta
taatctaacaatgaaatgtagaagaccaggaaataagacagttttaccagtcaccattatgtctggattggttttccactcacaaccaatcaatgataggccaaagcaggcatggtgttggtttggaggaaaatggaaggatgcaataaaa
gaggtgaaacagaccattgtcaaacatcccaggtatactggaactaacaatactgataaaatcaatttaacggctcctggaggaggagatccggaagttaccttcatgtggacaaattgcagaggagagttcctctactgtaaaatgaatt
ggtttctaaattgggtagaggatagggatgtaactacccagaggccaaaggaacggcatagaaggaattacgtgccgtgtcatattagacaaataatcaacacttggcataaagtaggcaaaaatgtttatttgcctccaagagaggga
gacctcacgtgtaactccacagtgaccagtctcatagcaaacatagattggactgatggaaaccaaactagtatcaccatgagtgcagaggtggcagaactgtatcgattggagttgggagattataaattagtagagatcactccgatt
ggcttggcccccacagatgtgaagaggtacactactggtggcacctcaagaaataaaagaggggtctttgtgctagggttcttgggttttctcgcaacggcaggttctgcaatgggcgcggcgtcgttgacgctgaccgctcagtccc
ggactttattggctgggatagtgcagcaacagcaacagctgttggacgtggtcaagagacaacaagaattgttgcgactgaccgtctggggaacaaagaacctccagactagggtcactgccatcgagaagtacttaaaggaccagg
cgcagctaaatgcttggggatgtgcgtttagacaagtctgccacactactgtaccatggccaaatgcaagtctaacaccagactggaacaatgatacttggcaagagtgggagcgaaaggttgacttcttggaggaaaatataacagcc
ctcctagaagaggcacaaattcaacaagagaagaacatgtatgaattacaaaagttgaatagctgggatgtgtttggcaattggtttgaccttgcttcttggataaagtatatacaatatggaatttatgtagttgtaggagtaatactgttaag
aatagtgatctatatagtacaaatgctagctaagttaaggcaggggtataggccagtgttctcttccccaccctcttatttccagtagactcatacccaacaggacccggcactgccaaccagagaaggcaaagaaggagacggtggag

```
aaggcggtggcaacagctcctggccttggcagatagaatatattcatttcctgatccgccaactgatacgcctcttgacttggctattcagcaactgcagaaccttgctatcgagagcataccagatcctccaaccaatactccagaggct
ctctgcgaccctacgaagggttcgagaagtcctcaggactgaactgacctacctacaatatgggtggagctatttccatgaggcggtccaagccggctggagatctgcgacagaaactcttgcgggcgcgtggagagacttatggga
gactcttaggagaggtggaagatggatcctcgcaatccctaggaggattaggcaagggcttgagctcacgctcttgtga'
```

7. 'SIVAGM677A Green monkey' 'M58410'

```
'atggggagattgcttataaaaatactaataatagcaatagggataagtataggaataggtaacctgtatgtgacagtgttttatggaatcccagtatggaaaaattcaacagttcaggcattttgcatgacgcccaataccaatatgtgggc
aaccaccaactgcataccagatgatcatgataatacagaggtgcctctaaacattacagaagctttcgaggcttgggataatccgctggtaaaacaagcagagagtaatatacatctactctttgaacaaacgatgaggccttgtgttaag
ctctcccccatatgtattaaaatgtcctgtgtagagctgaatggtacagccacgacaaaggccaccactactgcaactacaacaatgactaccccctgtcagaattgcagtacagagcagatagaaggagaaatggcagaggaaccag
catccaactgcacttttgcaattgcaggatatcaaagagatgtaaaaaagaattatagcatgacctggtatgatcaggagttagtctgcaataataaaacaggaagtgaaaagggaagtaaggattgttacatgatacattgtaatgattca
gtgataaaagaagcttgtgataaaacatattgggatactttaagagtaagatactgtgcaccagcagggtatgctttgctaaaatgtaatgataaggattatagaggctttgctccaaagtgcaagaatgtttcagtagtgcattgtactagat
taatcaatactactataactacagggataggattaaatggtagtagatcagaaaatagaacagagatatggcagaaaggaggaaatgataatgatacagttataataaagttgaataagttttacaacttgacagtgagatgccgaagacc
tggtaataaaacagtgttgccagtaacaatcatggcagggttagtatttcactctcagaaatataataccaggttaaaacaagcgtggtgccacttccaaggagattggaaaggggcatggaaagaagtcagagaagaagtaaagaaa
gtgaaaaatcttacagaagtaagcatagaaaatatacatctgagaaggatatggggagatccagaatcagcgaatttttggttcaattgtcaaggtgaatttttctattgtaagatggactggtttatcaattatctaaacaatcgaacagaag
atgcagaaggtactaataggacctgtgacaaagggaagccaggaccaggaccatgtgttcagagaacttatgttgcctgccatatacgacaagtagtaaatgattggtacactgtctctaaaaaggtatatgctccaccaagggaaggt
catttggagtgtaactcatcagtcacggcactatacgtggcaatagattataacaacaagtctggcccaataaatgtgaccctaagtcctcaggtacgcagcatatgggcgtacgaactgggagactataaattagtagagataacacca
attggctttgctcctacagatgtaagaagatatactggccccacaagagaaaaaagggtgccattcgtgctagggtttctaggcttcttgggagctgctggaactgcaatgggcgcagcggcaacaacgctgacagtccagtctcggc
atttgcttgctgggatattgcagcagcagaagaacttgctggcggctgtggaacagcaacaacagttgttgaagctgaccatttggggtgtgaaaaacctcaatgcccgcgtcacagctctcgagaagtacctagaggatcaggcacg
gctaaattcatggggatgtgcgtggaaacaagtatgtcacaccacagtgccatggaagtataataacactcctaagtgggacaatatgacttggttggagtgggagagacaaattaatgccttggaaggcaacataactcaactattgga
agaagcacaaaatcaggaatcaaagaatctggatctgtaccagaaattggatgattggtcagggttctggtcatggttctcactgtcaacttggttaggctatgttaaaataggatttttagtgatagtgattattctaggattaagatttgcatg
ggtattatggggatgtatcagaaatattaggcagggatataatcctctcccccagatccatatccacagttcagcggaacggccagacaacggaggagggcaagacagaggtggagaaagcagcagcagcaaattgataagattgc
aggaagagtcctcaacaccttcgaggatcaacaactggtggctcaacttcaagagctgcagcttgagaataaggacttggtgttacaacatctgcctgaccctcctcatattcatcaggacagcagtgggatacctgcagtatgggctcc
agcaactccaagaggcagcaacagggcttgctcaagctctggcgagggctgcgagggaagcctggggcagactgggtgctattgtccgatccgcttatcgggcagtcatcaacagtccaagaagagtgcggcaaggccttgaaa
aagtcctggggtaa'
```

8. 'SIVlhoest L''Hoest monkeys' 'AF075269'

```
'atggcatgtccaggattaggaattctgcttttgcttcttggcataatatggggaaaacaatatgtaacagtattttatggggtgccaaattgggatgataatgtttcagtgcccttgatttgtgcttcagccaatactagcctctgggtcacgac
atcttgcttgccagatttgcaatcttatgcagaggtacctatctataatatttcagaaaactttacaataccagtaaaagacaatcaggttatacagcaagcctggtctgctatgaatgctatggtagacagtattatgaaaccatgtgtaaaaat
caatccatattgtgtgagaatgcagtgtggggaagttacaaaaacaccaacaacaacaccaaaaactactacacagatgccttgttttatcaacgaacaggtaacagttaagaacccggggaatgaaacaagattagaggaagatctta
attgcacaagagggcttaatgagaccacagagaggaatgcagaatgccagtataatgtaacagggttatgcagagactgcaggacggaaataaaacaaagctttagatatgatgatgtaacgtgctcaggagagagggagaacaga
acctgttacatgactcattgtaatgattcaataataacacaagattgtaataaggggggtaatgcaaaatgcttattttaggctttgtgctccagcaggctatatgctattgagatgcaatgaacaattaaactttagtaaaaaatgtgaaaatata
acagctacaccatgtactggatatatgcttagtagtgtaagtagtttctttggctttaatggaactaatcataccagggatgagcttattccactaactccaaataagatggaagatctaaatggtgcaaagtttgtgtataaagtagcaggaaa
atggggtttaattattagatgcataagaaaaggaaatagatcagaagtatccacaattagctctacaggatatttattttattatggcctagaacatggcagcagattaagattagcacagtgtaaatttgaaggccaatggggaagaatgttt
aataatctaggcaaaatgctaaaggagttgaatgcagaagccatgaattatacagaagggacagggacatgtgattctaagaaaacaacttgtggacggaaattaaaaggcttaccaatagctaatatgactaggcatggagcagactt
agcaacagagatgttaatgcatacttgtggagaggaaatgttcttttgtaatgtaactaggatattccaagaatggaacaataaaaattcagataaatggtatccttgggccaactgccacataaaaagcatcattgatgattgggctacaat
aggaaagaagatttatctaccacctacttcaggtttcaacaatagaatcagatgtactcatagagtaacagaaatgttctttgagatggaaaaatgggaaccacatgaggacttaggaggcaacctaagtataaagttcttgcctccatcat
gggaaactaatcaatttgtggcagaagggtctaaatataaattgataaaattgaatccaattgggtttgctcctacagatgagcacaggtacgcacccaggggaaggcaaaccagagcagcgcctctggctttaggagccttgggactt
ctcagtgctgcaggcactgcaatgggcttagtatcgacgatactaactgtccaggcccaagcagtcttgcaagggatattgcagcagcagaagcagttgctggtgcttgttgagaaacagcaagaactgttaagactcaccatatgggg
agtgaagaatctccaagcacgcctgacagcccttgaggagtatgtgaaacatcaagcgctccttgcctcttgggggtgtcaatggaaacaggtttgtcatactaatgtggagtggacatataatatcacacccaattggaccaaggatac
ctggagagaatgggaatcaaaggtagcaatttatgacaagaacattacttctctgttacaagaagcatacaccacagagctagagaatcagaataaatttaagaagttacaagaatttaacttttggagttggttggatatctcacactggttt
acttatgttaagtatgcagtgttaattatacttgtaatcatagggttaagagtattgagctttataatacaaaatgtagtaaaaatgtgtagggggtatagggtgctctccccctctgtttatattgaacaggactacaagtgggagaaagaaga
gaaccaagaacagccagacagagaagaagagaaaggagcagatacagagactatctacatcaacttagagcagtgcaagaaagaatcttccaggccactgtggaacgtggactggaacgagcctttacaagactcgctgttagtg
actctcctgaagtggctcaaggaagggggaatactcctcctatcactagtgtggcagagcctcagttggctgtggcatttgttgatccttttcttccaaaatgggcaacgcctttggcagaccagcagcagatggatggtggaaaacgct
cagaagattcagagctggctcagggagaaatgcagaaggaacagaggacagttatcgagcactgatagaaaaaacatccaattgggcaaaaagaagagatggagactcagattcggaggaagaagtgggatttccagtgaagcc
acagagaccgctttgtag'
```

9. 'SIVcpz Chimpanzees Cameroon' 'AF115393'

```
'atgaaagtgatggagaagatcaaaggcagcttgatagagaaatacatgtttatgggtttgataatcccatgtttgactggtagtgatcagttgtgggcaacagtgtactatggggtacctgtgtggaaagaggcagacacaacactctttt
gtgcctcagaggcgtcagctcttaacaaagaggcccataatatttgggcctcacaagcctgtgttcctacagaccctaacccacaagaagtgcagatcccaaatgtaactgaaaactttaatatgtggaacaataccatggcagaacaa
atgaatactgacattattagtctgtgggatcaaagtttgaaaccatgtgttaagctgacgccaatatgtgtcaccatggaatgtagaaaggttactttcaatagcacttctaataggaataaaacttctactatgactactaacagtcctaatgaa
aaaaaagacagcacagtaaagaattgtacctttaatatgactactgaggtaagagataaagagaaaaaagtatattcccttttttatgtagatgatctagtattgattgataacgataccgatacatatagactaataaattgcaataccacag
ctattacacaggcgtgtcctaagacctcctttgagccaattccaatacattactgtgcaaccccaggttatgcaatcatgaaatgcaatatgccaaactttaatggaacaggaacaggaagatgtaacaatataagcacagtccaatgtact
catggcatcagaccagtggtaactacacaattgatccttaatggaagtgtagccgaaaacaagactattgctagaaggaacgggtacaacttcctgatacaatttcagaaaacggtcagcataaattgtacaagaccaggaaataacag
cagaggacaaatacagataggaccaggcatgaccttttataacatagaaaatattgtaggagataccagacgagcatattgtcagatcaataggacagtctgggacgaaagactaaatgaaactggtcaagccttgcgagagcttttta
caaaccttacacaagtcaactttacagtaagtccaggaggagacccagaagtaacaaatatgatgtttaattgtggaggagaattcttttattgcaatactactacattgtttaattatacgtggaaaaacaacaatataactaagggagataa
caccaccttcttcccatgtagaataaggcaaattgtaaactcctggatgagagtggggaagggaatttatgccccaccaattagaggagttataagttgcacatcaaacgttactgggattatacttgagacagggcatggcatcaacaac
agcatcaccaacatcacgctataccctacaggaggaaacatggtagacttgtggaggcttgaattgcataaatacaaggtagttagcatagaacccataggagtagcaccaagtaaagctaaaaggcacacagtatcaagagaaaaa
agagcagcctttggactgggtgcgctgtttcttgggttcttaggagcagcaggaagcactatgggcgcagcatcagtggtgctgacggtacaggcccggcaattattgtcagggatagtgcaacagcagaacaacctgctgcgagca
atagaggcgcaacagcatttgttgcaactctcagtttggggcattaaacagctccaagcaagagttcttgctgtagaaagatacctcagagatcagcaaatcctaggcctatggggctgctcaggaaaagccatttgttatactactgtgc
cttggaataatacttggagtgctaatacatcctttgatgagatctggaataatttaacatggcaagactgggacaaaagggtaaaaaattattcaggtgtcatttttagccttatagaacaagcacaagagcaacaaaacacaaatgagaaa
agtctattggaattggatcaatggtcaagtttatggaattggtttgatattacccgctggctgtggtacataaaattgtttataatgatagtagcaggcttagtaggcataagaatagtaggggctatcatatcatttgtagcaaaggttaggca
gggatactctcccctctcgttgcagacccttatcccgacaacaagggaaccagacaggccaggaggaacagaagaagacgttggagagccaggcaaaggcagatccattcgcttagtgagcggattcttagctcttatctgggaag
atctccggaacctcctgctcttctgttaccaccgattgagagacttactattgatactggggaggacattggaaaacctgggacagagcctcaacaagggactgcaacagctgaggaacttcagcagatacctgtggggagtgataacc
tactggggaagggagctacaaactagtgccataagcttattagatgcaacagccattgcagtagctgaaggaacagacagaattctagaagttgcacaaataataggtagaggcattctgcatataccaagaagaattagacagggttt
agaaagaagcttgctataa'
```

10. 'SIVmnd5440 Mandrillus sphinx' 'AY159322'

```
'atgttaagatatcttaggtatatagtcttaggaataatagtaagtgtaatagtaggagaacaatgggtgacagtgtactatggcacacccaaatggcataaggcaaggacacacctgttttgtgcaacagacaataattcattctgggtgac
cacaagctgtgtccccagtctgttgcactatgaagaacagcatatccccaatataacagagaattttacaggccctatagaagagaatgaaatagtgacacaagcttggggagcaatttcttccatgatagatgcagtattaaaaccatgt
gtaaggttaacaccttattgtgtgaagatgaaatgtacagaaggtcaaaatgaaacagagcaagcaacagctaagaccacaacacctgtacccacaacaacaaccccctccactaccactagctccagcacaaataagacaacaact
cctgtgttggttgttgagaaacaaaataatgagacaacaacacagcaaaatagagtatgtaaattcaacacgacaggattatgtagggactgcaaattggaaatagaggaaaactttagatatgaagatgtaacttgcacaaagcttaata
aaactggtagtgctactaatagtactgaacctgaatatgagtgttacatgacttcatgtaacgcaacagtaataacacaggaccgtaacaaagcttccacagatagaatgacatttaggttgtgtgcacccccagggtttgtcttactaaaat
gtaatgaaaaattgaacaagacaaaattatgtgggaatgtgtctgcagtgcagtgcacagcaccgctgccagccacaatctctacaatgtttggcttcaatggaactaagcatgattatgatgagcttattcagacaaacccacgaaaag
gaaaagatgagtttcatgaccataaatatgtatatagagtggataaaaaatggggattacaggtaaggtgtagaaggaagggaaatagatctatcatttccacaccaagttctacaggattgctgttctatcatgggttggagccaggaaa
gaatcttagaaagggcaagtgtcagttagaagggaaatggggacaggctttacatagtctatcattagagctaagaaagataaatgatagcatttacaaagacaaccacaatatgacatgtaaaagcagtaacaacaaaaagaacaca
acagggtgtcatcttaaaactataagtataagtgagtccaccgttaaaggcgaaccaggagctgaaactattatgctcctctgtggaggagaatatttcttttgtaattggactaaaatatggaaggcatggaatagcaaacaaagttcagtt
tggtacccttacatgtctcgcaatattagacaaatagtaggtgattggcataaagtaggcaagaaaatttatatgcctcctgtgtctgggtttaataatgaaataagatgtactaatgatgtgacagaaatgttctttgaggtccaaaaaactga
tgatgacaatggatatatcataaaatttataccacaagattggattcaaaatcaatacacagcagtgggggctcattacaaattggtgaaggtggatccgataggttttgcccccacagacatacacagacatcatttgcctaatacaaggc
aaaaaagaggagcggtcttacttggaatgctcggcctcctaggtttggcaggttccgcgatgggctcagtggcggtggccctgactgtccagtcccagactttactgaatgggattgtggagcagcagaaggttctgctgagcctgata
gatcagcactccgagttattaaaactaactatctggggtgtaaagaatcttcaggtccgcctcacagccttggaggaatacgtagcggaccaatcaaggctctcggtatggggatgttcattctcccaagtttgtcatactagtgtaaaatg
gcccaataatagtatagtgcccaattggacctcagagacatggctggaatgggacagaagagtaaacagcatagtgacaaatatgaccatagacttacagagggcatatgaattagaacagaggaacatctttgagttgcaaaaattag
gggacttaaacttccatgggctaactgggttcgaccttacatggtggcttaagtatgttaagataggactgttagtagtagtagttatcataggattaagaatgttagcttgcttatggtcagtattagggaagtttaggcagggatatcgccct
```

```
cttccttatgtcttcaagggagactatctccggccccacaacctcaaacggccagacagagaaggaggagaagagccagacttagaaaaacagaacatcaagtcagagagctccagacaagaatctaggaagccttggaagccag
agcaagtcaggagctggctgaaaaggtcgacgctttacatttggctgaagaacctccaagcagtaattgaatatgggtggcaagagctcaaagcagcgggagcagcattatatcaagtattacaaggctttgcgcagaggctatggag
cagagggtaccaacttggactatcaggtgttagaggagcagcagcctttggcaggggcatctggaacatccctcgccgcatcagacaaggtgctgaggccctactcaactga'
```

11. 'SIVAGM3 Green monkeys' 'M30931'

```
'atgaagctgacattactgatagggatactattaatagggataggagtagtgcttaatacaaggcaacaatgggtcacagtattttatggagtaccagtatggaaaaacagctcagtacaggctttctgcatgacacccaccaccagactat
gggcaactactaactcgataccagatgatcatgactacacagaggtaccattaaacatcactgaaccatttgaagcatgggctgacagaaaccccttagtagcacaagcaggaagtaatatacacctgctatttgagcagactctgaag
ccatgtgtaaaattatcacctttgtgcattaaaatgtcctgtgtagaattgaactcctctgagcctaccaccactcctaaaagtaccacggcctcaacaaccaatatcacagcctcaacaaccactttgccgtgtgtccagaacaagacaagt
actgtgttagaatcatgtaatgaaacaatcatagaaaaggaattaaatgaagagcctgcttctaattgtacatttgcaatggcagggtatgtaagagatcagaaaaagaagtattcagtggtgtggaatgatgcagaaatcatgtgtaagaa
gggtaacaattctaacagagaatgttatatgattcattgtaatgattcagttataaaagaagcctgtgataaaacatattgggatgagttaagattaaggtactgtgccccggcagggtttgctttattaaaatgcaacgattatgattatgcag
ggtttaagacaaactgttctaatgtttcagtggtgcattgtactaacttgataaatacaacagtgactactggactgttgttgaatgggagctactcagagaatcgaacccagatatggcagaaacatagagtaagcaatgactcagtgttag
tgttatttaataaacattacaatctaacagttacttgcaaaagaccaggaaacaaaacagtcttaccagtaacaatcatggcagggctagtgtttcattctcagaggtacaatacaaggctgagacaagcttggtgtcacttccagggcaac
tggagaggagcctggaaagaagtaaaaaatgaaatagtaaaattaccaaaagatagataccaaggaaccaatgatactgaagagatttatctgcagagactatttggagatccagaagcagcaaatttatggtttaattgtcaggggga
attcttctattgtaaaatggattggtttctaaattacctgaataatcgtacagtagatccggaccataatccgtgtaatggtacgaagggaaaaggtaaggcaccaggaccctgtgcacaaagaacatatgttgcttgccatatacgatctgt
cattaatgattggtacacactatcaaggaaaacctatgcaccgccaagagaagggcacttgcaatgcacatccacggtaacgggtatgtcagtggagctaaattacaatagtaagaacaggactaatgtaacattaagtccccagatag
aaaccatctgggcagcagaattgggcaggtacaaattagtagaaattacaccaattggcttcgcacccacagaagtaagaaggtatacgggaggtcatgacagaacaaagcgagtcccgttcgtgctagggttcctaggcttcttagg
agctgctgggactgcaatgggagcagcggcgacagccctgacggtccagtctcagcatttacttgctgggatactgcagcagcagaagaatctgctggcggctgtggaggctcaacagcagatgttgaagctgaccatttggggtgt
gaaaaacctcaatgcccgcgtcacagctcttgagaagtacctagaggaccaggcgcggttgaatgcttgggggtgcgcatggaagcaagtctgtcatacaaccgtaccgtggcagtggaataataggacccctgattggaataatat
gacttggctggaatgggaaagacagatatcgtatttggaaggtaacataacaacacaattagaggaagccagagcacaggaggagaagaatttggatgcataccaaaaattaagtagttggtcagatttctggtcttggttcgatttctca
aagtggctgaacattctaaaaataggatttttggatgtactaggtattataggattaagattgctttatacagtatattcttgcatagctagggttaggcagggttactctcctctttctccacagatccatatccacccgtggaagggacagcca
gacaacgcagaagggccaggagaaggtggagacaagcgcaagaacagctccgagccttggcagaaagaatctggcacagcagagtggaagagcaactggtgcaagcgattgaccaattggtgctcgatcagcagcatctggct
atacaacagttgcctgaccctcctagttcatcttaggagcgctttccagtacatacaatatgggcttggggaactcaaagccgcagcacaagaagcagttgtcgctttggcacgccttgcacaaaacgcgggctaccagatatggcttgc
ttgcagatccgcttatagggcaatcatcaactctccaagaagagtgcgacaaggccttgaaggaatccttaattag'
```

12. 'SIVMM239 Simian macaque' 'M33262'

```
'atgggatgtcttgggaatcagctgcttatcgccatcttgcttttaagtgtctatgggatctattgtactctatatgtcacagtcttttatggtgtaccagcttggaggaatgcgacaattcccctcttttgtgcaaccaagaatagggatacttgg
ggaacaactcagtgcctaccagataatggtgattattcagaagtggcccttaatgttacagaaagctttgatgcctggaataatacagtcacagaacaggcaatagaggatgtatggcaactctttgagacctcaataaagccttgtgtaa
aattatccccattatgcattactatgagatgcaataaaagtgagacagatagatggggattgacaaaatcaataacaacaacagcatcaacaacatcaacgacagcatcagcaaaagtagacatggtcaatgagactagttcttgtatag
cccaggataattgcacaggcttggaacaagagcaaatgataagctgtaaattcaacatgacagggttaaaaagagacaagaaaaaagagtacaatgaaacttggtactctgcagatttggtatgtgaacaagggaataacactggtaat
gaaagtagatgttacatgaaccactgtaacacttctgttatccaagagtcttgtgacaaacattattgggatgctattagatttaggtattgtgcacctccaggttatgctttgcttagatgtaatgacacaaattattcaggctttatgcctaaatg
ttctaaggtggtggtctcttcatgcacaaggatgatggagacacagacttctacttggtttggctttaatggaactagagcagaaaatagaacttatatttactggcatggtagggataataggactataattagtttaaataagtattataatct
aacaatgaaatgtagaagaccaggaaataagacagttttaccagtcaccattatgtctggattggttttccactcacaaccaatcaatgataggccaaagcaggcatggtgttggtttggaggaaaatggaaggatgcaataaaagaggt
gaagcagaccattgtcaaacatcccaggtatactggaactaacaatactgataaaatcaatttgacggctcctggaggaggagatccggaagttaccttcatgtggacaaattgcagaggagagttcctctactgtaaaatgaattggttt
ctaaattgggtagaagataggaatacagctaaccagaagccaaaggaacagcataaaaggaattacgtgccatgtcatattagacaaataatcaacacttggcataaagtaggcaaaaatgtttatttgcctccaagagagggagacct
cacgtgtaactccacagtgaccagtctcatagcaaacatagattggattgatggaaaccaaactaatatcaccatgagtgcagaggtggcagaactgtatcgattggaattgggagattataaattagtagagatcactccaattggcttg
gcccccacagatgtgaagaggtacactactggtggcacctcaagaaataaaagaggggtctttgtgctagggttcttgggttttctcgcaacggcaggttctgcaatgggcgcggcgtcgttgacgctgaccgctcagtcccgaacttt
attggctgggatagtgcagcaacagcaacagctgttggacgtggtcaagagacaacaagaattgttgcgactgaccgtctggggaacaaagaacctccagactagggtcactgccatcgagaagtacttaaaggaccaggcgcag
ctgaatgcttggggatgtgcgtttagacaagtctgccacactactgtaccatggccaaatgcaagtctaacaccaaagtggaacaatgagacttggcaagagtgggagcgaaaggttgacttcttggaagaaaatataacagccctcct
agaggaggcacaaattcaacaagagaagaacatgtatgaattacaaaagttgaatagctgggatgtgtttggcaattggtttgaccttgcttcttggataaagtatatacaatatggagtttatatagttgtaggagtaatactgttaagaata
gtgatctatatagtacaaatgctagctaagttaaggcaggggtataggccagtgttctcttccccaccctcttatttccagcagacccatatccaacaggacccggcactgccaaccagagaaggcaaagaaagagacggtggagaag
gcggtggcaacagctcctggccttggcagatagaatatattcatttcctgatccgccaactgatacgcctcttgacttggctattcagcaactgcagaaccttgctatcgagagtataccagatcctccaaccaatactccagaggctctct
gcgaccctacagaggattcgagaagtcctcaggactgaactgacctacctacaatatgggtggagctatttccatgaggcggtccaggccgtctggagatctgcgacagagactcttgcgggcgcgtggggagacttatgggagact
cttaggagaggtggaagatggatactcgcaatccccaggaggattagacaagggcttgagctcactctcttgtga'
```

13. 'CIVcpzUS Chimpanzee' 'AF103818'

```
'atgaaagtgatggagaagaagaagagactctggctaagttattgtcttctctcgagcttgataatcccaggattgtctagcttatgggctacagtatattatggggtacctgtatggagagatgtagagacaaccttattctgtgcctctgat
gcaaaggcatacaagcaggaggcccacaacatttgggccacacaggcatgtgtccctactgaccccaatccacaagaagtacatttgccaaatgtgactgaaaagtttgacatgtgggaaaataatatggcagaacaaatgcaggaa
gatattattagcctatgggaccagagcttaaaaccttgtataaagttaaccccattgtgtgttactatgacttgtcttaaccccgatagtaatagtagtgctgtaaatactactgatataatgagaaactgttcttttaatataactactgaattaag
agacaaaaagaaacaagtgtattccttattttatgtagatgatctagctcatatcaataataatacctatagactgataaattgcaacaccaccgctatcacacaggcttgtcctaagacctcctttgagccaattccaatacactattgtgcac
caccaggctttgccatcctaaaatgtaatgaaaaagatttcaaaggaaagggagagtgtaaaaatgttagtacagtgcagtgtactcatggcataaaaccagtggtgactacacagctcataataaatggcagtttagcaactaaaaatgt
tactgtaagaagtaaaaactttgcagacattattctagtacaattctcagagggagtcaatatgacttgtattagaccaggaaacaatacagtgggaaatgtacaactaggaccaggaatgactttttataacataccaaagatagtaggag
atgtaagagaagcacattgtaacatctcaaaactgacatgggagaaacagagaaaatacactttagagataataaaaaaggaagcaaacctgacaaaggtagagttaattccaaatgcaggaggagacccagaagtggtaaatatga
tgcttaattgtggaggagaattttttattgtaatacaattcccctatttaacatgacctacaacaataccgacaacaccactatcacacttaagtgtagaataagacaaattgtaaatcagtggatgagagtaggaaaaggaatctttgcccca
ccaatcaaaggtgtgctaagttgtaactcaaatataacgggaatgattcttgacataagcataagcgcagtcaataacgatagtaggaatataacagtgatgcctacaggaggagatatgacggctttatggaaaaatgaattacataagt
ataaggtggtcagcatagaacctataggagtggcaccaggtaaggccaaaaggcatacagtgaaaagagaaaaaagagcagccttcggactaggtgcactgttcctggggtttcttggagcagcaggaagcactatgggcgcagc
atcagtagtgctgacggtacaggcccgacaattattgtcagggatagtgcaacagcagaataatctgctaagagcaatagaggcccaacagcatttattgcaactatcagtttggggcattaaacagcttcaagccagagtacttgctgt
ggaaagataccttaaggatcagcagatcctaggtctgtggggctgctcaggaaaaacaatttgttataccactgtgccttggaatgatacctggagtaacaacctctcctatgatgctatttggggcaatctaacttggcaagaatgggac
agaaaagtaagaaactattcaggtactatttttagtcttatagaacaagcacaagaacagcagaatacaaatgaaaaatcactcttggaattggatcaatggtcaagtctatggaactggtttgatattaccaactggctgtggtatataaaaa
tatttttaatagtagtagcaagcttagtaggaatcagaattgtaggtgtgatattttcactagtagcaaaagttaggcagggttattctcccctctcgttacagacccttttcccaacaactcgggaaccagacaggccagaaggaacagaag
aaggcgctggaaaaacagacaacgtcagatcaacgagattagtcagcggattcttagcacttgtctgggaagatcttcggaacctcctgatcttcctctaccaccgattggaggacttactattaattctgaggaggacagtacaaatcct
gggacagaacatcaacaagggactgcaactgttgaatgaactcagagcacgctgctgggggtaatcgcttactgggcaagggagctaaaagttagtgctacaagcttattagatacaacagcaatagcagtagctgaaggaactga
tagaattatagaattaacaagaaggctctttctaggtattatacacataccaagaagaattaggcaaggcctagagagaagcttattataa'
```

14. 'SIVmon Cercopithecus Monkeys' 'AY340701'

```
'atgagagaaggagacatgactcaggggttgacacagaatctgaatcagagcaacatgaggaaactcatggctttgtcaaccctgtgtttaatgatgactttggtgaatgggtctaactggaccaccgtctactacggtgttccagtatgg
aaaccggcaactcctcctctcttttgtgcctctgatccaaattacgggtctaaagaggcagggaacaattggctggcgtcttcctgcctcccgacagaccctaccccgcaatctctatatttgaacatcactgaggagttcaatgcttatcaa
aactacatggtggaagaaatggtagaggatatgaagtctctgttctcgcaggctttaaagccttgtgtgaaattaacacccatgtgtgttaggatgctctgtgttgaagttaacaccgtttcgaatgccagtaccactcctgcccctagtacc
cccacaccctggggaaattggggagggaatggaacaggacagccagtatataattgctcatttaaccagactacagaattcagggataaaaagaagcaaatgtatagcttattttggaaggaagacataatgaaggaggaaggaagc
aatggcagtcactattacatcctgaattgcaacacatcatacataacccaagcttgtgagaagtccaattatgagccagttcctttgcattattgtgcacccccaggctatgctttattgagatgtgatgaccccgcctttacgggacaaggg
agttgttctaatgtttctgcagtaacatgtactcatgctatccaacctatagtagctacttggttccagttgaacagtacgggtaatgctcccaatacaacagtaatgatgaataagcagaaaaatgagtctatagtagttagattggctaagca
cctacatgtcaacattacttgcattaggccaggaaacaaaactattaggaatttacagataggagcaggcatgacattctattcgcaacttatagtgggaggtaatactcgtaaagcctactgcaaagttaataagacccaatgggagact
gccctacaggcagttcacgaggcagtaaaaactgagtgggaaaagaaaaacaatggaaccaatgtgaccacaatatcttggagatttcaaccccaaggggacaaggaggtccagactcactggtttaactgccagggagaattctttt
attgtaatgtctcagctttattcattaatagaagaaccaacaaaacggacgggatatctccattcgacgtgaataacaagcccaacaccacgtatcatggtgggtggctagcatgtaccataagacaaatggtgacacaatggggatatgt
gtcaaagagtatctacctacctccacggaaaggccatgtgcagtgtacatctaatatcactgccctcctgattacaggggagttgtatcaaaacaacgtgacccttgttccctcggcacaggtgagtgactcatggagatcagagctcagt
cgatataaggtggtggaaattgaccccttatctatggcacctacaacagcgcagagaaggactggggtacacagggaaaagagagcaattactttgggtatggcctttctgggcttcctcagcacggcaggaggcacaatgggcgca
gcggcgacagccctgacggtacagtctcggtcattactggctgggatagtgcagcagcaggagaacctgctgagggccgtaacagcccaacagagcttactacagcttactgtttggggagtcaagcagctacaggcccgcctgac
ggctgtagagaagttcataaaggaccaaacattgctaaatgcatggggatgtgccaataaagccgtgtgtcacacaacagtgccgtggaacaattcttgggcaaaaggtcacttccctgagtgggacaatatgacatggcaacagtgg
agtgagttagtagataatgacacaatgaccattcagcagctcttggaagctgcgcaagagcagcaaggcaaaaaccaacatgagttaatgaagccgggacaatgggacttcctgtggaattggtttgacatctccaaatggttgtggta
catcaaaatattcattatagtagtagcagctttgataggcttaagaatacttatgttcatactaggagttatcagtaggttagggcagggatactctcttctttcttctcagatccctatcccatcccacgcgggacagccaacgcccgacgga
acaggcgcaggcggtggcgacggcagcaactccagatcgccagcatatctgaaaggattttttaccattatctgggaagatctcaggaaccttgtcctctggacataccagatcttgaaagactcagtattagtgatctaccggatcctc
cagagagtgtcccagaggctgccaccccctcctgcacatacgcctgctccaactgtgggaaagccttagacgcttgctagcctactgccagtatgggatccaagaactccaagcagcagtcacctctctcctcgacgcccttgctaggtt
cacaatcgtctggacggatgcgctacttcatgctggaggacgactatggagagcaatcgtggctatccccagacgcatccgacagggagcggaaatattccttaactga'
```

15. 'SIVcpzTAN1 Chimpanzee' 'AF447763'

```
'atgaagaatttaattggaataactttgatcctcataattacaatcctagggattggatttagcacttattacaccacagtgttttatggagtacctgtttggaaagaggcccaaccaaccttgttttgtgcctctgatgctgatattactagtagag
ataaacacaacatatgggcaacacataactgtgtgcctttagatcccaatccttatgaagtaaccctagccaatgtgtcaataaggtttaatatggaagaaaattacatggtgcaagagatgaaagaagatatattatcactttttcaacaga
gttttaagccttgtgtaaaattaacaccattttgcataaagatgacatgtacaatgactaataccacaaataaaaccctgaattcggcaacaacaaccttaacaccaacagtaaatttgagttctatacctaactatgaggtgtataattgttcatt
taatcagacaactgagtttagagataagaaaaaacaaatatattccttgttttatagagaagatattgtaaaagaggatggtaacaataatagttattatttacataattgcaatacctcagtcattactcaagaatgtgataaatctacttttgaac
caattcccatcagatactgtgctccagcaggctttgccctgttaaaatgtagagatcagaatttcacagggaaaggacaatgctccaatgtctcagtagttcactgtacacatgggatttatcctatgatagccacagcattacacttaaatg
ggtccctggaagaagaagaaacaaaagcttactttgttaatacctcagttaatacacccttattagtaaaatttaatgtatcaataaatttaacgtgtgaaagaacaggaaacaatacaagaggtcaagtacagataggtccaggtatgacct
tttataatatagaaaatgtagtaggggacaccaggaaagcttattgttcagtcaatgcaacaacatggtacaggaacttagattgggctatggctgccataaacacaaccatgagggccagaaatgaaacggtacaacaaacgttccaat
ggcagagggatggagaccctgaggtcactagcttctggttcaattgtcaaggagaattcttttactgtaatctcacaaattggactaatacctggacagctaatagaaccaataatactcatggtactcttgttgcaccatgcagactgagg
cagatagtaaatcattggggtatagtgtcaaaaggggtttaccttcccccaaggaggggaacagtaaaatgtcactcaaacatcacaggacttatcatgacagcagaaaaagacaacaataatagttataccccccaattttctgctgtag
tagaagactattggaaagtagaattagcaagatataaagtggtggaaattcagcccttgtcagtggctccaaggccaggaaaaaggcctgaaattaaggccaatcatactaggtcaagaagagatgtgggcataggactgttgtttcttg
gatttcttagtgcagcaggaagtacaatgggcgcagcgtcaatagcgctgacggcacaggccagaggattactctctggtattgtacagcagcaacaaaacctgcttcaggccatagaagcgcaacaacacttgttgcagctctctgta
tggggcattaagcagctccaggccagaatgcttgcagtagagaaatacataagagaccaacagctcctaagcctctggggatgtgctaacaaattggtgtgtcacagtagtgtgccatggaacctcacctgggctgaagattctacaaa
gtgcaatcacagtgatgcaaagtactatgactgtatatggaacaatttgacttggcaggaatgggatcgattagtagaaaactctacaggaaccatatactccctgttagagaaagcacaaacacaacaggagaaaaacaaacaagagt
tgttagaattagacaaatggagcagtctttgggattggtttgatataacacaatggctgtggtatataaaaatagctataatcatagtagcaggattagtaggacttagaattctcatgtttatagttaatgtagttaagcaagttaggcagggtt
atacacccctattttcacagatccctacccaagcggagcaggatccagaacagccaggaggaatcgcaggaggaggtggaggcagagacaacatcaggtggacgccctcgccagcaggattcttcagtatcgtctgggaggacct
caggaacctcctcatctggatataccagacctttcaaaacttcatctggatcctctggatcagcctgcaagcactgaaacaggggataatcagcttggcacacagcctagtaatagtgcatagaactatcatagtaggagttagacagatc
attgagtggagcagtaatacttatgctagcttaagagttttgctaatacaagccatagacagacttgctaactttacagggtggtggacagatttaatcatagaaggagtggtttacatagccaggggaatcagaaatattcctagaagaatt
agacagggtctggaactagccttaaattaa'
```

16. 'SIVsmSL92b Sooty Mangabey' 'AF334679'

```
'atggcgtgtcctggacttcacctgcttatagacatcttgtttttaagtgtgttagggacctggtgtgcacagtatgtaacaatcttttatggtattcctgcatggaggaacgctacgatccccctcttctgtgcgacccagaatagagacacat
gggggaccgttcagtgcttgccagataatggagactattcagaattggcccttaatgttacagaggccttcgatgcttgggataatacagtaactgaacaagcaatagaagatgtttggaatctctttgaaacatctattaaaccttgtgtaa
agttaactccattatgtattgctatgaaatgtaataaaaatgagacagacagatgggggttgacaagagcagctactactactagctcaccaactactactagccccttaactgctgctagcccatcaggagaggaaatcgttaacgacac
tatgtcttgtacaaagaacaacaattgctctggcatagagcaggaaccaatgataggttgtcaatttaacatgacaggactaaaaagagaccagaaaaggcagtacaatgaaacctggtactccagagatctagtctgtgagcaaggag
gaaatgaaagcagtagatgttacatgaatcattgcaatacaagtgttatacaagagtcatgtgacaagcactattgggatgccatcagatataggtactgtgcaccaccaggctatgctttgcttagatgtaatgatacaaattattcaggctt
tgcacctaattgtagtaaggtagtagtatcatcatgtacaaggatgatggaaacacagacttctacatggtttggctttaatggtactagagcagagaatagaacatatatatattggcatggcggaagtaatagaacaataattagcttaaat
aagtattataatttgacaataaagtgcagaagaccaggaaataagacagtcctaccagtcaccattatgtcaggtttggtctttcattcgcaaccaataaatgagaggccaagacaagcatggtgctggtttggaggaaaatggagagaa
gctatgcaggaggttaagaaaaccatagtcaaacaccccaggtatactggaaccaatgatactaggaaaattaatctaacggccccgggaggaggggatccggaagtcacattcatgtggacaaattgcagaggagaattcctttact
gtaaaatgaattggttcctcaattgggtagaggataggaatacgagcagcccaaggtggacaactcaaaccaaaaaggaacagcacaaaagaaattatgtaccttgccacatcaggcaaataattaatacgtggcacagggtaggaa
aaaatgtctatttgcctccaagggaaggagacttaacttgtaactccacagtgaccagcctaatagcaaatattgattggattgacaacaatgagaccaatattaccatgagtgcagaagtggcagaattgtatcgattggaattgggggat
tataaattggtagagatcactccaattggcatggctcccacacatgtgaaaaggtacaccacaagtacctcaaagaataaaagaggagtctttgtgctagggttcttgggttttctcgcgacggcaggttctgcaatgggcgcagcgtcg
ctaacgctgactgctcagtcccggactttattggctgggatagtgcaacaacagcaacagctgttggacgtggtcaagagacaacaagaattgttgcgactgaccgtctggggaactaaaaatcttcagactagagtcaccgccatcga
gaaatacctaaaggatcaggcacagctaaattcatggggatgtgcatttaggcaagtctgccacactactgtactatggccaaatgacagtcttgtcccagactggaacaatatgacatggcaagaatgggagaaaaaggttgaattcct
agaggcaaatataactcaaatgttggaggaagcacgactacagcaagagaaaaacatgtatgaattacagaaattgaatagctgggatgtgtttggaaattggtttgaccttacctcctgggtaagatacatacagtatggagtctttctag
tcataggaatagtattgttaagaatagttatctatgtagtacaaatgttaagtaggttaaggcagggttataggcctgttttctcctccccccccatcttatcatcagcagatccatatccagcgggaccaggaactgccagccaaagaaggaa
cagaaggagaaggtggaaacggaggtggatacagattgtggccttggcagatagaatatattcatttcctgatccgccagctgatacggatattgacttggctatacaacaatttgactcgcttagcatccagagcctaccagaacctcc
aacaactgtgccagagactctcagagatcagtcagccgattagagagcttgtcagaagagaagcaggctatatccgttatgggtggaattacttcatcgaagcctgccaggaagcatggagatctgcgcaagaagctattgtcggcgc
gtggggactcatatgggagactctgggacgggttggaagagggatcgcagcaatcccaaggaggatcagacagggccttgaactcatgcttaactag'
```

4.5.4 Multiple Sequence Aligment Results For GAG

The following command for progressive multiple sequence alignment is used via MATLAB;

```
>> gagaligned=multialign(gag.'terminalGapAdjust'.true)
```

Table 1. Only the aligned result of the first 99 nucleotide sites were listed for demonstration.

LOCUS NAME	VERSION	MULTIPLE SEQUENCE ALIGNMENT
'HIV-1 (Zaire)'	**'K03454'**	**'atgggtgcgagagcgtcag---tattaagcgggggaaaattagataaatgggaaaaaattcggttacggccaggaggaaagaaaaaatatagactaaaa**
'HIV1-NDK (Zaire)'	**'M27323'**	**'atgggtgcgagagcgtcag---tattaagcgggggaaaattagatacatgggaaagaattcggttacggccaggaggaaagaaaaaatatgcactaaaa**
'HIV-2 (Senegal)'	**'M15390'**	**'atgggcgcgagaaactccg---tcttgagagggaaaaaagcagatgaattagaaagaatcaggttacggcccggcggaaagaaaaagtacaggctaaaa**
'HIV2-MCN13'	**'AY509259'**	**'atgggcgcgagaaactccg---tcttgaaagggaaaaaagcagacgaattagaaacaattaggttacggcccggcggaaagaaaaaatacaggctaaag**
'HIV-2UC1 (IvoryCoast)'	**'L07625'**	**'atgggcgcgagaagctccg---tcttgtcagggaagaaaacagatgaattagaaaaagttaggttacggcccggcggaaagaaaaggtattgtctaaaa**
'SIVMM251 Macaque'	**'M19499'**	**'atgggcgcgagaaactccg---tcttgtcagggaagaaagcagatgaattagaaaaaattaggctacgacccggcggaaagaaaaagtacatgttgaag**
'SIVAGM677A Green monkey'	**'M58410'**	**'atgggcgggggtcactcag---cactgtcagggagaagcctcgacacgttcgagaagattaggctacgtccgaacgggaaaaagaagtaccaaattaaa**
'SIVlhoest L''Hoest monkeys'	**'AF075269'**	**'atgggatcgggtaactctg---tcctcagtaggcaaattgaaaaagatttttgtagtgtcaggctaagacctgggagtaaaaagacttatcaaaaaaga**
'SIVcpz Chimpanzees Cameroon'	**'AF115393'**	**'atgggtgcgagagcgtcag---tattaacagggggacgattagatgcttgggaaaaaattaggctaaggccagggggaaagaaaaaatatatgatgaaa**
'SIVmnd5440 Mandrillus sphinx'	**'AY159322'**	**'atgggcgcgagcgcgtcag---gacttaggggagaaaaattggatgaactggaaaagattaggttacggccctccggaaagaaaaagtaccagttaaaa**
'SIVAGM3 Green monkeys'	**'M30931'**	**'atgggggcggctacctcag---cactaaataggagacaattagacaaatttgagcatatacgacttcgcccgaccggaaagaaaaagtaccaaattaaa**

'SIVMM239 Simian macaque'	**'M33262'**	**'atgggcgtgagaaactccg---tcttgtcagggaagaaagcagatgaattagaaaaaattaggctacgacccaacggaaagaaaaagtacatgttgaag**
'CIVcpzUS Chimpanzee'	**'AF103818'**	**'atgggtgcgagagcgtcag---ttctgacagggggacgattagatgcttgggaaaagattcggcttagacctgggggaaagaaaaagtatatgatgaaa**
'SIVmon Cercopithecus Monkeys'	**'AY340701'**	**'atgggggcgaggcactcggcgatgcttagtggtactaaattagacaagtatgagaaggtgcggcttagaccgcgcggtaagaagaagtacctgattaag**
'SIVcpzTAN1 Chimpanzee'	**'AF447763'**	**'atgggtgcgagagcgtcag---tgttgaggggagataagctggatacatgggaatccataaggcttaaatccagaggcaggaaaaaatatttaataaaa**
'SIVsmSL92b Sooty Mangabey'	**'AF334679'**	**'atgggcgcgagaggctccg---tcttgtcagggaaaaaagcagatgaattagaaaaagttaggttacggcccggcggaaggaaaaagtacatgcttaag**

4.5.5 Distance Tables of GAG protein coding gene

Table 2. P-distance (gag)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	0.00000															
2	0.04685	0.00000														
3	0.38277	0.37855	0.00000													
4	0.39202	0.39355	0.07014	0.00000												
5	0.38806	0.38885	0.17145	0.17149	0.00000											
6	0.39217	0.39298	0.18079	0.18425	0.18031	0.00000										
7	0.40312	0.40442	0.35477	0.35427	0.34926	0.35199	0.00000									
8	0.41720	0.41961	0.41614	0.42109	0.42170	0.40588	0.41319	0.00000								
9	0.27058	0.27150	0.36735	0.37571	0.37811	0.37678	0.38783	0.42617	0.00000							
10	0.38862	0.38885	0.35266	0.35554	0.35255	0.35057	0.35644	0.43074	0.38132	0.00000						
11	0.39518	0.39130	0.34553	0.34904	0.36003	0.34953	0.28029	0.43064	0.39887	0.35418	0.00000					
12	0.39375	0.39457	0.18783	0.19133	0.18526	0.00986	0.35385	0.40988	0.38380	0.35843	0.35074	0.00000				
13	0.26514	0.26166	0.38717	0.40214	0.38964	0.39134	0.38092	0.42777	0.18717	0.38920	0.38630	0.39532	0.00000			
14	0.41500	0.41744	0.39944	0.39817	0.40028	0.39662	0.38717	0.44615	0.40510	0.39483	0.37692	0.40294	0.40385	0.00000		
15	0.30934	0.30341	0.38338	0.39256	0.38456	0.38920	0.38402	0.41484	0.32584	0.38947	0.39889	0.39343	0.30922	0.40636	0.00000	
16	0.39301	0.39455	0.22492	0.21860	0.22813	0.20590	0.36339	0.42157	0.38363	0.36203	0.36764	0.21242	0.39588	0.39776	0.40196	0.00000

Table 3. Alignment score (gag)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	0.00000															
2	0.00711	0.00000														
3	0.47470	0.46430	0.00000													
4	0.49791	0.50181	0.01594	0.00000												
5	0.48791	0.48990	0.09524	0.09528	0.00000											
6	0.49831	0.50038	0.10590	0.10999	0.10533	0.00000										
7	0.52653	0.52993	0.40780	0.40663	0.39522	0.40142	0.00000									
8	0.56393	0.57046	0.56108	0.57451	0.57618	0.53375	0.55316	0.00000								
9	0.23721	0.23882	0.43722	0.45735	0.46321	0.45996	0.48733	0.58846	0.00000							
10	0.48931	0.48990	0.40296	0.40957	0.40270	0.39820	0.41163	0.60113	0.47111	0.00000						
11	0.50599	0.49611	0.38682	0.39473	0.41997	0.39583	0.25454	0.60087	0.51547	0.40644	0.00000					
12	0.50232	0.50441	0.11431	0.11861	0.11120	0.00032	0.40567	0.54433	0.47725	0.41624	0.39857	0.00000				
13	0.22777	0.22184	0.48567	0.52395	0.49189	0.49618	0.47012	0.59288	0.11350	0.49079	0.48350	0.50633	0.00000			
14	0.55801	0.56459	0.51696	0.51368	0.51913	0.50969	0.48567	0.64493	0.53170	0.50508	0.46031	0.52606	0.52842	0.00000		
15	0.31004	0.29827	0.47622	0.49929	0.47916	0.49079	0.47781	0.55758	0.34399	0.49146	0.51553	0.50151	0.30979	0.53501	0.00000	
16	0.50044	0.50436	0.16390	0.15483	0.16862	0.13736	0.42785	0.57581	0.47684	0.42466	0.43791	0.14620	0.50777	0.51261	0.52349	0.00000

Table 4. Hasegawa (gag)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	0.00000															
2	0.05368	0.00000														
3	1.14095	1.07420	0.00000													
4	1.48551	1.73390	0.08075	0.00000												
5	1.27535	1.30356	0.24595	0.24629	0.00000											
6	1.51237	1.61249	0.26190	0.26638	0.26660	0.00000										
7	1.21318	1.18499	0.83418	0.83704	0.80556	0.82151	0.00000									
8	1.07549	1.06001	1.09045	1.05992	1.05201	1.20293	1.10736	0.00000								
9	0.47197	0.47236	0.95364	1.04116	1.07317	1.06060	1.26289	1.03954	0.00000							
10	1.28092	1.28833	0.82819	0.84439	0.82874	0.80952	0.84894	1.02016	1.11676	0.00000						
11	1.62666	1.42097	0.77636	0.79748	0.87070	0.81179	0.51948	1.01803	1.34783	0.83537	0.00000					
12	1.80745	1.84647	0.27318	0.27790	0.27470	0.01029	0.83199	1.15174	1.17129	0.85928	0.81795	0.00000				
13	0.45579	0.44521	1.24930	1.24940	1.33733	1.44083	1.11131	1.03840	0.27244	1.31647	1.22723	1.63029	0.00000			
14	1.08780	1.06599	1.32725	1.38037	1.29519	1.47598	1.25471	1.00293	1.20093	1.70528	1.05085	1.23439	1.22224	0.00000		
15	0.61660	0.59178	1.16017	1.54860	1.18618	1.32802	1.16840	1.09781	0.69414	1.31930	1.34781	1.70320	0.62216	1.17851	0.00000	
16	1.61140	1.83470	0.35943	0.34766	0.36494	0.31688	0.89488	1.06181	1.15903	0.88624	0.93690	0.32905	1.54204	1.39052	1.26178	0.00000

Table 5. Jukes-Cantor (gag)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	0.00000															
2	0.04838	0.00000														
3	0.53556	0.52700	0.00000													
4	0.55469	0.55791	0.07364	0.00000												
5	0.54645	0.54808	0.19466	0.19471	0.00000											
6	0.55501	0.55672	0.20687	0.21143	0.20623	0.00000										
7	0.57833	0.58114	0.48046	0.47950	0.47007	0.47520	0.00000									
8	0.60939	0.61484	0.60701	0.61822	0.61962	0.58431	0.60043	0.00000								
9	0.33562	0.33706	0.50471	0.52128	0.52611	0.52343	0.54596	0.62990	0.00000							
10	0.54760	0.54808	0.47647	0.48192	0.47625	0.47253	0.48362	0.64054	0.53261	0.00000						
11	0.56136	0.55320	0.46312	0.46967	0.49050	0.47057	0.35097	0.64032	0.56918	0.47934	0.00000					
12	0.55832	0.56005	0.21620	0.22089	0.21278	0.00993	0.47870	0.59308	0.53766	0.48743	0.47284	0.00000				
13	0.32716	0.32180	0.54459	0.57619	0.54972	0.55326	0.53179	0.63361	0.21532	0.54881	0.54281	0.56163	0.00000			
14	0.60446	0.60994	0.57041	0.56770	0.57221	0.56441	0.54459	0.67766	0.58261	0.56060	0.52372	0.57794	0.57989	0.00000		
15	0.39885	0.38882	0.53681	0.55582	0.53923	0.54881	0.53813	0.60410	0.42746	0.54937	0.56923	0.55766	0.39864	0.58535	0.00000	
16	0.55677	0.56001	0.26739	0.25842	0.27199	0.24071	0.49699	0.61931	0.53732	0.49437	0.50528	0.24975	0.56282	0.56682	0.57582	0.00000

Table 6. Kimura (gag)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	0.00000															
2	0.04881	0.00000														
3	0.54097	0.53135	0.00000													
4	0.56112	0.56474	0.07487	0.00000												
5	0.55166	0.55350	0.19805	0.19804	0.00000											
6	0.55751	0.55988	0.21101	0.21628	0.20908	0.00000										
7	0.59044	0.59530	0.48838	0.48535	0.47562	0.48116	0.00000									
8	0.61599	0.62198	0.61135	0.62319	0.62599	0.58623	0.60591	0.00000								
9	0.34357	0.34588	0.50726	0.52433	0.52920	0.52543	0.55268	0.63460	0.00000							
10	0.55832	0.55958	0.48177	0.48820	0.48118	0.47940	0.49046	0.64654	0.53736	0.00000						
11	0.57073	0.56249	0.47082	0.47738	0.49899	0.47488	0.35403	0.64704	0.57523	0.48551	0.00000					
12	0.56122	0.56367	0.22126	0.22675	0.21610	0.00995	0.48523	0.59538	0.54062	0.49569	0.47752	0.00000				
13	0.33509	0.32990	0.54904	0.58277	0.55497	0.55625	0.53616	0.63726	0.22020	0.55464	0.54657	0.56542	0.00000			
14	0.61174	0.61927	0.57623	0.57309	0.57943	0.56923	0.54770	0.68099	0.58727	0.57059	0.52821	0.58389	0.58451	0.00000		
15	0.40319	0.39371	0.54034	0.56014	0.54258	0.55152	0.54298	0.60882	0.43006	0.55845	0.57504	0.56100	0.40158	0.59100	0.00000	
16	0.56223	0.56602	0.27192	0.26210	0.27717	0.24470	0.50651	0.62325	0.54249	0.50293	0.51308	0.25446	0.56851	0.57512	0.58018	0.00000

Table 7. Nei-Timura (gag)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	0.00000															
2	0.05357	0.00000														
3	0.87682	0.86545	0.00000													
4	0.92474	0.93301	0.08047	0.00000												
5	0.90312	0.90767	0.24281	0.24313	0.00000											
6	0.93516	0.93838	0.25828	0.26287	0.26310	0.00000										
7	0.97605	0.97208	0.73469	0.73973	0.71891	0.72854	0.00000									
8	1.07063	1.09116	1.09085	1.11707	1.10363	1.01537	1.06802	0.00000								
9	0.45475	0.45471	0.81371	0.84915	0.85757	0.86646	0.89319	1.14411	0.00000							
10	0.88957	0.89372	0.74037	0.74589	0.74472	0.72490	0.75946	1.23116	0.88753	0.00000						
11	0.93970	0.91671	0.69767	0.71262	0.75642	0.72858	0.50032	1.17872	0.95638	0.73836	0.00000					
12	0.94118	0.94441	0.26907	0.27390	0.27092	0.01029	0.73433	1.03736	0.89532	0.75219	0.73206	0.00000				
13	0.44052	0.43069	0.90450	0.98759	0.90933	0.93752	0.87140	1.16412	0.26920	0.91639	0.90172	0.95531	0.00000			
14	1.07575	1.11521	0.98285	0.96900	0.97259	0.96103	0.92890	1.68479	1.00820	0.93151	0.86041	0.99773	1.02354	0.00000		
15	0.58222	0.56239	0.88342	0.92783	0.88696	0.91903	0.88421	1.07553	0.64652	0.90664	0.96412	0.93856	0.58816	1.02572	0.00000	
16	0.92895	0.94338	0.35173	0.34084	0.35663	0.31108	0.76649	1.12015	0.87925	0.77319	0.79857	0.32257	0.94232	0.94962	0.97810	0.00000

Table 8. Tajima-Nei (gag)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	0.00000															
2	0.04843	0.00000														
3	0.477255	0.47025	0.00000													
4	0.491630	0.49403	0.07396	0.00000												
5	0.485572	0.48647	0.19047	0.19058	0.00000											
6	0.491199	0.49223	0.20299	0.20817	0.20074	0.00000										
7	0.511904	0.51441	0.43521	0.43270	0.42580	0.42867	0.00000									
8	0.529734	0.53360	0.52658	0.53558	0.53626	0.50830	0.52173	0.00000								
9	0.320707	0.32199	0.45103	0.46385	0.46722	0.46615	0.48308	0.54531	0.00000							
10	0.490379	0.49125	0.43077	0.43479	0.43020	0.42868	0.43781	0.55113	0.47395	0.00000						
11	0.497677	0.49109	0.42170	0.42631	0.44389	0.42510	0.32735	0.55191	0.50053	0.43256	0.00000					
12	0.493614	0.49471	0.21257	0.21790	0.20726	0.00994	0.43167	0.51491	0.47763	0.44121	0.42717	0.00000				
13	0.313168	0.30830	0.48154	0.50615	0.48605	0.48773	0.47091	0.54457	0.21169	0.48646	0.47988	0.49435	0.00000			
14	0.534980	0.53929	0.50605	0.50370	0.50703	0.50068	0.48541	0.58102	0.51206	0.49832	0.47173	0.51179	0.50988	0.00000		
15	0.367833	0.36042	0.47649	0.49087	0.47769	0.48469	0.47750	0.52520	0.38890	0.48935	0.50150	0.49168	0.36575	0.51548	0.00000	
16	0.494487	0.49734	0.25699	0.24842	0.26148	0.23293	0.45026	0.53577	0.47808	0.44775	0.45708	0.24185	0.49631	0.50438	0.50705	0.00000

4.5.6 Translated GAG protein-coding gene (Amino Acid Sequence) (designated by AAGAG)

1. 'HIV-1 (Zaire)"K03454'

```
'MGARASVLSGGKLDKWEKIRLRPGGKKKYRLKHIVWASRELERYALNPGLLETSEGCKQIIGQLQPAIQTGTEELRSLYNTVATLYCVHKGIDVKD
TKEALEKMEEEQNKSKKKAQQAAADTGNNSQVSQNYPIVQNLQGQMVHQAISPRTLNAWVKVIEEKAFSPEVIPMFSALSEGATPQDLNTMLNTV
GGHQAAMQMLKETINEEAAEWDRLHPVHAGPIAPGQMREPRGSDIAGTTSTLQEQIAWMTSNPPIPVGEIYKRWIIVGLNKIVRMYSPVSILDIRQGP
KEPFRDYVDRFYKTLRAEQASQDVKNWMTETLLVQNANPDCKTILKALGPQATLEEMMTACQGVGGPSHKARVLAEAMSQATNSVTTAMMQRG
NFKGPRKIIKCFNCGKEGHIAKNCRAPRKKGCWRCGKEGHQLKDCTERQANFLGRIWPSHKGRPGNFLQSRPEPTAPPAESFGFGEEITPSQKQEQKD
KELYPLTSLKSLFGNDPLSQ*'
```

2. 'HIV1-NDK (Zaire)"M27323'

```
'MGARASVLSGGKLDTWERIRLRPGGKKKYALKHLIWASRELERFTLNPGLLETSEGCKQIIGQLQPSIQTGSEEIRSLYNTVATLYCVHERIEVKDTK
EAVEKMEEEQNKSKKKTQQAAADSSQVSQNYPIVQNLQGQMVHQAISPRTLNAWVKVIEEKAFSPEVIPMFSALSEGATPQDLNTMLNTVGGHQA
AMQMLKETINDEAAEWDRLHPVHAGPVAPGQMREPRGSDIAGTTSTLQEQIAWMTSNPPIPVGEIYKRWIILGLNKIVRMYSPVSILDIRQGPKEPFR
DYVDRFYKTLRAEQASQDVKNWMTETLLVQNANPDCKTILKALGPQATLEEMMTACQGVGGPGHKARVLAEAMSQVTGSATAVMMQRGNFKG
PRKSIKCFNCGKEGHTAKNCRAPRKKGCWKCGREGHQMKDCTERQANFLGKIWPSHKGRPGNFLQSRPEPTAPPAESFGFGEEITPSQKQEQKDKEL
YPLASLKSLFGNDPSSQ*'
```

3. 'HIV-2 (Senegal)"M15390'

```
'MGARNSVLRGKKADELERIRLRPGGKKKYRLKHIVWAANKLDRFGLAESLLESKEGCQKILTVLDPMVPTGSENLKSLFNTVCVIWCIHAEEKVKD
TEGAKQIVRRHLVAETGTAEKMPSTSRPTAPSSEKGGNYPVQHVGGNYTHIPLSPRTLNAWVKLVEEKKFGAEVVPGFQALSEGCTPYDINQMLNC
VGDHQAAMQIIREIINEEAAEWDVQHPIPGPLPAGQLREPRGSDIAGTTSTVEEQIQWMFRPQNPVPVGNIYRRWIQIGLQKCVRMYNPTNILDIKQGP
KEPFQSYVDRFYKSLRAEQTDPAVKNWMTQTLLVQNANPDCKLVLKGLGMNPTLEEMLTACQGVGGPGQKARLMAEALKEVIGPAPIPFAAAQQ
RKAFKCWNCGKEGHSARQCRAPRRQGCWKCGKPGHIMTNCPDRQAGFLGLGPWGKKPRNFPVAQVPQGLTPTAPPVDPAVDLLEKYMQQGKRQ
REQRERPYKEVTEDLLHLEQGETPYREPPTEDLLHLNSLFGKDQ*'
```

4. 'HIV2-MCN13"AY509259'

```
'MGARNSVLKGKKADELETIRLRPGGKKKYRLKHIVWAANELDRFGLAESLLESKEGCQRILTVLGPLVPTGSENLKSLFNTVCVIWCIHAEEKVKDT
EGAKQIVQRHLAAETGTAEKMPNTSRPTAPPSGKGGNFPVQQVGGNYTHVPLSPRTLNAWVKLVEEKKFGAEVVPGFQALSEGCTPYDINQMLNC
VGDHQAAMQIIREIVNEEAADWDVQHPIPGPLPAGQLREPRGSDIAGTTSTVDEQIQWMFRPQNPVPVGNIYRRWIQIGLQKCVRMYNPTNILDIKQG
PKEPFQSYVDRFYKSLRAEQTDPAVKNWMTQTLLVQNANPDCKLVLKGLGMNPTLEEMLTACQGVGGPGQKARLMAEALKEALTPAPIPFAAAQ
QKRTIKCWNCGKEGHSARQCRAPRRQGCWKCGKPGHVIANCPDRQVGFLGMGPRGKKPRNFPVAQVPQGLTPTAPPVDPAVDLLENYMQQGKRQ
REQRERPYKEVTEDLLHLEQGEAPCRETTEDLLHLNSLF*'
```

5. 'HIV-2UC1 (IvoryCoast)' 'L07625'

```
'MGARSSVLSGKKTDELEKVRLRPGGKKRYCLKHIIWAVNELDRFGLAESLLESKEGCHKILTVLAPLVPTGSENLKSLFNTVCVIYCLHAEEKVKDT
EEAKKIAQRHLAADTEKMPATSRPTAPPSGGNYPVQQIAGNYVHMPLSPRTLNAWVKLVEEKKFGAEVVPGFQALSEGCTPYDINQMLNCVGDHQ
AAMQIIREIINEEAADWDQQHPIPGPLPAGQLRDPRGSDIAGTTSTVEEQIQWMYRAQNPVPVGNIYRRWIQIGLQKCVRMYNPTNILDIKQGPKEPFQ
SYVDRFYKSLRAEQTDPAVKNWMTQTLLIQNANPDCKLVLKGLGMNPTLEEMLTACQGIGGPGQKARLMAEALKEALTPAPIPFAAAQQKAGKRG
TVTCWNCGKVGHTAKQCRAPRRQGCWKCGKQGHIMSKCPERQAGFLGLGPWGKKPRNFPMTQVPQGVTPSAPPMDPAEGMTPRGATPSAPPADP
AVEMLKSYMKMGRQQRESRERPYKEVTEDLLHLNSLFGEDQ*'
```

6. 'SIVMM251 Macaque"M19499'

```
'MGARNSVLSGKKADELEKIRLRPGGKKKYMLKHVVWAANELDRFGLAESLLENKEGCQKILSVLAPLVPTGSENLKSLYNTVCVIWCIHAEEKVK
HTEEAKQIVQRHLVVETGTAETMPKTSRPTAPSSGRGGNYPVQQIGGNYVHLPLSPRTLNAWVKLIEEKKFGAEVVPGFQALSEGCTPYDINQMLNC
VGDHQAAMQIIRDIINEEAADWDLQHPQPAPQQGQLREPSGSDIAGTTSSVDEQIQWMYRQQNPIPVGNIYRRWIQLGLQKCVRMYNPTNILDVKQG
PKEPFQSYVDRFYKSLRAEQTDAAVKNWMTQTLLIQNANPDCKLVLKGLGVNPTLEEMLTACQGVGGPGQKARLMAEALKEALAPVPIPFAAAQK
RGPRKPIKCWNCGKEGHSARQCRAPRRQGCWKCGKMDHVMAKCPDRQAGFLGLGPWGKKPRNFPMAQVHQGLTPTAPPEDPAVDLLKNYMQL
GKQQRESREKPYKEVTEDLLHLNSLFGGDQ*'
```

7. 'SIVAGM677A Green monkey"M58410'

```
'MGGGHSALSGRSLDTFEKIRLRPNGKKKYQIKHLIWAGKEMERFGLHEKLLETKEGCQKIIEVLTPLEPTGSEGLKALFNLCCVIWCIHAEQKVKDTE
EAVVTVKQHYHLVDKNEKAAKKKNETTAPPGGESRNYPVVNQNNAWVHQPLSPRTLNAWVKCVEEKRWGAEVVPMFQALSEGCLSYDVNQML
NVIGDHQGALQILKEVINEEAAEWDRTHRPPAGPLPAGQLRDPTGSDIAGTTSSIQEQIEWTFNANPRIDVGAQYRKWVILGLQKVVQMYNPQKVLD
IRQGPKEPFQDYVDRFYKALRAEQAPQDVKNWMTQTLLIQNANPDCKLILKGLGMNPTLEEMLIACQGVGGPQHKAKLMVEMMSNGQNMVQVG
PQKKGPRGPLKCFNCGKFGHMQRECKAPRQIKCFKCGKIGHMAKDCKNGQANFLGYGHWGGAKPRNFVQYRGDTVGLEPTAPPMETAYDPAKKL
LQQYAEKGQRLREEREQTRKQKEKEVEDVSLSSLFGGDQ*'
```

8. 'SIVlhoest L"Hoest monkeys"AF075269'

```
'MGSGNSVLSRQIEKDFCSVRLRPGSKKTYQKRHVEWATKELDRFGLGSQLLETAEGCKKILSVCWPLYATGSKNLKALVGTVCVICCCHLGVRISD
TQEAISKVKIEPAPEAAGKKQQTGGNYPLIRENQRWVHTPLSPRTIQTWVKIVEDRGWKPETVAMFSALTEKALPDDLNVMLNAIGDHQGAMQIIKD
HIVEEGAEWDRQHPQQQPAQPGGGLRTPNATDVAGVTSTVEEQLAWTTADTPVDVGKIYKEWVIQAMEKVVRIHQPVSVMDIKQGPKEPFKEYAD
RFFKALRAEGGSHEVKEWMKEKMLVQNANPDCRLVIKALGEGASLEEMMKACQGVGGPAHKGKILAEAMASAIQQQVRQNMVQVTPLRNAQGR
FVRTGGGGPRKPLTCFNCGKPGHTARMCRQPRQEGCWNCGSKEHRFAQCPKPKGKVNFLGYGPWSSKPPGNYPLLGGAAGRIPSAPPMESAPTKAE
RALETYRTLGQQLKRQQQQVPQKCVDEPCLNSLFPDDQ*'
```

9. 'SIVcpz Chimpanzees Cameroon"AF115393'

```
'MGARASVLTGGRLDAWEKIRLRPGGKKKYMMKHLVWASRELDRFACNPGLMETAEGCEQLLRQLEPALKTGSEGLRSLFNTLAVLWCVHKRISV
EDTQQALTKLKEAVTASREQEVAQPQQQQQDSAVSRNYPVVQNAQGQLVHQPMSPRTLNAWVKVIEEKNFNPEVIPMFMALSEGATPQDVNTML
NAIGGHQGAMQVLKEVINEEAAEWDRTHPVHAGPVAPGQLREARGSDIAGTTSTLVEQIAWMTANPPVPVGEIYRRWVVLGLNKVVRMYCPVSIL
DIKQGPKEPFRDYVDRFYKTLRAEQATQEVKNWMTDTLLVQNANPDCKNILRALGPGATLEEMMTACQGVGGPAHKARVLAEAMSQLQNPTGVF
```

```
LQRGNNGKPTRKIKCFNCGKEGHLARNCRAPRRKGCWKCGQEGHQMKNCPAGERQANFLGKGWSPWSGGSKRPGNFLQGRKEPTAPPLESYGYQ
EATGKRQEEEREKEKEPLYPPLTSLKSLFGSDPSSE*'
```

10. 'SIVmnd5440 Mandrillus sphinx"AY159322'

```
'MGASASGLRGEKLDELEKIRLRPSGKKKYQLKHVIWVSKELDRFGLHEKLLESQEGCEKILSVLFPLVPTGSENLISLYNTCCCIWCVHAKVKVTDTE
EAKEKVKQRYHLVVERENAASEEEKGATATPAVRSKNYPIQVINQTPVHQGISPRTLNAWVKCIEEKKFSPEIVPMFIALSEGCIPYDLNGMLNAIGD
HQGALQIVKDVINEEAADWDLRHPPVGPMPQGVLRNPTGNDIAGTTSSIEEQIEWTTRQQDQVNVGGIYKQWIVLGLQKCVSMYNPVNILDIKQGP
KEPFKDYVDRFYKALRAERTDPQVKTWMTQTLLIQNANPDCKATLKGLGMNPTLEEMLLACQGVGGPKYKAQMMAEAMQQAQAAVMMQNSG
GPPRGPPRQPPRNPRCPNCGKFGHVLRDCRAPRKRGCFKCGDPGHLMRNCPKMVNFLGNTPWGSGKPRNFPAMPLTPSAPPMPGLEDPAEKMLLD
YMKKGQQQRAAAGAQKEEKKGPYEAAYNSLSSLFGTDQLQ*'
```

11. 'SIVAGM3 Green monkeys"M30931'

```
'MGAATSALNRRQLDKFEHIRLRPTGKKKYQIKHLIWAGKEMERFGLHERLLESEEGCKKIIEVLYPLEPTGSEGLKSLFNLVCVLFCVHKDKEVKDT
EEAVAIVRQCCHLVEKERNAERNTTETSSGQKKNDKGVTVPPGGSQNFPAQQQGNAWIHVPLSPRTLNAWVKAVEEKKFGAEIVPMFQALSEGCTP
YDINQMLNVLGDHQGALQIVKEIINEEAAQWDIAHPPPAGPLPAGQLRDPRGSDIAGTTSTVQEQLEWIYTANPRVDVGAIYRRWIILGLQKCVKMY
NPVSVLDIRQGPKEAFKDYVDRFYKAIRAEQASGEVKQWMTESLLIQNANPDCKVILKGLGMHPTLEEMLTACQGVGGPSYKAKVMAEMMQNMQ
SQNMMQQGGQRGRPRPPVKCYNCGKFGHMQRQCPEPRKMRCLKCGKPGHLAKDCRGQVNFLGYGRWMGAKPRNFPAATLGVEPTAPPPPSPYD
PAKKLLQQYADKGKQLREQRKKPPAVNPDWTEGYSLNSLFGEDQ*'
```

12. 'SIVMM239 Simian macaque"M33262'

```
'MGVRNSVLSGKKADELEKIRLRPNGKKKYMLKHVVWAANELDRFGLAESLLENKEGCQKILSVLAPLVPTGSENLKSLYNTVCVIWCIHAEEKVK
HTEEAKQIVQRHLVVETGTTETMPKTSRPTAPSSGRGGNYPVQQIGGNYVHLPLSPRTLNAWVKLIEEKKFGAEVVPGFQALSEGCTPYDINQMLNC
VGDHQAAMQIIRDIINEEAADWDLQHPQPAPQQGQLREPSGSDIAGTTSSVDEQIQWMYRQQNPIPVGNIYRRWIQLGLQKCVRMYNPTNILDVKQG
PKEPFQSYVDRFYKSLRAEQTDAAVKNWMTQTLLIQNANPDCKLVLKGLGVNPTLEEMLTACQGVGGPGQKARLMAEALKEALAPVPIPFAAAQQ
RGPRKPIKCWNCGKEGHSARQCRAPRRQGCWKCGKMDHVMAKCPDRQAGFLGLGPWGKKPRNFPMAQVHQGLMPTAPPEDPAVDLLKNYMQL
GKQQREKQRESREKPYKEVTEDLLHLNSLFGGDQ*'
```

13. 'CIVcpzUS Chimpanzee"AF103818'

```
'MGARASVLTGGRLDAWEKIRLRPGGKKKYMMKHLVWASRELERFACNPGLMETADGCLQLLKQLEPALKTGSEGLRSLFNTLAVLWCVHSRVT
VEDTQQALVKLKEVVQLQKQQEEKEQQQQEASGSNIGSSNYPVIQNAQGQMVHQAMSPRTLNAWVKAVEEKAFNPEVIPMFMALSEGATPQDVN
TMLNAIGGHQGAMQVLKEAINEEAAD*DRTHPLHAGPIAPGQMREPRGSDIAGTTSTLQEQVGWMTANPPIPVGDIYRRWVVLGLNKVVKMYCPV
GILDIKQGPKEPFRDYVDRFYKTLRAEQASPEVKTWMTETLLVQNANPDCKTILRALGPGATLEEMVTACQGVGGPAHKARVLAEAMCQMKNPSS
VFLQKGNAGKPGRKIKCFNCGKEGHLARNCRAPRRKGCWKCGQEGHQMKDCTAGNRQANFLGKHWSPSWSGGSKRPGNFLENRKEPTAPPIEDF
GYQEETVTQEKQGKEKEPFQLTSLKSLFGSDPSSE*'
```

14. 'SIVmon Cercopithecus Monkeys"AY340701'

```
'MGARHSAMLSGTKLDKYEKVRLRPRGKKKYLIKHIVWAAKELDRFGLSDSLLETQDGCKKILEVILPLQPTGSESIKSLFGIASVLYCIHAGIEIEDTE
QAKQQVKIRCHLAGEQGEQKAAAAAAPPTGGVPSGNYPVVRTQGGGFQHQAVEPRLLKTWVQVIEEKKFAPEVVALFQALSEGMIPYDINQLLNAI
GDHQGALQIIKDVINEEAANWDLVHPQPPQPQPNAGLGDPTGADIAGVSSTPQQQIEWITRANNPVQVSDIYRKWVILGLQRCVKMYNPVNILDIKQ
GPKEPFKDYVDRFYKCLRAEQTDQAVKNWMTQTLLVQNANPDCKLILKAMPGASLEDMLQACQGVGGPMHKSRILAEAMAGAIANMPMNMVQ
ARGPPQRKGQPKCFNCGKFGHMAKNCKAPQRRKCYNCGQPGHLAKDCPQPPKQNKGVNFLGNPFGPKKGPRNFPLTSVQPSAPTLPAGTAEPVNL
NQEIGSPKTGTPKETRRDLYPSLASLFGEDQ*'
```

15. 'SIVcpzTAN1 Chimpanzee"AF447763'

```
'MGARASVLRGDKLDTWESIRLKSRGRKKYLIKHLVWAGSELQRFAMNPGLMENVEGCWKIILQLQPSVDIGSPEIISLFNTICVLYCVHAGERVQDT
EEAVKIVKMKLTVQKNNSTATSSGQRQNAGEKEETVPPSGNTGNTGRATETPSGSRLYPVITDAQGVARHQPISPRTLNAWVRVIEEKGFNPEVIPMF
SALSEGATPYDLNSMLNAVGEHQAAMQMLKEVINEEAAEWDRAHPAHAGPQQAGMLREPTGADIAGTTSTLQEQVLWMTTPQAQGGVPVGDIYK
RWIILGLNKLVRMYSPVSILDIKQGPKEPFRDYVDRFYKTIRAEQASQPVKTWMTETLLVQNANPDCKHILKALGQGATLEEMLTACQGVGGPSHK
AKILAEAMASATAGGVNMLQGGKRPPLKKGQLQCFNCGKVGHTARNCRAPRKKGCWRCGQEGHQMKDCTTRNNSTGVNFLGKRTPLWGCRPGN
FVQNTPEKGKAQEQETAQTPVVPTAPPLEMTMKGGFSLKSIFGSDQ*'
```

16. 'SIVsmSL92b Sooty Mangabey"AF334679'

```
'MGARGSVLSGKKADELEKVRLRPGGRKKYMLKHIIWAARELDRFGSAESLLESKEGCQRILAVLAPLMPTGSENLKSLFSTVCVVWCLHAEMKVK
DTEEAKKTVQSHLVVESGTAEKLPAQSRPTAPPSGGNYPVQQVGNNYVHTPLSPRTLNAWVKLVEEKKFGAEVVPGFQALSEGCTPYDINQMLNCV
GEHQAAMQIIREIINEEAADWDLQHPRGQQPAQPAGGLREPSGSDIAGTTSTPSEQIEWMYRAQNPVPVGDIYRRWIQLGLQKCVRMYNPTNILDVK
QGPKEPFQSYVDRFYKSLRAEQTDPAVKNWMTQTLLIQNANPDCKLVLKGLGMNPTLEEMLTACQGVGGPGQKARLMAEAMKDALTGSLVAAQF
RGAAKGQGNKPIIRCFNCGKTGHSARQCRAPRRKGCWKCGEEGRIQANCPNQKAGFLGLGPWGKKPRNFPMQTTSLTPSAPPDPAARIVKEYLEKA
QREKTRRSRPYKEVTEDLLHLNSLFGEDQ*'
```

4.5.7 Distance Tables of Translated GAG protein-coding gene (Amino Acid Sequence) (designated by AAGAG)

Table 9. P-Distance (AAGAG)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	0.00000															
2	0.06627	0.00000														
3	0.46107	0.45155	0.00000													
4	0.46281	0.46362	0.06757	0.00000												
5	0.46107	0.45567	0.13000	0.12500	0.00000											
6	0.45399	0.45267	0.12698	0.12200	0.14741	0.00000										
7	0.47862	0.46516	0.42063	0.42800	0.41551	0.41235	0.00000									
8	0.54845	0.54772	0.53170	0.53608	0.52130	0.52245	0.52121	0.00000								
9	0.27800	0.27163	0.44286	0.44444	0.45510	0.44399	0.46045	0.54303	0.00000							
10	0.49798	0.49491	0.42315	0.42656	0.40719	0.39563	0.43028	0.55332	0.47984	0.00000						
11	0.47667	0.47143	0.41865	0.41200	0.40954	0.42231	0.28824	0.53737	0.48283	0.43426	0.00000					
12	0.45918	0.45791	0.13294	0.12800	0.15139	0.00986	0.41235	0.52546	0.44919	0.39842	0.42430	0.00000				
13	0.29659	0.29234	0.47336	0.47107	0.48255	0.46830	0.46436	0.55992	0.15873	0.49798	0.47465	0.47143	0.00000			
14	0.50305	0.50410	0.47686	0.47870	0.46371	0.48193	0.48600	0.54694	0.49293	0.49398	0.49004	0.48193	0.49696	0.00000		
15	0.37778	0.36789	0.49284	0.49897	0.49180	0.49184	0.48577	0.55165	0.39113	0.49696	0.48907	0.49287	0.39960	0.50000	0.00000	
16	0.46217	0.45885	0.21800	0.21169	0.20758	0.21315	0.42600	0.53252	0.45214	0.43141	0.43200	0.22112	0.46516	0.45968	0.48057	0.00000

Table 9. Gamma Distance (AAGAG)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	0.00000															
2	0.06975	0.00000														
3	0.72434	0.70060	0.00000													
4	0.72876	0.73082	0.07120	0.00000												
5	0.72434	0.71081	0.14423	0.13809	0.00000											
6	0.70663	0.70338	0.14052	0.13443	0.16601	0.00000										
7	0.76981	0.73476	0.62757	0.64443	0.61602	0.60898	0.00000									
8	0.97631	0.97389	0.92258	0.93636	0.89066	0.89414	0.89040	0.00000								
9	0.35376	0.34344	0.67946	0.68328	0.70940	0.68219	0.72278	0.95861	0.00000							
10	0.82272	0.81413	0.63330	0.64111	0.59759	0.57263	0.64971	0.99248	0.77307	0.00000						
11	0.76467	0.75092	0.62308	0.60820	0.60277	0.63138	0.37062	0.94046	0.78107	0.65903	0.00000					
12	0.71960	0.71639	0.14785	0.14176	0.17109	0.00994	0.60898	0.90331	0.69481	0.57860	0.63592	0.00000				
13	0.38466	0.37748	0.75596	0.75000	0.78032	0.74282	0.73271	1.01483	0.18053	0.82272	0.75933	0.75092	0.00000			
14	0.83711	0.84009	0.76517	0.77005	0.73105	0.77866	0.78964	0.97133	0.80864	0.81154	0.80067	0.77866	0.81986	0.00000		
15	0.53546	0.51555	0.80840	0.82552	0.80552	0.80562	0.78902	0.98691	0.56311	0.81986	0.79800	0.80848	0.58113	0.82843	0.00000	
16	0.72713	0.71876	0.26166	0.25259	0.24674	0.25467	0.63982	0.92515	0.70206	0.65235	0.65372	0.26618	0.73476	0.72084	0.77503	0.00000

Table 10. Poisson Distance (AAGAG)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	0.00000															
2	0.06856	0.00000														
3	0.61816	0.60065	0.00000													
4	0.62140	0.62291	0.06996	0.00000												
5	0.61816	0.60820	0.13926	0.13353	0.00000											
6	0.60511	0.60271	0.13580	0.13011	0.15948	0.00000										
7	0.65127	0.62579	0.54582	0.55862	0.53701	0.53162	0.00000									
8	0.79508	0.79345	0.75864	0.76805	0.73668	0.73908	0.73650	0.00000								
9	0.32573	0.31695	0.58493	0.58779	0.60716	0.58697	0.61701	0.78314	0.00000							
10	0.68911	0.68302	0.55018	0.55610	0.52287	0.50356	0.56261	0.80591	0.65362	0.00000						
11	0.64755	0.63758	0.54240	0.53103	0.52686	0.54872	0.34001	0.77084	0.65938	0.56963	0.00000					
12	0.61468	0.61232	0.14264	0.13697	0.16416	0.00991	0.53162	0.74541	0.59636	0.50820	0.55217	0.00000				
13	0.35182	0.34579	0.64124	0.63691	0.65884	0.63168	0.62429	0.82079	0.17284	0.68911	0.64368	0.63758	0.00000			
14	0.69928	0.70138	0.64791	0.65143	0.62308	0.65764	0.66553	0.79173	0.67910	0.68117	0.67342	0.65764	0.68708	0.00000		
15	0.47446	0.45869	0.67893	0.69109	0.67689	0.67695	0.66509	0.80219	0.49615	0.68708	0.67151	0.67899	0.51016	0.69315	0.00000	
16	0.62021	0.61405	0.24590	0.23787	0.23267	0.23971	0.55513	0.76040	0.60173	0.56460	0.56563	0.24989	0.62579	0.61559	0.65503	0.00000

Table 11. Alignment-Score Distance (AAGAG)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	0.00000															
2	0.00221	0.00000														
3	0.20867	0.19827	0.00000													
4	0.20814	0.20500	0.00306	0.00000												
5	0.20402	0.19795	0.00984	0.00884	0.00000											
6	0.21109	0.20524	0.01106	0.01036	0.01280	0.00000										
7	0.22774	0.21858	0.16359	0.17095	0.15910	0.16148	0.00000									
8	0.28486	0.28486	0.27997	0.27900	0.26250	0.26234	0.27566	0.00000								
9	0.05810	0.05418	0.19145	0.19272	0.19632	0.19546	0.21265	0.28638	0.00000							
10	0.24957	0.25113	0.16973	0.16927	0.16484	0.16069	0.18529	0.33555	0.24077	0.00000						
11	0.23335	0.23296	0.16395	0.16296	0.15735	0.17025	0.06893	0.28847	0.23832	0.18986	0.00000					
12	0.21597	0.21010	0.01229	0.01156	0.01379	0.51880	0.16208	0.26639	0.20003	0.16265	0.17110	0.00000				

13	0.06424	0.06018	0.21443	0.21286	0.21824	0.21547	0.21819	0.31176	0.01806	0.25519	0.23288	0.21858	0.00000			
14	0.23771	0.23999	0.22065	0.22444	0.20336	0.22247	0.23430	0.30568	0.24466	0.24298	0.23022	0.22207	0.23671	0.00000		
15	0.13834	0.13177	0.25016	0.24915	0.23738	0.24509	0.24234	0.31559	0.15270	0.27340	0.25438	0.24671	0.16206	0.25225	0.00000	
16	0.21405	0.20870	0.03662	0.03485	0.03489	0.03607	0.16767	0.27049	0.20096	0.18427	0.16801	0.03884	0.21673	0.20129	0.23183	0.00000

Table 12. Jukes-Cantor Distance (AAGAG)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	0.00000															
2	0.06869	0.00000														
3	0.63102	0.61270	0.00000													
4	0.63442	0.63599	0.07009	0.00000												
5	0.63102	0.62060	0.13980	0.13402	0.00000											
6	0.61737	0.61486	0.13631	0.13058	0.16019	0.00000										
7	0.66575	0.63902	0.55554	0.56885	0.54639	0.54079	0.00000									
8	0.81808	0.81634	0.77924	0.78926	0.75592	0.75847	0.75572	0.00000								
9	0.32889	0.31993	0.59629	0.59926	0.61950	0.59841	0.62982	0.80534	0.00000							
10	0.70559	0.69917	0.56008	0.56624	0.53171	0.51169	0.57301	0.82967	0.66822	0.00000						
11	0.66184	0.65137	0.55199	0.54017	0.53584	0.55856	0.34347	0.79222	0.67428	0.58032	0.00000					
12	0.62737	0.62490	0.14321	0.13749	0.16492	0.00991	0.54079	0.76518	0.60822	0.51649	0.56215	0.00000				
13	0.35555	0.34938	0.65522	0.65067	0.67370	0.64519	0.63744	0.84560	0.17368	0.70559	0.65778	0.65137	0.00000			
14	0.71633	0.71855	0.66222	0.66592	0.63617	0.67245	0.68075	0.81450	0.69504	0.69722	0.68906	0.67245	0.70345	0.00000		
15	0.48159	0.46531	0.69486	0.70768	0.69271	0.69277	0.68028	0.82568	0.50402	0.70345	0.68705	0.69492	0.51852	0.70985	0.00000	
16	0.63317	0.62672	0.24765	0.23950	0.23423	0.24137	0.56522	0.78111	0.61383	0.57508	0.57616	0.25170	0.63902	0.62833	0.66970	0.00000

4.5.8 POL protein-coding genes for 16 species

1. 'HIV-1 (Zaire)' 'K03454'

```
'atgaatttgccaggaaaatggaaaccaaaaatgatagggggaattggaggttttatcaaagtaagacagtatgatcaaatacccatagaaatctgtggacagaaagctataggtacagtattagtaggacctacgcctgtcaacataatcg
gaagaaatttgttgacccagattggctgcactttaaattttccaattagtcctattgaaactgtaccagtaaaattaaagccaggaatggatggcccaaaagttaaacaatggccattgacagaagaaaaaataaaagcattaacagaaatttgt
acagatatggaaaaggaaggaaaaatttcaagaattgggcctgaaaatccatacaatactccaatatttgccataaagaaaaaagacagtaccaagtggagaaaattagtagatttcagagaacttaataagagaactcaagatttctggga
agttcaattaggaataccgcatcctgcagggctgaaaaagaaaaaatcagtaacagtactggatgtgggtgatgcatatttttcagttcccttagatgaagattttaggaaatataccgcctttaccatatctagtataaacaatgagacaccag
ggattagatatcagtacaatgtgcttccacagggatggaaaggatcaccggcaatattccaaagtagcatgacaaaaatcttagagccctttagaaaacaaaatccagaaatggttatctatcaatacatggatgatttgtatgtaggatctga
cttagaaatagggcagcataggacaaaaatagagaaattaagagaacatctattgaggtggggatttaccagaccagataaaaaacatcagaaagaacccccatttctttggatgggttatgaactccatcctgataaatggacagtacagt
ctataaaactgccagaaaaggagagctggactgtcaatgatatacagaacttagtggagagattaaactgggcaagccagatttatccaggaattaaagtaagacaattatgtaaactccttaggggaaccaaagcactaacagaagtaat
accactaacagaagaagcagaattagaactggcagaaaacagggaaattttaaaagaaccagtacatggagtgtattatgacccatcaaaagacttaatagcagaaatacagaaacaagggcacggccaatggacataccaaatttatc
aagaaccatttaaaaatctgaaaacaggaaagtatgcaagaatgaggggtgcccacactaatgatgtaaagcaattagcagaggcagtgcaaagaatatccacagaaagcatagtgatatggggaaggactcctaaatttagactaccca
tacaaaaggaaacatgggaaacatggtgggcagagtattggcaagccacttggattcctgagtgggaatttgtcaatacccctcctttagtaaaattatggtaccagttagagaaggaacccataataggagcagaaactttctatgtagatg
ggcagctaatagagagactaaattaggaaaagcaggatatgttactgacagaggaagacagaaagttgtccctttgactgacacgacaaatcagaagactgagttacaagcaattaatctagccttgcaggattcgggattagaagtaaa
catagtaacagattcacaatatgcattaggaatcattcaagcacaaccagataagagtgaatcagagttagtcaatcaaataatagagcagttaataaaaaaggaaaaggtttacctggcatgggtaccagcacacaaaggaattggagga
aatgaacaagtagataaattagtcagtcaaggaatcaggaaagtactatttttggatggaatagataaggctcaagaagaacatgagaaatatcacaacaattggagagcaatggctagtgattttaacctaccacccgtggtagcaaaaga
aatagtagctagctgtgataaatgtcagctaaaaggagaagccatgcatggacaagtagactgtagtccaggaatatggcaattagattgtacacacttagaaggaaaagttatcctggtagcagttcatgtagccagtggctatatagaag
cagaagttattccagcagaaacagggcaggaaacagcatattttcttttaaaattagcaggaagatggccagtaaaagtagtacatacagacaatggcagcaatttcaccagtgctgcagttaaggccgcctgttggtgggcaggtatcaa
acaggaatttggaattccctacaatccccaaagtcaaggagtagtagaatctatgaataaagaattaaagaaaattataggacaggtaagagatcaagctgaacatcttaagacagcagtacaaatggcagtattcatccacaattttaaaag
aagaaggggggattggggggatacagtgcaggggaaagaataatagacataatagcaacagacatacaaactaaagaattacaaaaacaaattataaaaattcaaaattttcgggtttattacagagacagcagagatccaatttggaaagg
accagcaaagctcctctggaaaggtgaaggggcagtagtaatacaagacaagagtgacataaaggtagtaccaagaagaaaagtaaagattattagggattatggaaaacagatggcaggtgatgattgtgtggcaagtagacaggat
gaggattaa'
```

2. 'HIV1-NDK (Zaire)' 'M27323'

```
'ttttttagggaagatttggccttcccacaagggaaggccggggaattttcttcagagcagaccagagccaacagccccaccagcagagagcttcgggtttggggaggagataacccccctctcagaaacaggagcagaaagacaagga
actgtatcctttagcttccctcaaatcactctttggcaacgacccctcgtcacaataaagataggggggacagctaaaggaagctctattagatacaggagcagatgatacagtattagaagaaataaatttgccaggaaaatggaagccaaa
aatgatagggggaattggaggttttatcaaagtaagacagtatgatcaaatactcatagaaatctgtggatataaagctatgggtacagtattagtaggacctacacctgtcaacataattggaagaaatttgttgacccagattggctgcacttt
aaattttccaattagtcctattgaaactgtaccagtaaaattaaagccaggaatggatggcccaaaagttaaacaatggccattgacagaagaaaaaataaaagcattaacagaaatttgtacagaaatggaaaaggaaggaaaaatttcaa
gaattgggcctgaaaatccatataatactccaatatttgccataaagaaaaaagacagtaccaagtggagaaaattagtagatttcagagaacttaataagagaactcaagatttctgggaggttcaattaggaataccgcatcctgcagggc
tgaaaaagaaaaaatcagtaacagtactggatgtgggtgatgcatatttctcagttcccttagatgaagattttaggaaatataccgcatttaccatacctagtataaacaatgagacaccagggattagatatcagtacaatgtgctcccacag
ggatggaaaggatcaccggcaatattccaaagtagcatgacaaaaatcttagagccctttagaaaacaaaatccagaaatagttatctatcaatacatggatgatttgtatgtaggatctgacttagaaatagggcagcatagaacaaaaata
gaggaattaagagaacatctattgaggtggggatttaccacaccagataaaaaacatcagaaagaacctccatttctttggatgggttatgaactccatcctgataaatggacagtacagcctataaacctgccagaaaaagaaagctggac
tgtcaatgatatacagaagttagtggggaaattaaactgggcaagccagatttatgcaggaattaaagtaaagcaattatgtaaactccttaggggaaccaaagcactaacagaagtagtaccactaacagaagaagcagaattagaactg
gcagaaaacagggaaattctaaaagaaccagtacatggagtgtattatgacccatcaaaagacttaatagcagaactacagaaacaaggggacggccaatggacataccaaatttatcaagaaccatttaaaaatctaaaaacaggaaag
tatgcaagaacgaggggtgcccacactaatgatgtaaaacaattaacagaggcagtgcaaaaaatagccacagaaagcatagtgatatggggaaagactcctaaatttaaactacccatacaaaaggaaacatgggaaacatggtggat
agagtattggcaagccacctggattcctgagtgggaatttgtcaatacccctcctttagtaaaattatggtaccagttagagaaggaacccataataggagcagaaactttctatgtagatggggcagctaatagagagactaaattaggaaa
agcaggatatgttactgacagaggaagacagaaagttgtccctttcactgacacgacaaatcagaagactgagttacaagcaattaatctagctttacaggattcgggattagaagtaaacatagtaacagattcacaatatgcactaggaat
cattcaagcacaaccagataagagtgaatcagagttagtcagtcaaataatagagcagctaataaaaaaggaaaaggtttacctggcatgggtaccagcacacaaaggaattggaggaaatgaacaagtagataaattagtcagtcagg
gaatcaggaaagtactatttttggatggaatagataaggctcaggaagaacatgagaaatatcacaacaattggagagcaatggctagtgattttaacctaccacctgtggtagcgaaagaaatagtagctagctgtgataaatgtcagctaa
aaggagaagccatgcatggacaagtagactgtagtccaggaatatggcaattagattgtacacatctggaaggaaaagttatcctggtagcagttcatgtagccagtggctatatagaagcagaagttattccagcagaaacggggcaag
aaacagcatactttctcttaaaattagcaggaagatggccagtaaaagtagtacatacagataatggcagcaatttcaccagtgctacagttaaggccgcctgttggtgggcagggatcaaacaggaatttggaattccctacaatccccaa
agtcaaggagtagtagaatctatgaataaagaattaaagaaaattataggacaggtaagagatcaagctgaacatcttaagacagcagtacaaatggcagtatttatccacaattttaaaagaaaaggggggattggggggatacagtgcag
gggaaagaataatagacataatagcaacagacatacaaactagagaattacaaaaacaaatcataaaaattcaaaattttcgggtttattacagggacagcagagatccaatttggaaaggaccagcaaagcttctctggaaaggtgaagg
ggcagtagtaatacaagacaatagtgacataaaggtagtaccaagaagaaaagtaaagatcattagggattatggaaaacagatggcaggtgatgattgtgtggcaagtagacaggatgaggattaa'
```

3. 'HIV-2 (Senegal)' 'M15390'

```
'atgacaggcgacaccccaatcaacatttttggcagaaatattctgacagccttaggcatgtcattaaatctaccagtcgccaaagtagagccaataaaaataatgctaaagccagggaaagatggaccaaaactgagacaatggccctta
acaaaagaaaaaatagaagcactaaaagaaatctgtgaaaaaatggaaaaagaaggccagctagaggaagcacctccaactaatccttataatacccccacatttgcaatcaagaaaaaggacaaaaacaaatggaggatgctaatag
atttcagagaactaaacaaggtaactcaagatttcacagaaattcagttaggaattccacacccagcagggttggccaagaagagaagaattactgtactagatgtaggggatgcttacttttccataccactacatgaggactttagaccata
tactgcatttactctaccatcagtgaacaatgcagaaccaggaaaaagatacatatataaagtcttgccacagggatggaagggatcaccagcaatttttcaacacacaatgagacaggtattagaaccattcagaaaagcaaacaaggat
gtcattatcattcagtacatggatgatatcttaatagctagtgacaggacagatttagaacatgatagggtagtcctgcagctcaaggaacttctaaatggcctaggattttctaccccagatgagaagttccaaaaagaccctccataccactg
gatgggctatgaactatggccaactaaatggaagttgcagaaaatacagttgccccaaaaagaaatatggacagtcaatgacatccagaagctagtgggtgtcctaaattgggcagcacaactctacccagggataaagaccaaacactt
atgtaggttaatcagaggaaaaatgacactcacagaagaagtacagtggacagaattagcagaagcagagctagaagaaaacagaattatcctaagccaggaacaagagggacactattaccaagaagaaaaagagctagaagcaa
```

```
cagtccaaaaggatcaagagaatcagtggacatataaaatacaccaggaagaaaaaattctaaaagtaggaaaatatgcaaaggtgaaaaacacccataccaatggaatcagattgttagcacaggtagttcagaaaataggaaaagaa
gcactagtcatttggggacgaataccaaaatttcacctaccagtagagagagaaatctgggagcagtggtgggataactactggcaagtgacatggatcccagactgggacttcgtgtctaccccaccactggtcaggttagcgtttaacc
tggtaggggatcctataccaggtgcagagaccttctacacagatggatcctgcaataggcaatcaaaagaaggaaaagcaggatatgtaacagatagagggaaagacaaggtaaagaaactagagcaaactaccaatcagcaagcag
aactagaagcctttgcgatggcactaacagactcgggtccaaaagttaatattatagtagactcacagtatgtaatggggatcagtgcaagccaaccaacagagtcagaaagtaaaatagtgaaccagatcatagaagaaatgataaaaaa
ggaagcaatctatgttgcatgggtcccagcccacaaaggcataggggaaaccaggaagtagatcatttagtgagtcagggtatcagacaagtgttgttcctggaaaaaatagagcccgctcaggaagaacatgaaaaatatcatagca
atgtaaaagaactgtctcataaatttggaatacccaatttagtggcaaggcaaatagtaaactcatgtgcccaatgtcaacagaaaggggaagctatacatgggcaagtaaatgcagaactaggcacttggcaaatggactgcacacattta
gaaggaaagatcattatagtagcagtacatgttgcaagtggatttatagaagcagaagtcatcccacaggaatcaggaagacaaacagcactcttcctattgaaactggcaagtaggtggccaataacacacttgcatacagataatggtg
ccaacttcacttcacaggaggtgaagatggtagcatggtggataggtatagaacaatcctttggagtaccttacaatccacagagccaaggagtagtagaagcaatgaatcaccatctaaaaaaccaaataagtagaatcagagaacagg
caaatacaatagaaacaatagtactaatggcaattcattgcatgaattttaaaagaagggggggaataggggatatgactccatcagaaagattaatcaatatgatcaccacagaacaagagatacaattcctccaagccaaaaattcaaaa
ttaaaagattttcgggtctatttcagagaaggcagagatcagttgtggaaaggacctggggaactactgtggaaaggagaaggagcagtcctagtcaaggtaggaacagacataaaaataataccaagaaggaaagccaagatcatcag
agactatggaggaagacaagagatggatagtggttcccacctggagggtgccagggaggatggagaaatggcatag'
```

4. 'HIV2-MCN13' 'AY509259'

```
'acaggtgggtttttttagggatgggcccccggggaaagaagccccgcaacttccccgtggcccaagtcccgcaggggctaacaccaacagcacccccagtagatccagcagtggacctactggagaattatatgcagcaaggaaaaa
gacaaagagaacagagagagagaccatacaaagaagtgacagaggacttactgcacctcgagcagggagaggcaccatgcagagagacgacagaggacttgctgcacctcaattctctcttttgaaaagaccagtagtcacggcata
cgtcgagggccagccagtagaagttctgctagacacgggggctgacgactcaatagtagcagggatagagttagggagcaattatagtccaaagatagtaggaggaatagggggattcataaataccaaggaatataaaaatgtaaaa
atagaagttttaggtaaaaaggtaagggccaccataatgacaggtgacaccccaatcaacatttttggcagaaatattctgacagccttaggcatgtcattaaatttaccagtcgccaaaatagaaccaataaaaataatgttaaagccagga
aaagatggaccaaaactgaggcaatggcccttaacaaaagaaaaaatagaggcactaaaagaaatctgtgaaaaaatggaaagagaaggccagctagaggaagcgcctccaactaatccttataacacccccacatttgcaatcaaga
aaaaggacaaaaataaatggaggatgctaatagattttagagaactaaacaaggtaactcaagatttcacagaaattcagttaggaattccacacccagcaggattggccaagaaaaaaagaattactgtactagatataggggatgcttac
ttttccataccactacatgaagactttagacagtatactgcatttactttaccatcaataaacaatgcagaaccaggaaaaagatatatatataaggtcctgcctcagggatggaaggggtcaccagcaatttttcaatacacaatgaggcaggt
cttagaaccattcagaaaagcaaacctagatgtcattatcattcagtacatggatgatatcctaatagctagtgacaggacagatctagaacatgacaaggtggtcctgcagctaaaggaacttctaaataacctaggattttctaccccagat
gagaagttccaaaaggaccctccataccactggatgggctatgaactgtggccaactaagtggaagctgcagaagatacagttgccccaaaaagatgtatggacagtaaatgacatccaaaagttagtgggtgtcttaaactgggcagca
caaatctacccagggataaaaaccagacacttatgtaagctaattagaggaaaaatgacactcacagaagaagtacagtggacagaactagcagaggcggagttagaagagaacaagattatcttaagccaggagcaagagggacac
tattaccaagaagaaaaagagttagaagcaacagtccaaaaggatcaagacaatcagtggacatataaagtacaccagggagagaaaattctaaaagtagggaaatatgcaaagataaaaaatacccataccaatggggtcagattgtta
gcacaagtagttcaaaagataggaaaagaagcactaatcatttggggacgaataccaaaatttcacctaccagtagaaagagagacatgggaacagtggtgggatgactactggcaggtgacatggatccctgactgggacttcgtatct
accccgccgctggtcagactagcatttaacctggtaaaagatcctataccaagaacagagactttctacacagatggatcctgcaataggcaatcaaaggaaggaaaagcaggatatgtaacagatagagggagagacaaggtaagga
tgctagaacaaactaccaatcagcaagcagaattagaagcctttgcaatggcactaacagactcaggtccaaaagccaatattatagtagactcacagtatgtaatggggatagtagcaggccagccaacagaatcagagagtagaatag
taaatcaaatcatagaggagatgataaaaaaggaagcaatctatgttgcatgggtcccagcccataaaggcataggagggaatcaggaggtagatcagttagtaagtcagggcatcagacaagtgttgttcctggaaaaaatagagccc
gctcaggaagaacatgagaaataccatagcaatgtaaaagaactatcccataaatttggattgcccaaattagtagcaagacaaatagtaaacacatgtgcccaatgtcaacagaaaggggaggctatacatgggcaagtagatgcagaa
ttaggcacttggcaaatggactgcacacacttagaaggaaagatcattatagtagcagtacatgttgcaagtggattcatagaagcagaagtcatcccacaggaatcaggaaggcagacagcactcttcctattaaaactggccagtaggt
ggccaataacacacttgcacacagataatggtgccaacttcacttcacaggaagtaaaaatggtagcatggtgggtaggtatagaacaatctttcggagtaccttacaatccacaaagccaaggagtagtagaagcaatgaatcaccacct
aaaaaatcagataagtagaattagagaacaggcaaatacagtagaaacaatagtactgatggcaacacactgcatgaattttaaaagaaggggaggaataggggatatgaccccagcagaaagactaatcaatatgatcaccacagaac
aagaaatacaattcctccacgccaaaaattcaaaattaaaaaattttcgggtctatttcagagaaggcagagatcagctgtggaaaggacccggggaactactgtggaagggagacggagcagtcatagtcaaggtagggacagacata
aaagtagtaccaaggaggaaagccaagatcatcaaagactatggaggaaggcaagaactggatagtggttcccacttggagggtgccagggaggatggagaaatggcatag'
```

5. 'HIV-2UC1 (IvoryCoast)' 'L07625'

```
'atgctggaaatgtggaaaacaaggacacatcatgtcaaaatgcccagaaagacaggctggttttttagggttaggaccctggggaaagaagcctcgcaacttccccatgacccaagtgcctcagggagtgacaccatctgcacccccg
atggacccagcagagggcatgacacctcggggggcgacaccatctgcgccccctgcagatccagcagtggagatgctgaaaagctacatgaagatggggagacagcagagagagagccgagagagaccctacaaggaggtgac
agaggatttgctgcacctcaattctctctttggagaagaccagtagtcagagcatgtatcgagggtcagccagtggaagtattactagacacaggagctgacgactcaatagtagcaggaatagaattaggtagcaattacaccccaaaaa
tagtagggggggataggagggttcataaataccaaagaatacaaagatgtagaaatagaagtagtgggaaaaagagtaagggcaactataatgacaggagacaccccaataaacatttttggcagaaatattttaaatacattaggcatgac
tctaaatttcccagtagcaaagatagaaccagtaaaagtcaaactaaagcctgggaaagatggaccaaaaatcagacaatggcctctatccaaagaaaagatattggccctcaaggaaatctgtgaaaagatggaaaaggagggacagc
tagaagaagcgccccctactaatccatacaacacacccacctttgccataaaaaagagagataaaaacaaatggagaatgctaatagattttagagaattaaacaaggtgacccaagactttacagaagtccaactgggtattccccaccc
ggcagggttggcagaaaaaaggagaataacagtactagatgtgggagatgcctacttcagtatcccactagatccaaacttcagacagtatacagcattcaccttgccatcaataaacaatgcagagccaggaaagagatacatttataaa
gttctaccacaaggatggaaggggtccccagcaatttttcaatactccatgaggaaggtattagatcctttcagaaaggccaacagcgatgtcattataattcagtacatggatgacatccttatagcaagtgacagaagtgatctggagcat
gacagggtagtgtcccaactaaaagagctattaaatgacatgggattctctaccccagaagaaaagttccaaaaagaccctccgttcaaatggatgggttatgagctctggccaaaaaggtggaaactgcaaaaaatacaactgccagaa
aaggaagtttggacagtgaatgacattcaaaaactagtgggagtattaaactgggcagctcaactctttcctggaattaagacaaggcacatatgtaaactaattaggggaaagatgaccctaacagaagaagtacaatggacagaattag
cagaggcagagttacaggagaacaaaatcattttagaacaggaacaagaaggatcctactacaaggaagggtaccactagaagcaacagtacagaaaaacctagcaaatcagtggacatacaaaattcatcagggaaatagaatttta
aaagtaggaaaatatgcaaaggttaaaaacacccacaccaacggggtaagactactggcacatgtagttcagaaaataggaaaagaagccctagtcatctggggagagataccagtattccatctgccagtagaaagagagacatggg
accagtggtggacagattactggcaagtaacttggatcccagaatgggattttgtctcaaccccaccattagtaagactagcctacaacctagtcaaagacccccctagaaaaggtggaaacctactacacagatggatcctgcaacagagc
ctcaaaggagggaaaggcaggatatgtcacagacaggggaaaagataaggttaaagtgctagagcagacaacaaatcaacaagcagaacttgaagcatttgcactagcactacaggattcaggaccacaagtcaacatcatagtaga
ctcacaatatgtcatgggaatagtagctggacagccaacagaaacggagtcaccactagtaaatcaaataattgaagaaatgatcaaaaaagaagcaatatatgtaggatgggtgccagctcacagaggactaggtggtaatcaggaagt
agaccacctagtaagtcaaggaattaggcaggttctattcctagaaaagatagaaccagcacaagaagaacatgaaaagtaccatggcaatgtaaaagaattagttcataaatttggactcccacagttagtggcaaaacaaatagtaaact
cctgtgataaatgccaacaaaaggggaagctgttcatggacaggtaaatgcagaactaggaacatggcagatggactgtacacacttagaaggaaaggtcataatagtggcagtccatgtagccagtggatttatagaggcagaggta
ataccccaagaaacaggaagacaaacagctctcttcctgttaaaattggccagcagatggcctatcacacacctgcacacagacaacggcgccaacttcacctcacaagacgtgaagatggcggcctggtggataggatagaacaaa
ccttcggggtaccctataacccacaaagtcagggagtagtagaagcaatgaaccatcacctaaaaaaccaaatagatagaatcagggaccaggcagtatcaatagagacagttgtactaatggcagctcactgcatgaattttaaaagaa
ggggaggaataggggatatgacccctgcagagagactagttaacatgataaccacagaacaagaaatacagttcttccaagcaaaaaatttaaaatttcaaaatttccaggtctattacagagaaggcagagatcaactctggaaaggac
ctggtgagctattgtggaaaggggaaggagcagtcctcataaaggtagggacagagatcaaagtaatacccagaaggaaagcaaagatcataaggcactatggagggggaaaagaattggattgtggtaccgacatggaggatacca
ggcaggctagagaaatggcacagtctggtcaagtacctgaagcatag'
```

6. 'SIVMM251 Macaque' 'M19499'

```
'gtgttggaattgtgggaaggagggacactctgcaaggcaatgcagagccccaagaagacagggatgctggaaatgtggaaaaatggaccatgttatggccaaatgcccagacagacaggcgggtttttttaggccttggtccatgggg
aaagaagccccgcaatttccccatggctcaagtgcatcaggggctgacgccaactgctcccccagaggacccagctgtggatctgctaaagaactacatgcagttgggcaagcagcagagagaaagcagagagaagccttacaagg
aggtgacagaggatttgctgcacctcaattctctctttggaggagaccagtagtcactgctcatattgaaggacagcctgtagaagtattattggatacaggggctgatgattctattgtaacaggaatagagttaggtccacattataccccaa
aaatagtaggaggaataggaggttttattaatactaaagaatacaaaaatgtaaaaatagaagttttaggcaaaaggattaaagggacaatcatgacaggggacactccgattaacatttttggtaggaatttgctaacagctctggggatgtc
tctaaatcttcccatagctaaggtagagcctgtaaaagtcaccttaaagccaggaaaggttggaccaaaattgaagcagtggccattatcaaaagaaaagatagttgcattaagagaaatctgtgaaaagatggaaaaggatggtcagttg
gaggaagctcccccgaccaatccatacaacacccccacatttgccataaagaaaaaagataagaacaaatggagaatgctgatagattttagggaactaaatagggtcactcaggactttacagaagtccaattaggaataccacaccct
gcaggactagcaaaaaggaaaaggattacagtactggatataggtgatgcatatttctccatacctctagatgaagaatttaggcagtacactgcctttactttaccatcagtaaataatgcagagccaggaaaacgatacatttataaggttct
gcctcagggatggaaggggtcaccagccatcttccaatacactatgagacatgtgctagaacccttcaggaaggcaaatccagatgtgaccttagtccagtatatggatgacatcttaatagctagtgacaggacagacctggaacatgac
agggtagttttacagctaaaggaactcttaaatagcatagggttctctaccccagaagagaaattccaaaaagatcccccatttcaatggatggggtacgaattgtggccgacaaaatggaagttgcaaaagatagagttgccacaaagag
agacctggacagtgaatgatatacagaagttagtaggagtattaaattgggcagctcaaatttatccaggtataaaaaccaaacatctctgtaggttaattagaggaaaaatgactctaacagaggaagttcagtggactgagatggcagaa
gcagaatatgaggaaaataagataattctcagtcaggaacaagaaggatgttattaccaagaaggcaagccattagaagccacggtaataaagagtcaggacaatcagtggtcttataaaattcaccaagaagacaaaatactgaaagta
ggaaaatttgcaaagataaagaatacacataccaatggagttagactattagcacatgtaatacagaaaataggaaaggaagcaatagtgatctggggacaggtcccaaaattccacttaccagttgagagggatgtatgggaacagtggt
ggacagactattggcaggtaacctggataccggagtgggattttatctcaacgccaccactagtaagattagtcttcaatctagtgaaggaccctatagagggagaagaaacctattatacagatggatcatgtaataaacagtcaaaagaa
gggaaagcaggatatatcacagataggggcaaagacaaagtaaaagtgttagaacagactactaatcaacaagcagaattagaagcatttctcatggcattgacagactcagggccaaagacaaatattatagtagattcacaatatgttat
gggaataataacaggatgccctacagaatcagagagcaggctagttaaccaaataatagaagaaatgattaaaaagtcagaaatttatgtagcatgggtaccagcacacaaaggtataggaggaaaccaagaaatagaccacctagtta
gtcaggggattagacaagttctcttcttggaaaagatagagccagcacaagaagaacatgataaataccatagtaatgtaaaagaattggtattcaaatttggattacccagaatagtggccagacagatagtagacacctgtgataaatgtc
atcagaaaggagaagctatacatgggcaggtaaattcagatctagggacttggcaaatggactgtacccatctagaaggaaaaatagtcatagttgcagtacatgtagctagtggattcatagaagcagaagtaattccacaagagacag
gaagacagacagcactatttctgttaaaattggcaggcagatggcctattacacatctacacacagataatggtgctaactttgcctcgcaagaagtaaagatggttgcatggtgggcagggatagagcacacctttggggtaccatacaat
ccacagagtcagggagtagtggaagcaatgaatcaccacctgaaaaatcaaatagatagaatcagggaacaagcaaattcagtagaaaccatagtattaatggcagttcattgcatgaattttaaaagaaggggaggaataggggatatg
actccagcagaaagattaattaacatgatcactacagaacaagaaatacaatttcaacaatcaaaaaactcaaaatttaaaaattttcgggtctattacagagaaggcagagatcaactgtggaagggacccggtgagctattgtggaaagg
ggaaggagcagtcatcttaaaggtagggacagacattaaggtagtacccagaagaaaggctaaaattatcaaagattatggaggaggaaaagaggtggatagcagttcccacatggaggataccggagaggctagagaggtggcata
g'
```

7. 'SIVAGM677A Green monkey' 'M58410'

```
'gtgcggcaaaattggccatatggcaaaagactgcaagaatggacaggcaaatttttttagggtatggccattggggaggagcgaaaccaagaaattttgtgcaatacagaggagacacagttggtctggaaccaacagcccccccaatg
gaaacagcttacgatccagcaaagaagctcctccagcagtatgcagagaagggacagcgcctgagagaggagagagaacagacaaggaaacagaaggagaaagaagtggaggatgtttccttgagctccctctttggaggagacc
aatgaaacgagtcatcatagaaggaacgccagtgcaagccttgttagatacaggagcagatgacactataattcaagaaaaggacttgcactttcccccacataaaccatggcgttccaaggtagtaggaggtataggaggagggattca
tgtcaaagaatatcaggggggtacaagtacaattggaggataaaatcatcaccggctcaattctaataggaagtacaccaatcaatattataggaagaaatattttagctcaggcaggcatgaaattagttatgggagttctatctagtcagattg
aggaaacaaaagtacaactaaaagaagggaaagatggacctaaattgaaacaatggcccttatcaagagaaaaaattgaagctttaacagaaatatgcaaacaaatggaagaggagggaaaattatctaggataggaggagaaaatcc
ttataatacaccagtgtttgccataaagaaaaaggataaaacacaatggagaatgcttgtagatttcagggaactaaacaaagctactcaagactttttttgaggttcagctgggaattcctcacccagcgggccttcagaaaaagaagcaaat
cacagtaatagacataggggatgcctattattcaataccattatgcaaggaattcagaaaatatacagcatttaccatcccctcagtaaataatacagggccagggataaggtatcagttcaattgtctgcctcagggatggaaaggatctcct
acaattttccagaatacggcagcaaacattttagaggagatcaaaaggcacactcctgggttagaaattgtccaatacatggacgatttgtggttggcgtcagaccatgatgagactagacataatcaacaggtagacatagtaagaaagat
gctgctagaaaaaggtctagaaaccccagacaagaaagtccaaagagaaccgccatgggaatggatggggtataaattgcatccgaataaatggaccattaacaaaatagaattacccccctagaaggagaatggacagtaaacaaa
atacagaaggtagtaggagttctaaattgggcaagtcaaatttatccaggaattaaaaccaaacatacctgtgccatgttgagagggaaaaagaacctcctagaagaaatagtatggacagaagaggcagaggcagaatataagaacaat
caagggatagtgcaggaaacacaagaaggaacatactatgaccctctcaaagaattaatagcaacagttcaaaagcaaggagaagggcaatggacataccaattcacccaagaaggggcagtattaaaggtgggaagatatgccaag
caaagagaaactcatactaatgatctaaggactctagcacaccttgtccaaaaaatctgtaaggaagcacttaccatttggggaagacttccacgagtacaactcccagtagacaagaaaacatgggatatgtggtggcaggactattggc
aagtatcctggataccagaatgggagtttgttagcacaccactcctagtaaaactgtggtattccttagtaaaagaaccaatcaaaggagaagatgtttattatgtggatggggcagcatccaaagtgaccaaattaggtaaggcaggatatc
tgtcagagagaggaaaaagtagaattagggaattagaaaacaccactaaccaacaagcagaattaacagcagttaagatggcattggaggacagtggagaaaatgtaaatatagtcacagattctcaatatgtaatgaacatcttgacagc
atgtccacaggaaagtaactcaccccttagtggaacagataatacaagccctaatgaaaaagaggcaggtctacttacaatgggtaccagctcataaggggataggaggcaatacagaaatagataaattagtaagcaaaggaataagac
agatcctcttcttagatagaatagaagaagcacaagatgaccatgcaaagtaccataacaattggagaagtatggtacaggaatttggattacctaatatagtagcaaaagagatagtagcggcatgtcccaaatgccaaataagaggaga
acctaagcatggacaggtagacgcctccattgaaacttggcagatggactgcacccatttagaaggaaaagttataatagtagcagtacatgtagccagtggattcatagaagcagaggtgatcccaagagaaactgggaaggagacag
cacactttctgctgaaactgttagcaagatggccagtgaaacatctacacactgataatggcccaaactttacctctcagaatgtggcagcggtgtgctggtggggtaatatagagcacaccactggaataccttataacccacagtcacag
ggtagtgtagaaagcatgaacagacagctcaaggaaatcatctctcaaataagagatgattgtgagagattggagacagcagtgcaaatggctacgcatatccacaattttaaaagaaagggaggaataggggggtatctctagtgcagaa
agattggttaatatgctaacaacacaactagaactaaatactctacaaaaccaaatccaaaaaattttgaattttaaggtctactacagagaaggtagagatccagtgtggaaaggaccagcgcgactcatctggaaaggagaaggcgcgg
tggtaattaaagagggggaagacatcaaggtagtccccaggagaaaggctaagattatcaaagattatggagagagaaaaacaatggatagtgagggtagtatggagggtgtcagagaggcaaataagcagatggagggggatagt
gacttacaagatcaggaataa'
```

8. 'SIVlhoest L''Hoest monkeys'   'AF075269'

```
'atttgcacagtgcccaaaaccaaaaggaaaggtgaattttttaggatatggtccctggagcagcaaacctccaggcaattatcctctactgggaggagctgcagggagaattccatcagctcccccaatggagagtgctccaacaaaagc
agagagagcactggagacttacaggactttaggacaacagctgaaaaggcagcagcagcaagtcccacagaaatgtgtggacgagccctgcctgaattcactctttccagacgaccaatagaaagtatattggtcaatgggcaaccagt
gtcagcactgttggacactggggctgatgacacaatttttcagaaaattcagtaagaatagaaggaccctatactccaagagtagtagggggcatagggggggcaaattagggtaaaagaatatagagatgtctttatagaaatagcagga
aaaactacagtagggacagtcttattaggacctaccccagtagatattgtagggaggaatattttaactgcaatgggggccaaattgattttagctcagttaagtgataaaattcctattacaaaggtatctttaaaacctgggtgtgatggacca
agagtaaagcagtggcctctatcaaaagagaaaatagaaggcctacaagctatttgtgataggttagaaaaagaaggaaaaatatctccagtggatccagggaacccatacaatactccaatatttgctatcaagaagaaagacaaaaatg
aatggagaaagttaattgatttcagaaagctcaatgaattaactcaagactttcatgagctacaattaggtattccacacccagcaggtatcaagaaatgcaaacagattacagtagtagacataggagatgcctatttcagtatccctttagatc
caaattacagaaagtatacagcatttactataccttctctcaataatcaagagccaggaaaaagatatcagtacaatgtgctgccacaagggtggaaggggagcccttgtatatttcagggaacagtagcaggacttctctcagagtttagga
aattaaatccagacatgatcatttaccaatatatggatgatttattcataggatcagatagagagagaaaaggacatgatcaggcagtaaaagaactcagagaacttcttatgacatggaacttagagacaccagaaaagaagttccaagca
gagccaccctatcattggatgggttatgtactgcatcctgataggtgggaaatagaaaagattaaattaccagagatggatctaacaaaaactacagtaaatcaaatacagaaactggtgggagtacttaattgggcagctcaattgtatgat
ggtattaggacaaaagaactctgcaaattaataaggggagtaaaacccttggaagaaatcataaactggacagaggaagccttggaagaatatgggcagaacaaagaggtacttaaagaaaagatgcagggagcctattatgacccag
aaaaggaacttattgtcagggtacagcaaaacaaaaaggggataattactttccagtggagacaaggaaataacatcttaagagctgggaggtatcaaagacagaaggcagcacacacaaatcccctacagaaattagtagaagctattc
agaagataggaaaagaaagcatagtcatctggggctttgtgccaaaaattcaagtcccagtgactagagaagtctgggagcactggtggagcgaccactggcaggttacatggattccagacttagaattcatttctaccccgcaattaga
acaagagtggtacatttgggaggcagaacccataataggggtagacacctactatgtagatggagcagcagaaaaggtaggaaaaacaggaaaagcaggatatataacacaatcagggaaagagaaagtaaaggagttaactgaca
ccacaaatcagcaagcagaactagaggcagttctaatggcattacaagatagcaatagtaaagtaaatatagtaacagattcacaatatgtaatgaaaatattgtcacaaagaccaacagaaacagaacaccccatagtaaaagacatcat
agaacaatgcaagcagaaagatcaagtttatctaggatgggtgcctgctcataaaggaataggaggtaatcaagaggtagaccacctagtaagtaaaggtataagacagaaacaggtcatgtttctagaaaagatagaacctgcagtaga
agaacatagcaaattccataacaatgcaaaggatctagaagaaaaatttaacctaccccccatggtagccaaacaaattgtcaatgactgtgcaaactgtcaaaagaaaggagaagccataacaggacaagtggatgtctcagtgggtatt
tggcagctagactgtactcacttagaaggacaggtcattataaatgcagtccatgtggcctcaggcttcatggttgcagaagttataccagatgagacaggaaaaacaacgtcacatttcctgttaaaattatgcagtagatggcctgtaaaac
aaatacatacagacaatggtccaaattttgtaagtaaggaggtacaggcagtaacctggtggatagggatagaacatacaactgggataccttataaccctcagagtcaaggggttgtagaagcaaaaaacaaagttttaaaaagtattata
gaaagagtaagagaagatgcacagcaactaaaaacagcagtactaatggcagtccacattcataattttaaacaaaggggaggactagggggcttacaccagcagagaggtttattaatatgattaatgcagaactagaaacacaatat
ctacaaaaattaaattcaaaaatttttaaaatttaaggtttattacagacaaggaagagatcctcagtggaaaggaccagcgcaacttctgtggaaaggagaaggtgctgtagtggtaaaagaaggggagaacatcttctcagtccccagaag
aaaagcaaaactagtaaaagattatggagagggaccaaaggatagtgagagtagcctggataacaactag'
```

9. 'SIVcpz Chimpanzees Cameroon' 'AF115393'

```
'ttttttagggaaaggttggtctccgtggagcgggggatcaaagagaccagggaacttcctgcagggcaggaaggaacccacagctccgcccctagagagttatgggtatcaggaagcaacaggaaagagacaggaggaggagag
ggagaaggaaaaggaaccccctgtatcctccgttaacctccctcaaatcactctttggcagcgacccctcctcggagtaaaaatagaagggcagctaatagaagctctgctagatacaggagctgatgatacagtaatagataatgtacaac
taacaggaaaatggagaccaaaaatgataggggggaattggaggttttatcaaagttaaacaatatgacaacatagcaatagaaattgagggaagaaaaacaacaggcacggtattggtgggaccaacaccagtgaatataattggcaga
aatatattgacacaaattggctgtactttaaattttcctataagtcccattgaaacagtgccagtaaaattaaagccaggaatggatggccctagagtaaaacaatggcctttaacagcagaaaaaattaaagcattaacagaaatttgtcagga
aatggaaaaagaaggaaagatttcaagaatagggcccgaaaatccatataatacaccaatttttgctataaaaaagaaagatagtacaaaatggagaaaactggtagatttcagggaattaaataaaagaacacaggacttttgggaagtc
cagttaggcataccacacccggcaggtttaaaaaagaaaaagtcagtaacagtgctagatgtgggagatgcttatttttcctgcccattagatgaaaattttaggaaatatacagcatttactatacccagtgtaaataatgaaacaccaggaat
tagatatcaatataatgtgttaccacaaggatggaaagggtcaccagcaatttttcaaagcagcatgataaaaattctaaccccattcagacagcagcatccagacttaataatttaccagtatatggatgatctttatgtgggatctgatttggaa
ctaagagaacatagagaaaaggtagagctactcagacagcatctacttacttgggggtttactaccccctgataaaaagcatcagaaggagccacctttttttgtggatgggctatgagctccatcctgataagtggacagtgcagccaataca
gttaccacagaaagagacctggactgtcaatgatattcagaaactagtgggaaaactaaattgggcaagtcaaatatatccaggaataaaagtaaaacagttatgcaagctgataaaagggacaaaagctctaacagacatagtgaccatg
acacaggaagcagaaatggaattagaagagaacagagagattttaaaggatccagtgcatggggtgtactatgacccagaaaaggaattaatagcagaaatacagaaacagggaaatagtcagtggacttatcaaattttccaggaaca
gcataagaacctaaaaacaggcaaatatgctaggcaaagatcagcacatactaatgacataagacaattagcagaggtagtacaaaaaatagctactgaaagtatagtaatctggggaaaaacaccaaaattcagattacctgttcaaaag
gaggtctgggaaacctggtggtcagagtattggcaagccacctggatcccagactgggaattcgtgaataccccctccccttgtaaaattatggtaccaattggaaacagaacccataccaggagcagaaacgttctatgtggatggggca
gctaacagagagacaaagaaaggaaaagcaggctacgtcactgatagaggcagacagaaaatagtaaacttagaaaataccactaatcagaaagcagaactgacagctgtttatctagcattacaagatgcagaacatatagttaatata
gtcactgattcccaatatgtgttagggataattcactcccagccagatcaaagtgagtcagaattggttaatctaataatagaagagttaataaagaaggaaaaaagttacctttcatgggtgccagcacataaaggagttggaggaaatgaa
caagtagataaattagttagttctgggaatcagaaaagttctctttttagatggtatagattaagcacaaggaagaacatgaggaaatatcataacaattggagagctatggcaagtgatttcaatataccacatatagtggccaaagaactagt
agcccgttgtgataaatgccaactaaaaggagaggccatgcatgggcaagtagactgcagcccaggcatatggcaattagactgcacacacctagaaggaaaagttatcctggtggcagtacatgtagccagtgggtacctagaggca
gaggtcattcctgcagagactggacaggaaacagcttattttattctaaaattagcaggaagatggcctgtaaaagtaatccacactgataatgggcctaactttaccagtaatacagttaaggcagcctgctggtgggcaggcatccaaca
ggaatttggaattccctataatccacaaagtcaaggagtggtagaatcaatgaataaggaattaaaaaagatcataggacaaattagagaacaagcagaacacttgaggacagcagtccagatggcagtgttcattcacaattttaaaagaa
aagggggggattgggggggtacactgcaggagaaagaattatagacatcatagcaacagacatacaaacaactaatctacaaaaacaaattttaaaagttcaaaattttcgggtctattacagggacagcagagatcccatttggaaaggacc
tgccagacttctgtggaaaggtgaaggggcagtggtaattaaagagagagaggaagttaaagtaatacccagaaggaaagcaaaaataattagggattatggaaaacagatggctggtgatgatagtatggcaggtggacaggatgag
agtcaaggcctggaatag'
```

10. 'SIVmnd5440 Mandrillus sphinx' 'AY159322'

```
ttttttagggaatacccccctggggcagtggcaagcccaggaacttccctgccatgccattgaccccatcagcaccccccgatgccagggttggaagacccggcagaaaagatgctgttagactatatgaagaaagggcaacaacagagg
gcagcagcaggagcccagaaagaggaaaagaagggtccctacgaggcagcttacaactccctcagttctctctttggaacagaccaactacagtagtagaaatagaggggcaaaaggtagaggcacttctagatactggagcagacg
atacagtaattaaagatttagagttaacaggaaattggaagccacgaattataggaggaattggaggagcaatcagagtaagacagtattttaattgtaaagtagaaatagcaggaaaaattactcatgctgcagttctggtgggccctaccc
ctgtaaatattataggtagaaatgtacttaagaagttaggatgtactttaaattttcctattagtaagatagagacagtaaaagtaacactaaaaccaggaacagatggaccgagaataaaacagtggccactgtctaaagaaaagatcctagc
cttacaagaaatatgcaatcaaatggaaaaagaaggcaagatctctagagtaggtccagaaaatccatacaacaccccagtattctgtataaaaaagaaagatggaaccagctggagaaaattagtagattttagacaactgaacaaagtc
actcaggatttctttgaggtgcagttaggaatcacacatccaggaggcctaaagcaatgtgagcaaattacagtattggacataggagatgcctatttttcatgtcccttggatgtggactttagaaagtatactgcgttcaccattccatcggtg
aacaaccagggcccaggaatcagataccagtataatgtcctaccacagggatggaaaggctctccagcaattttccaggcaacagcagacaaaattctaaaggcttttaaggaaaagcatccagaagtcttaatctatcagtatatggatg
acttgtttgtggggagtgatctaagtgcctctgaacacagtaaaatgatagtagaattaagggaacatttgcagttctgggggctcgagaccccagacaaaaagttccaaaaggaacctccctttgaatggatgggatatgtgctgcaccca
aagaaatggacagttcagaaaatacagctaccagaaaaagaaagctggacagtaaatgatattcagaaattagtagggaaacttaattgggcaagccaaatctattccggaattaaaacaaaagagctctgtaaattgataagggggca
agacccctagataaaatagtggaatggactagagaagcagaactagaatatgaagaaaacaagctaatagtgcaggaggaagtacatggagtgtactaccagccagaaaaacctttaatggcaaaggtgcaaaagttaacacaaggac
```

agtggagttaccaaatagaacaggaagacaacaaacctctcaaggtaggaaagtatgccaggacaaagaatgcccacacaaatgaattgagggtacttgcagggttggtacaaaaggtagctaaggagtgtctagtaatctggggaaa
gctaccgaaattctatctcccgttagagagggaggtatgggaccaatggtggcatgactattggcaagtaacatggatcccagagtgggacttcatttcaacaccaccattaataaggttatggtacaacctgctaaaagacccaataccag
gggaagatgtatattatgtagatggggcagcaaacagaaactccaaggaaggcaaggcaggctattatacctctagaggtaaaagcaaggtagtagccttggaagaaacgaccaatcaaaaggcagaattgacagctataaagatggc
tttgcaagattcaggacccagggtgaatatagtcactgattcccagtatgcactaggcatactttcagcagccccagatcagtcagacaaccccatagttagggaaataatagaactcatgataggcaaggaagagatttatctctcatgggt
cccagcccacaagggtataggaggtaatgaacagatagataaattagtaagtaagggaataagacaggtactatttttagagggaatagataaagcccaggaagaacatgacaaatatcacaataattggagagcgttgtctcaggacttt
aacctaccccccatagtggcaaaagagatagttgcacaatgtccaaaatgtcagacaaagggagaacctgttcatgggcaggtaaatgcagatccagggacatggcagatggactgcacccatttagaaggaaaaatcatcatagtagc
agtccatgtagccagtggatatttagaagcagaggtaataccagcagagacaggaaaagaaacagcgcacttcctgttaaagctagcaggtaggtggccagtaaaacatctacacactgacaatggtcctaactttgtcagtgaaaaggt
ggccacagtctgctggtgggctcaaatagagcacaccacaggagtaccttacaacccacagagtcaaggagtagtagaagcaaaaaatcatcatcttaagaagatcatagaacaagttagggatcaagctgaaaggctagaaacagca
gtgcaaatggcagtgctaattcacaattttaaaagaaaagggggatagggggagtacagtcctggagaaagaatagtagacataataaccacagacatcctaacaaccaaattacgacaaaatatttcaaaaattcaaaattttcaggtttatt
acagagaaggaagggaccaacagtggaaaggaccagcagagctcatttggaaaggagaaggcgctgtggtaatcaaagaaggagttgatttaaaggtggttcctagaagaaaagcaaaaataattaaggactatggaaaaacagtgg
atagcaatccccacatgggaagatag'

11.'SIVAGM3 Green monkeys' 'M30931'

'accaagaaagatgagatgcttgaaatgtgggaaaccagggcatttagcaaaagattgcagaggacaggtaaattttttagggtatggccggtggatgggagcgaaacccagaaattttcccgccgctactcttggggtggagccaactg
cgccccctccaccgagtccatacgaccctgcaaagaagctcctgcagcaatatgcagacaaggggaagcagttgagggaacaaaggaaaaaaccaccagcagtgaatcccgattggacagagggatattctttgaactccctctttgg
agaagaccaataaaaacagtttacatagaaggggtccccatcagagcattattagatacgggggcagatgataccattataaaagaagcagatttacaattatcaggaacatggaaaccaaaaataatagggggcattggagggggactc
aatgtaaaagagtatagtgatagggaagtaagattggaagacaaaattttgagagggaccatattgataggaagcactcccataaacataattggaagaaatatattagcaccagcaggagccaaattagtaatgggtcaactgtcagaac
aaattcccattacccctgtgaaattaaaagaaggggctagaggacctttcttaaaacaatggcccctctccaaagaaaaaataaaagccttacaggaaatatgtgaccaattagagaaagaaggaaaaattagcaagataggaggagaga
atgcatacaacactccagtgttttgcataaagaaaaaagacaagtcacaatggagaatgttagtagattttagggaactaaacaaagcaacacaagattttttcgaagtacagttaggcatacctcatccatcagggttcgaaaagatgacgg
aaataacagtattagacataggggatgcctattattcaataccattagacccagagtttagaaagtataccgcttttaccattccatcagtaaataatcaagggccaggtactagatatcagttcaactgtcttccacaaggatggaagggatcc
ccaactatttttcagaacacagcagcttccattctagaagaaataaaaaaggagttaaaacccctaaccattgtgcaatacatggatgacctatgggtagggtctcaggaagatgaatacacgcatgatcggttggtagaacaactaagaat
gaaattaagtgcctggggattagaaacaccagacaagaaagtacagaaaaaaccaccttatgagtggatgggatacaaattgtggccacacaagtggcagataagcagcatagaattagaagacaaagaagaatggactgtaaatgat
atacaaagactagtggggaaactaaattgggcagcacagctttacccaggactcagaactaaaaacttgtgtaaattaatcagaggaaaaaagaacttactagaaacagtaacctggacagaggaagcagaagcagaatatgcagaaaa
caaagagatcttaaaaacggaacaggaagggacctactacaaaccaggaagacccatcagagcagcagtgcaaaaactagaaggaggtcaatggagttaccaattcaagcaagagggacaagtattaaaagtaggtaaatacacaaa
gcagaaaaacactcataccaatgagttccgtgtattggcaggattagtacaaaaactttgtaaagaatctttagttatatggggagagttgccagtccttgaactcccaatagagagggaagtatgggaacaatggtgggctgattactggca
ggtaagttggattccagactgggaatttgtcagtaccccacccctagtaaaattatggtataccctgacaaaagaacccataccaaaggaagatgtctactatgtggatggagcttgtaatagaaattcaagggaaggaaaagcaggatata
tcacacaatatgggaaacaaagggtggaaaaattagaaaatacaacaaaccagcaagcagaattaatggccataaaaatggcactagaagatagtgggcctaatgtaaacatagtaacagattcacaatatgcaatgggaatattaactg
cccaacccacacagagtgactcacccttaatagaacaaattatagcactaatggtacaaaaacatcagatatacttgcaatgggtaccagcagacaaagggataggaggcaatgaagagatagataaactagtaagtcaagggatgagg
aaaattttatttttagaaaaaatagaagaagcccaggaggaacatgaaaggtaccataataattggaggaacttagcagacacttatgggctaccacaaattgtggcaaaagaaatagtagccatgtgtccaaaatgtcagataaaagggg
aaccagtccatgggcaagtagatgcctcgccaggggtatggcaaatggactgtacacatttagaaggcaaggtaatcatagtagcagtccatgtagccagtggattcatagaagcagaagttatacctagagaaacagggaaagaaaca
gcaaaatttttattaaagatactaagtagatggcccataacccaactgcatacagacaatggacccaattttacgtctcaagaagtagcagcaatgtgttggtggggaaaaatagaacacaccacaggtgtaccctataaccctcaatcacaa
ggctctatagagagtatgaataaacagttaaaagaaataattgggaaaataagagatgactgtcaatacacagaaacagcagtacttatggcatgccacatccacaattttaaaagaaagggaggaatagggggggttaacaccggcagag
agattaatcaatatgattactacacaattagaattacaacacctacaaaccaaaattcaaaaaattttaaattttagagtctactacagagaagggagagatcctgtctggaaaggaccaggacagttaatttggaaagggaaggtgcagtg
gtcatcaaaggaggtgtggaattaaaagaatacccaagaaggaaagcaaaaattataaaggattatgaaccaagaaaaagaatgggtgatgagagtaacttggaaggtgccggaggagctgataactaa'

12.Simian macaque' 'M33262'

'gtgttggaattgtgggaaagagggacactctgcaaggcaatgcagagccccaagaagacagggatgctggaaatgtggaaaaatggaccatgttatggccaaatgcccagacagacaggcgggtttttttaggccttggtccatgggg
aaagaagccccgcaatttccccatggctcaagtgcatcaggggctgatgccaactgctcccccagaggacccagctgtggatctgctaaagaactacatgcagttgggcaagcagcagagagaaaagcagagagaaagcagagaga
agccttacaaggaggtgacagaggatttgctgcacctcaattctctctttggaggagaccagtagtcactgctcatattgaaggacagcctgtagaagtattactggatacaggggctgatgattctattgtaacaggaatagagttaggtcc
acattataccccaaaaatagtaggaggaataggaggttttattaatactaaagaatacaaaaatgtagaaatagaagttttaggcaaaaggattaaagggacaatcatgacaggggacaccccgattaacatttttggtagaaatttgctaaca
gctctggggatgtctctaaattttcccatagctaaagtagagcctgtaaaagtcgccttaaagccaggaaaggatggaccaaaattgaagcagtggccattatcaaaagaaaagatagttgcattaagagaaatctgtgaaaagatggaaaa
ggatggtcagttggaggaagctcccccgaccaatccatacaacacccccacatttgctataaagaaaaaggataagaacaaatggagaatgctgatagattttagggaactaaatagggtcactcaggactttacggaagtccaattagga
ataccacaccctgcaggactagcaaaaaggaaaagaattacagtactggatataggtgatgcatatttctccatacctctagatgaagaatttaggcagtacactgcctttactttaccatcagtaaataatgcagagccaggaaaacgatac
atttataaggttctgcctcagggatggaaggggtcaccagccatcttccaatacactatgagacatgtgctagaacccttcaggaaggcaaatccagatgtgaccttagtccagtatatggatgacatcttaatagctagtgacaggacagac
ctggaacatgacagggtagttttacagtcaaaggaactcttgaatagcatagggttttctaccccagaagagaaattccaaaaagatcccccatttcaatggatggggtacgaattgtggccaacaaaatggaagttgcaaaagatagagtt
gccacaaagagagacctggacagtgaatgatatacagaagttagtaggagtattaaattgggcagctcaaatttatccaggtataaaaaccaaacatctctgtaggttaattagaggaaaaatgactctaacagaggaagttcagtggactg
agatggcagaagcagaatatgaggaaaataaaataattctcagtcaggaacaagaaggatgttattaccaagaaggcaagccattagaagccacggtaataaagagtcaggacaatcagtggtcttataaaattcaccaagaagacaaa
atactgaaagtaggaaaatttgcaaagataaagaatacacataccaatggagtgagactattagcacatgtaatacagaaaataggaaaggaagcaatagtgatctggggacaggtcccaaaattccacttaccagttgagaaggatgtat
gggaacagtggtggacagactattggcaggtaacctggataccggaatgggattttatctcaacaccaccgctagtaagattagtcttcaatctagtgaaggaccctatagagggagaagaaacctattatacagatggatcatgtaataaa
cagtcaaaagaagggaaagcaggatatatcacagataggggcaaagacaaagtaaaagtgttagaacagactactaatcaacaagcagaattggaagcatttctcatggcattgacagactcagggccaaaggcaaatattatagtaga
ttcacaatatgttatgggaataataacaggatgccctacagaatcagagagcaggctagttaatcaaataatagaagaaatgattaaaaagtcagaaatttatgtagcatgggtaccagcacacaaaggtataggaggaaaccaagaaata
gaccacctagttagtcaagggattagacaagttctcttcttggaaaagatagagccagcacaagaagaacatgataaataccatagtaatgtaaaagaattggtattcaaatttggattacccagaatagtggccagacagatagtagacac
ctgtgataaatgtcatcagaaaggagaggctatacatgggcaggcaaattcagatctagggacttggcaaatggattgtacccatctagagggaaaaataatcatagttgcagtacatgtagctagtggattcatagaagcagaggtaattc
cacaagagacaggaagacagacagcactatttctgttaaaattggcaggcagatggcctattacacatctacacacagataatggtgctaactttgcttcgcaagaagtaaagatggttgcatggtgggcagggatagagcacacctttgg
ggtaccatacaatccacagagtcagggagtagtggaagcaatgaatcaccacctgaaaaatcaaatagatagaatcagggaacaagcaaattcagtagaaaccatagtattaatggcagttcattgcatgaattttaaaagaaggggagg
aataggggatatgactccagcagaaagattaattaacatgatcactacagaacaagagatacaatttcaacaatcaaaaaactcaaaatttaaaaattttcgggtctattacagagaaggcagagatcaactgtggaagggacccggtgag
ctattgtggaaaggggaaggagcagtcatcttaaaggtagggacagacattaaggtagtacccagaagaaaggctaaaattatcaaagattatggaggaggaaaagaggtggatagcagttcccacatggaggataccggagaggct
agagaggtggcatag'

13.'CIVcpzUS Chimpanzee' 'AF103818'

'ttttttagggaaacattggtccccatcgtggagcgggggatcaaagagaccagggaacttcctggaaaacaggaaggagcccacagctccaccaatagaggatttcgggtatcaagaagagacagtgacacaggagaaacagggaa
aggaaaaggagccctttcagctaacttccctcaaatcactctttggcagcgacccctcctcagaataaaagtagcagggcagatagtagaagctcttttagacacaggagctgatgatacagtactagacaacatacaaattgaagggaca
tggagaccaaaaatgatgggggaattggaggttttataaaagtaaaacaatatgatcacgtcaatatagaaatagagggaagaaaagcacagggttcagttttagtgggaccaacaccagtaaatattattggcaggaatattttaacaca
aattggatgtactttaaattttcctatcagctctattaaaactgtaccagtaagattaaaaccaggaatggatggccctagagtaaagcaatggccattgacagcagaaaaaattaaggcattaacagaaatttgccaagaaatggaaaaagaa
ggaaagattacaagagtagggccagaaaatccttacaatacaccaatttttgctatcaaaaagaaagatagcactaaatggagaaaattagtagactttagagaattaaataaaaggacacaagacttctgggaagttcagttaggcatacc
acacccagcaggattaaagaagaaaagatcagtaacagtcctagatgtaggggatgcctatttttcttgtccactggataaggaatttagaaaatacacagcatttaccatccctagtgtaaacaatgagaccccaggaattagataccagta
taatgtgttaccacaaggatggaaagggtcaccagcaattttccagagcagtatgacaaaaattctagatccttttagaaaacaacatccagatgttataatctatcaatatatggatgatctctatgtagggtcagatctaaacttagaaaagca
tagggaaaaggtagaactgctcagacaatatttgcttacttggggattcactacccctgataagaagcatcaggaggaaccaccattcttatggatgggctatgaactccatccagataagtggacagtgcagcccatacagttaccacaaa
aagaaatttggacagtcaatgatattcagaaattagtagggaaactgaactgggcaagtcagatatatccaggaataaaaataaagcagttatgcaaattaataaaaggagctaaagctctaactgaagttgtaaattttacacatgaggcag
aaatggagttagaagaaaacagagaaattctaaaggaaccagtacatggggtctattatgacccagaaaaagaattagtagcagaagtacaaaaacaaggaaggagtcagtggacataccaaatttttcaagaacggcataagaatctaa
agacaggaaaatatgccagacaaagatcagcacatactaatgatatcagacagctagttgaagtggtgcaaaaaatagctactgagagcattgtcatttggggaaaggtaccaaaatttaaattgctagtgcagaaggaagtctgggaaac
atggtggtcagaatattggcaggccacctggattccagattgggaatttgtcaatacccctcccctcgtaaaactttggtataagttagagacagaggccatagagggggcagaaacattttatgtggatggagcagcccaacgagaaacc
aagaaaggaaaggcaggatatgttactgataggggtagacaaaaaattataactttagaaaatactactaatcaaaaggcagagctcacagcggtatacttagcattaaaagattcagagaatacagttaatgtagtcactggctcccaatat
gtcttaggaatcatccactctcagccagaccaaagtgaatcagaattggtcaatcagataatagaagaattaataaagaaagaaaaaagttatatctcatgggtaccagcacataaaggaattggaggaaatgaacaaatagataaactagt
cagttcagggatcagaaaagttcttttcctagatggcatagataaagcacaggaagaccatgacaaatatcatagtaattggacagctatggccagtgattttaacctgccaccagtggtcgctaaggagatagtggccagctgtgataaat
gtcagccaaaaggagaggccatacatgggcaggtagattgcagtccaggtatctggcagctagattgtacacatttagaagggaaaatcattatagtggcagtacatgtggccagtggatacctagaagcagaggtcattcctgcagaaa
caggacaagagacagcctattttatcttaaaattagcaggaagatggcctgtaaaagtgattcatactgataatggatctaactttacaagtagtacagttagagcagcttgctggtgggcaggcatacaacaagagtttggaattccatacaa
tccacaaagtcaaggagtggtagaatccatgaataaagaattaaagaaaatcataggacaaatcagggatcaagcagagcatttaaagacagctgtacagatggcagtattcattcacaattttaaaagaaaaggggggattgggggtat
actgcaggagaaagaatcatagacatcatagcatcagaactacaaacagacttattacaaaaacaaattttaaaagttcaaaattttcgggtctattacagggacagcagagatccaatttggaaaggaccagccaaacttctgtggaaagg
tgaaggggcagtagtaatcaaggaaaacgaggaggttaaagtagtacccagaagaaaagcaaaaattataaaagactatggaaaacagatggcaggtgctgatagtatggcaggtagacaggatgagagttaa'

14.'SIVmon Cercopithecus Monkeys' 'AY340701'

'ttttttagggaatcctttcggtcccaagaaggggccaaggaacttccctctcacctctgtgcagcccagcgcaccgactctgccagcagggactgcggagccagtcaatctcaaccaagagattggttccccgaagacggggaccccaa
aggagacgaggcgggatctttatccctccttagcatctctctttggggaagaccaataagatctgtgctgatagggggacaaaaggtgtccatgctattggatacaggtgccgacgacacaattatacaagatcagcacgtaaatttaggg
gagaattggacacctaaagtagtaggggggaatagggggaatgatagaagttaaacaatacaaagcaatacaagtgatatttgaagacagagaggtttgggcaacagtattagttggaccaactcctataaatattttagggagaaatgtctt
agctaagatgggggtaaccctcaacatggctgctgggggagacctccaacccccatctttagaggtcactttaaaagcaggaaaagagggaccaaagttgaaacagtggccattatccagagaaaagatagaggctcttacacaaatta
ctcaagagatgcttaaattaggacaattggaaccaaccgagcccaataacccatataactcaccctgttttgccataaagaaaaaggacaagagcaaatggagaatgctcatagatttcagagaacttaacaaggcaactcaggacttcttt
gaggtgcagttgggaatcccacacccagcgggacttgagaagatggaccatgtgacaatagtagacatgaaggacgccttctacagcattcccttgtgggaaccgttcagaaaatacacggcgttctcaatacсctctttaaacaatgctg
aacctgctaaaaggtatcagttcaaggtcctcccgcagggctggaaagggtccccaactgtgtttcaacatacagcggcagagctaccacaggagatcaggaaaaaatacccagaggtaactttaatacagtacatggatgacctactga
taggatcaaatcactctctgccagaacacaggaaaatagtaggggtcataagagcgaccctactaaggaaagggattcaaacacccccagagaaatttcaggatgactacccggtgcaatggttagggtatgagctacacccatcccatt
ggaaaattgccaaaatagagctaccagaccaggaacaatggactgtcaatgagatccaaaagttgataggaaagcttaattgggcagcacaaatctattcaggaataaaaaccaaaaacctgtgcaaaagtatcagaggggcacggag
cctgacagatacagtggtcctctcagatttagctcaagctgagctagcggaaaatagggagattctcaaggaggaatccaccgggggatattacgaccccaacgagccattaaaggtggagttaacctcactagcagaaaaccaatggg
gctacagattcttccaagaaaaatcagtattgaagacaggaaaatttgccaaaattagaagtacccactccaatacctatcagcagttagcagatgcgctagccaaattgggaaaagaggccatagtcacatggggacgacttcccatttttc
gcctaccagtggtgaaagagcaatgggacgcatggtgggcagataattggcaaatcaattgggtaccagacatagaggcggtgtattccccacacctgctcagacagtggtatactctggtacaagaccccatagaaggggctccaaca
tactacgtggatggggccgccaataggacatcaaaggaagggaaagcaggttatgtaaccaactttggggcaagaaaggccgtcacattagaacagaccaccaaccaaaaggcagagctagaggccgtgttactagccctcagagat
ggaccacctcagatgaacataattacagattctcagtatgtactgggcattctagcctcctgcccggagatctcagaaagtcccatagtggaggccatcatacaggaactcttaaaaaaggaccaggtattcttgtcatgggtaccagctcac
aaaggaataggaggaaatgaagaagtagatcgcctagtgggccagggaattaggcaagtcctgtttatggaaaatttggaacccgcaagggaagatcatgaaaagtatcatagtaattggaaatacttaagagatacctaccacattccta
ctctgttggccaaagaaatagtcaaccattgtcacaaatgtcaaaaccatggagaacccaaatcaggacaggtaaacgcagaagtaggagtatggcagatggactgcacccatctagagggaagcataattttggtagcagtacatgtag
ccagtggttatgtatgggccaagatcctaccaagggaaacaggaaaatgtacagggatagccctgttggagctggccgctatgtggccggtcactcagatacatacagataatgggccaaatttcaccagccaagaattcgaagcagct
gcatggtgggcaaatattcatcacaccacaggggtaccctacaacccacaaagccaaggggttgtagagaatatgaatagacaactaaaagagaccattaagaaaatcagggatgaggttgaatacctcccaacagcattagcacaagc
actgttcatcttaaattttaaaagaaagggaggaataggggatatggcccctgtagacagatttataaacatgatacacacagaattagaactacaaacctcaaacaaccaaattacaaaattttcaaaatttcgggtttatttcaggacaggtg
ccgaccctcagtggaaaggaccagcgcgcctactctggaaaggtgaaggtgccgtagtcataaaaactgaggagggcgatatcctgacagtccctaggagaaaggcaaagataatcaaaccctatggcaccacggaaaatgtgggtg
gtgacaccaaccaatataatctcagaaagcaagatggattggctaattag'

15.'SIVcpzTAN1 Chimpanzee' 'AF447763'

'ttttttagggaaacgcacccccttgtgggggtgcagaccagggaactttgtgcagaacaccccagagaaagggaaggctcaggagcaggagacagcacagacaccagtggtgccaactgccccaccactggagatgacgatgaaa
ggcgggttctccctcaagtcaatctttggcagcgaccaatgatgacagtaaaagtccagggacaagtctgtcaagctcttttagatactggagcagatgacagtgtttttttgtaacatcaaattaaagggacagtggacaccaaaaaccatag
gaggaataggaggatttgtaccagttagtgagtactataatattccagtacaaattggcaataaagaagtcagagccactgtcctagtgggagaaaccccccattaatataataggtagaaatattttaaagcaattaggatgtaccttaaattttc
ctattagcccaatagaggtagtaaaagtacaattaaaagaaggaatggatgggccaaaagtaaagcagtggcccctctccaaggagaaaattgaggcattaacagaaatatgtaagacattggaaaaggaaggaaaaatttctgcagttg
gaccagaaaacccatataacacaccaatttttgccattaagaaaaaggatacctctaaatggagaaaattagtagatttcagagaactgaataaaagaactcaagatttttgggagttacagctaggaataccccatccggcagggttaaga
aaaagaaatatggtgacagtactggatgtaggggatgcctacttttccattcccctggatccagacttcagaaagtatacagcttttaccatacccagtctcaataataacacaccagggaaaagatttcagtataacgtgttacctcaaggttg
gaagggatctccagcaatttttcagagcagtatgacaaaaatcctagatcctttcagaaaagaacacccagatgtggacatttaccaatatatggatgatctttacataggttcagatcttaatgaagaggaacataggaaactgataaagaag
ctgagacagcatctgttaacatggggattagagaccccctgacaaaaagtatcaggaaaaacctccattcatgtggatgggctatgagctacatccaaataaatggacagttcaaaatatcacattaccagaaccagagcagtggacagtga
atcatatccagaagttggtaggcaaacttaattgggccagtcaaatttatcatggaataaaaactaaagaactatgcaaattgattagaggagtaaaaggattaactgagccagtagaaatgaccagggaagcagaattggagttagaagaa
aataagcagattctaaaagaaaaggttcaaggagcatactatgatcctaaattacctctgcaagcagcaatacagaagcaggggcaaggacagtggacatatcagatatatcaggaagaagggaaaaatttaaaaacaggaaaatatgc
aaaatcaccaggtacccacaccaatgagataagacaattagcaggactgatacagaaaataggcaatgagagcataataatttggggtattgtgcctaaatttttattacctgtatccaaagagacatggagccagtggtggactgattactg
gcaagttacctgggtacctgagtgggaatttattaacaccccaccactaatcaggctatggtacaatctgttgtctgaccccatcccagaagcagaaaccttttatgtagatggggcagcaaacagagacagtaaaaagggaagagcagg
atatgtaacaaacagaggcagatacaggtcaaaggacttagagaacaccactaatcaacaagcagaattatgggcagtagatctagccttaaaagactcaggagcacaggtaaatatagtcacagattcccaatatgttatgggagttttac
agggattaccagatcaaagtgactcccccatagtagagcaaattattcaaaagttaacacaaaagacagcaatttatctagcatgggtaccagcccataaaggtataggggggtaatgaagaagtagacaaattggttagtaaaaatattaga
aaaatattattcctggatggaattaatgaagcacaggaagaccatgataaatatcacagtaattggaaagctttagctgatgaatataatctgcccccagttgtggctaaagaaattattgctcagtgtccaaaatgccatataaaggagagg
ctatacatggacaggtggactacagtccagaaatctggcaaatagactgtacccacctagaaggaaaggtcatcatagtagcagtgcatgtagctagtggtttcatagaagcagaagtcataccagaagaaacaggaagagaaaccgctt
acttcatcctaaaattggcaggaagatggcctgtaaagaaaatacatacagataatggaccaaattttactagtacagcagtgaaggcagcctgctggtgggcacaaattcaacatgaatttgggattccatataatcctcaaagtcaaggag
tagtagaatctatgaataaacaattaaagcaaattatagagcaagtcagggaccaagcagagcaactgaggacagcagtaatcatggcagtgtatatccacaattttaaaagaaaaggggggattggggagtacactgcaggggaaaga
ctattagacatactaactacaaatatacagacaaaacaattacaaaaacaaattttaaaagttcaaaattttcgggtttattatagggacgccagagatccaatttggaagggaccagcgcgactactgtggaaaggtgaaggggcagtagta
ataaaagaaggagaagacattaaagtagtacccaggagaaaagcaaaaatcataaaagagtatggaaaacagatggcaggtgcaggtggtatggatgatagacagaatgagacttag'

16.'SIVsmSL92b Sooty Mangabey' 'AF334679'

'ctgtggaaagacaggacactctgcgagacagtgcagagctcccagaaggaagggatgctggaaatgtggggaagaaggacgtattcaggccaactgcccaaaccagaaggcgggtttttttaggccttggcccatggggaaagaag
cctcgcaacttccccatgcagacaacatcattgacaccgtcagcacccccagatccagcagcccggatagtcaaggagtatctagagaaagcacagagggaaaagacaaggaggagcaggccttacaaggaggtgacggaggattt
gctgcacctcaattctctcttcggagaagaccagtagtcagagcactaatagagggtcagcccctagaggttctactggacacaggggcagacgactccatagtagcagggatagagttaggtccagggtacacaccaagaattgtagg
ggggattggaggattcataaatacaaaagaatatagaacagtaaaaatagaagttttaggaaaggtaataagagaaacactaatgacaggagataccccgataaatatttttggcagaaatattctagcagcattaggggtctctttaaattac
ccagtggcaaaggtagaatatacaaaagtaaagttgaaggagggaatggatgggccaagattaaaacaatggcccttatctaaggaaaagatacaagcattaacagagatctgcgagaaaatggaaaaggagggacaattggaacga
gcccctcctacaaatccttataatacaccgacctttgctattaagaaaaagaataaggacaagtggagaatgctcatagatttcagagaattaaacaaaatgacacaagaattcacagaagtacagctgggtatcccacatccggcaggcct
aaaggaaatggaaaggataacagtgctagatataggggatgcctatttttcagtacccctagacccggaatttagacaatacacagcattcactataccatcagtgaacaaccaagaaccaggaaagaggtatatttacaaggtcctgcca
caaggatggaagggatctcctgcaattttccaggcaaccatgaggcaggtattagaacccttcagaaaagctaacccagatgtgcttctcgtccagtatatggatgatttactaattggcagtaacagaggcctaacagagcatgacaaaat
ggtaacacagttaagagacatgctcaacaatctggggttcagtaccccagaagacaagtttcaaaagaatccaccattacagtggatgggatatttgttataccccaagaagtggaaattacagaaaatagaactccctgaaaaggagcga
tggacagtaaatgacatacaaaaactagtgggagttctaaattgggcagcacagatatatccagggattaaaacaaaaaacctgtgcaaaatgatcagagggaaaatgactctgacagaagaagttcagtggacagaattagcagaagc
agaattggcagagaatagaattattctaaatcaagagcaggaaggaagatattacagagaagatgaaccgctagaagcaacggttctaaagaatcaagacaatcagtggagttataagattcatcaaggtgacaggattctcaaggtagg
gaagtttgcaaaaataaaaaatacacatacaaatggaataaggttactggctaatgtagttcaaaagataggaaaagaaagtctagtgatatggggaaagacaccatttttttcaccttccagtagaaagggaggtttgggatcaatggtggac
agattattggcaagctacttggattccagagtgggacttcatctccacaccaccgttaatcaggttagtgtttaatttagtaaaagatccaatagaaaaggaggaagtatattacattgatggctcctgcaatagaaacagtaaagaagggaaa
gcaggatatgtcacagacaggggaaaagaaaaggtcctgccattagaacaggctaccaatcagcaagcagagctgcaggccctgttattggccttgaaagattctccaagtaaggtcaatgtagtaacggactcacagtatgtcttgaac
ataataacaggacagccatcagaatcggattcagatatagtggcacaaattattgaacaactagttcagaaagaagcagtatacataggatgggttccagctcataaaggcataggagggaacaatgaggtagaccgtttagttagtcaag
gaataagacaggtcctgttcctagaaagtatagagccagcacaagaagaccatgacaaatatcatagcaatgtaaaggagttagctcagaaatataacatcccacaattagtagctaaacagatagtaaatgcttgtaacaaatgccaaca
gaagggagaagcgatacatggtcagacaaatgcagaggtaggcacatggcagatggactgcacccacttagaaggaaaggttatcatagtagcagtacatgtggctagtggctttatagaagcagaggtaatacctagggaaacagga
agacagacagcattgttcctcttaaaaatagcatcaagatggcctataaagcatctgcatacagacaacggtgccaatttcacctcccaggaggtgaagatggtagcatggtggttaggagtggaacagtcctttggagtaccgtacaacc
cacaaagtcagggagtagtagaagccatggacttacatcttaagaaaaacatagataaaataagggaacaagcagagtcagtagagaccctagtactaatggcagcacattgcatgaattttaaaagaaggggaggaataggggatatg
actcctgcagaaagaatagtaaatatgatcaccacagaactagaaacccaatacttaaattcacaaaattcaaaatttcaaaattttcgggtttattacagagaaggaagagaccaactgtggaagggaccagcagagctcctctggaaagg
agaaggggctgttgtaattaaggtagggacagaaataaaagtggtaccaagaagaaaagcaaaaatcatcaaagattatggaggaaggcaagagatgggtagcagtgccagtatggaggatcagtag'

4.5.9 Multiple Sequence Aligment Results For POL

The following command for progressive multiple sequence alignment is used via MATLAB;

```
>> polaligned=multialign(pol.'terminalGapAdjust'.true)
```

Table 13. Only the aligned result of the first 99 nucleotide sites were listed for demonstration.

LOCUS NAME	VERSION	MULTIPLE SEQUENCE ALIGMENT
'HIV-1 (Zaire)'	**'K03454'**	'--
'HIV1-NDK (Zaire)'	**'M27323'**	'--
'HIV-2 (Senegal)'	**'M15390'**	'--
'HIV2-MCN13'	**'AY509259'**	'--
'HIV-2UC1 (IvoryCoast)'	**'L07625'**	'--**atgctggaaatgtggaaaacaaggac**
'SIVMM251 Macaque'	**'M19499'**	**'gtgttggaattgtgggaaggagggacactctgcaaggcaatgcagagccccaagaagacagggatgctggaa**
'SIVAGM677A Green monkey'	**'M58410'**	'---**gtgcggcaaaattggccatat**
'SIVlhoest L''Hoest monkeys'	**'AF075269'**	'--**atttgc**
'SIVcpz Chimpanzees Cameroon'	**'AF115393'**	'--
'SIVmnd5440 Mandrillus sphinx'	**'AY159322'**	'--
'SIVAGM3 Green monkeys'	**'M30931'**	'--**accaagaaagatgagatgcttgaaatgtgggaaacca**
'SIVMM239 Simian macaque'	**'M33262'**	**'gtgttggaattgtgggaaagagggacactctgcaaggcaatgcagagccccaagaagacagggatgctggaa**
'CIVcpzUS Chimpanzee'	**'AF103818'**	'--
'SIVmon Cercopithecus Monkeys'	**'AY340701'**	'--
'SIVcpzTAN1 Chimpanzee'	**'AF447763'**	'--
'SIVsmSL92b Sooty Mangabey'	**'AF334679'**	'---------**ctgtggaaagacaggacactctgcgagacagtgcagagctcccagaaggaagggatgctggaaat**

4.5.10 Distance Tables of POL

Table 14. Alignment Score (gag)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	0.00000															
2	0.00236	0.00000														
3	0.41567	0.40509	0.00000													
4	0.42577	0.44237	0.01780	0.00000												
5	0.41977	0.43991	0.13649	0.12740	0.00000											
6	0.41806	0.43353	0.13098	0.13164	0.14655	0.00000										
7	0.41484	0.42278	0.42396	0.45203	0.45670	0.42948	0.00000									
8	0.44177	0.46347	0.43454	0.45119	0.43683	0.44095	0.46994	0.00000								
9	0.12687	0.12995	0.40098	0.41812	0.42807	0.42539	0.40731	0.45470	0.00000							
10	0.30175	0.31492	0.38050	0.42979	0.41233	0.38091	0.38255	0.46287	0.31152	0.00000						
11	0.40577	0.43038	0.39729	0.42133	0.38786	0.41816	0.28879	0.43098	0.42263	0.40153	0.00000					
12	0.41891	0.43821	0.13247	0.13232	0.14788	0.00044	0.42850	0.44619	0.42101	0.38458	0.42333	0.00000				
13	0.14690	0.14535	0.39929	0.41494	0.42296	0.42451	0.42478	0.45071	0.07509	0.31721	0.44687	0.42358	0.00000			
14	0.51688	0.51910	0.47564	0.49820	0.50431	0.49640	0.50897	0.56394	0.53018	0.51631	0.51129	0.50237	0.54309	0.00000		
15	0.23731	0.24344	0.41156	0.44216	0.42402	0.42261	0.40912	0.42712	0.24110	0.33808	0.44441	0.42162	0.24223	0.52087	0.00000	
16	0.41071	0.43563	0.20907	0.20885	0.19477	0.21145	0.44687	0.45735	0.42807	0.38840	0.45173	0.21064	0.41231	0.49585	0.44090	0.00000

Table 15. Nei-Tamura (gag)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	0.00000															
2	0.03011	0.00000														
3	0.76791	0.74940	0.00000													
4	0.77888	0.81116	0.08574	0.00000												
5	0.77193	0.81566	0.31037	0.29670	0.00000											
6	0.76998	0.80207	0.29888	0.29921	0.32435	0.00000										
7	0.76176	0.77672	0.77896	0.82810	0.84721	0.79400	0.00000									
8	0.81664	0.85917	0.81015	0.84783	0.81513	0.83545	0.86186	0.00000								
9	0.29685	0.30236	0.74418	0.77238	0.79552	0.78589	0.75204	0.84043	0.00000							
10	0.57422	0.59959	0.69815	0.78117	0.75230	0.71040	0.70514	0.85045	0.59693	0.00000						
11	0.74829	0.79612	0.73679	0.77524	0.72015	0.77116	0.54645	0.79297	0.78375	0.73627	0.00000					
12	0.76964	0.80809	0.30093	0.29982	0.32586	0.01212	0.79414	0.84142	0.77823	0.71476	0.78072	0.00000				
13	0.32751	0.32654	0.73976	0.76859	0.77920	0.78942	0.78514	0.83912	0.20547	0.60687	0.83422	0.78783	0.00000			
14	0.98975	1.00268	0.90975	0.95423	0.96145	0.96417	0.97859	1.12078	1.05640	1.02225	0.96368	0.98480	1.13058	0.00000		
15	0.46916	0.48085	0.76643	0.82270	0.78468	0.79108	0.75421	0.78826	0.48147	0.63301	0.81789	0.79126	0.47991	1.03165	0.00000	
16	0.75350	0.79863	0.42268	0.42296	0.39930	0.42298	0.82096	0.86243	0.79325	0.71868	0.83389	0.42139	0.76478	0.93358	0.82045	0.00000

Table 16. Jukes Cantor (gag)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	0.00000															
2	0.02752	0.00000														
3	0.48695	0.47822	0.00000													
4	0.49528	0.50895	0.07804	0.00000												
5	0.49033	0.50692	0.23980	0.23030	0.00000											
6	0.48893	0.50167	0.23407	0.23476	0.25011	0.00000										
7	0.48627	0.49282	0.49379	0.51690	0.52075	0.49834	0.00000									
8	0.50845	0.52632	0.50250	0.51621	0.50439	0.50778	0.53164	0.00000								
9	0.22973	0.23298	0.47483	0.48897	0.49718	0.49496	0.48006	0.51910	0.00000							
10	0.39180	0.40299	0.45788	0.49859	0.48420	0.45822	0.45958	0.52583	0.40011	0.00000						
11	0.47879	0.49908	0.47178	0.49162	0.46398	0.48901	0.38071	0.49957	0.49269	0.47529	0.00000					
12	0.48963	0.50552	0.23563	0.23547	0.25145	0.01176	0.49753	0.51209	0.49135	0.46126	0.49327	0.00000				
13	0.25046	0.24889	0.47343	0.48635	0.49296	0.49424	0.49447	0.51582	0.17016	0.40493	0.51266	0.49347	0.00000			
14	0.57035	0.57219	0.53634	0.55493	0.55997	0.55344	0.56381	0.60940	0.58136	0.56988	0.56573	0.55837	0.59205	0.00000		
15	0.33571	0.34116	0.48356	0.50878	0.49384	0.49268	0.48155	0.49639	0.33908	0.42251	0.51063	0.49186	0.34009	0.57365	0.00000	
16	0.48286	0.50340	0.31016	0.30996	0.29691	0.31234	0.51266	0.52128	0.49717	0.46442	0.51666	0.31160	0.48418	0.55299	0.50774	0.00000

Table 17. P Distance (gag)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	0.00000															
2	0.02702	0.00000														
3	0.35818	0.35359	0.00000													
4	0.36250	0.36951	0.07412	0.00000												
5	0.35994	0.36847	0.20525	0.19830	0.00000											
6	0.35921	0.36580	0.20106	0.20157	0.21268	0.00000										
7	0.35782	0.36123	0.36173	0.37352	0.37544	0.36408	0.00000									
8	0.36925	0.37821	0.36622	0.37317	0.36718	0.36891	0.38084	0.00000								
9	0.19788	0.20027	0.35180	0.35923	0.36348	0.36234	0.35456	0.37462	0.00000							
10	0.30518	0.31177	0.34269	0.36421	0.35674	0.34288	0.34362	0.37797	0.31008	0.00000						
11	0.35389	0.36446	0.35017	0.36061	0.34599	0.35925	0.29855	0.36472	0.36117	0.35204	0.00000					
12	0.35958	0.36776	0.20220	0.20209	0.21364	0.01167	0.36367	0.37110	0.36047	0.34453	0.36147	0.00000				
13	0.21293	0.21181	0.35105	0.35787	0.36131	0.36197	0.36209	0.37297	0.15224	0.31290	0.37138	0.36157	0.00000			
14	0.39941	0.40027	0.38315	0.39213	0.39453	0.39142	0.39635	0.41720	0.40452	0.39919	0.39725	0.39377	0.40941	0.00000		
15	0.27064	0.27411	0.35641	0.36942	0.36176	0.36116	0.35535	0.36308	0.27279	0.32303	0.37036	0.36073	0.27343	0.40095	0.00000	
16	0.35604	0.36668	0.25402	0.25389	0.24518	0.25546	0.37138	0.37571	0.36348	0.34623	0.37339	0.25498	0.35673	0.39120	0.36889	0.00000

Table 18. Kimura (gag)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	0.00000															
2	0.02766	0.00000														
3	0.49270	0.48315	0.00000													
4	0.50237	0.51668	0.07939	0.00000												
5	0.49670	0.51377	0.24383	0.23390	0.00000											
6	0.49524	0.50769	0.23905	0.23996	0.25487	0.00000										
7	0.49267	0.49884	0.49980	0.52415	0.52721	0.50315	0.00000									
8	0.51279	0.53114	0.50746	0.52114	0.50852	0.51091	0.53941	0.00000								
9	0.23317	0.23636	0.47908	0.49432	0.50319	0.49935	0.48419	0.52356	0.00000							
10	0.39564	0.40646	0.46374	0.50733	0.49179	0.46231	0.46541	0.53307	0.40333	0.00000						
11	0.48342	0.50394	0.47594	0.49734	0.46870	0.49381	0.38569	0.50435	0.49667	0.48152	0.00000					
12	0.49620	0.51203	0.24075	0.24073	0.25645	0.01180	0.50240	0.51570	0.49561	0.46566	0.49851	0.00000				
13	0.25432	0.25246	0.47722	0.49036	0.49860	0.49797	0.49991	0.51940	0.17381	0.40780	0.51683	0.49716	0.00000			
14	0.57702	0.57805	0.54012	0.55980	0.56594	0.55873	0.56929	0.61814	0.58755	0.57372	0.57309	0.56412	0.59758	0.00000		
15	0.34042	0.34508	0.48779	0.51375	0.49867	0.49583	0.48636	0.50096	0.34172	0.42766	0.51653	0.49501	0.34374	0.57900	0.00000	
16	0.48806	0.50980	0.31467	0.31440	0.30187	0.31807	0.51856	0.52572	0.50209	0.46901	0.52427	0.31740	0.48826	0.55889	0.51298	0.00000

Table 19. Hasegawa (gag)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	0.00000															
2	0.03015	0.00000														
3	0.91953	0.88018	0.00000													
4	0.96203	1.05890	0.08602	0.00000												
5	0.93562	1.04385	0.31657	0.30262	0.00000											
6	0.92818	1.00651	0.30354	0.30391	0.33100	0.00000										
7	0.91409	0.95053	0.95618	1.14202	1.19689	0.98776	0.00000									
8	1.06396	1.31050	1.01551	1.13945	1.03214	1.06271	1.54663	0.00000								
9	0.30238	0.30796	0.86758	0.93134	0.97668	0.96832	0.89132	1.17774	0.00000							
10	0.61297	0.64198	0.79572	0.97876	0.90045	0.80300	0.80198	1.29414	0.63589	0.00000						
11	0.88373	0.99242	0.85504	0.94468	0.82199	0.93332	0.58265	0.99596	0.95653	0.86304	0.00000					
12	0.93122	1.03357	0.30562	0.30488	0.33248	0.01212	0.98241	1.09966	0.94800	0.81299	0.95549	0.00000				
13	0.33504	0.33363	0.86336	0.92231	0.95252	0.96645	0.96190	1.13913	0.20775	0.64957	1.10327	0.96214	0.00000			
14	1.05890	1.05449	1.56058	1.15284	1.11321	1.16337	1.09327	0.95985	1.02308	1.07192	1.07703	1.12426	0.99922	0.00000		
15	0.48744	0.50144	0.90761	1.06473	0.96013	0.95969	0.89590	0.97636	0.50293	0.68720	1.07866	0.95509	0.50039	1.05095	0.00000	
16	0.90088	1.01778	0.43862	0.43850	0.41242	0.43853	1.09817	1.21021	0.98001	0.82419	1.13822	0.43694	0.91118	1.16484	1.05503	0.00000

Table 20. Tajima-Nei (gag)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	0.00000															
2	0.02756	0.00000														
3	0.44277	0.43536	0.00000													
4	0.44964	0.46099	0.07835	0.00000												
5	0.44644	0.45995	0.23372	0.22413	0.00000											
6	0.44195	0.45204	0.23011	0.23096	0.24454	0.00000										
7	0.44060	0.44503	0.44684	0.46553	0.46850	0.44807	0.00000									
8	0.45575	0.46967	0.45342	0.46395	0.45443	0.45436	0.47672	0.00000								
9	0.22327	0.22610	0.43141	0.44351	0.45096	0.44520	0.43378	0.46430	0.00000							
10	0.36422	0.37340	0.41892	0.45232	0.44050	0.41612	0.41924	0.47136	0.37034	0.00000						
11	0.43279	0.44875	0.42866	0.44496	0.42318	0.44109	0.35520	0.44873	0.44345	0.43281	0.00000					
12	0.44268	0.45527	0.23185	0.23162	0.24624	0.01178	0.44735	0.45784	0.44235	0.41871	0.44472	0.00000				
13	0.24228	0.24061	0.43049	0.44100	0.44879	0.44477	0.44625	0.46120	0.16909	0.37413	0.45900	0.44414	0.00000			
14	0.51048	0.51168	0.48109	0.49469	0.49802	0.49332	0.50246	0.53828	0.51639	0.50442	0.50674	0.49753	0.52647	0.00000		
15	0.31744	0.32099	0.43821	0.45842	0.44718	0.44277	0.43507	0.44635	0.31745	0.38983	0.45859	0.44221	0.31974	0.51000	0.00000	
16	0.43680	0.45382	0.29561	0.29547	0.28511	0.29854	0.46061	0.46615	0.44869	0.42144	0.46559	0.29797	0.43802	0.49326	0.45726	0.00000

4.5.11 Translated POL protein-coding gene (Amino Acid Sequence) (designated by AAPOL)

1. 'HIV-1 (Zaire)"K03454'

```
'MRARGIERNCQNWWKWGIMLLGILMTCSAADNLWVTVYYGVPVWKEATTTLFCASDAKSYETEAHNIWATHACVPTDPNPQEIALENVTENFNM
WKNNMVEQMHEDIISLWDQSLKPCVKLTPLCVTLNCSDELRNNGTMGNNVTTEEKGMKNCSFNVTTVLKDKKQQVYALFYRLDIVPIDNDSSTNS
TNYRLINCNTSAITQACPKVSFEPIPIHYCAPAGFAILKCRDKKFNGTGPCTNVSTVQCTHGIRPVVSTQLLLNGSLAEEEVIIRSENLTNNAKNIIAHLN
ESVKITCARPYQNTRQRTPIGLGQSLYTTRSRSIIGQAHCNISRAQWSKTLQQVARKLGTLLNKTIIKFKPSSGGDPEITTHSFNCGGEFFYCNTSGLFNS
TWNISAWNNITESNNSTNTNITLQCRIKQIIKMVAGRKAIYAPPIERNILCSSNITGLLLTRDGGINNSTNETFRPGGGDMRDNWRSELYKYKVVQIEPL
GVAPTRAKRRVVEREKRAIGLGAMFLGFLGAAGSTMGARSVTLTVQARQLMSGIVQQQNNLLRAIEAQQHLLQLTVWGIKQLQARILAVERYLKD
QQLLGIWGCSGKHICTTNVPWNSSWSNRSLNEIWQNMTWMEWEREIDNYTGLIYSLIEESQTQQEKNEKELLELDKWASLWNWFSITQWLWYIKIFI
MIIGGLIGLRIVFAVLSLVNRVRQGYSPLSFQTLLPAPRGPDRPEGTEEEGGERGRDRSVRLLNGFSALIWDDLRSLCLFSYHRLRDLILIAVRIVELLGR
RGWDILKYLWNLLQYWSQELRNSASSLFDAIAIAVAEGTDRVIEIIQRACRAVLNIPRRIRQGLERSLL*'
```

2. 'HIV1-NDK (Zaire)"M27323'

```
'MRAREKERNCQNLWKWGIMLLGMLMTCSAAEDLWVTVYYGVPIWKEATTTLFCASDAKAYKKEAHNIWATHACVPTDPNPQEIELENVTENFN
MWKNNMVEQMHEDIISLWDQSLKPCVKLTPLCVTLNCTDELRNSKGNGKVEEEEKRKNCSFNVRDKREQVYALFYKLDIVPIDNNNRTNSTNYRLI
NCDTSTITQACPKISFEPIPIHFCAPAGFAILKCRDKKFNGTGPCSNVSTVQCTHGIRPVVSTQLLLNGSLAEEEIIIRSENLTNNVKTIIVQLNASIVINCT
RPYKYTRQRTSIGLRQSLYTITGKKKKTGYIGQAHCKISRAEWNKALQQVATKLGNLLNKTTITFKPSSGGDPEITSHMLNCGGDFFYCNTSRLFNST
WNQTNSTGFNNGTVTLPCRIKQIVNLWQRVGKAMYAPPIEGLIKCSSNITGLLLTRDGGANNSSHETIRPGGGDMRDNWRSELYKYKVVKIEPIGVA
PTKARRRVVEREKRAIGLGAVFLGFLGAAGSTMGAASVTLTVQARQLMSGIVHQQNNLLRAIEAQQHLLQLTVWGIKQLQARVLAVERYLRDQQL
LGIWGCSGRHICTTNVPWNSSWSNRSLDEIWQNMTWMEWEREIDNYTGLIYSLIEESQIQQEKNEKELLELDKWASLWNWFSITKWLWYIKLFIMIV
GGLIGLRIVFAVLSVVNRVRQGYSPLSFQTLLPVPRGPDRPEEIEEEGGERGRDRSIRLVNGLFALFWDDLRNLCLFSYHRLRDSILIAARIVELLGRRG
WEALKYLWNLLQYWSQELRNSASSLLDTIAIAVAERTDRVIEVVQRACRAILNVPRRIRQGLERLLL*'
```

3. 'HIV-2 (Senegal)"M15390'

```
'MMNQLLIAILLASACLVYCTQYVTVFYGVPTWKNATIPLFCATRNRDTWGTIQCLPDNDDYQEITLNVTEAFDAWNNTVTEQAIEDVWHLFETSIKP
CVKLTPLCVAMKCSSTESSTGNNTTSKSTSTTTTTPTDQEQEISEDTPCARADNCSGLGEEETINCQFNMTGLERDKKKQYNETWYSKDVVCETNNS
TNQTQCYMNHCNTSVITESCDKHYWDAIRFRYCAPPGYALLRCNDTNYSGFAPNCSKVVASTCTRMMETQTSTWFGFNGTRAENRTYIYWHGRDN
RTIISLNKYYNLSLHCKRPGNKIVKQIMLMSGHVFHSHYQPINKRPRQAWCWFKGKWKDAMQEVKETLAKHPRYRGTNDTRNISFAAPGKGSDPEV
AYMWTNCRGEFLYCNMTWFLNWIENKTHRNYAPCHIKQIINTWHKVGRNVYLPPREGELSCNSTVTSIIANIDWQNNNQTNITFSAEVAELYRLELG
DYKLVEITPIGFAPTKEKRYSSAHGRHTRGVFVLGFLGFLATAGSAMGAASLTVSAQSRTLLAGIVQQQQLLDVVKRQQELLRLTVWGTKNLQAR
VTAIEKYLQDQARLNSWGCAFRQVCHTTVPWVNDSLAPDWDNMTWQEWEKQVRYLEANISKSLEQAQIQQEKNMYELQKLNSWDIFGNWFDLT
```

```
SWVKYIQYGVLIIVAVIALRIVIYVVQMLSRLRKGYRPVFSSPPGYIQQIHIHKDRGQPANEETEEDGGSNGGDRYWPWPIAYIHFLIRQLIRLLTRLYSI
CRDLLSRSFLTLQLIYQNLRDWLRLRTAFLQYGCEWIQEAFQAAARATRETLAGACRGLWRVLERIGRGILAVPRRIRQGAEIALL*'
```

4. 'HIV2-MCN13"AY509259'

```
'MMGGRNQLLVAILLTSTCLIYCTNYVTVFYGIPAWRNASIPLFCATKNRDTWGTIQCLPDNDDYQEITLNVTEAFDAWDNTVTEQAIEDVWNLFETS
IKPCVKLTPLCVAMRCNNTDARNTTTPTTASPRTIKPVTEISENSSCIRANNCSGLGEEEVVNCQFNMTGLERDKKKQYSETWYSKDVVCEGNGTTD
TCYMNHCNTSVITESCDKHYWDAMRFRYCAPPGFALLRCNDTNYSGFAPNCSKVVAATCTRMMETQTSTWFGFNGTRAENRTFIYWHGRDNRTII
SLNKYYNLTIHCKRPGNKTVVPITLMSGLRFHSQPVINKRPRQAWCWFKGEWKGAMQEVKETLAKHPRYKGTNETKNINFTAPGKGSDPEVAYMW
TNCRGEFLYCNMTWFLNWIENKTHRNYVPCHIRQIINTWHKVGKNVYLPPREGELTCNSTVTSIIANIDANGNNTNITFSAEVAELYRLELGDYKLVE
ITPIGFAPTAEKRYSSTPMRNKRGVFVLGFLGFLATAGSAMGAASLTLSAQSRTLLAGIVQQQQQLLDVVKRQQEMLRLTVWGTKNLQARVTAIEK
YLKDQAQLNSWGCAFRQVCHTTVPWVNDTLTPEWNNMTWQEWEGKIRDLEANISQQLEQAQIQQEKNMYELQKLNSWDVFGNWFDLTSWIKYI
QYGVYIIIGIVVLRIVIYIVQMLSRLRKGYRPVFSSPPGYLQQIHIHKDWEQPAREETEEDVGNNVGDSSWPWPIRYIHFLIHQLIRLLAGLYNICRNLLS
RISLTLRPVFQSLQRALTAIRDWLRTDAAYLQYGCEWIQGAFQAFARATRETLAGTWRDLWGALQRIGRGILAVPRRIRQGAEIALL*'
```

5. 'HIV-2UC1 (IvoryCoast)"L07625'

```
'MAHTSNHLFILLLLLISVYGFLGHKKNYVTVFYGIPAWRNATVPLFCATTNRDTWGTVQCLPDNGDYTEISVNITEAFDAWNNTVTEQAVDDVWSLF
ETSIKPCVKLTPLCVAMRCNNTGTNTTTKPITTPITTTKPSENLLNDTSPCIKNDTCPGIGLENTVDCYFNMTGLRRDEKKQYKDTWYEKDLECNGNS
TSTICYMRTCNTSVIQESCDKHYWDSLRFRYCAPPGYALLRCNDTNYSGFMPKCSKVVVSSCTRMMETQTSTWFGFNGTRTENRTYMYWHSKDNR
TIISLNKYYNLTMHCRRPGNKTVIPITIMSGLNFHSQPLNTRPRQAWCWFKGNWIEAIREVKETIIKHPRYKGTNNTERIRLVGPSAGSDPEVRHMWT
NCRGEFFYCNMTWFLNWVENRTGTTQKNYVTCHIKQIVNTWHKVGKYVYLPPREGTLSCNSSVTSLIANIDVYYDGNDTKTNITMSAEVGELYRLE
LGDYKLVEITPIGFAPTEIKRYSSTTPRNKRGVMVLGFLGLLAMAGSAMGATSLTLSAQSRTLLAGIVQQQQQLLDVVKRQQELLRLTVWGTKNLQT
RVTAIEKYLKDQALLNSWGCAFRQVCHTTVPWPNETLTPDWENMTWQQWEKRVNFLDANITALLEEAQIQQERNMYELQKLNSWDVFGNWFDFT
SWMAYIRLGLYVVAGLIVLRIVIYIMQMLARLRKGYRPVFSSPPSYTQQIPIRKHRGQPANEETEDEGGNEGAYRSWPWQIEYAHFLIRQLRNLLIWL
YNGCRNLLLKTSQILQPALQPLRLSLAYLQYGISWFQEAIQAATRAARETLANTGRALWKALRRTAEAIIAIPRRIRQGLELALL*'
```

6. 'SIVMM251 Macaque"M19499'

```
'MGCLGNQLLIAILLLSVYGIYCTQYVTVFYGVPAWRNATIPLFCATKNRDTWGTTQCLPDNGDYSELALNVTESFDAWENTVTEQAIEDVWQLFET
SIKPCVKLSPLCITMRCNKSETDRWGLTKSSTTITTAAPTSAPVSEKIDMVNETSSCIAQNNCTGLEQEQMISCKFTMTGLKRDKTKEYNETWYSTDL
VCEQGNSTDNESRCYMNHCNTSVIQESCDKHYWDTIRFRYCAPPGYALLRCNDTNYSGFMPKCSKVVVSSCTRMMETQTSTWFGFNGTRAENRTY
IYWHGRDNRTIISLNKYYNLTMKCRRPGNKTVLPVTIMSGLVFHSQPINDRPKQAWCWFGGKWKDAIKEVKQTIVKHPRYTGTNNTDKINLTAPGG
GDPEVTFMWTNCRGEFLYCKMNWFLNWVEDRDVTTQRPKERHRRNYVPCHIRQIINTWHKVGKNVYLPPREGDLTCNSTVTSLIANIDWTDGNQT
SITMSAEVAELYRLELGDYKLVEITPIGLAPTDVKRYTTGGTSRNKRGVFVLGFLGFLATAGSAMGAASLTLTAQSRTLLAGIVQQQQQLLDVVKRQ
QELLRLTVWGTKNLQTRVTAIEKYLKDQAQLNAWGCAFRQVCHTTVPWPNASLTPDWNNDTWQEWERKVDFLEENITALLEEAQIQQEKNMYEL
QKLNSWDVFGNWFDLASWIKYIQYGIYVVVGVILLRIVIYIVQMLAKLRQGYRPVFSSPPSYFQ*THTQQDPALPTREGKEGDGGEGGGNSSWPWQI
EYIHFLIRQLIRLLTWLFSNCRTLLSRAYQILQPILQRLSATLRRVREVLRTELTYLQYGWSYFHEAVQAGWRSATETLAGAWRDLWETLRRGGRWI
LAIPRRIRQGLELTLL*'
```

7. 'SIVAGM677A Green monkey"M58410'

```
'MGRLLIKILIIAIGISIGIGNLYVTVFYGIPVWKNSTVQAFCMTPNTNMWATTNCIPDDHDNTEVPLNITEAFEAWDNPLVKQAESNIHLLFEQTMRPC
VKLSPICIKMSCVELNGTATTKATTTATTTMTTPCQNCSTEQIEGEMAEEPASNCTFAIAGYQRDVKKNYSMTWYDQELVCNNKTGSEKGSKDCYM
IHCNDSVIKEACDKTYWDTLRVRYCAPAGYALLKCNDKDYRGFAPKCKNVSVVHCTRLINTTITTGIGLNGSRSENRTEIWQKGGNDNDTVIIKLNK
FYNLTVRCRRPGNKTVLPVTIMAGLVFHSQKYNTRLKQAWCHFQGDWKGAWKEVREEVKKVKNLTEVSIENIHLRRIWGDPESANFWFNCQGEFF
YCKMDWFINYLNNRTEDAEGTNRTCDKGKPGPGPCVQRTYVACHIRQVVNDWYTVSKKVYAPPREGHLECNSSVTALYVAIDYNNKSGPINVTLS
PQVRSIWAYELGDYKLVEITPIGFAPTDVRRYTGPTREKRVPFVLGFLGFLGAAGTAMGAAATTLTVQSRHLLAGILQQQKNLLAAVEQQQQLLKLT
IWGVKNLNARVTALEKYLEDQARLNSWGCAWKQVCHTTVPWKYNNTPKWDNMTWLEWERQINALEGNITQLLEEAQNQESKNLDLYQKLDDW
SGFWSWFSLSTWLGYVKIGFLVIVIILGLRFAWVLWGCIRNIRQGYNPLPQIHIHSSAERPDNGGGQDRGGESSSSKLIRLQEESSTPSRINNWWLNFKS
CSLRIRTWCYNICLTLLIFIRTAVGYLQYGLQQLQEAATGLAQALARAAREAWGRLGAIVRSAYRAVINSPRRVRQGLEKVLG*'
```

8. 'SIVlhoest L"Hoest monkeys"AF075269'

```
'MACPGLGILLLLLGIIWGKQYVTVFYGVPNWDDNVSVPLICASANTSLWVTTSCLPDLQSYAEVPIYNISENFTIPVKDNQVIQQAWSAMNAMVDSI
MKPCVKINPYCVRMQCGEVTKTPTTTPKTTTQMPCFINEQVTVKNPGNETRLEEDLNCTRGLNETTERNAECQYNVTGLCRDCRTEIKQSFRYDDVT
CSGERENRTCYMTHCNDSIITQDCNKGVMQNAYFRLCAPAGYMLLRCNEQLNFSKKCENITATPCTGYMLSSVSSFFGFNGTNHTRDELIPLTPNKM
EDLNGAKFVYKVAGKWGLIIRCIRKGNRSEVSTISSTGYLFYYGLEHGSRLRLAQCKFEGQWGRMFNNLGKMLKELNAEAMNYTEGTGTCDSKKT
TCGRKLKGLPIANMTRHGADLATEMLMHTCGEEMFFCNVTRIFQEWNNKNSDKWYPWANCHIKSIIDDWATIGKKIYLPPTSGFNNRIRCTHRVTE
MFFEMEKWEPHEDLGGNLSIKFLPPSWETNQFVAEGSKYKLIKLNPIGFAPTDEHRYAPRGRQTRAAPLALGALGLLSAAGTAMGLVSTILTVQAQA
VLQGILQQQKQLLVLVEKQQELLRLTIWGVKNLQARLTALEEYVKHQALLASWGCQWKQVCHTNVEWTYNITPNWTKDTWREWESKVAIYDKNI
TSLLQEAYTTELENQNKFKKLQEFNFWSWLDISHWFTYVKYAVLIILVIIGLRVLSFIIQNVVKMCRGYRVLSPSVYIEQDYKWEKEENQEQPDREEE
KGADTETIYINLEQCKKESSRPLWNVDWNEPLQDSLLVTLLKWLKEGGILLLSLVWQSLSWLWHLLILFFQNGQRLWQTSSRWMVENAQKIQSWL
REKCRRNRGQLSSTDRKNIQLGKKKRWRLRFGGRSGISSEATETAL*'
```

9. 'SIVcpz Chimpanzees Cameroon"AF115393'

```
'MKVMEKIKGSLIEKYMFMGLIIPCLTGSDQLWATVYYGVPVWKEADTTLFCASEASALNKEAHNIWASQACVPTDPNPQEVQIPNVTENFNMWNN
TMAEQMNTDIISLWDQSLKPCVKLTPICVTMECRKVTFNSTSNRNKTSTMTTNSPNEKKDSTVKNCTFNMTTEVRDKEKKVYSLFYVDDLVLIDND
TDTYRLINCNTTAITQACPKTSFEPIPIHYCATPGYAIMKCNMPNFNGTGTGRCNNISTVQCTHGIRPVVTTQLILNGSVAENKTIARRNGYNFLIQFQK
TVSINCTRPGNNSRGQIQIGPGMTFYNIENIVGDTRRAYCQINRTVWDERLNETGQALRELFTNLTQVNFTVSPGGDPEVTNMMFNCGGEFFYCNTTT
LFNYTWKNNNITKGDNTTFFPCRIRQIVNSWMRVGKGIYAPPIRGVISCTSNVTGIILETGHGINNSITNITLYPTGGNMVDLWRLELHKYKVVSIEPIG
VAPSKAKRHTVSREKRAAFGLGALFLGFLGAAGSTMGAASVVLTVQARQLLSGIVQQQNNLLRAIEAQQHLLQLSVWGIKQLQARVLAVERYLRD
QQILGLWGCSGKAICYTTVPWNNTWSANTSFDEIWNNLTWQDWDKRVKNYSGVIFSLIEQAQEQQNTNEKSLLELDQWSSLWNWFDITRWLWYIK
LFIMIVAGLVGIRIVGAIISFVAKVRQGYSPLSLQTLIPTTREPDRPGGTEEDVGEPGKGRSIRLVSGFLALIWEDLRNLLLFCYHRLRDLLLILGRTLEN
LGQSLNKGLQQLRNFSRYLWGVITYWGRELQTSAISLLDATAIAVAEGTDRILEVAQIIGRGILHIPRRIRQGLERSLL*'
```

10. 'SIVmnd5440 Mandrillus sphinx"AY159322'

```
'MLRYLRYIVLGIIVSVIVGEQWVTVYYGTPKWHKARTHLFCATDNNSFWVTTSCVPSLLHYEEQHIPNITENFTGPIEENEIVTQAWGAISSMIDAVL
KPCVRLTPYCVKMKCTEGQNETEQATAKTTTPVPTTTTPSTTTSSSTNKTTTPVLVVEKQNNETTTQQNRVCKFNTTGLCRDCKLEIEENFRYEDVTC
TKLNKTGSATNSTEPEYECYMTSCNATVITQDRNKASTDRMTFRLCAPPGFVLLKCNEKLNKTKLCGNVSAVQCTAPLPATISTMFGFNGTKHDYD
ELIQTNPRKGKDEFHDHKYVYRVDKKWGLQVRCRRKGNRSIISTPSSTGLLFYHGLEPGKNLRKGKCQLEGKWGQALHSLSLELRKINDSIYKDNH
NMTCKSSNNKKNTTGCHLKTISISESTVKGEPGAETIMLLCGGEYFFCNWTKIWKAWNSKQSSVWYPYMSRNIRQIVGDWHKVGKKIYMPPVSGFN
NEIRCTNDVTEMFFEVQKTDDDNGYIIKFIPQDWIQNQYTAVGAHYKLVKVDPIGFAPTDIHRHHLPNTRQKRGAVLLGMLGLLGLAGSAMGSVAV
```

```
ALTVQSQTLLNGIVEQQKVLLSLIDQHSELLKLTIWGVKNLQVRLTALEEYVADQSRLSVWGCSFSQVCHTSVKWPNNSIVPNWTSETWLEWDRRV
NSIVTNMTIDLQRAYELEQRNIFELQKLGDLNFHGLTGFDLTWWLKYVKIGLLVVVVIIGLRMLACLWSVLGKFRQGYRPLPYVFKGDYLRPHNLK
RPDREGGEEPDLEKQNIKSESSRQESRKPWKPEQVRSWLKRSTLYIWLKNLQAVIEYGWQELKAAGAALYQVLQGFAQRLWSRGYQLGLSGVRGA
AAFGRGIWNIPRRIRQGAEALLN*'
```

11. 'SIVAGM3 Green monkeys"M30931'

```
'MKLTLLIGILLIGIGVVLNTRQQWVTVFYGVPVWKNSSVQAFCMTPTTRLWATTNSIPDDHDYTEVPLNITEPFEAWADRNPLVAQAGSNIHLLFEQ
TLKPCVKLSPLCIKMSCVELNSSEPTTTPKSTTASTTNITASTTTLPCVQNKTSTVLESCNETIIEKELNEEPASNCTFAMAGYVRDQKKKYSVVWNDA
EIMCKKGNNSNRECYMIHCNDSVIKEACDKTYWDELRLRYCAPAGFALLKCNDYDYAGFKTNCSNVSVVHCTNLINTTVTTGLLLNGSYSENRTQI
WQKHRVSNDSVLVLFNKHYNLTVTCKRPGNKTVLPVTIMAGLVFHSQRYNTRLRQAWCHFQGNWRGAWKEVKNEIVKLPKDRYQGTNDTEEIYL
QRLFGDPEAANLWFNCQGEFFYCKMDWFLNYLNNRTVDPDHNPCNGTKGKGKAPGPCAQRTYVACHIRSVINDWYTLSRKTYAPPREGHLQCTST
VTGMSVELNYNSKNRTNVTLSPQIETIWAAELGRYKLVEITPIGFAPTEVRRYTGGHDRTKRVPFVLGFLGFLGAAGTAMGAAATALTVQSQHLLAG
ILQQQKNLLAAVEAQQQMLKLTIWGVKNLNARVTALEKYLEDQARLNAWGCAWKQVCHTTVPWQWNNRTPDWNNMTWLEWERQISYLEGNIT
TQLEEARAQEEKNLDAYQKLSSWSDFWSWFDFSKWLNILKIGFLDVLGIIGLRLLYTVYSCIARVRQGYSPLSPQIHIHPWKGQPDNAEGPGEGGDK
RKNSSEPWQKESGTAEWKSNWCKRLTNWCSISSIWLYNSCLTLLVHLRSAFQYIQYGLGELKAAAQEAVVALARLAQNAGYQIWLACRSAYRAIIN
SPRRVRQGLEGILN*'
```

12. 'SIVMM239 Simian macaque"M33262'

```
'MGCLGNQLLIAILLLSVYGIYCTLYVTVFYGVPAWRNATIPLFCATKNRDTWGTTQCLPDNGDYSEVALNVTESFDAWNNTVTEQAIEDVWQLFET
SIKPCVKLSPLCITMRCNKSETDRWGLTKSITTTASTTSTTASAKVDMVNETSSCIAQDNCTGLEQEQMISCKFNMTGLKRDKKKEYNETWYSADLV
CEQGNNTGNESRCYMNHCNTSVIQESCDKHYWDAIRFRYCAPPGYALLRCNDTNYSGFMPKCSKVVVSSCTRMMETQTSTWFGFNGTRAENRTYI
YWHGRDNRTIISLNKYYNLTMKCRRPGNKTVLPVTIMSGLVFHSQPINDRPKQAWCWFGGKWKDAIKEVKQTIVKHPRYTGTNNTDKINLTAPGG
GDPEVTFMWTNCRGEFLYCKMNWFLNWVEDRNTANQKPKEQHKRNYVPCHIRQIINTWHKVGKNVYLPPREGDLTCNSTVTSLIANIDWIDGNQT
NITMSAEVAELYRLELGDYKLVEITPIGLAPTDVKRYTTGGTSRNKRGVFVLGFLGFLATAGSAMGAASLTLTAQSRTLLAGIVQQQQQLLDVVKRQ
QELLRLTVWGTKNLQTRVTAIEKYLKDQAQLNAWGCAFRQVCHTTVPWPNASLTPKWNNETWQEWERKVDFLEENITALLEEAQIQQEKNMYEL
QKLNSWDVFGNWFDLASWIKYIQYGVYIVVGVILLRIVIYIVQMLAKLRQGYRPVFSSPPSYFQQTHIQQDPALPTREGKERDGGEGGGNSSWPWQI
EYIHFLIRQLIRLLTWLFSNCRTLLSRVYQILQPILQRLSATLQRIREVLRTELTYLQYGWSYFHEAVQAVWRSATETLAGAWGDLWETLRRGGRWIL
AIPRRIRQGLELTLL*'
```

13. 'CIVcpzUS Chimpanzee"AF103818'

```
'MKVMEKKKRLWLSYCLLSSLIIPGLSSLWATVYYGVPVWRDVETTLFCASDAKAYKQEAHNIWATQACVPTDPNPQEVHLPNVTEKFDMWENNM
AEQMQEDIISLWDQSLKPCIKLTPLCVTMTCLNPDSNSSAVNTTDIMRNCSFNITTELRDKKKQVYSLFYVDDLAHINNNTYRLINCNTTAITQACPKT
SFEPIPIHYCAPPGFAILKCNEKDFKGKGECKNVSTVQCTHGIKPVVTTQLIINGSLATKNVTVRSKNFADIILVQFSEGVNMTCIRPGNNTVGNVQLGP
GMTFYNIPKIVGDVREAHCNISKLTWEKQRKYTLEIIKKEANLTKVELIPNAGGDPEVVNMMLNCGGEFFYCNTIPLFNMTYNNTDNTTITLKCRIRQ
IVNQWMRVGKGIFAPPIKGVLSCNSNITGMILDISISAVNNDSRNITVMPTGGDMTALWKNELHKYKVVSIEPIGVAPGKAKRHTVKREKRAAFGLG
ALFLGFLGAAGSTMGAASVVLTVQARQLLSGIVQQQNNLLRAIEAQQHLLQLSVWGIKQLQARVLAVERYLKDQQILGLWGCSGKTICYTTVPWN
DTWSNNLSYDAIWGNLTWQEWDRKVRNYSGTIFSLIEQAQEQQNTNEKSLLELDQWSSLWNWFDITNWLWYIKIFLIVVASLVGIRIVGVIFSLVAK
VRQGYSPLSLQTLFPTTREPDRPEGTEEGAGKTDNVRSTRLVSGFLALVWEDLRNLLIFLYHRLEDLLLILRRTVQILGQNINKGLQLLNELRARCWG
VIAYWARELKVSATSLLDTTAIAVAEGTDRIIELTRRLFLGIIHIPRRIRQGLERSLL*'
```

14. 'SIVmon Cercopithecus Monkeys"AY340701'

```
'MREGDMTQGLTQNLNQSNMRKLMALSTLCLMMTLVNGSNWTTVYYGVPVWKPATPPLFCASDPNYGSKEAGNNWLASSCLPTDPTPQSLYLNIT
EEFNAYQNYMVEEMVEDMKSLFSQALKPCVKLTPMCVRMLCVEVNTVSNASTTPAPSTPTPWGNWGGNGTGQPVYNCSFNQTTEFRDKKKQMYS
LFWKEDIMKEEGSNGSHYYILNCNTSYITQACEKSNYEPVPLHYCAPPGYALLRCDDPAFTGQGSCSNVSAVTCTHAIQPIVATWFQLNSTGNAPNTT
VMMNKQKNESIVVRLAKHLHVNITCIRPGNKTIRNLQIGAGMTFYSQLIVGGNTRKAYCKVNKTQWETALQAVHEAVKTEWEKKNNGTNVTTISW
RFQPQGDKEVQTHWFNCQGEFFYCNVSALFINRRTNKTDGISPFDVNNKPNTTYHGGWLACTIRQMVTQWGYVSKSIYLPPRKGHVQCTSNITALLI
TGELYQNNVTLVPSAQVSDSWRSELSRYKVVEIDPLSMAPTTAQRRTGVHREKRAITLGMAFLGFLSTAGGTMGAAATALTVQSRSLLAGIVQQQE
NLLRAVTAQQSLLQLTVWGVKQLQARLTAVEKFIKDQTLLNAWGCANKAVCHTTVPWNNSWAKGHFPEWDNMTWQQWSELVDNDTMTIQQLL
EAAQEQQGKNQHELMKPGQWDFLWNWFDISKWLWYIKIFIIVVAALIGLRILMFILGVISRLGQGYSLLSSQIPIPSHAGQPTPDGTGAGGGDGSNSRS
PAYLKGFFTIIWEDLRNLVLWTYQILKDSVLVIYRILQRVSQRLPPLLHIRLLQLWESLRRLLAYCQYGIQELQAAVTSLLDALARFTIVWTDALLHAG
GRLWRAIVAIPRRIRQGAEIFLN*'
```

15. 'SIVcpzTAN1 Chimpanzee"AF447763'

```
'MKNLIGITLILIITILGIGFSTYYTTVFYGVPVWKEAQPTLFCASDADITSRDKHNIWATHNCVPLDPNPYEVTLANVSIRFNMEENYMVQEMKEDILS
LFQQSFKPCVKLTPFCIKMTCTMTNTTNKTLNSATTTLTPTVNLSSIPNYEVYNCSFNQTTEFRDKKKQIYSLFYREDIVKEDGNNNSYYLHNCNTSVI
TQECDKSTFEPIPIRYCAPAGFALLKCRDQNFTGKGQCSNVSVVHCTHGIYPMIATALHLNGSLEEEETKAYFVNTSVNTPLLVKFNVSINLTCERTGN
NTRGQVQIGPGMTFYNIENVVGDTRKAYCSVNATTWYRNLDWAMAAINTTMRARNETVQQTFQWQRDGDPEVTSFWFNCQGEFFYCNLTNWTN
TWTANRTNNTHGTLVAPCRLRQIVNHWGIVSKGVYLPPRRGTVKCHSNITGLIMTAEKDNNNSYTPQFSAVVEDYWKVELARYKVVEIQPLSVAPR
PGKRPEIKANHTRSRRDVGIGLLFLGFLSAAGSTMGAASIALTAQARGLLSGIVQQQQNLLQAIEAQQHLLQLSVWGIKQLQARMLAVEKYIRDQQL
LSLWGCANKLVCHSSVPWNLTWAEDSTKCNHSDAKYYDCIWNNLTWQEWDRLVENSTGTIYSLLEKAQTQQEKNKQELLELDKWSSLWDWFDIT
QWLWYIKIAIIIVAGLVGLRILMFIVNVVKQVRQGYTPLFSQIPTQAEQDPEQPGGIAGGGGGRDNIRWTPSPAGFFSIVWEDLRNLLIWIYQTFQNFIW
ILWISLQALKQGIISLAHSLVIVHRTIIVGVRQIIEWSSNTYASLRVLLIQAIDRLANFTGWWTDLIIEGVVYIARGIRNIPRRIRQGLELALN*'
```

16. 'SIVsmSL92b Sooty Mangabey"AF334679'

```
'MACPGLHLLIDILFLSVLGTWCAQYVTIFYGIPAWRNATIPLFCATQNRDTWGTVQCLPDNGDYSELALNVTEAFDAWDNTVTEQAIEDVWNLFET
SIKPCVKLTPLCIAMKCNKNETDRWGLTRAATTTSSPTTTSPLTAASPSGEEIVNDTMSCTKNNNCSGIEQEPMIGCQFNMTGLKRDQKRQYNETWY
SRDLVCEQGGNESSRCYMNHCNTSVIQESCDKHYWDAIRYRYCAPPGYALLRCNDTNYSGFAPNCSKVVVSSCTRMMETQTSTWFGFNGTRAENR
TYIYWHGGSNRTIISLNKYYNLTIKCRRPGNKTVLPVTIMSGLVFHSQPINERPRQAWCWFGGKWREAMQEVKKTIVKHPRYTGTNDTRKINLTAPG
GGDPEVTFMWTNCRGEFLYCKMNWFLNWVEDRNTSSPRWTTQTKKEQHKRNYVPCHIRQIINTWHRVGKNVYLPPREGDLTCNSTVTSLIANIDWI
DNNETNITMSAEVAELYRLELGDYKLVEITPIGMAPTHVKRYTTSTSKNKRGVFVLGFLGFLATAGSAMGAASLTLTAQSRTLLAGIVQQQQQLLDV
VKRQQELLRLTVWGTKNLQTRVTAIEKYLKDQAQLNSWGCAFRQVCHTTVLWPNDSLVPDWNNMTWQEWEKKVEFLEANITQMLEEARLQQEK
NMYELQKLNSWDVFGNWFDLTSWVRYIQYGVFLVIGIVLLRIVIYVVQMLSRLRQGYRPVFSSPPSYHQQIHIQRDQELPAKEGTEGEGGNGGGYRL
WPWQIEYIHFLIRQLIRILTWLYNNLTRLASRAYQNLQQLCQRLSEISQPIRELVRREAGYIRYGWNYFIEACQEAWRSAQEAIVGAWGLIWETLGRV
GRGIAAIPRRIRQGLELMLN*'
```

4.5.12 Distance Tables of Translated POL protein-coding gene (Amino Acid Sequence) (designated by AAPOL)

Table 21. P-Distance (AAPOL)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	0.00000															
2	0.03614	0.00000														
3	0.40252	0.40023	0.00000													
4	0.41009	0.43713	0.05815	0.00000												
5	0.40680	0.43114	0.14481	0.18340	0.00000											
6	0.41886	0.43404	0.14139	0.16192	0.18243	0.00000										
7	0.39145	0.41916	0.40365	0.44791	0.45853	0.43689	0.00000									
8	0.42354	0.45045	0.41082	0.44926	0.45605	0.43929	0.44801	0.00000								
9	0.16539	0.18943	0.40068	0.44862	0.44960	0.45255	0.42885	0.45771	0.00000							
10	0.31718	0.33467	0.37832	0.41741	0.40653	0.39900	0.39268	0.43198	0.35159	0.00000						
11	0.41009	0.43143	0.38743	0.43542	0.42678	0.42539	0.30221	0.45020	0.44048	0.38393	0.00000					
12	0.41667	0.43273	0.14367	0.16618	0.18365	0.00946	0.43659	0.44193	0.45219	0.40279	0.42802	0.00000				
13	0.17634	0.18862	0.39220	0.44279	0.44279	0.43932	0.42189	0.44611	0.13532	0.35465	0.44311	0.43800	0.00000			
14	0.46820	0.47629	0.43558	0.48269	0.48123	0.47347	0.46542	0.50250	0.47295	0.47183	0.47421	0.47557	0.46667	0.00000		
15	0.27711	0.30461	0.39908	0.43043	0.42342	0.42409	0.40140	0.43876	0.30800	0.34573	0.41466	0.42381	0.29288	0.46559	0.00000	
16	0.41209	0.42656	0.22414	0.24876	0.24561	0.23727	0.43206	0.44488	0.44132	0.39400	0.42162	0.23946	0.42685	0.45399	0.41860	0.00000

Table 22. Jukes-Cantor Distance (AAPOL)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	0.00000															
2	0.03685	0.00000														
3	0.52358	0.51961	0.00000													
4	0.53680	0.58561	0.06001	0.00000												
5	0.53103	0.57458	0.15712	0.20377	0.00000											
6	0.55236	0.57991	0.15309	0.17752	0.20257	0.00000										
7	0.50456	0.55290	0.52554	0.60579	0.62611	0.58518	0.00000									
8	0.56078	0.61062	0.53809	0.60836	0.62134	0.58963	0.60598	0.00000								
9	0.18171	0.21128	0.52040	0.60714	0.60901	0.61461	0.57041	0.62453	0.00000							
10	0.38596	0.41259	0.48248	0.54977	0.53056	0.51749	0.50666	0.57612	0.43908	0.00000						
11	0.53680	0.57512	0.49775	0.58246	0.56664	0.56411	0.36376	0.61013	0.59184	0.49185	0.00000					
12	0.54845	0.57750	0.15577	0.18267	0.20409	0.00951	0.58462	0.59455	0.61393	0.52405	0.56889	0.00000				
13	0.19507	0.21027	0.50584	0.59615	0.59615	0.58968	0.55780	0.60239	0.14599	0.44394	0.59677	0.58723	0.00000			
14	0.64499	0.66107	0.58274	0.67400	0.67102	0.65544	0.63951	0.71515	0.65439	0.65217	0.65690	0.65964	0.64197	0.00000		
15	0.32763	0.36728	0.51763	0.57329	0.56056	0.56176	0.52164	0.58863	0.37228	0.42982	0.54488	0.56127	0.35016	0.63985	0.00000	
16	0.54033	0.56624	0.25565	0.28843	0.28418	0.27299	0.57628	0.60009	0.59342	0.50891	0.55731	0.27592	0.56677	0.61738	0.55191	0.00000

Table 23. Alignment Score Distance (AAPOL)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	0.00000															
2	0.00095	0.00000														
3	0.13367	0.13137	0.00000													
4	0.13891	0.16632	0.00145	0.00000												
5	0.14264	0.16983	0.01344	0.02639	0.00000											
6	0.14578	0.16444	0.01157	0.01888	0.02318	0.00000										
7	0.12749	0.15978	0.13150	0.18314	0.19860	0.17720	0.00000									
8	0.15682	0.18727	0.14800	0.19055	0.19606	0.18804	0.18751	0.00000								
9	0.01924	0.02986	0.14147	0.18153	0.19258	0.18708	0.16905	0.19472	0.00000							
10	0.08059	0.09934	0.11460	0.15373	0.15485	0.14640	0.14449	0.17425	0.10819	0.00000						
11	0.14311	0.17152	0.12739	0.17479	0.17366	0.16992	0.07787	0.18676	0.18223	0.13834	0.00000					
12	0.14473	0.16391	0.01194	0.01983	0.02392	0.17657	0.17843	0.19045	0.18817	0.14866	0.17318	0.00000				
13	0.01908	0.02622	0.12983	0.17457	0.17805	0.17327	0.15844	0.18772	0.01386	0.10580	0.17581	0.17305	0.00000			
14	0.19312	0.20872	0.16606	0.21732	0.22081	0.20578	0.19948	0.23246	0.21401	0.20350	0.20841	0.21005	0.20078	0.00000		
15	0.05568	0.07337	0.13364	0.16567	0.16195	0.16287	0.14852	0.17930	0.08080	0.10139	0.15538	0.16357	0.06974	0.20051	0.00000	
16	0.13787	0.15845	0.03234	0.04408	0.04451	0.04362	0.17628	0.18553	0.17560	0.13814	0.16183	0.04426	0.16288	0.18794	0.15362	0.00000

Table 24. Poisson Distance (AAPOL)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	0.00000															
2	0.03681	0.00000														
3	0.51504	0.51121	0.00000													
4	0.52778	0.57470	0.05991	0.00000												
5	0.52222	0.56412	0.15643	0.20260	0.00000											
6	0.54276	0.56923	0.15244	0.17665	0.20142	0.00000										
7	0.49667	0.54328	0.51693	0.59404	0.61347	0.57429	0.00000									
8	0.55085	0.59866	0.52902	0.59649	0.60891	0.57855	0.59422	0.00000								
9	0.18079	0.21002	0.51197	0.59532	0.59712	0.60248	0.56011	0.61196	0.00000							
10	0.38152	0.40747	0.47532	0.54027	0.52177	0.50916	0.49870	0.56559	0.43324	0.00000						

11	0.52778	0.56463	0.49009	0.57167	0.55649	0.55406	0.35984	0.59819	0.58067	0.48439	0.00000					
12	0.53900	0.56692	0.15510	0.18174	0.20292	0.00951	0.57375	0.58327	0.60183	0.51549	0.55865	0.00000				
13	0.19400	0.20902	0.49791	0.58481	0.58481	0.57860	0.54799	0.59079	0.14540	0.43796	0.58539	0.57625	0.00000			
14	0.63149	0.64681	0.57195	0.65911	0.65629	0.64145	0.62626	0.69816	0.64045	0.63834	0.64285	0.64545	0.62861	0.00000		
15	0.32450	0.36328	0.50930	0.56287	0.55065	0.55180	0.51316	0.57760	0.36817	0.42423	0.53556	0.55133	0.34655	0.62659	0.00000	
16	0.53118	0.55610	0.25378	0.28603	0.28185	0.27085	0.56574	0.58857	0.58219	0.50088	0.54752	0.27373	0.55661	0.60513	0.54232	0.00000

Table 25. Gamma Distance (AAPOL)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	0.00000															
2	0.03715	0.00000														
3	0.58743	0.58248	0.00000													
4	0.60397	0.66578	0.06082	0.00000												
5	0.59674	0.65171	0.16271	0.21322	0.00000											
6	0.62355	0.65850	0.15840	0.18468	0.21191	0.00000										
7	0.56378	0.62423	0.58988	0.69168	0.71796	0.66523	0.00000									
8	0.63419	0.69790	0.60558	0.69499	0.71177	0.67092	0.69193	0.00000								
9	0.18921	0.22144	0.58346	0.69341	0.69583	0.70307	0.64641	0.71591	0.00000							
10	0.42034	0.45195	0.53656	0.62028	0.59615	0.57984	0.56638	0.65367	0.48374	0.00000						
11	0.60397	0.65240	0.55536	0.66175	0.64162	0.63841	0.39425	0.69728	0.67375	0.54809	0.00000					
12	0.61861	0.65543	0.16127	0.19025	0.21357	0.00953	0.66452	0.67723	0.70219	0.58802	0.64447	0.00000				
13	0.20372	0.22034	0.56537	0.67929	0.67929	0.67099	0.63042	0.68731	0.15082	0.48961	0.68007	0.66785	0.00000			
14	0.74256	0.76365	0.66212	0.78070	0.77677	0.75626	0.73540	0.83553	0.75488	0.75197	0.75818	0.76177	0.73861	0.00000		
15	0.35230	0.39837	0.58002	0.65007	0.63391	0.63544	0.58501	0.66965	0.40424	0.47258	0.61412	0.63481	0.37839	0.73585	0.00000	
16	0.60840	0.64111	0.27058	0.30750	0.30268	0.29005	0.65387	0.68434	0.67578	0.56917	0.62980	0.29335	0.64178	0.70664	0.62298	0.00000

4.5.13 ENV protein-coding genes for 16 species

1. 'HIV-1 (Zaire)''K03454'

```
'atgagagcgagggggatagagagaaattgtcaaaactggtggaaatggggcatcatgctccttgggatattgatgacctgtagtgctgcagacaatctgtgggtcacagtttattatggggtgcctgtatggaaggaagcaaccaccact
ctattttgtgcatcagatgctaaatcatatgaaacagaggcacataatatctgggccacacatgcctgtgtacccacggaccccaacccacaagaaatagcactggaaaatgtgacagaaaactttaacatgtggaaaaataacatggtgga
acagatgcatgaggatataatcagtttatgggatcaaagcctaaaaccatgtgtaaaattaaccccactctgtgtcactttaaactgtagtgatgaattgaggaacaatggcactatggggaacaatgtcactacagaggagaaaggaatga
aaaactgctctttcaatgtaaccacagtactaaaagataagaagcagcaagtatatgcactttttttatagacttgatatagtaccaatagacaatgatagtagtaccaatagtaccaattataggttaataaattgtaatacctcagccattacaca
ggcttgtccaaaggtatcctttgagccaattcccatacattattgtgccccagctggttttgcgattctaaagtgtagagataagaagttcaatggaacaggcccatgcacaaatgtcagcacagtacaatgtacacatggaattaggccagtg
gtgtcaactcaactgctgttgaatggcagtctagcagaagaagaggtcataattagatccgaaaatctcacaaacaatgctaaaaacataatagcacatcttaatgaatctgtaaaaattacctgtgcaaggccctatcaaaatacaagacaa
agaacacctataggactagggcaatcactctatactacaagatcaagatcaataataggacaagcacattgtaatattagtagagcacaatggagtaaaactttacaacaagtagctagaaaattaggaacccttcttaacaaaacaataata
aagtttaaaccatcctcaggaggggacccagaaattacaacacacagttttaattgtggaggggaattcttctactgtaatacatcaggactgtttaatagtacatggaatattagtgcatggaataatattacagagtcaaataatagcacaaa
cacaaacatcacactccaatgcagaataaaacaaattataaagatggtggcaggcaggaaagcaatatatgcccctcctatcgaaagaaacattctatgttcatcaaatattacagggctactattgacaagagatggtggtataaataatag
tactaacgagacctttagacctggaggaggagatatgagggacaattggagaagtgaattatataaatataaggtagtacaaattgaaccactaggagtagcacccaccagggcaaagagaagagtggtggaaagagaaaaaaagagc
aataggattaggagctatgttccttgggttcttgggagcagcaggaagcacgatgggcgcacggtcagtgacgctgacggtacaggccagacaattaatgtctggtatagtgcaacagcaaaacaatttgctgagggctatagaggcgc
aacagcatctgttgcaactcacggtctggggcattaaacagctccaggcaagaatcctggctgtggaaagatacctaaaggatcaacagctcctaggaatttggggttgctctggaaaacacatttgcaccactaatgtgccctggaactct
agttggagtaatagatctctaaatgagatttggcagaacatgacctggatggagtgggaaagagaaattgacaattacacaggcttaatatatagcttaattgaggaatcgcagacccagcaagaaaagaatgaaaaagaattgttggaatt
ggacaagtgggcaagtttgtggaattggtttagcataacacaatggctgtggtatataaaaatattcataatgataataggaggcttgataggtttaagaatagtttttgctgtgctttctttagtaaatagagttaggcagggatactcacctctgt
cgtttcagaccctcctcccagccccgaggggacccgacaggcccgaaggaacagaagaagaaggtggagagcgaggcagagacagatccgtgagattgctgaacggattctcggcacttatctgggacgacctgcggagcctgtg
cctcttcagctaccaccgcttgagagacttaatcttaattgcagtgaggattgtagaacttctgggacgcagggggtgggacatcctcaaatatctgtggaatctcctacagtattggagtcaggaactgaggaacagtgctagtagcttgttt
gatgccatagcaatagcagtagctgaggggacagatagagttatagaaataatacaaagagcttgcagagctgttcttaacatacccagaagaataagacagggcttagaaaggtctttactttaa'
```

2. 'HIV1-NDK (Zaire)''M27323'

```
'atgagagcgagggagaaagagaggaattgtcaaaacttgtggaaatggggcatcatgctccttgggatgttgatgacctgtagtgctgcagaagatttgtgggtcacagtttattatggggtgcctatatggaaggaagcaactaccactc
tattttgtgcatcagatgctaaagcatataaaaaagaggcacataatatctgggccacacatgcctgtgtacccacggaccccaacccacaagaaatagaattggaaaatgtgacagaaaactttaacatgtggaaaaataacatggtggaa
caaatgcatgaggatataatcagtttatgggatcaaagcctaaaaccatgtgtaaaattaactccactctgtgtcactttaaactgcactgatgaattgaggaacagcaagggcaatgggaaggtagaagaggaggaaaaaaggaaaaact
gctctttcaatgtaagggataagagggagcaagtatatgcactttttttataaacttgatatagtgccaatagacaataataataggaccaatagtactaattataggttaataaattgtgatacctcaaccattacacaggcttgtccaaagatatcc
tttgaaccaattcccatacatttttgtgccccagctggttttgcaattctaaagtgtagagataagaagttcaatgggacaggcccatgctcaaatgtcagcacagtacaatgtacacatggaattagaccagtggtgtcaactcaactgctgttg
aatggcagtctagcagaagaagagatcataattagatctgaaaatctcacaaacaatgttaaaaccataatagtacagcttaatgcatctatagtaattaattgtacaaggccctacaaatatacaaggcaaaggacatcgataggactaagg
caatcactctatacaataacaggaaaaaagaagaaaacaggatacataggacaagcacattgtaaaattagcagagcagaatggaataaagctttacaacaggtagctacaaaactaggaaaccttcttaacaaaacaacaataactttta
agccatcctcaggaggggacccagaaattacatcacacatgcttaattgtggaggggacttcttctactgtaatacatcaagactgtttaatagtacatggaatcagactaatagtacagggttcaataatggcacagtcacactcccatgcag
aataaaacaaattgtaaacttgtggcagagagtaggaaaagcaatgtatgcccctcccatcgaaggactaattaaatgttcatcaaatattacagggctactattaacaagagatggtggtgcaaataatagttctcatgagaccatcagacct
ggaggaggagatatgagggacaattggagaagtgaattatataaatataaagtagtaaaaattgaaccaataggagtagcacccaccaaggcaaggagaagagtggtggaaagagaaaaaagagcaataggactaggagctgtgttc
cttgggttcttgggagcagcaggaagcacgatgggcgcagcgtcagtgacgctgacggtacaggccagacagttaatgtctggtatagtgcaccagcaaaacaatttgctgagggctatagaggcgcaacaacatctgttgcaactcac
ggtctggggcattaaacagctccaggcaagagtcctggctgtggaaagatacctaagggatcaacagctcctaggaatttggggttgctctggaaggcacatttgcaccactaatgtaccctggaactctagttggagtaatagatctctag
atgagatttggcagaacatgacctggatggagtgggaaagagaaattgacaattacacaggcttaatatacagcttaattgaagaatcgcagatccagcaagaaaagaatgaaaaagaattattggaattggacaaatgggcaagtttgtg
gaattggtttagcataacaaaatggctgtggtatataaaattattcataatgatagtaggaggcctgataggtttaagaatagtttttgctgtgctttctgtagtaaatagagttaggcagggatactcacctctgtcatttcagaccctcctcccagt
cccgaggggacccgacaggcccgaagaaatagaagaagaaggtggagagcgaggcagagacagatccattcgattggtgaacggattatttgcacttttctgggacgacctgaggaacctgtgcctcttcagctaccaccgcttgaga
gactcaatcttaattgcagcgaggattgtggaacttctgggacgcagggggtgggaagccctcaaatacctgtggaacctcctgcagtattggagtcaggaactgaggaatagtgctagtagcttgcttgataccatagcaatagcagtag
ctgagaggacagatagggttatagaagtagtacaaagagcttgcagagctattcttaacgtacccagaagaataagacagggcttggaaaggcttttgctataa'
```

3. 'HIV-2 (Senegal)''M15390'

```
'atgatgaatcagctgcttattgccattttattagctagtgcttgcttagtatattgcacccaatatgtaactgttttctatggcgtacccacgtggaaaaatgcaaccattcccctcttttgtgcaaccagaaatagggatacttggggaaccataca
gtgcttgcctgacaatgatgattatcaggaaataactttgaatgtaacagaggcttttgatgcatggaataatacagtaacagaacaagcaatagaagatgtctggcatctattcgagacatcaataaaaccatgtgtcaaactaacacctttat
gtgtagcaatgaaatgcagcagcacagagagcagcacagggaacaacacaacctcaaagagcacaagcacaaccacaaccacacccacagaccaggagcaagagataagtgaggatactccatgcgcacgcgcagacaactgct
caggattgggagaggaagaaacgatcaattgccagttcaatatgacaggattagaaagagataagaaaaaacagtataatgaaacatggtactcaaaagatgtggtttgtgagacaaataatagcacaaatcagacccagtgttacatgaa
ccattgcaacacatcagtcatcacagaatcatgtgacaagcactattgggatgctataaggtttagatactgtgcaccaccgggttatgccctattaagatgtaatgataccaattattcaggctttgcacccaactgttctaaagtagtagcttct
acatgcaccaggatgatggaaacgcaaacttccacatggtttggctttaatggcactagagcagagaatagaacatatatctattggcatggcagagataatagaactatcatcagcttaaacaaatattataatctcagtttgcattgtaagag
gccagggaataagatagtgaaacaaataatgcttatgtcaggacatgtgtttcactcccactaccagccgatcaataaaagacccagacaagcatggtgctggttcaaaggcaaatggaaagacgccatgcaggaggtgaaggaaacc
cttgcaaaacatcccaggtatagaggaaccaatgacacaaggaatattagctttgcagcgccaggaaaaggctcagacccagaagtagcatacatgtggactaactgcagaggagagtttctctactgcaacatgacttggttcctcaatt
```

ggatagagaataagacacaccgcaattatgcaccgtgccatataaagcaaataattaacacatggcataaggtagggagaaatgtatatttgcctcccagggaaggggagctgtcctgcaactcaacagtaaccagcataattgctaacat
tgactggcaaaacaataatcagacaaacattacctttagtgcagaggtggcagaactatacagattggagttgggagattataaattggtagaaataacaccaattggcttcgcacctacaaaagaaaaaagatactcctctgctcacggga
gacatacaagaggtgtgttcgtgctagggttcttgggttttctcgcaacagcaggttctgcaatgggcgcggcgtccctgaccgtgtcggctcagtcccggactttactggccgggatagtgcagcaacagcaacagctgttggacgtggt
caagagacaacaagaactgttgcgactgaccgtctggggaacgaaaaacctccaggcaagagtcactgctatagagaagtacctacaggaccaggcgcggctaaattcatggggatgtgcgtttagacaagtctgccacactactgtac
catgggttaatgattccttagcacctgactgggacaatatgacgtggcaggaatgggaaaaacaagtccgctacctggaggcaaatatcagtaaaagtttagaacaggcacaaattcagcaagagaaaaatatgtatgaactacaaaaatt
aaatagctgggatatttttggcaattggtttgacttaacctcctgggtcaagtatattcaatatggagtgcttataatagtagcagtaatagctttaagaatagtgatatatgtagtacaaatgttaagtaggcttagaaagggctataggcctgtttt
ctcttcccccccggttatatccaacagatccatatccacaaggaccggggacagccagccaacgaagaaacagaagaagacggtggaagcaacggtggagacagatactggccctggccgatagcatatatacatttcctgatccgc
cagctgattcgcctcttgaccagactatacagcatctgcagggacttactatccaggagcttcctgaccctccaactcatctaccagaatctcagagactggctgagacttagaacagccttcttgcaatatgggtgcgagtggatccaagaa
gcattccaggccgccgcgagggctacaagagagactcttgcgggcgcgtgcaggggcttgtggagggtattggaacgaatcgggaggggaatactcgcggttccaagaaggatcagacagggagcagaaatcgccctcctgtga'

4. 'HIV2-MCN13''AY509259'

'atgatgggtggtagaaatcagctgcttgttgccattttgctaactagtacttgcttgatatattgcaccaattatgtgactgttttctatggcatacccgcgtggagaaatgcatccattcccctcttttgtgcaaccaagaatagggatacttgggg
aaccatacagtgcttgccagacaatgatgattatcaggagataactttgaatgtgacagaggctttcgatgcatgggataatacagtaacagaacaagcaatagaagatgtctggaatctatttgagacatcaataaaaccatgtgtcaaatta
acgcctttatgtgtagcaatgagatgtaacaacacagatgcaaggaacacaacccacacccacaacagcatccccgcgtacaataaaacccgtgacagagataagtgagaattcctcatgcatacgcgcaaacaactgctcaggattggg
agaagaagaggtggtcaattgtcaattcaatatgacaggattagagagagataagaaaaagcaatatagtgagacatggtactcgaaggatgtagtttgtgaaggaaatggcaccacagatacatgttacatgaaccattgcaacacatcg
gtcatcacagagtcatgtgacaagcactattgggatgctatgaggtttagatactgtgcaccaccaggttttgccctactaagatgcaatgataccaattattcaggctttgcgcccaattgctctaaggtagtagctgctacatgcaccagaat
gatggaaacgcaaacttctacatggtttggctttaatggcactagagcagaaaatagaacatttatctattggcatggtagggataacagaactatcatcagcttaaacaaatattataatctcactatacattgtaagaggccaggaaataaga
cagtggtaccaataacacttatgtcagggttaaggtttcactcccagccggtcatcaataaaagacccagacaagcatggtgttggttcaaaggtgaatggaagggagccatgcaggaggtgaaggaaacccttgcaaaacatcccaggt
ataaaggaaccaatgaaacaaagaatattaactttacagcaccaggaaagggctcagacccagaggtggcatacatgtggactaactgcagaggagaatttctctactgcaacatgacttggttcctcaattggatagaaaataagacaca
ccgcaattatgtaccgtgccatataagacaaataattaacacctggcataaggtagggaaaaatgtatatttgcctcccagggaaggggagttgacctgcaactcaacagtaactagcataattgctaacattgatgcaaatggaaataatac
aaatattacctttagtgcagaggtggcagaactataccgattagagttgggagattataaattggtagaaataacaccaattggcttcgcacctacagcagaaaaaagatactcctctactccaatgaggaacaagagaggtgtgttcgtgct
agggttcttgggttttctcgcaacagcaggctctgcaatgggcgcggcgtccttaacgctgtcggctcagtctcggactttactggccgggatagtgcagcaacagcaacagctgttggacgtggtcaagagacaacaggaaatgttgcg
actgaccgtctggggaacaaaaaatctccaggcaagagtcactgctatcgagaagtacttaaaggaccaggcgcaactaaattcatggggatgtgcatttagacaagtctgccacactactgtaccatgggtaaatgataccttaacgcct
gagtggaacaatatgacgtggcaagaatgggaaggcaaaatccgcgacctggaggcaaatatcagtcaacaattagaacaagcacaaattcagcaagagaagaatatgtatgaactacaaaagttaaatagctgggatgtttttggtaac
tggtttgacttaacctcctggatcaagtatattcaatatggagtttatataataataggaatagtagttcttagaatagtaatatatatagtacagatgttaagtagacttagaaagggctataggcctgttttctcttcccccccggttacctccaac
agatccatatccacaaggactgggaacagccagccagagaagaaacagaagaagacgttggaaacaacgttggagacagctcgtggccttggccgataagatatatacatttcctgatccaccagctgattcgcctcttggccggactat
acaacatctgcaggaacttactatccaggatctccctgaccctccgaccagttttccagagtcttcagagggcactgacagcaatcagagactggctaagaactgacgcagcctacttgcagtatgggtgcgagtggatccaaggagcgtt
ccaggccttcgcaagggctacgagagagactcttgcgggcacgtggagagacttgtggggggcactgcagcggatcgggaggggaatacttgcagtcccaagaagaatcaggcagggagcagagatcgccctcctatga'

5. 'HIV-2UC1 (IvoryCoast)''L07625'

'atggcacacactagcaatcacctgtttattttgctcctacttataagtgtctatgggtttctgggtcataagaaaaactatgtcactgtcttctatggcatacctgcatggaggaacgcaacggttcctctcttctgtgcaacccacaaacagggata
cctggggaactgtacagtgcctcccagacaatggggattacaccgagatcagtgtaaatataacagaggcttttgatgcatggaataatacagtaacagaacaggcagtagatgatgtgtggagtctttttgaaacctccataaaaccatgt
gttaagctgacgccgctgtgcgtggcaatgagatgtaataacacaggaaccaatactactactaaacctattactacacctattactactactaaaccatcagaaaacttactcaacgacacaagtccatgtattaaaaatgacacctgcccgg
gaataggactagaaaacacagttgactgttactttaacatgacaggactaagaagagatgagaaaaaacaatataaggacacctggtacgaaaaagatttagagtgcaacggcaacagcaccagcaccatatgttacatgagaacttgta
atacctcagtgatccaagagtcatgtgataaacattactgggacagcttaagatttaggtactgtgctcccccaggatatgctctgctaagatgtaatgacaccaactattcaggctttatgccaaaatgtagtaaagtagtagtgtcctcttgca
caagaatgatggagacacagacctctacatggtttggcttcaatggtacgaggacagaaaataggacatatatgtattggcatagtaaagataataggaccatcataagcttgaataagtattataatctaacaatgcattgcagaaggccag
ggaacaagacagtcataccaataacaatcatgtcaggactgaatttccattcgcagcctctcaataccagacccagacaggcttggtgctggtttaagggaaactggatagaagccataagggaggtaaaggagaccatcataaaacatc
ccaggtataaaggaacaaataacaccgagaggataagattggtagggccctcggcaggttcagacccagaggtgagacatatgtggactaactgtaggggagaattcttctattgtaatatgacctggtttttaaactgggtggagaacag
aacaggtacaactcagaagaattatgtgacgtgccacatcaaacagatagtcaacacgtggcacaaggttggaaaatacgtatatttgcctccaagggagggtacgctctcctgtaattccagtgtcaccagtcttatagccaacattgacgt
gtattatgatggcaatgataccaagaccaatattaccatgagtgcagaagtgggagaactgtacagattggaactgggggactacaaattagtagaaataacaccgattggctttgcacctacagagataaaaagatattcctcaacgacac
cgaggaataaaagaggtgtaatggtgctagggttcttgggacttcttgcaatggcaggttctgcaatgggcgcaacgtccttgacgctgtcagctcagtcccggactttactggctgggatagtgcagcagcagcaacagctgctggacgt
ggtcaaaagacaacaggaactgttgcggctgaccgtctggggaacgaaaaacctccagacaagagtcactgccatcgagaaatacctaaaggaccaggcactactaaattcgtggggatgtgcttttaggcaggtttgtcacacgactgt
accatggccaaatgaaactctcacaccagactgggaaaacatgacatggcaacagtgggaaaagcgagtcaatttcttggatgcaaatataacagccctgttagaagaggctcagatacaacaagaaaggaacatgtatgaattacaga
aattaaatagctgggatgtttttggtaattggtttgacttcacctcctggatggcatacatcaggttaggactatatgtagtagcaggattaatagtgttaagaatagtaatatacattatgcagatgctagcaagacttaggaagggctataggcc
agtattctcctcccctccctcttatactcaacagatccctatccgcaaacaccggggacagccagccaacgaagaaacagaagacgaaggtggaaacgaaggggcttacagatcttggccctggcagatagaatacgctcactttctgatt
cgccaactgaggaacctcttgatttggctgtacaacggctgcaggaacttactgttgaagacctcccaaatcctccaaccagcactccaaccgctcaggctttcacttgcatacctccagtatgggatcagctggttccaagaagcaatcca
agcagcaacgagggctgcgagagagactcttgcgaacacaggaagagccctatggaaagctctcagaaggacagcggaagcaatcatcgcgatccccaggaggatcagacaaggacttgaactcgccctcttgtga'

6. 'SIVMM251 Macaque''M19499'

'atgggatgtcttgggaatcagctgcttatcgccatcttgcttttaagtgtctatgggatctattgtactcaatatgtcacagtcttttatggtgtaccagcttggaggaatgcgacaattcccctcttctgtgcaaccaagaatagggatacttggg
gaacaactcagtgcctaccagataatggtgattattcagaattggcccttaatgttacagaaagctttgatgcttgggagaatacagtcacagaacaggcaatagaggacgtatggcaactctttgagacctcaataaagccttgtgtaaaatt
atccccattatgcattactatgagatgcaataaaagtgagacagatagatggggattgacaaaatcatcaacaacaataacaacagcagcaccaacatcagcaccagtatcagaaaaaatagacatggtcaatgagactagttcttgtatag
ctcagaataattgcacaggcttggaacaagagcaaatgataagctgtaaattcaccatgacagggttaaaaagagacaagacaaaggagtacaatgaaacttggtactctacagatttggtttgtgaacaagggaatagcactgataatga
aagcagatgctacatgaatcactgtaacacttctgttatccaagaatcttgtgacaaacattattgggatactattagatttaggtattgtgcacctccaggttatgctttgcttagatgtaatgacacaaattattcaggctttatgcctaaatgttcta
aggtggtggtctcttcatgcacaaggatgatggagacacagacttctacttggtttggctttaatggaactagagcagaaaatagaacttatatttactggcatggtagggataataggactataattagtttaaataagtattataatctaacaat
gaaatgtagaagaccaggaaataagacagttttaccagtcaccattatgtctggattggttttccactcacaaccaatcaatgataggccaaagcaggcatggtgttggtttggaggaaaatggaaggatgcaataaaagaggtgaaacag
accattgtcaaacatcccaggtatactggaactaacaatactgataaaatcaatttaacggctcctggaggaggagatccggaagttaccttcatgtggacaaattgcagaggagagttcctctactgtaaaatgaattggtttctaaattgggt
agaggatagggatgtaactacccagaggccaaaggaacggcatagaaggaattacgtgccgtgtcatattagacaaataatcaacacttggcataaagtaggcaaaaatgtttatttgcctccaagagagggagacctcacgtgtaactc
cacagtgaccagtctcatagcaaacatagattggactgatggaaaccaaactagtatcaccatgagtgcagaggtggcagaactgtatcgattggagttgggagattataaattagtagagatcactccgattggcttggcccccacagatg
tgaagaggtacactactggtggcacctcaagaaataaaagaggggtctttgtgctagggttcttgggttttctcgcaacggcaggttctgcaatgggcgcggcgtcgttgacgctgaccgctcagtcccggactttattggctgggatagtg
cagcaacagcaacagctgttggacgtggtcaagagacaacaagaattgttgcgactgaccgtctggggaacaaagaacctccagactagggtcactgccatcgagaagtacttaaaggaccaggcgcagctaaatgcttggggatgtg
cgtttagacaagtctgccacactactgtaccatggccaaatgcaagtctaacaccagactggaacaatgatacttggcaagagtgggagcgaaaggttgacttcttggaggaaaatataacagccctcctagaagaggcacaaattcaac
aagagaagaacatgtatgaattacaaaagttgaatagctgggatgtgtttggcaattggtttgaccttgcttcttggataaagtatatacaatatggaatttatgtagttgtaggagtaatactgttaagaatagtgatctatatagtacaaatgcta
gctaagttaaggcaggggtataggccagtgttctcttccccaccctcttatttccagtagactcatacccaacaggacccggcactgccaaccagagaaggcaaagaaggagacggtggagaaggcggtggcaacagctcctggcctt
ggcagatagaatatattcatttcctgatccgccaactgatacgcctcttgacttggctattcagcaactgcagaaccttgctatcgagagcataccagatcctccaaccaatactccagaggctctctgcgaccctacgaagggttcgagaag
tcctcaggactgaactgacctacctacaatatgggtggagctatttccatgaggcggtccaagccggctggagatctgcgacagaaactcttgcgggcgcgtggagagacttatgggagactcttaggagaggtggaagatggatcctc
gcaatccctaggaggattaggcaagggcttgagctcacgctcttgtga'

7. 'SIVAGM677A Green monkey''M58410'

'atggggagattgcttataaaaatactaataatagcaatagggataagtataggaataggtaacctgtatgtgacagtgttttatggaatcccagtatggaaaaattcaacagttcaggcattttgcatgacgcccaataccaatatgtgggcaa
ccaccaactgcataccagatgatcatgataatacagaggtgcctctaaacattacagaagctttcgaggcttgggataatccgctggtaaaacaagcagagagtaatatacatctactctttgaacaaacgatgaggccttgtgttaagctctc
ccccatatgtattaaaatgtcctgtgtagagctgaatggtacagccacgacaaaggccaccactactgcaactacaacaatgactaccccctgtcagaattgcagtacagagcagatagaaggagaaatggcagaggaaccagcatcca
actgcacttttgcaattgcaggatatcaaagagatgtaaaaaagaattatagcatgacctggtatgatcaggagttagtctgcaataataaaacaggaagtgaaaagggaagtaaggattgttacatgatacattgtaatgattcagtgataaa
agaagcttgtgataaaacatattgggatactttaagagtaagatactgtgcaccagcagggtatgctttgctaaaatgtaatgataaggattatagaggctttgctccaaagtgcaagaatgtttcagtagtgcattgtactagattaatcaatact
actataactacagggataggattaaatggtagtagatcagaaaatagaacagagatatggcagaaaggaggaaatgataatgatacagttataataaagttgaataagttttacaacttgacagtgagatgccgaagacctggtaataaaac
agtgttgccagtaacaatcatggcagggttagtatttcactctcagaaatataataccaggttaaaacaagcgtggtgccacttccaaggagattggaaaggggcatggaaagaagtcagagaagaagtaaagaaagtgaaaaatcttaca
gaagtaagcatagaaaatatacatctgagaaggatatggggagatccagaatcagcgaatttttggttcaattgtcaaggtgaatttttctattgtaagatggactggtttatcaattatctaaacaatcgaacagaagatgcagaaggtactaat
aggacctgtgacaaagggaagccaggaccaggaccatgtgttcagagaacttatgttgcctgccatatacgacaagtagtaaatgattggtacactgtctctaaaaaggtatatgctccaccaagggaaggtcatttggagtgtaactcatc
agtcacggcactatacgtggcaatagattataacaacaagtctggcccaataaatgtgaccctaagtcctcaggtacgcagcatatgggcgtacgaactgggagactataaattagtagagataacaccaattggctttgctcctacagatgt
aagaagatatactggccccacaagagaaaaaagggtgccattcgtgctagggtttctaggcttcttgggagctgctggaactgcaatgggcgcagcggcaacaacgctgacagtccagtctcggcatttgcttgctgggatattgcagca
gcagaagaacttgctggcggctgtggaacagcaacaacagttgttgaagctgaccatttggggtgtgaaaaacctcaatgcccgcgtcacagctctcgagaagtacctagaggatcaggcacggctaaattcatggggatgtgcgtgga
aacaagtatgtcacaccacagtgccatggaagtataataacactcctaagtgggacaatatgacttggttggagtgggagagacaaattaatgccttggaaggcaacataactcaactattggaagaagcacaaaatcaggaatcaaaga
atctggatctgtaccagaaattggatgattggtcagggttctggtcatggttctcactgtcaacttggttaggctatgttaaaataggatttttagtgatagtgattattctaggattaagatttgcatgggtattatggggatgtatcagaaatattag
gcagggatataatcctctcccccagatccatatccacagttcagcggaacggccagacaacggaggagggcaagacagaggtggagaaagcagcagcagcaaattgataagattgcaggaagagtcctcaacaccttcgaggatca

acaactggtggctcaacttcaagagctgcagcttgagaataaggacttggtgttacaacatctgcctgaccctcctcatattcatcaggacagcagtgggatacctgcagtatgggctccagcaactccaagaggcagcaacagggcttgc
tcaagctctggcgagggctgcgagggaagcctggggcagactgggtgctattgtccgatccgcttatcgggcagtcatcaacagtccaagaagagtgcggcaaggccttgaaaaagtcctggggtaa'

8. 'SIVlhoest L''Hoest monkeys''AF075269'

'atggcatgtccaggattaggaattctgcttttgcttcttggcataatatggggaaaacaatatgtaacagtattttatggggtgccaaattgggatgataatgtttcagtgcccttgatttgtgcttcagccaatactagcctctgggtcacgacatc
ttgcttgccagatttgcaatcttatgcagaggtacctatctataatatttcagaaaactttacaataccagtaaaagacaatcaggttatacagcaagcctggtctgctatgaatgctatggtagacagtattatgaaaccatgtgtaaaaatcaatc
catattgtgtgagaatgcagtgtggggaagttacaaaaacaccaacaacaacaccaaaaactactacacagatgccttgttttatcaacgaacaggtaacagttaagaacccggggaatgaaacaagattagaggaagatcttaattgcac
aagagggcttaatgagaccacagagaggaatgcagaatgccagtataatgtaacagggttatgcagagactgcaggacggaaataaaacaaagctttagatatgatgatgtaacgtgctcaggagagagggagaacagaacctgttac
atgactcattgtaatgattcaataataacacaagattgtaataagggggtaatgcaaaatgcttattttaggctttgtgctccagcaggctatatgctattgagatgcaatgaacaattaaactttagtaaaaaatgtgaaaatataacagctacac
catgtactggatatatgcttagtagtgtaagtagtttctttggctttaatggaactaatcataccagggatgagcttattccactaactccaaataagatggaagatctaaatggtgcaaagtttgtgtataaagtagcaggaaaatggggtttaatt
attagatgcataagaaaaggaaatagatcagaagtatccacaattagctctacaggatatttattttattatggcctagaacatggcagcagattaagattagcacagtgtaaatttgaaggccaatggggaagaatgtttaataatctaggcaa
aatgctaaaggagttgaatgcagaagccatgaattatacagaagggacagggacatgtgattctaagaaaacaacttgtggacggaaattaaaaggcttaccaatagctaatatgactaggcatggagcagacttagcaacagagatgtta
atgcatacttgtggagaggaaatgttcttttgtaatgtaactaggatattccaagaatggaacaataaaaattcagataaatggtatccttgggccaactgccacataaaaagcatcattgatgattgggctacaataggaaagaagatttatcta
ccacctacttcaggtttcaacaatagaatcagatgtactcatagagtaacagaaatgttctttgagatggaaaaatgggaaccacatgaggacttaggaggcaacctaagtataaagttcttgcctccatcatgggaaactaatcaatttgtgg
cagaagggtctaaatataaattgataaaattgaatccaattgggtttgctcctacagatgagcacaggtacgcacccaggggaaggcaaaccagagcagcgcctctggctttaggagccttgggacttctcagtgctgcaggcactgcaat
gggcttagtatcgacgatactaactgtccaggcccaagcagtcttgcaagggatattgcagcagcagaagcagttgctggtgcttgttgagaaacagcaagaactgttaagactcaccatatggggagtgaagaatctccaagcacgcct
gacagcccttgaggagtatgtgaaacatcaagcgctccttgcctcttgggggtgtcaatggaaacaggtttgtcatactaatgtggagtggacatataatatcacacccaattggaccaaggatacctggagagaatgggaatcaaaggta
gcaatttatgacaagaacattacttctctgttacaagaagcatacaccacagagctagagaatcagaataaatttaagaagttacaagaatttaacttttggagttggttggatatctcacactggtttacttatgttaagtatgcagtgttaattata
cttgtaatcatagggttaagagtattgagctttataatacaaaatgtagtaaaaatgtgtagggggtatagggtgctctccccctctgtttatattgaacaggactacaagtgggagaaagaagagaaccaagaacagccagacagagaaga
agagaaaggagcagatacagagactatctacatcaacttagagcagtgcaagaaagaatcttccaggccactgtggaacgtggactggaacgagcctttacaagactcgctgttagtgactctcctgaagtggctcaaggaaggggaa
tactcctcctatcactagtgtggcagagcctcagttggctgtggcatttgttgatccttttcttccaaaatgggcaacgcctttggcagaccagcagcagatggatggtggaaaacgctcagaagattcagagctggctcagggagaaatgc
agaaggaacagaggacagttatcgagcactgatagaaaaaacatccaattgggcaaaaagaagagatggagactcagattcggaggaagaagtgggatttccagtgaagccacagagaccgctttgtag'

9. 'SIVcpz Chimpanzees Cameroon''AF115393'

'atgaaagtgatggagaagatcaaaggcagcttgatagagaaatacatgtttatgggtttgataatcccatgtttgactggtagtgatcagttgtgggcaacagtgtactatggggtacctgtgtggaaagaggcagacacaacactcttttgt
gcctcagaggcgtcagctcttaacaaagaggcccataatatttgggcctcacaagcctgtgttcctacagaccctaacccacaagaagtgcagatcccaaatgtaactgaaaactttaatatgtggaacaataccatggcagaacaaatga
atactgacattattagtctgtgggatcaaagtttgaaaccatgtgttaagctgacgccaatatgtgtcaccatggaatgtagaaaggttactttcaatagcacttctaataggaataaaacttctactatgactactaacagtcctaatgaaaaaa
agacagcacagtaaagaattgtacctttaatatgactactgaggtaagagataaagagaaaaaagtatattcccttttttatgtagatgatctagtattgattgataacgataccgatacatatagactaataaattgcaataccacagctattaca
caggcgtgtcctaagacctcctttgagccaattccaatacattactgtgcaaccccaggttatgcaatcatgaaatgcaatatgccaaactttaatggaacaggaacaggaagatgtaacaatataagcacagtccaatgtactcatggcatc
agaccagtggtaactacacaattgatccttaatggaagtgtagccgaaaacaagactattgctagaaggaacgggtacaacttcctgatacaatttcagaaaacggtcagcataaattgtacaagaccaggaaataacagcagaggacaa
atacagataggaccaggcatgaccttttataacatagaaaatattgtaggagataccagacgagcatattgtcagatcaataggacagtctgggacgaaagactaaatgaaactggtcaagccttgcgagagctttttacaaaccttacaca
agtcaactttacagtaagtccaggaggagacccagaagtaacaaatatgatgtttaattgtggaggagaattcttttattgcaatactactacattgtttaattatacgtggaaaaacaacaatataactaagggagataacaccaccttcttccc
atgtagaataaggcaaattgtaaactcctggatgagagtggggaagggaatttatgccccaccaattagaggagttataagttgcacatcaaacgttactgggattatacttgagacagggcatggcatcaacaacagcatcaccaacatca
cgctataccctacaggaggaaacatggtagacttgtggaggcttgaattgcataaatacaaggtagttagcatagaacccataggagtagcaccaagtaaagctaaaaggcacacagtatcaagagaaaaaagagcagcctttggactg
ggtgcgctgtttcttgggttcttaggagcagcaggaagcactatgggcgcagcatcagtggtgctgacggtacaggcccggcaattattgtcagggatagtgcaacagcagaacaacctgctgcgagcaatagaggcgcaacagcattt
gttgcaactctcagtttggggcattaaacagctccaagcaagagttcttgctgtagaaagatacctcagagatcagcaaatcctaggcctatggggctgctcaggaaaagccatttgttatactactgtgccttggaataatacttggagtgct
aatacatcctttgatgagatctggaataatttaacatggcaagactgggacaaaagggtaaaaaattattcaggtgtcatttttagccttatagaacaagcacaagagcaacaaaacacaaatgagaaaagtctattggaattggatcaatggt
caagtttatggaattggtttgatattacccgctggctgtggtacataaaattgtttataatgatagtagcaggcttagtaggcataagaatagtaggggctatcatatcatttgtagcaaaggttaggcagggatactctcccctctcgttgcagac
ccttatcccgacaacaagggaaccagacaggccaggaggaacagaagaagacgttggagagccaggcaaaggcagatccattcgcttagtgagcggattcttagctcttatctgggaagatctccggaacctcctgctcttctgttacca
ccgattgagagacttactattgatactggggaggacattggaaaacctgggacagagcctcaacaaggactgcaacagctgaggaacttcagcagatacctgtggggagtgataacctactggggaagggagctacaaactagtgcc
ataagcttattagatgcaacagccattgcagtagctgaaggaacagacagaattctagaagttgcacaaataataggtagaggcattctgcatataccaagaagaattagacagggtttagaaagaagcttgctataa'

10. 'SIVmnd5440 Mandrillus sphinx''AY159322'

'atgttaagatatcttaggtatatagtcttaggaataatagtaagtgtaatagtaggagaacaatgggtgacagtgtactatggcacacccaaatggcataaggcaaggacacacctgttttgtgcaacagacaataattc
attctgggtgaccacaagctgtgtccccagtctgttgcactatgaagaacagcatatccccaatataacagagaattttacaggccctatagaagagaatgaaatagtgacacaagcttggggagcaatttcttccatgatagatgcagtatta
aaaccatgtgtaaggttaacaccttattgtgtgaagatgaaatgtacagaaggtcaaaatgaaacagagcaagcaacagctaagaccacaacacctgtacccacaacaacaaccccctccactaccactagctccagcacaaataagac
aacaactcctgtgttggttgttgagaaacaaaataatgagacaacaacacagcaaaatagagtatgtaaattcaacacgacaggattatgtagggactgcaaattggaaatagaggaaaactttagatatgaagatgtaacttgcacaaagc
ttaataaaactggtagtgctactaatagtactgaacctgaatatgagtgttacatgacttcatgtaacgcaacagtaataacacaggaccgtaacaaagcttccacagatagaatgacatttaggttgtgtgcacccccagggtttgtcttactaa
aatgtaatgaaaaattgaacaagacaaaattatgtgggaatgtgtctgcagtgcagtgcacagcaccgctgccagccacaatctctacaatgtttggcttcaatggaactaagcatgattatgatgagcttattcagacaaacccacgaaaag
gaaaagatgagtttcatgaccataaatatgtatatagagtggataaaaaatggggattacaggtaaggtgtagaaggaagggaaatagatctatcatttccacaccaagttctacaggattgctgttctatcatgggttggagccaggaaaga
atcttagaaagggcaagtgtcagttagaagggaaatggggacaggctttacatagtctatcattagagctaagaaagataaatgatagcatttacaaagacaaccacaatatgacatgtaaaagcagtaacaacaaaaagaacacaacag
ggtgtcatcttaaaactataagtataagtgagtccaccgttaaaggcgaaccaggagctgaaactattatgctcctctgtggaggagaatatttcttttgtaattggactaaaatatggaaggcatggaatagcaaacaaagttcagtttggtac
ccttacatgtctcgcaatattagacaaatagtaggtgattggcataaagtaggcaagaaaatttatatgcctcctgtgtctgggtttaataatgaaataagatgtactaatgatgtgacagaaatgttctttgaggtccaaaaaactgatgatgaca
atggatatatcataaaatttataccacaagattggattcaaaatcaatacacagcagtgggggctcattacaaattggtgaaggtggatccgataggttttgcccccacagacatacacagacatcatttgcctaatacaaggcaaaaaagag
gagcggtcttacttggaatgctcggcctcctaggtttggcaggttccgcgatgggctcagtggcggtggccctgactgtccagtcccagactttactgaatgggattgtggagcagcagaaggttctgctgagcctgatagatcagcactc
cgagttattaaaactaactatctggggtgtaaagaatcttcaggtccgcctcacagccttggaggaatacgtagcggaccaatcaaggctctcggtatggggatgttcattctcccaagtttgtcatactagtgtaaaatggcccaataatagta
tagtgcccaattggacctcagagacatggctggaatgggacagaagagtaaacagcatagtgacaaatatgaccatagacttacagagggcatatgaattagaacagaggaacatctttgagttgcaaaaattaggggacttaaacttcca
tgggctaactgggttcgaccttacatggtggcttaagtatgttaagataggactgttagtagtagtagttatcataggattaagaatgttagcttgcttatggtcagtattagggaagtttaggcagggatatcgccctcttccttatgtcttcaagg
gagactatctccggccccacaacctcaaacggccagacagagaaggaggagaagagccagacttagaaaaacagaacatcaagtcagagagctccagacaagaatctaggaagccttggaagccagagcaagtcaggagctggc
tgaaaaggtcgacgctttacatttggctgaagaacctccaagcagtaattgaatatgggtggcaagagctcaaagcagcgggagcagcattatatcaagtattacaaggctttgcgcagaggctatggagcagagggtaccaacttggac
tatcaggtgttagaggagcagcagcctttggcaggggcatctggaacatccctcgccgcatcagacaaggtgctgaggccctactcaactga'

11. 'SIVAGM3 Green monkeys''M30931'

'atgaagctgacattactgatagggatactattaatagggataggagtagtgcttaatacaaggcaacaatgggtcacagtattttatggagtaccagtatggaaaaacagctcagtacaggctttctgcatgacacccaccaccagactatg
ggcaactactaactcgataccagatgatcatgactacacagaggtaccattaaacatcactgaaccatttgaagcatgggctgacagaaaccccttagtagcacaagcaggaagtaatatacacctgctatttgagcagactctgaagccat
gtgtaaaattatcacctttgtgcattaaaatgtcctgtgtagaattgaactcctctgagcctaccaccactcctaaaagtaccacggcctcaacaaccaatatcacagcctcaacaaccactttgccgtgtgtccagaacaagacaagtactgtg
ttagaatcatgtaatgaaacaatcatagaaaaggaattaaatgaagagcctgcttctaattgtacatttgcaatggcagggtatgtaagagatcagaaaaagaagtattcagtggtgtggaatgatgcagaaatcatgtgtaagaagggtaac
aattctaacagagaatgttatatgattcattgtaatgattcagttataaaagaagcctgtgataaaacatattgggatgagttaagattaaggtactgtgccccggcagggtttgctttattaaaatgcaacgattatgattatgcagggtttaagac
aaactgttctaatgtttcagtggtgcattgtactaacttgataaatacaacagtgactactggactgttgttgaatgggagctactcagagaatcgaacccagatatggcagaaacatagagtaagcaatgactcagtgttagtgttatttaataa
acattacaatctaacagttacttgcaaaagaccaggaaacaaaacagtcttaccagtaacaatcatggcagggctagtgtttcattctcagaggtacaatacaaggctgagacaagcttggtgtcacttccagggcaactggagaggagcc
tggaaagaagtaaaaaatgaaatagtaaaattaccaaaagatagataccaaggaaccaatgatactgaagagatttatctgcagagactatttggagatccagaagcagcaaatttatggtttaattgtcaggggaattcttctattgtaaaat
ggattggtttctaaattacctgaataatcgtacagtagatccggaccataatccgtgtaatggtacgaagggaaaaggtaaggcaccaggaccctgtgcacaaagaacatatgttgcttgccatatacgatctgtcattaatgattggtacaca
ctatcaaggaaaacctatgcaccgccaagagaagggcacttgcaatgcacatccacggtaacgggtatgtcagtggagctaaattacaatagtaagaacaggactaatgtaacattaagtccccagatagaaaccatctgggcagcaga
attgggcaggtacaaattagtagaaattacaccaattggcttcgcacccacagaagtaagaaggtatacgggaggtcatgacagaacaaagcgagtcccgttcgtgctagggttcctaggcttcttaggagctgctgggactgcaatggg
agcagcggcgacagccctgacggtccagtctcagcatttacttgctgggatactgcagcagcagaagaatctgctggcggctgtggaggctcaacagcagatgttgaagctgaccatttggggtgtgaaaaacctcaatgcccgcgtca
cagctcttgagaagtacctagaggaccaggcgcggttgaatgcttgggggtgcgcatggaagcaagtctgtcatacaaccgtaccgtggcagtggaataataggacccctgattggaataatatgacttggctggaatgggaaagacag
atatcgtatttggaaggtaacataacaacacaattagaggaagccagagcacaggaggagaagaatttggatgcataccaaaaattaagtagttggtcagatttctggtcttggttcgatttctcaaagtggctgaacattctaaaaataggatt
tttggatgtactaggtattataggattaagattgctttatacagtatattcttgcatagctagggttaggcagggttactctcctctttctccacagatccatatccacccgtggaagggacagccagacaacgcagaagggccaggagaaggt
ggagacaagcgcaagaacagctccgagccttggcagaaagaatctggcacagcagagtggaagagcaactggtgcaagcgattgaccaattggtgctcgatcagcagcatctggctatacaacagttgcctgaccctcctagttcatct
taggagcgctttccagtacatacaatatgggcttggggaactcaaagccgcagcacaagaagcagttgtcgctttggcacgccttgcacaaaacgcgggctaccagatatggcttgcttgcagatccgcttatagggcaatcatcaactct
ccaagaagagtgcgacaaggccttgaaggaatccttaattag'

12. 'SIVMM239 Simian macaque''M33262'

'atgggatgtcttgggaatcagctgcttatcgccatcttgcttttaagtgtctatgggatctattgtactctatatgtcacagtcttttatggtgtaccagcttggaggaatgcgacaattcccctcttttgtgcaaccaagaataggatacttgggg
aacaactcagtgcctaccagataatggtgattattcagaagtggcccttaatgttacagaaagctttgatgcctggaataatacagtcacagaacaggcaatagaggatgtatggcaactctttgagacctcaataaagccttgtgtaaaattat

ccccattatgcattactatgagatgcaataaaagtgagacagatagatggggattgacaaaatcaataacaacaacagcatcaacaacatcaacgacagcatcagcaaaagtagacatggtcaatgagactagttcttgtatagcccagga
taattgcacaggcttggaacaagagcaaatgataagctgtaaattcaacatgacagggttaaaaagagacaagaaaaaagagtacaatgaaacttggtactctgcagatttggtatgtgaacaagggaataacactggtaatgaaagtaga
tgttacatgaaccactgtaacacttctgttatccaagagtcttgtgacaaacattattgggatgctattagatttaggtattgtgcacctccaggttatgctttgcttagatgtaatgacacaaattattcaggctttatgcctaaatgttctaaggtggt
ggtctcttcatgcacaaggatgatggagacacagacttctacttggtttggctttaatggaactagagcagaaaatagaacttatatttactggcatggtagggataataggactataattagtttaaataagtattatatctaacaatgaaatgt
agaagaccaggaaataagacagttttaccagtcaccattatgtctggattggttttccactcacaaccaatcaatgataggccaaagcaggcatggtgttggtttggaggaaaatggaaggatgcaataaaagaggtgaagcagaccattgt
caaacatcccaggtatactggaactaacaatactgataaaatcaatttgacggctcctggaggaggagatccggaagttaccttcatgtggacaaattgcagaggagagttcctctactgtaaaatgaattggtttctaaattgggtagaagat
aggaatacagctaaccagaagccaaaggaacagcataaaaggaattacgtgccatgtcatattagacaaataatcaacacttggcataaagtaggcaaaaatgtttatttgcctccaagagagggagacctcacgtgtaactccacagtga
ccagtctcatagcaaacatagattggattgatggaaaccaaactaatatcaccatgagtgcagaggtggcagaactgtatcgattggaattgggagattataaattagtagagatcactccaattggcttggcccccacagatgtgaagagg
tacactactggtggcacctcaagaaataaaagaggggtctttgtgctagggttcttgggttttctcgcaacggcaggttctgcaatgggcgcggcgtcgttgacgctgaccgctcagtcccgaactttattggctgggatagtgcagcaaca
gcaacagctgttggacgtggtcaagagacaacaagaattgttgcgactgaccgtctggggaacaaagaacctccagactagggtcactgccatcgagaagtacttaaaggaccaggcgcagctgaatgcttggggatgtgcgtttagac
aagtctgccacactactgtaccatggccaaatgcaagtctaacaccaaagtggaacaatgagacttggcaagagtgggagcgaaaggttgacttcttggaagaaaatataacagccctcctagaggaggcacaaattcaacaagagaag
aacatgtatgaattacaaaagttgaatagctgggatgtgtttggcaattggtttgaccttgcttcttggataaagtatatacaatatggagtttatatagttgtaggagtaatactgttaagaatagtgatctatatagtacaaatgctagctaagttaa
ggcaggggtataggccagtgttctcttccccaccctcttatttccagcagacccatatccaacaggacccggcactgccaaccagagaaggcaaagaaagagacggtggagaaggcggtggcaacagctcctggccttggcagatag
aatatattcatttcctgatccgccaactgatacgcctcttgacttggctattcagcaactgcagaaccttgctatcgagagtataccagatcctccaaccaatactccagaggctctctgcgaccctacagaggattcgagaagtcctcaggac
tgaactgacctacctacaatatgggtggagctatttccatgaggcggtccaggccgtctggagatctgcgacagagactcttgcgggcgcgtggggagacttatgggagactcttaggagaggtggaagatggatactcgcaatcccca
ggaggattagacaagggcttgagctcactctcttgtga'

13. 'CIVcpzUS Chimpanzee''AF103818'

'atgaaagtgatggagaagaagaagagactctggctaagttattgtcttctctcgagcttgataatcccaggattgtctagcttatgggctacagtatattatggggtacctgtatggagagatgtagagacaaccttattctgtgcctctgatgc
aaaggcatacaagcaggaggcccacaacatttgggccacacaggcatgtgtccctactgaccccaatccacaagaagtacatttgccaaatgtgactgaaaagtttgacatgtgggaaaataatatggcagaacaaatgcaggaagatat
tattagcctatgggaccagagcttaaaaccttgtataaagttaaccccattgtgtgttactatgacttgtcttaaccccgatagtaatagtagtgctgtaaatactactgatataatgagaaactgttcttttaatataactactgaattaagagacaaa
aagaaacaagtgtattccttattttatgtagatgatctagctcatatcaataataatacctatagactgataaattgcaacaccaccgctatcacacaggcttgtcctaagacctcctttgagccaattccaatacactattgtgcaccaccaggctt
tgccatcctaaaatgtaatgaaaaagatttcaaaggaaagggagagtgtaaaaatgttagtacagtgcagtgtactcatggcataaaaccagtggtgactacacagctcataataaatggcagtttagcaactaaaaatgttactgtaagaag
taaaaactttgcagacattattctagtacaattctcagagggagtcaatatgacttgtattagaccaggaaacaatacagtgggaaatgtacaactaggaccaggaatgactttttataacataccaaagatagtaggagatgtaagagaagca
cattgtaacatctcaaaactgacatgggagaaacagagaaaatacactttagagataataaaaaaggaagcaaacctgacaaaggtagagttaattccaaatgcaggaggagacccagaagtggtaaatatgatgcttaattgtggagga
gaatttttttattgtaatacaattcccctatttaacatgacctacaacaataccgacaacaccactatcacacttaagtgtagaataagacaaattgtaaatcagtggatgagagtaggaaaaggaatctttgccccaccaatcaaaggtgtgcta
agttgtaactcaaatataacgggaatgattcttgacataagcataagcgcagtcaataacgatagtaggaatataacagtgatgcctacaggaggagatatgacggctttatggaaaaatgaattacataagtataaggtggtcagcatagaa
cctataggagtggcaccaggtaaggccaaaaggcatacagtgaaaagagaaaaaagagcagccttcggactaggtgcactgttcctggggtttcttggagcagcaggaagcactatgggcgcagcatcagtagtgctgacggtacag
gcccgacaattattgtcagggatagtgcaacagcagaataatctgctaagagcaatagaggcccaacagcatttattgcaactatcagtttggggcattaaacagcttcaagccagagtacttgctgtggaaagataccttaaggatcagca
gatcctaggtctgtggggctgctcaggaaaaacaatttgttataccactgtgccttggaatgatacctggagtaacaacctctcctatgatgctatttggggcaatctaacttggcaagaatgggacagaaaagtaagaaactattcaggtact
atttttagtcttatagaacaagcacaagaacagcagaatacaaatgaaaaatcactcttggaattggatcaatggtcaagtctatggaactggtttgatattaccaactggctgtggtatataaaaatatttttaatagtagtagcaagcttagtag
gaatcagaattgtaggtgtgatattttcactagtagcaaaagttaggcagggttattctcccctctcgttacagacccttttcccaacaactcgggaaccagacaggccagaaggaacagaagaaggcgctggaaaaacagacaacgtca
gatcaacgagattagtcagcggattcttagcacttgtctgggaagatcttcggaacctcctgatcttcctctaccaccgattggaggacttactattaattctgaggaggacagtacaaatcctgggacagaacatcaacaagggactgcaac
tgttgaatgaactcagagcacgctgctggggggtaatcgcttactgggcaagggagctaaaagttagtgctacaagcttattagatacaacagcaatagcagtagctgaaggaactgatagaattatagaattaacaagaaggctctttcta
ggtattatacacataccaagaagaattaggcaaggcctagagagaagcttattataa'

14. 'SIVmon Cercopithecus Monkeys''AY340701'

'atgagagaaggagacatgactcaggggttgacacagaatctgaatcagagcaacatgaggaaactcatggctttgtcaaccctgtgtttaatgatgactttggtgaatgggtctaactggaccaccgtctactacggtgttccagtatggaa
accggcaactcctcctctcttttgtgcctctgatccaaattacgggtctaaagaggcagggaacaattggctggcgtcttcctgcctcccgacagaccctaccccgcaatctctatatttgaacatcactgaggagttcaatgcttatcaaaact
acatggtggaagaaatggtagaggatatgaagtctctgttctcgcaggctttaaagccttgtgtgaaattaacacccatgtgtgttaggatgctctgtgttgaagttaacaccgtttcgaatgccagtaccactcctgcccctagtacccccaca
ccctggggaaattggggagggaatggaacaggacagccagtatataattgctcatttaaccagactacagaattcagggataaaaagaagcaaatgtatagcttattttggaaggaagacataatgaaggaggaaggaagcaatggcag
tcactattacatcctgaattgcaacacatcatacataacccaagcttgtgagaagtccaattatgagccagttcctttgcattattgtgcacccccaggctatgctttattgagatgtgatgaccccgcctttacgggacaagggagttgttctaat
gtttctgcagtaacatgtactcatgctatccaacctatagtagctacttggttccagttgaacagtacgggtaatgctcccaatacaacagtaatgatgaataagcagaaaaatgagtctatagtagttagattggctaagcacctacatgtcaac
attacttgcattaggccaggaaacaaaactattaggaatttacagataggagcaggcatgacattctattcgcaacttatagtgggaggtaatactcgtaaagcctactgcaaagttaataagacccaatgggagactgccctacaggcagtt
cacgaggcagtaaaaactgagtgggaaaagaaaaacaatggaaccaatgtgaccacaatatcttggagatttcaaccccaaggggacaaggaggtccagactcactggtttaactgccagggagaattcttttattgtaatgtctcagcttt
attcattaatagaagaaccaacaaaacggacgggatatctccattcgacgtgaataacaagcccaacaccacgtatcatggtgggtggctagcatgtaccataagacaaatggtgacacaatggggatatgtgtcaaagagtatctaccta
cctccacggaaaggccatgtgcagtgtacatctaatatcactgccctcctgattacaggggagttgtatcaaaacaacgtgacccttgttccctcggcacaggtgagtgactcatggagatcagagctcagtcgatataaggtggtggaaat
tgacccccttatctatggcacctacaacagcgcagagaaggactggggtacacagggaaaagagagcaattactttgggtatggcctttctgggcttcctcagcacggcaggaggcacaatgggcgcagcggcgacagccctgacggt
acagtctcggtcattactggctgggatagtgcagcagcaggagaacctgctgagggccgtaacagcccaacagagcttactacagcttactgtttggggagtcaagcagctacaggcccgcctgacggctgtagagaagttcataaagg
accaaacattgctaaatgcatggggatgtgccaataaagccgtgtgtcacacaacagtgccgtggaacaattcttgggcaaaaggtcacttccctgagtgggacaatatgacatggcaacagtggagtgagttagtagataatgacacaat
gaccattcagcagctcttggaagctgcgcaagagcagcaaggcaaaaaccaacatgagttaatgaagccgggacaatgggacttcctgtggaattggtttgacatctccaaatggttgtggtacatcaaaatattcattatagtagtagcag
ctttgataggcttaagaatacttatgttcatactaggagttatcagtaggttagggcagggatactctcttctttcttctcagatccctatcccatcccacgcgggacagccaacgcccgacggaacaggcgcaggcggtggcgacggcagc
aactccagatcgccagcatatctgaaaggatttttttaccattatctgggaagatctcaggaaccttgtcctctggacataccagatcttgaaagactcagtattagtgatctaccggatcctccagagagtgtcccagaggctgccacccctcc
tgcacatacgcctgctccaactgtgggaaagccttagacgcttgctagcctactgccagtatgggatccaagaactccaagcagcagtcacctctctcctcgacgcccttgctaggttcacaatcgtctggacggatgcgctacttcatgct
ggaggacgactatggagagcaatcgtggctatccccagacgcatccgacagggagcggaaatattccttaactga'

15. 'SIVcpzTAN1 Chimpanzee''AF447763'

'atgaagaatttaattggaataactttgatcctcataattacaatcctagggattggatttagcacttattacaccacagtgttttatggagtacctgtttggaaagaggcccaaccaaccttgttttgtgcctctgatgctgatattactagtagagat
aaacacaacatatgggcaacacataactgtgtgcctttagatcccaatccttatgaagtaaccctagccaatgtgtcaataaggtttaatatggaagaaaattacatggtgcaagagatgaaagaagatatattatcactttttcaacagagtttt
aagccttgtgtaaaattaacaccattttgcataaagatgacatgtacaatgactaataccacaaataaaaccctgaattcggcaacaacaaccttaacaccaacagtaaatttgagttctatacctaactatgaggtgtataattgttcatttaatca
gacaactgagtttagagataagaaaaaacaaatatattccttgttttatagagaagatattgtaaaagaggatggtaacaataatagttattatttacataattgcaatacctcagtcattactcaagaatgtgataaatctactttgaaccaattccc
atcagatactgtgctccagcaggctttgccctgttaaaatgtagagatcagaatttcacagggaaaggacaatgctccaatgtctcagtagttcactgtacacatgggatttatcctatgatagccacagcattacacttaaatgggtccctgga
agaagaagaaacaaaagcttactttgttaatacctcagttaatacacccttattagtaaaatttaatgtatcaataaatttaacgtgtgaaagaacaggaaacaatacaagaggtcaagtacagataggtccaggtatgaccttttataatataga
aaatgtagtaggggacaccaggaaagcttattgttcagtcaatgcaacaacatggtacaggaacttagattgggctatggctgccataaacacaaccatgagggccagaaatgaaacggtacaacaaacgttccaatggcagagggatg
gagaccctgaggtcactagcttctggttcaattgtcaaggagaattcttttactgtaatctcacaaattggactaatacctggacagctaatagaaccaataatactcatggtactcttgttgcaccatgcagactgaggcagatagtaaatcatt
ggggtatagtgtcaaaaggggtttaccttcccccaaggaggggaacagtaaaatgtcactcaaacatcacaggacttatcatgacagcagaaaaagacaacaataatagttatacccccccaattttctgctgtagtagaagactattggaaa
gtagaattagcaagatataaagtggtggaaattcagcccttgtcagtggctccaaggccaggaaaaaggcctgaaattaaggccaatcatactaggtcaagaagagatgtgggcataggactgttgtttcttggatttcttagtgcagcagg
aagtacaatgggcgcagcgtcaatagcgctgacggcacaggccagaggattactctctggtattgtacagcagcaacaaaacctgcttcaggccatagaagcgcaacaacacttgttgcagctctctgtatggggcattaagcagctcca
ggccagaatgcttgcagtagagaaatacataagagaccaacagctcctaagcctctggggatgtgctaacaaattggtgtgtcacagtagtgtgccatggaacctcacctgggctgaagattctacaaagtgcaatcacagtgatgcaaa
gtactatgactgtatatggaacaatttgacttggcaggaatgggatcgattagtagaaaactctacaggaaccatatactccctgttagagaaagcacaaacacaacaggagaaaaacaaacaagagttgttagaattagacaaatggagc
agtctttgggattggtttgatataacacaatggctgtggtatataaaaatagctataatcatagtagcaggattagtaggacttagaattctcatgtttatagttaatgtagttaagcaagttaggcagggttatacacccctattttcacagatccct
acccaagcggagcaggatccagaacagccaggaggaatcgcaggaggaggtggaggcagagacaacatcaggtggacgccctcgccagcaggattcttcagtatcgtctgggaggacctcaggaacctcctcatctggatatacca
gacctttcaaaacttcatctggatcctctggatcagcctgcaagcactgaaacaggggataatcagcttggcacacagcctagtaatagtgcatagaactatcatagtaggagttagacagatcattgagtggagcagtaatacttatgctag
cttaagagttttgctaatacaagccatagacagacttgctaactttacagggtggtggacagatttaatcatagaaggagtggtttacatagccaggggaatcagaaatattcctagaagaattagacagggtctggaactagccttaaattaa'

16. 'SIVsmSL92b Sooty Mangabey''AF334679'

'atggcgtgtcctggacttcacctgcttatagacatcttgtttttaagtgtgttagggacctggtgtgcacagtatgtaacaatcttttatggtattcctgcatggaggaacgctacgatccccctcttctgtgcgacccagaatagagacacatgg
gggaccgttcagtgcttgccagataatggagactattcagaattggcccttaatgttacagaggccttcgatgcttgggataatacagtaactgaacaagcaatagaagatgtttggaatctctttgaaacatctattaaaccttgtgtaaagtta
actccattatgtattgctatgaaatgtaataaaaatgagacagacagatgggggttgacaagagcagctactactactagctcaccaactactactagccccttaactgctgctagcccatcaggagaggaaatcgttaacgacactatgtctt
gtacaaagaacaacaattgctctggcatagagcaggaaccaatgataggttgtcaatttaacatgacaggactaaaaagagaccagaaaaggcagtacaatgaaacctggtactccagagatctagtctgtgagcaaggaggaaatgaa
agcagtagatgttacatgaatcattgcaatacaagtgttatacaagagtcatgtgacaagcactattgggatgccatcagatataggtactgtgcaccaccaggctatgctttgcttagatgtaatgatacaaattattcaggctttgcacctaatt
gtagtaaggtagtagtatcatcatgtacaaggatgatggaaacacagacttctacatggtttggctttaatggtactagagcagagaatagaacatatatatattggcatggcggaagtaatagaacaataattagcttaaataagtattataattt
gacaataaagtgcagaagaccaggaaataagacagtcctaccagtcaccattatgtcaggtttggtctttcattcgcaaccaataaatgagaggccaagacaagcatggtgctggtttggaggaaaatggagagaagctatgcaggaggt
taagaaaaccatagtcaaacaccccaggtatactggaaccaatgatactaggaaaattaatctaacggccccgggaggaggggatccggaagtcacattcatgtggacaaattgcagaggagaattcctttactgtaaaatgaattggttc

```
ctcaattgggtagaggataggaatacgagcagcccaaggtggacaactcaaaccaaaaaggaacagcacaaaagaaattatgtaccttgccacatcaggcaaataattaatacgtggcacagggtaggaaaaaatgtctatttgcctcca
agggaaggagacttaacttgtaactccacagtgaccagcctaatagcaaatattgattggattgacaacaatgagaccaatattaccatgagtgcagaagtggcagaattgtatcgattggaattgggggattataaattggtagagatcact
ccaattggcatggctcccacacatgtgaaaaggtacaccacaagtacctcaaagaataaaagaggagtctttgtgctagggttcttgggttttctcgcgacggcaggttctgcaatgggcgcagcgtcgctaacgctgactgctcagtccc
ggactttattggctgggatagtgcaacaacagcaacagctgttggacgtggtcaagagacaacaagaattgttgcgactgaccgtctggggaactaaaaatcttcagactagagtcaccgccatcgagaaatacctaaaggatcaggcac
agctaaattcatggggatgtgcatttaggcaagtctgccacactactgtactatggccaaatgacagtcttgtcccagactggaacaatatgacatggcaagaatgggagaaaaaggttgaattcctagaggcaaatataactcaaatgttgg
aggaagcacgactacagcaagagaaaaacatgtatgaattacagaaattgaatagctgggatgtgtttggaaattggtttgaccttacctcctgggtaagatacatacagtatggagtctttctagtcataggaatagtattgttaagaatagtt
atctatgtagtacaaatgttaagtaggttaaggcagggttataggcctgttttctcctcccccccatcttatcatcagcagatccatatccagcgggaccaggaactgccagccaaagaaggaacagaaggagaaggtggaaacggaggt
ggatacagattgtggccttggcagatagaatatattcatttcctgatccgccagctgatacggatattgacttggctatacaacaatttgactcgcttagcatccagagcctaccagaacctccaacaactgtgccagagactctcagagatca
gtcagccgattagagagcttgtcagaagagaagcaggctatatccgttatgggtggaattacttcatcgaagcctgccaggaagcatggagatctgcgcaagaagctattgtcggcgcgtggggactcatatgggagactctgggacgg
gttggaagagggatcgcagcaatcccaaggaggatcagacagggccttgaactcatgcttaactag'
```

4.5.14. Multiple Sequence Aligment Results For ENV

The following command for progressive multiple sequence alignment is used via MATLAB;

>> envaligned=multialign(env.'terminalGapAdjust'.true)

Table 26. Only the aligned result of the first 99 nucleotide sites were listed for demonstration.

LOCUS NAME	VERSION	MULTIPLE SEQUENCE ALIGMENT
'HIV-1 (Zaire)'	'K03454'	'atgagagcgagggggatagaga---gaaattgtcaaaactggtggaaatggggcatcatg------ctccttgggatattgatgacctgtagtgctgc------------agacaa
'HIV1-NDK (Zaire)'	'M27323'	'atgagagcgagggagaaagaga---ggaattgtcaaaacttgtggaaatggggcatcatg------ctccttgggatgttgatgacctgtagtgctgc------------agaaga
'HIV-2 (Senegal)'	'M15390'	'atgatg---------aatcagct-------------gcttattgccattttattag---------------------ctagtgcttgcttagtatattgc------------acccaatatgtaactgt
'HIV2-MCN13'	'AY509259'	'atgatgggtggtagaaatcagct-------------gcttgttgccattttgctaa---------------------ctagtacttgcttgatatattgc------------accaattatgtgac
'HIV-2UC1 (IvoryCoast)'	'L07625'	'atggcacacactagcaatcacct-------------gtttattttgctcctactta---------------------taagtgtctatgggtttctgggt------cataagaaaaactatgtca
'SIVMM251 Macaque'	'M19499'	'atgggatgtcttgggaatcagct-------------gcttatcgccatcttgcttt---------------------taagtgtctatgggatctattgt------------actcaatatgtcaca
'SIVAGM677A Green monkey'	'M58410'	'--------atggggagat----t-------------gcttataaaaatactaataa---------------------tagcaatagggataagtatagga------ataggtaacctgtatgtga
'SIVlhoest L''Hoest monkeys'	'AF075269'	'atggcatgtccagga-------------------------ttaggaattctgcttt---------------------tgcttcttggcataatatgggga------------aaacaatatgtaacag
'SIVcpz Chimpanzees Cameroon'	'AF115393'	'atgaaagtgatggagaagatca------------aaggcagcttgatagagaaatacatg------tttatgggtttgataatcccatgtttgactggtag---------tgatcagttgt
'SIVmnd5440 Mandrillus sphinx'	'AY159322'	'-----atgttaagatatc----------------------ttaggtatatagtcttagga----------------ataatagtaagtgtaatagtagga------------gaacaatgggtgac
'SIVAGM3 Green monkeys'	'M30931'	'-----atgaagctgacat----t-------------actgatagggatactattaa---------------------tagggataggagtagtgcttaat------acaaggcaacaatgggtc
'SIVMM239 Simian macaque'	'M33262'	'atgggatgtcttgggaatcagct-------------gcttatcgccatcttgcttt---------------------taagtgtctatgggatctattgt------------actctatatgtcacag
'CIVcpzUS Chimpanzee'	'AF103818'	'atgaaagtgatggagaagaaga------------agagactctggctaagttattgtctt------ctctcgagcttgataatcccaggattgtctag-----------------cttatggg
'SIVmon Cercopithecus Monkeys'	'AY340701'	'atgagag--aaggagacatgactcaggggttgacacagaatct-gaatcagagcaacatgaggaaactcatggctttgtcaaccctgtgtttaatgatgactttggtgaatggg
'SIVcpzTAN1 Chimpanzee'	'AF447763'	'---------atgaagaatttaa-------ttggaataactt----------------tg------atcctcataattacaatcctagggattggattta------------gcacttattacaccaca
'SIVsmSL92b Sooty Mangabey'	'AF334679'	'atggcgtgtcctggacttcacct-------------gcttatagacatcttgtttt---------------------taagtgtgttagggacctggtgt------------gcacagtatgtaaca

4.5.15 Distance Tables of ENV protein coding gene

Table 27. Alignment Score (env)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	0.00000															
2	0.01917	0.00000														
3	0.79405	0.76644	0.00000													
4	0.81489	0.79811	0.03934	0.00000												
5	0.84224	0.82827	0.22862	0.23233	0.00000											
6	0.82799	0.81486	0.21664	0.21095	0.23054	0.00000										
7	0.81141	0.81000	0.56917	0.54168	0.54547	0.56141	0.00000									
8	0.95719	0.96633	0.78416	0.77768	0.83838	0.77877	0.75747	0.00000								
9	0.31742	0.32306	0.74565	0.76186	0.78950	0.82980	0.79461	0.95579	0.00000							
10	0.96326	0.93964	0.75884	0.75884	0.74769	0.72964	0.72322	0.50012	0.92031	0.00000						
11	0.82184	0.81842	0.58113	0.58024	0.56925	0.58107	0.26533	0.78353	0.80724	0.75662	0.00000					
12	0.80039	0.79001	0.20985	0.19982	0.22689	0.00241	0.55867	0.77683	0.81341	0.72439	0.58383	0.00000				
13	0.35266	0.34945	0.76378	0.75524	0.79733	0.81768	0.75750	0.91740	0.16946	0.88040	0.78104	0.80303	0.00000			
14	0.66093	0.65137	0.75492	0.74098	0.75649	0.80340	0.75710	0.95899	0.60499	0.86585	0.74686	0.79810	0.61681	0.00000		
15	0.52473	0.51692	0.78067	0.76445	0.80588	0.78232	0.79405	0.93713	0.50188	0.87836	0.78539	0.76695	0.48262	0.57847	0.00000	
16	0.83689	0.82873	0.21227	0.21242	0.24473	0.12482	0.56632	0.78731	0.79989	0.73055	0.59807	0.12059	0.80237	0.76200	0.77842	0.00000

Table 28. Hasegawa (env)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	0.00000															
2	0.09814	0.00000														
3	1.22946	1.21802	0.00000													
4	1.24130	1.22449	0.14220	0.00000												
5	1.26235	1.24538	0.45067	0.45360	0.00000											
6	1.24573	1.23098	0.43899	0.43197	0.45907	0.00000										
7	1.22405	1.22085	1.35338	1.17749	1.19442	1.28896	0.00000									
8	1.51941	1.55272	1.21005	1.20865	1.24946	1.21161	1.21154	0.00000								
9	0.60929	0.61734	1.20489	1.20244	1.21434	1.24640	1.22929	1.51740	0.00000							
10	1.53716	1.45261	1.19020	1.19869	1.20233	1.20661	1.20290	1.01977	1.39552	0.00000						
11	1.22978	1.22687	1.48937	1.47689	1.35378	1.49104	0.52476	1.21543	1.23850	1.20914	0.00000					
12	1.22840	1.21870	0.42907	0.41570	0.45221	0.03105	1.26848	1.21005	1.23239	1.20971	1.53674	0.00000				
13	0.66650	0.66000	1.20333	1.20463	1.21970	1.23940	1.20683	1.39703	0.36782	1.33352	1.21388	1.23009	0.00000			
14	1.26325	1.28961	1.19579	1.20259	1.20164	1.22228	1.21065	1.53109	1.65286	1.29568	1.20296	1.22120	1.46424	0.00000		
15	1.09778	1.06784	1.21872	1.21259	1.22970	1.22759	1.23221	1.44792	1.02366	1.32781	1.22665	1.22526	0.96493	1.45809	0.00000	
16	1.25978	1.24810	0.43769	0.43880	0.48086	0.29824	1.32376	1.21981	1.22664	1.19651	1.91535	0.29226	1.22383	1.20888	1.22304	0.00000

Table 29. Jukes Cantor (env)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	0.00000															
2	0.08116	0.00000														
3	0.80927	0.78414	0.00000													
4	0.82850	0.81299	0.11918	0.00000												
5	0.85412	0.84098	0.32793	0.33125	0.00000											
6	0.84071	0.82847	0.31708	0.31188	0.32965	0.00000										
7	0.82527	0.82396	0.61376	0.59089	0.59403	0.60729	0.00000									
8	0.96745	0.97692	0.80022	0.79432	0.85048	0.79532	0.77607	0.00000								
9	0.40511	0.40987	0.76548	0.78001	0.80510	0.84241	0.80977	0.96601	0.00000							
10	0.97373	0.94949	0.77730	0.77730	0.76731	0.75125	0.74557	0.55651	0.92999	0.00000						
11	0.83496	0.83178	0.62376	0.62301	0.61383	0.62370	0.36041	0.79965	0.82140	0.77530	0.00000					
12	0.81509	0.80556	0.31087	0.30161	0.32636	0.02778	0.60501	0.79355	0.82712	0.74661	0.62602	0.00000				
13	0.43472	0.43204	0.78175	0.77407	0.81227	0.83109	0.77610	0.92709	0.27281	0.89067	0.79738	0.81752	0.00000			
14	0.69136	0.68316	0.77378	0.76132	0.77519	0.81786	0.77573	0.96931	0.64380	0.87661	0.76656	0.81298	0.65377	0.00000		
15	0.57684	0.57038	0.79704	0.78235	0.82015	0.79854	0.80927	0.94694	0.55796	0.88869	0.80134	0.78461	0.54209	0.62153	0.00000	
16	0.84907	0.84140	0.31309	0.31323	0.34231	0.22756	0.61138	0.80309	0.81463	0.75206	0.63797	0.22303	0.81691	0.78014	0.79499	0.00000

Table 30. Kimura (env)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	0.00000															
2	0.08163	0.00000														
3	0.81305	0.78773	0.00000													
4	0.83301	0.81878	0.12075	0.00000												
5	0.86025	0.84786	0.33463	0.33901	0.00000											
6	0.84738	0.83617	0.32151	0.31578	0.33468	0.00000										
7	0.83454	0.83409	0.61907	0.59507	0.59898	0.61341	0.00000									
8	0.97129	0.98294	0.80815	0.80170	0.85971	0.80187	0.78071	0.00000								
9	0.40859	0.41379	0.77161	0.78771	0.81251	0.84945	0.81367	0.96889	0.00000							
10	0.98037	0.95482	0.78965	0.78610	0.77428	0.75719	0.75316	0.55918	0.93639	0.00000						
11	0.84552	0.84230	0.63075	0.62930	0.61928	0.62881	0.36317	0.80569	0.82512	0.78055	0.00000					
12	0.82018	0.81163	0.31506	0.30515	0.33169	0.02790	0.61169	0.80034	0.83392	0.75200	0.63173	0.00000				
13	0.43975	0.43724	0.78930	0.78058	0.81944	0.83680	0.78206	0.93084	0.27618	0.89235	0.80354	0.82264	0.00000			
14	0.69927	0.68977	0.78342	0.76816	0.78276	0.82554	0.78059	0.97196	0.64608	0.88178	0.77331	0.81982	0.65864	0.00000		
15	0.58235	0.57579	0.80176	0.78714	0.82597	0.80137	0.81243	0.95131	0.56104	0.89077	0.80468	0.78678	0.54508	0.62485	0.00000	
16	0.85420	0.84754	0.31615	0.31610	0.34766	0.23021	0.61900	0.80840	0.82013	0.76136	0.64469	0.22545	0.82382	0.78579	0.79841	0.00000

Table 31. Nei Tamura (env)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	0.00000															
2	0.09807	0.00000														
3	1.84202	2.02732	0.00000													
4	2.13972	2.43702	0.14249	0.00000												
5	2.29896	2.88011	0.44572	0.44849	0.00000											
6	2.46113	2.46901	0.43352	0.42675	0.45301	0.00000										
7	2.66815	2.52592	1.09095	1.02339	1.02806	1.06447	0.00000									
8	1.44052	1.40957	2.04683	1.97390	2.07304	2.00841	2.17656	0.00000								
9	0.59585	0.60305	2.10085	2.15537	1.82193	2.06358	2.06869	1.48843	0.00000							
10	1.41265	1.47994	1.86214	2.33979	2.29466	1.86340	1.72081	0.93495	1.67280	0.00000						
11	2.26507	2.37356	1.14471	1.13988	1.09213	1.13473	0.51641	1.70637	2.32542	1.72325	0.00000					
12	2.26745	2.28999	0.42392	0.41113	0.44629	0.03107	1.05859	2.06956	1.95822	1.77781	1.13752	0.00000				
13	0.65004	0.64358	1.94213	1.91504	1.80954	1.86641	2.22871	1.58818	0.36576	1.94359	1.85901	1.88113	0.00000			
14	1.45104	1.38257	1.73241	2.77154	2.06225	1.84049	2.26538	1.50817	1.22476	2.24671	2.31044	1.94191	1.25034	0.00000		
15	0.98275	0.96406	1.79791	1.86507	2.03778	1.75420	1.98642	1.52004	0.93736	1.97992	1.82428	1.73305	0.89731	1.15622	0.00000	
16	2.21332	2.42351	0.43274	0.43367	0.47414	0.29653	1.07417	2.65791	1.74255	1.84856	1.16907	0.29071	1.81139	2.04481	1.90318	0.00000

Table 32. P Distance (env)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	0.0000															
2	0.07692	0.00000														
3	0.49505	0.48637	0.00000													
4	0.50151	0.49632	0.11019	0.00000												
5	0.50985	0.50561	0.26564	0.26778	0.00000											
6	0.50552	0.50150	0.25858	0.25516	0.26675	0.00000										
7	0.50044	0.50000	0.41913	0.40888	0.41031	0.41626	0.00000									
8	0.54353	0.54612	0.49196	0.48992	0.50868	0.49027	0.48352	0.00000								
9	0.31300	0.31577	0.47973	0.48491	0.49363	0.50607	0.49523	0.54314	0.00000							
10	0.54526	0.53853	0.48395	0.48395	0.48038	0.47455	0.47246	0.39289	0.53296	0.00000						
11	0.50364	0.50259	0.42351	0.42319	0.41916	0.42349	0.28617	0.49176	0.49915	0.48324	0.00000					
12	0.49703	0.49379	0.25450	0.24834	0.26463	0.02727	0.41525	0.48966	0.50105	0.47284	0.42449	0.00000				
13	0.32992	0.32841	0.48553	0.48280	0.49607	0.50237	0.48353	0.53212	0.22870	0.52128	0.49098	0.49784	0.00000			
14	0.45165	0.44837	0.48270	0.47822	0.48320	0.49796	0.48340	0.54404	0.43212	0.51695	0.48012	0.49631	0.43632	0.00000		
15	0.40243	0.39943	0.49086	0.48574	0.49873	0.49138	0.49505	0.53781	0.39357	0.52067	0.49235	0.48653	0.38595	0.42254	0.00000	
16	0.50823	0.50575	0.25596	0.25605	0.27483	0.19628	0.41808	0.49295	0.49687	0.47485	0.42964	0.19292	0.49764	0.48496	0.49016	0.00000

Table 33. Tajima-Nei (env)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	0.00000															
2	0.08046	0.00000														
3	0.66175	0.64635	0.00000													
4	0.67348	0.66534	0.11840	0.00000												
5	0.69074	0.68355	0.31225	0.31606	0.00000											
6	0.68128	0.67427	0.30117	0.29528	0.31226	0.00000										
7	0.67182	0.67141	0.53329	0.51571	0.51908	0.52717	0.00000									
8	0.75207	0.75748	0.65839	0.65292	0.68918	0.65202	0.63982	0.00000								
9	0.37269	0.37703	0.63445	0.64431	0.66056	0.68086	0.65998	0.75090	0.00000							
10	0.75758	0.74422	0.64736	0.64428	0.63694	0.62435	0.62125	0.48836	0.73110	0.00000						
11	0.67978	0.67732	0.54322	0.54062	0.53310	0.53840	0.33456	0.65515	0.66743	0.63968	0.00000					
12	0.66402	0.65904	0.29555	0.28617	0.30996	0.02779	0.52657	0.65065	0.67157	0.62074	0.54057	0.00000				
13	0.39778	0.39587	0.64710	0.64124	0.66688	0.67551	0.64040	0.72855	0.26112	0.70925	0.65471	0.66658	0.00000			
14	0.59243	0.58604	0.64581	0.63483	0.64471	0.67002	0.64413	0.75736	0.55569	0.70414	0.63832	0.66716	0.56590	0.00000		
15	0.50615	0.50150	0.65591	0.64520	0.67106	0.65294	0.66079	0.73969	0.49114	0.70831	0.65641	0.64386	0.47862	0.54010	0.00000	
16	0.68470	0.68139	0.29636	0.29529	0.32309	0.21991	0.53112	0.65610	0.66419	0.62599	0.54971	0.21563	0.66712	0.64679	0.65274	0.00000

4.5.16 Translated ENV protein-coding gene (Amino Acid Sequence) (designated by AAENV)

1. 'HIV-1 (Zaire)'- 'K03454'

```
'MRARGIERNCQNWWKWGIMLLGILMTCSAADNLWVTVYYGVPVWKEATTTLFCASDAKSYETEAHNIWATHACVPTDPNPQEIALENVTENFNM
WKNNMVEQMHEDIISLWDQSLKPCVKLTPLCVTLNCSDELRNNGTMGNNVTTEEKGMKNCSFNVTTVLKDKKQQVYALFYRLDIVPIDNDSSTNS
TNYRLINCNTSAITQACPKVSFEPIPIHYCAPAGFAILKCRDKKFNGTGPCTNVSTVQCTHGIRPVVSTQLLLNGSLAEEEVIIRSENLTNNAKNIIAHLN
ESVKITCARPYQNTRQRTPIGLGQSLYTTRSRSIIGQAHCNISRAQWSKTLQQVARKLGTLLNKTIIKFKPSSGGDPEITTHSFNCGGEFFYCNTSGLFNS
TWNISAWNNITESNNSTNTNITLQCRIKQIIKMVAGRKAIYAPPIERNILCSSNITGLLLTRDGGINNSTNETFRPGGGDMRDNWRSELYKYKVVQIEPL
GVAPTRAKRRVVEREKRAIGLGAMFLGFLGAAGSTMGARSVTLTVQARQLMSGIVQQQNNLLRAIEAQQHLLQLTVWGIKQLQARILAVERYLKD
QQLLGIWGCSGKHICTTNVPWNSSWSNRSLNEIWQNMTWMEWEREIDNYTGLIYSLIEESQTQQEKNEKELLELDKWASLWNWFSITQWLWYIKIFI
MIIGGLIGLRIVFAVLSLVNRVRQGYSPLSFQTLLPAPRGPDRPEGTEEEGGERGRDRSVRLLNGFSALIWDDLRSLCLFSYHRLRDLILIAVRIVELLGR
RGWDILKYLWNLLQYWSQELRNSASSLFDAIAIAVAEGTDRVIEIIQRACRAVLNIPRRIRQGLERSLL*'
```

2. 'HIV1-NDK (Zaire)'- 'M27323'

```
MRAREKERNCQNLWKWGIMLLGMLMTCSAAEDLWVTVYYGVPIWKEATTTLFCASDAKAYKKEAHNIWATHACVPTDPNPQEIELENVTENFN
MWKNNMVEQMHEDIISLWDQSLKPCVKLTPLCVTLNCTDELRNSKGNGKVEEEEKRKNCSFNVRDKREQVYALFYKLDIVPIDNNNRTNSTNYRLI
NCDTSTITQACPKISFEPIPIHFCAPAGFAILKCRDKKFNGTGPCSNVSTVQCTHGIRPVVSTQLLLNGSLAEEEIIIRSENLTNNVKTIIVQLNASIVINCT
RPYKYTRQRTSIGLRQSLYTITGKKKKTGYIGQAHCKISRAEWNKALQQVATKLGNLLNKTTITFKPSSGGDPEITSHMLNCGGDFFYCNTSRLFNST
WNQTNSTGFNNGTVTLPCRIKQIVNLWQRVGKAMYAPPIEGLIKCSSNITGLLLTRDGGANNSSHETIRPGGGDMRDNWRSELYKYKVVKIEPIGVA
PTKARRRVVEREKRAIGLGAVFLGFLGAAGSTMGAASVTLTVQARQLMSGIVHQQNNLLRAIEAQQHLLQLTVWGIKQLQARVLAVERYLRDQQL
LGIWGCSGRHICTTNVPWNSSWSNRSLDEIWQNMTWMEWEREIDNYTGLIYSLIEESQIQQEKNEKELLELDKWASLWNWFSITKWLWYIKLFIMIV
GGLIGLRIVFAVLSVVNRVRQGYSPLSFQTLLPVPRGPDRPEEIEEEGGERGRDRSIRLVNGLFALFWDDLRNLCLFSYHRLRDSILIAARIVELLGRRG
WEALKYLWNLLQYWSQELRNSASSLLDTIAIAVAERTDRVIEVVQRACRAILNVPRRIRQGLERLLL
```

3. 'HIV-2 (Senegal)'- 'M15390'

```
'MMNQLLIAILLASACLVYCTQYVTVFYGVPTWKNATIPLFCATRNRDTWGTIQCLPDNDDYQEITLNVTEAFDAWNNTVTEQAIEDVWHLFETSIKP
CVKLTPLCVAMKCSSTESSTGNNTTSKSTSTTTTTPTDQEQEISEDTPCARADNCSGLGEEETINCQFNMTGLERDKKKQYNETWYSKDVVCETNNS
TNQTQCYMNHCNTSVITESCDKHYWDAIRFRYCAPPGYALLRCNDTNYSGFAPNCSKVVASTCTRMMETQTSTWFGFNGTRAENRTYIYWHGRDN
RTIISLNKYYNLSLHCKRPGNKIVKQIMLMSGHVFHSHYQPINKRPRQAWCWFKGKWKDAMQEVKETLAKHPRYRGTNDTRNISFAAPGKGSDPEV
AYMWTNCRGEFLYCNMTWFLNWIENKTHRNYAPCHIKQIINTWHKVGRNVYLPPREGELSCNSTVTSIIANIDWQNNNQTNITFSAEVAELYRLELG
DYKLVEITPIGFAPTKEKRYSSAHGRHTRGVFVLGFLGFLATAGSAMGAASLTVSAQSRTLLAGIVQQQQQLLDVVKRQQELLRLTVWGTKNLQAR
VTAIEKYLQDQARLNSWGCAFRQVCHTTVPWVNDSLAPDWDNMTWQEWEKQVRYLEANISKSLEQAQIQQEKNMYELQKLNSWDIFGNWFDLT
SWVKYIQYGVLIIVAVIALRIVIYVVQMLSRLRKGYRPVFSSPPGYIQQIHIHKDRGQPANEETEEDGGSNGGDRYWPWPIAYIHFLIRQLIRLLTRLYSI
CRDLLSRSFLTLQLIYQNLRDWLRLRTAFLQYGCEWIQEAFQAAARATRETLAGACRGLWRVLERIGRGILAVPRRIRQGAEIALL*'
```

4. 'HIV2-MCN13'-‘AY509259'

```
'MMGGRNQLLVAILLTSTCLIYCTNYVTVFYGIPAWRNASIPLFCATKNRDTWGTIQCLPDNDDYQEITLNVTEAFDAWDNTVTEQAIEDVWNLFETS
IKPCVKLTPLCVAMRCNNTDARNTTTPTTASPRTIKPVTEISENSSCIRANNCSGLGEEEVVNCQFNMTGLERDKKKQYSETWYSKDVVCEGNGTTD
TCYMNHCNTSVITESCDKHYWDAMRFRYCAPPGFALLRCNDTNYSGFAPNCSKVVAATCTRMMETQTSTWFGFNGTRAENRTFIYWHGRDNRTII
SLNKYYNLTIHCKRPGNKTVVPITLMSGLRFHSQPVINKRPRQAWCWFKGEWKGAMQEVKETLAKHPRYKGTNETKNINFTAPGKGSDPEVAYMW
TNCRGEFLYCNMTWFLNWIENKTHRNYVPCHIRQIINTWHKVGKNVYLPPREGELTCNSTVTSIIANIDANGNNTNITFSAEVAELYRLELGDYKLVE
ITPIGFAPTAEKRYSSTPMRNKRGVFVLGFLGFLATAGSAMGAASLTLSAQSRTLLAGIVQQQQQLLDVVKRQQEMLRLTVWGTKNLQARVTAIEK
YLKDQAQLNSWGCAFRQVCHTTVPWVNDTLTPEWNNMTWQEWEGKIRDLEANISQQLEQAQIQQEKNMYELQKLNSWDVFGNWFDLTSWIKYI
QYGVYIIIGIVVLRIVIYIVQMLSRLRKGYRPVFSSPPGYLQQIHIHKDWEQPAREETEEDVGNNVGDSSWPWPIRYIHFLIHQLIRLLAGLYNICRNLLS
RISLTLRPVFQSLQRALTAIRDWLRTDAAYLQYGCEWIQGAFQAFARATRETLAGTWRDLWGALQRIGRGILAVPRRIRQGAEIALL*'
```

5. 'HIV-2UC1 (IvoryCoast)'-'L07625'

```
'MAHTSNHLFILLLLISVYGFLGHKKNYVTVFYGIPAWRNATVPLFCATTNRDTWGTVQCLPDNGDYTEISVNITEAFDAWNNTVTEQAVDDVWSLF
ETSIKPCVKLTPLCVAMRCNNTGTNTTTKPITTPITTTKPSENLLNDTSPCIKNDTCPGIGLENTVDCYFNMTGLRRDEKKQYKDTWYEKDLECNGNS
TSTICYMRTCNTSVIQESCDKHYWDSLRFRYCAPPGYALLRCNDTNYSGFMPKCSKVVVSSCTRMMETQTSTWFGFNGTRTENRTYMYWHSKDNR
TIISLNKYYNLTMHCRRPGNKTVIPITIMSGLNFHSQPLNTRPRQAWCWFKGNWIEAIREVKETIIKHPRYKGTNNTERIRLVGPSAGSDPEVRHMWT
NCRGEFFYCNMTWFLNWVENRTGTTQKNYVTCHIKQIVNTWHKVGKYVYLPPREGTLSCNSSVTSLIANIDVYYDGNDTKTNITMSAEVGELYRLE
LGDYKLVEITPIGFAPTEIKRYSSTTPRNKRGVMVLGFLGLLAMAGSAMGATSLTLSAQSRTLLAGIVQQQQQLLDVVKRQQELLRLTVWGTKNLQT
RVTAIEKYLKDQALLNSWGCAFRQVCHTTVPWPNETLTPDWENMTWQQWEKRVNFLDANITALLEEAQIQQERNMYELQKLNSWDVFGNWFDFT
SWMAYIRLGLYVVAGLIVLRIVIYIMQMLARLRKGYRPVFSSPPSYTQQIPIRKHRGQPANEETEDEGGNEGAYRSWPWQIEYAHFLIRQLRNLLIWL
YNGCRNLLLKTSQILQPALQPLRLSLAYLQYGISWFQEAIQAATRAARETLANTGRALWKALRRTAEAIIAIPRRIRQGLELALL*'
```

6. 'SIVMM251 Macaque'-'M19499'

```
'MGCLGNQLLIAILLLSVYGIYCTQYVTVFYGVPAWRNATIPLFCATKNRDTWGTTQCLPDNGDYSELALNVTESFDAWENTVTEQAIEDVWQLFET
SIKPCVKLSPLCITMRCNKSETDRWGLTKSSTTITTAAPTSAPVSEKIDMVNETSSCIAQNNCTGLEQEQMISCKFTMTGLKRDKTKEYNETWYSTDL
VCEQGNSTDNESRCYMNHCNTSVIQESCDKHYWDTIRFRYCAPPGYALLRCNDTNYSGFMPKCSKVVVSSCTRMMETQTSTWFGFNGTRAENRTY
IYWHGRDNRTIISLNKYYNLTMKCRRPGNKTVLPVTIMSGLVFHSQPINDRPKQAWCWFGGKWKDAIKEVKQTIVKHPRYTGTNNTDKINLTAPGG
GDPEVTFMWTNCRGEFLYCKMNWFLNWVEDRDVTTQRPKERHRRNYVPCHIRQIINTWHKVGKNVYLPPREGDLTCNSTVTSLIANIDWTDGNQT
SITMSAEVAELYRLELGDYKLVEITPIGLAPTDVKRYTTGGTSRNKRGVFVLGFLGFLATAGSAMGAASLTLTAQSRTLLAGIVQQQQQLLDVVKRQ
QELLRLTVWGTKNLQTRVTAIEKYLKDQAQLNAWGCAFRQVCHTTVPWPNASLTPDWNNDTWQEWERKVDFLEENITALLEEAQIQQEKNMYEL
QKLNSWDVFGNWFDLASWIKYIQYGIYVVVGVILLRIVIYIVQMLAKLRQGYRPVFSSPPSYFQ*THTQQDPALPTREGKEGDGGEGGGNSSWPWQI
EYIHFLIRQLIRLLTWLFSNCRTLLSRAYQILQPILQRLSATLRRVREVLRTELTYLQYGWSYFHEAVQAGWRSATETLAGAWRDLWETLRRGGRWI
LAIPRRIRQGLELTLL*'
```

7. 'SIVAGM677A Green monkey'-'M58410'

```
'MGRLLIKILIIAIGISIGIGNLYVTVFYGIPVWKNSTVQAFCMTPNTNMWATTNCIPDDHDNTEVPLNITEAFEAWDNPLVKQAESNIHLLFEQTMRPC
VKLSPICIKMSCVELNGTATTKATTTATTTMTTPCQNCSTEQIEGEMAEEPASNCTFAIAGYQRDVKKNYSMTWYDQELVCNNKTGSEKGSKDCYM
IHCNDSVIKEACDKTYWDTLRVRYCAPAGYALLKCNDKDYRGFAPKCKNVSVVHCTRLINTTITTGIGLNGSRSENRTEIWQKGGNDNDTVIIKLNK
FYNLTVRCRRPGNKTVLPVTIMAGLVFHSQKYNTRLKQAWCHFQGDWKGAWKEVREEVKKVKNLTEVSIENIHLRRIWGDPESANFWFNCQGEFF
YCKMDWFINYLNNRTEDAEGTNRTCDKGKPGPGPCVQRTYVACHIRQVVNDWYTVSKKVYAPPREGHLECNSSVTALYVAIDYNNKSGPINVTLS
PQVRSIWAYELGDYKLVEITPIGFAPTDVRRYTGPTREKRVPFVLGFLGFLGAAGTAMGAAATTLTVQSRHLLAGILQQQKNLLAAVEQQQQLLKLT
IWGVKNLNARVTALEKYLEDQARLNSWGCAWKQVCHTTVPWKYNNTPKWDNMTWLEWERQINALEGNITQLLEEAQNQESKNLDLYQKLDDW
SGFWSWFSLSTWLGYVKIGFLVIVIILGLRFAWVLWGCIRNIRQGYNPLPQIHIHSSAERPDNGGGQDRGGESSSSKLIRLQEESSTPSRINNWWLNFKS
CSLRIRTWCYNICLTLLIFIRTAVGYLQYGLQQLQEAATGLAQALARAAREAWGRLGAIVRSAYRAVINSPRRVRQGLEKVLG*'
```

8. 'SIVlhoest L"Hoest monkeys'   'AF075269'

```
'MACPGLGILLLLLLGIIWGKQYVTVFYGVPNWDDNVSVPLICASANTSLWVTTSCLPDLQSYAEVPIYNISENFTIPVKDNQVIQQAWSAMNAMVDSI
MKPCVKINPYCVRMQCGEVTKTPTTTPKTTTQMPCFINEQVTVKNPGNETRLEEDLNCTRGLNETTERNAECQYNVTGLCRDCRTEIKQSFRYDDVT
CSGERENRTCYMTHCNDSIITQDCNKGVMQNAYFRLCAPAGYMLLRCNEQLNFSKKCENITATPCTGYMLSSVSSFFGFNGTNHTRDELIPLTPNKM
EDLNGAKFVYKVAGKWGLIIRCIRKGNRSEVSTISSTGYLFYYGLEHGSRLRLAQCKFEGQWGRMFNNLGKMLKELNAEAMNYTEGTGTCDSKKT
TCGRKLKGLPIANMTRHGADLATEMLMHTCGEEMFFCNVTRIFQEWNNKNSDKWYPWANCHIKSIIDDWATIGKKIYLPPTSGFNNRIRCTHRVTE
MFFEMEKWEPHEDLGGNLSIKFLPPSWETNQFVAEGSKYKLIKLNPIGFAPTDEHRYAPRGRQTRAAPLALGALGLLSAAGTAMGLVSTILTVQAQA
VLQGILQQQKQLLVLVEKQQELLRLTIWGVKNLQARLTALEEYVKHQALLASWGCQWKQVCHTNVEWTYNITPNWTKDTWREWESKVAIYDKNI
TSLLQEAYTTELENQNKFKKLQEFNFWSWLDISHWFTYVKYAVLIILVIIGLRVLSFIIQNVVKMCRGYRVLSPSVYIEQDYKWEKEENQEQPDREEE
KGADTETIYINLEQCKKESSRPLWNVDWNEPLQDSLLVTLLKWLKEGGILLLSLVWQSLSWLWHLLILFFQNGQRLWQTSSRWMVENAQKIQSWL
REKCRRNRGQLSSTDRKNIQLGKKKRWRLRFGGRSGISSEATETAL*'
```

9. 'SIVcpz Chimpanzees Cameroon'-'AF115393'

```
'MKVMEKIKGSLIEKYMFMGLIIPCLTGSDQLWATVYYGVPVWKEADTTLFCASEASALNKEAHNIWASQACVPTDPNPQEVQIPNVTENFNMWNN
TMAEQMNTDIISLWDQSLKPCVKLTPICVTMECRKVTFNSTSNRNKTSTMTTNSPNEKKDSTVKNCTFNMTTEVRDKEKKVYSLFYVDDLVLIDND
TDTYRLINCNTTAITQACPKTSFEPIPIHYCATPGYAIMKCNMPNFNGTGTGRCNNISTVQCTHGIRPVVTTQLILNGSVAENKTIARRNGYNFLIQFQK
TVSINCTRPGNNSRGQIQIGPGMTFYNIENIVGDTRRAYCQINRTVWDERLNETGQALRELFTNLTQVNFTVSPGGDPEVTNMMFNCGGEFFYCNTTT
LFNYTWKNNNITKGDNTTFFPCRIRQIVNSWMRVGKGIYAPPIRGVISCTSNVTGIILETGHGINNSITNITLYPTGGNMVDLWRLELHKYKVVSIEPIG
VAPSKAKRHTVSREKRAAFGLGALFLGFLGAAGSTMGAASVVLTVQARQLLSGIVQQQNNLLRAIEAQQHLLQLSVWGIKQLQARVLAVERYLRD
QQILGLWGCSGKAICYTTVPWNNTWSANTSFDEIWNNLTWQDWDKRVKNYSGVIFSLIEQAQEQQNTNEKSLLELDQWSSLWNWFDITRWLWYIK
LFIMIVAGLVGIRIVGAIISFVAKVRQGYSPLSLQTLIPTTREPDRPGGTEEDVGEPGKGRSIRLVSGFLALIWEDLRNLLLFCYHRLRDLLLILGRTLEN
LGQSLNKGLQQLRNFSRYLWGVITYWGRELQTSAISLLDATAIAVAEGTDRILEVAQIIGRGILHIPRRIRQGLERSLL*'
```

10. 'SIVmnd5440 Mandrillus sphinx'-'AY159322'

```
'MLRYLRYIVLGIIVSVIVGEQWVTVYYGTPKWHKARTHLFCATDNNSFWVTTSCVPSLLHYEEQHIPNITENFTGPIEENEIVTQAWGAISSMIDAVL
KPCVRLTPYCVKMKCTEGQNETEQATAKTTTPVPTTTTPSTTTSSSTNKTTTPVLVVEKQNNETTTQQNRVCKFNTTGLCRDCKLEIEENFRYEDVTC
TKLNKTGSATNSTEPEYECYMTSCNATVITQDRNKASTDRMTFRLCAPPGFVLLKCNEKLNKTKLCGNVSAVQCTAPLPATISTMFGFNGTKHDYD
ELIQTNPRKGKDEFHDHKYVYRVDKKWGLQVRCRRKGNRSIISTPSSTGLLFYHGLEPGKNLRKGKCQLEGKWGQALHSLSLELRKINDSIYKDNH
NMTCKSSNNKKNTTGCHLKTISISESTVKGEPGAETIMLLCGGEYFFCNWTKIWKAWNSKQSSVWYPYMSRNIRQIVGDWHKVGKKIYMPPVSGFN
NEIRCTNDVTEMFFEVQKTDDDNGYIIKFIPQDWIQNQYTAVGAHYKLVKVDPIGFAPTDIHRHHLPNTRQKRGAVLLGMLGLLGLAGSAMGSVAV
ALTVQSQTLLNGIVEQQKVLLSLIDQHSELLKLTIWGVKNLQVRLTALEEYVADQSRLSVWGCSFSQVCHTSVKWPNNSIVPNWTSETWLEWDRRV
NSIVTNMTIDLQRAYELEQRNIFELQKLGDLNFHGLTGFDLTWWLKYVKIGLLVVVVIIGLRMLACLWSVLGKFRQGYRPLPYVFKGDYLRPHNLK
RPDREGGEEPDLEKQNIKSESSRQESRKPWKPEQVRSWLKRSTLYIWLKNLQAVIEYGWQELKAAGAALYQVLQGFAQRLWSRGYQLGLSGVRGA
AAFGRGIWNIPRRIRQGAEALLN*'
```

11. 'SIVAGM3 Green monkeys' 'M30931'

'MKLTLLIGILLIGIGVVLNTRQQWVTVFYGVPVWKNSSVQAFCMTPTTRLWATTNSIPDDHDYTEVPLNITEPFEAWADRNPLVAQAGSNIHLLFEQ
TLKPCVKLSPLCIKMSCVELNSSEPTTTPKSTTASTTNITASTTTLPCVQNKTSTVLESCNETIIEKELNEEPASNCTFAMAGYVRDQKKKYSVVWNDA
EIMCKKGNNSNRECYMIHCNDSVIKEACDKTYWDELRLRYCAPAGFALLKCNDYDYAGFKTNCSNVSVVHCTNLINTTVTTGLLLNGSYSENRTQI
WQKHRVSNDSVLVLFNKHYNLTVTCKRPGNKTVLPVTIMAGLVFHSQRYNTRLRQAWCHFQGNWRGAWKEVKNEIVKLPKDRYQGTNDTEEIYL
QRLFGDPEAANLWFNCQGEFFYCKMDWFLNYLNNRTVDPDHNPCNGTKGKGKAPGPCAQRTYVACHIRSVINDWYTLSRKTYAPPREGHLQCTST
VTGMSVELNYNSKNRTNVTLSPQIETIWAAELGRYKLVEITPIGFAPTEVRRYTGGHDRTKRVPFVLGFLGFLGAAGTAMGAAATALTVQSQHLLAG
ILQQQKNLLAAVEAQQQMLKLTIWGVKNLNARVTALEKYLEDQARLNAWGCAWKQVCHTTVPWQWNNRTPDWNNMTWLEWERQISYLEGNIT
TQLEEARAQEEKNLDAYQKLSSWSDFWSWFDFSKWLNILKIGFLDVLGIIGLRLLYTVYSCIARVRQGYSPLSPQIHIHPWKGQPDNAEGPGEGGDK
RKNSSEPWQKESGTAEWKSNWCKRLTNWCSISSIWLYNSCLTLLVHLRSAFQYIQYGLGELKAAAQEAVVALARLAQNAGYQIWLACRSAYRAIIN
SPRRVRQGLEGILN*'

12. 'SIVMM239 Simian macaque'-'M33262'

'MGCLGNQLLIAILLLSVYGIYCTLYVTVFYGVPAWRNATIPLFCATKNRDTWGTTQCLPDNGDYSEVALNVTESFDAWNNTVTEQAIEDVWQLFET
SIKPCVKLSPLCITMRCNKSETDRWGLTKSITTTASTTSTTASAKVDMVNETSSCIAQDNCTGLEQEQMISCKFNMTGLKRDKKKEYNETWYSADLV
CEQGNNTGNESRCYMNHCNTSVIQESCDKHYWDAIRFRYCAPPGYALLRCNDTNYSGFMPKCSKVVVSSCTRMMETQTSTWFGFNGTRAENRTYI
YWHGRDNRTIISLNKYYNLTMKCRRPGNKTVLPVTIMSGLVFHSQPINDRPKQAWCWFGGKWKDAIKEVKQTIVKHPRYTGTNNTDKINLTAPGG
GDPEVTFMWTNCRGEFLYCKMNWFLNWVEDRNTANQKPKEQHKRNYVPCHIRQIINTWHKVGKNVYLPPREGDLTCNSTVTSLIANIDWIDGNQT
NITMSAEVAELYRLELGDYKLVEITPIGLAPTDVKRYTTGGTSRNKRGVFVLGFLGFLATAGSAMGAASLTLTAQSRTLLAGIVQQQQQLLDVVKRQ
QELLRLTVWGTKNLQTRVTAIEKYLKDQAQLNAWGCAFRQVCHTTVPWPNASLTPKWNNETWQEWERKVDFLEENITALLEEAQIQQEKNMYEL
QKLNSWDVFGNWFDLASWIKYIQYGVYIVVGVILLRIVIYIVQMLAKLRQGYRPVFSSPPSYFQQTHIQQDPALPTREGKERDGGEGGGNSSWPWQI
EYIHFLIRQLIRLLTWLFSNCRTLLSRVYQILQPILQRLSATLQRIREVLRTELTYLQYGWSYFHEAVQAVWRSATETLAGAWGDLWETLRRGGRWIL
AIPRRIRQGLELTLL*'

13. 'CIVcpzUS Chimpanzee'-'AF103818'

'MKVMEKKKRLWLSYCLLSSLIIPGLSSLWATVYYGVPVWRDVETTLFCASDAKAYKQEAHNIWATQACVPTDPNPQEVHLPNVTEKFDMWENNM
AEQMQEDIISLWDQSLKPCIKLTPLCVTMTCLNPDSNSSAVNTTDIMRNCSFNITTELRDKKKQVYSLFYVDDLAHINNNTYRLINCNTTAITQACPKT
SFEPIPIHYCAPPGFAILKCNEKDFKGKGECKNVSTVQCTHGIKPVVTTQLIINGSLATKNVTVRSKNFADIILVQFSEGVNMTCIRPGNNTVGNVQLGP
GMTFYNIPKIVGDVREAHCNISKLTWEKQRKYTLEIIKKEANLTKVELIPNAGGDPEVVNMMLNCGGEFFYCNTIPLFNMTYNNTDNTTITLKCRIRQ
IVNQWMRVGKGIFAPPIKGVLSCNSNITGMILDISISAVNNDSRNITVMPTGGDMTALWKNELHKYKVVSIEPIGVAPGKAKRHTVKREKRAAFGLG
ALFLGFLGAAGSTMGAASVVLTVQARQLLSGIVQQQNNLLRAIEAQQHLLQLSVWGIKQLQARVLAVERYLKDQQILGLWGCSGKTICYTTVPWN
DTWSNNLSYDAIWGNLTWQEWDRKVRNYSGTIFSLIEQAQEQQNTNEKSLLELDQWSSLWNWFDITNWLWYIKIFLIVVASLVGIRIVGVIFSLVAK
VRQGYSPLSLQTLFPTTREPDRPEGTEEGAGKTDNVRSTRLVSGFLALVWEDLRNLLIFLYHRLEDLLLILRRTVQILGQNINKGLQLLNELRARCWG
VIAYWARELKVSATSLLDTTAIAVAEGTDRIIELTRRLFLGIIHIPRRIRQGLERSLL*'

14. 'SIVmon Cercopithecus Monkeys'-'AY340701'

'MREGDMTQGLTQNLNQSNMRKLMALSTLCLMMTLVNGSNWTTVYYGVPVWKPATPPLFCASDPNYGSKEAGNNWLASSCLPTDPTPQSLYLNIT
EEFNAYQNYMVEEMVEDMKSLFSQALKPCVKLTPMCVRMLCVEVNTVSNASTTPAPSTPTPWGNWGGNGTGQPVYNCSFNQTTEFRDKKKQMYS
LFWKEDIMKEEGSNGSHYYILNCNTSYITQACEKSNYEPVPLHYCAPPGYALLRCDDPAFTGQGSCSNVSAVTCTHAIQPIVATWFQLNSTGNAPNTT
VMMNKQKNESIVVRLAKHLHVNITCIRPGNKTIRNLQIGAGMTFYSQLIVGGNTRKAYCKVNKTQWETALQAVHEAVKTEWEKKNNGTNVTTISW
RFQPQGDKEVQTHWFNCQGEFFYCNVSALFINRRTNKTDGISPFDVNNKPNTTYHGGWLACTIRQMVTQWGYVSKSIYLPPRKGHVQCTSNITALLI
TGELYQNNVTLVPSAQVSDSWRSELSRYKVVEIDPLSMAPTTAQRRTGVHREKRAITLGMAFLGFLSTAGGTMGAAATALTVQSRSLLAGIVQQQE
NLLRAVTAQQSLLQLTVWGVKQLQARLTAVEKFIKDQTLLNAWGCANKAVCHTTVPWNNSWAKGHFPEWDNMTWQQWSELVDNDTMTIQQLL
EAAQEQQGKNQHELMKPGQWDFLWNWFDISKWLWYIKIFIIVVAALIGLRILMFILGVISRLGQGYSLLSSQIPIPSHAGQPTPDGTGAGGGDGSNSRS
PAYLKGFFTIIWEDLRNLVLWTYQILKDSVLVIYRILQRVSQRLPPLLHIRLLQLWESLRRLLAYCQYGIQELQAAVTSLLDALARFTIVWTDALLHAG
GRLWRAIVAIPRRIRQGAEIFLN*'

15. 'SIVcpzTAN1 Chimpanzee'-'AF447763'

'MKNLIGITLILIITILGIGFSTYYTTVFYGVPVWKEAQPTLFCASDADITSRDKHNIWATHNCVPLDPNPYEVTLANVSIRFNMEENYMVQEMKEDILS
LFQQSFKPCVKLTPFCIKMTCTMTNTTNKTLNSATTTLTPTVNLSSIPNYEVYNCSFNQTTEFRDKKKQIYSLFYREDIVKEDGNNNSYYLHNCNTSVI
TQECDKSTFEPIPIRYCAPAGFALLKCRDQNFTGKGQCSNVSVVHCTHGIYPMIATALHLNGSLEEEETKAYFVNTSVNTPLLVKFNVSINLTCERTGN
NTRGQVQIGPGMTFYNIENVVGDTRKAYCSVNATTWYRNLDWAMAAINTTMRARNETVQQTFQWQRDGDPEVTSFWFNCQGEFFYCNLTNWTN
TWTANRTNNTHGTLVAPCRLRQIVNHWGIVSKGVYLPPRRGTVKCHSNITGLIMTAEKDNNNSYTPQFSAVVEDYWKVELARYKVVEIQPLSVAPR
PGKRPEIKANHTRSRRDVGIGLLFLGFLSAAGSTMGAASIALTAQARGLLSGIVQQQQNLLQAIEAQQHLLQLSVWGIKQLQARMLAVEKYIRDQQL
LSLWGCANKLVCHSSVPWNLTWAEDSTKCNHSDAKYYDCIWNNLTWQEWDRLVENSTGTIYSLLEKAQTQQEKNKQELLELDKWSSLWDWFDIT
QWLWYIKIAIIIVAGLVGLRILMFIVNVVKQVRQGYTPLFSQIPTQAEQDPEQPGGIAGGGGGRDNIRWTPSPAGFFSIVWEDLRNLLIWIYQTFQNFIW
ILWISLQALKQGIISLAHSLVIVHRTIIVGVRQIIEWSSNTYASLRVLLIQAIDRLANFTGWWTDLIIEGVVYIARGIRNIPRRIRQGLELALN*'

16. 'SIVsmSL92b Sooty Mangabey'-'AF334679'

'MACPGLHLLIDILFLSVLGTWCAQYVTIFYGIPAWRNATIPLFCATQNRDTWGTVQCLPDNGDYSELALNVTEAFDAWDNTVTEQAIEDVWNLFET
SIKPCVKLTPLCIAMKCNKNETDRWGLTRAATTTSSPTTTSPLTAASPSGEEIVNDTMSCTKNNNCSGIEQEPMIGCQFNMTGLKRDQKRQYNETWY
SRDLVCEQGGNESSRCYMNHCNTSVIQESCDKHYWDAIRYRYCAPPGYALLRCNDTNYSGFAPNCSKVVVSSCTRMMETQTSTWFGFNGTRAENR
TYIYWHGGSNRTIISLNKYYNLTIKCRRPGNKTVLPVTIMSGLVFHSQPINERPRQAWCWFGGKWREAMQEVKKTIVKHPRYTGTNDTRKINLTAPG
GGDPEVTFMWTNCRGEFLYCKMNWFLNWVEDRNTSSPRWTTQTKKEQHKRNYVPCHIRQIINTWHRVGKNVYLPPREGDLTCNSTVTSLIANIDWI
DNNETNITMSAEVAELYRLELGDYKLVEITPIGMAPTHVKRYTTSTSKNKRGVFVLGFLGFLATAGSAMGAASLTLTAQSRTLLAGIVQQQQQLLDV
VKRQQELLRLTVWGTKNLQTRVTAIEKYLKDQAQLNSWGCAFRQVCHTTVLWPNDSLVPDWNNMTWQEWEKKVEFLEANITQMLEEARLQQEK
NMYELQKLNSWDVFGNWFDLTSWVRYIQYGVFLVIGIVLLRIVIYVVQMLSRLRQGYRPVFSSPPSYHQQIHIQRDQELPAKEGTEGEGGNGGGYRL
WPWQIEYIHFLIRQLIRILTWLYNNLTRLASRAYQNLQQLCQRLSEISQPIRELVRREAGYIRYGWNYFIEACQEAWRSAQEAIVGAWGLIWETLGRV
GRGIAAIPRRIRQGLELMLN*'

4.5.17. Distance Tables of Translated ENV protein-coding gene (Amino Acid Sequence) (designated by AAENV)

Table 34. Alignment-Score Distance

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	0.00000															
2	0.01682	0.00000														
3	0.47126	0.45818	0.00000													
4	0.50089	0.48739	0.02048	0.00000												
5	0.48277	0.46881	0.06695	0.06128	0.00000											
6	0.48756	0.48203	0.05863	0.05858	0.06392	0.00000										
7	0.45538	0.45173	0.27393	0.26919	0.27231	0.27991	0.00000									
8	0.65543	0.64876	0.51323	0.51732	0.52982	0.53869	0.48029	0.00000								
9	0.13818	0.12859	0.44576	0.44955	0.45572	0.47295	0.45076	0.63388	0.00000							
10	0.63095	0.60857	0.51741	0.52053	0.52856	0.51628	0.49957	0.33880	0.58529	0.00000						
11	0.47328	0.45917	0.28475	0.28984	0.28766	0.29101	0.10044	0.51843	0.45876	0.51675	0.00000					
12	0.48311	0.47606	0.05445	0.05336	0.06032	0.00139	0.27058	0.53076	0.46331	0.50819	0.28826	0.00000				
13	0.13251	0.13304	0.44900	0.45206	0.43477	0.46934	0.42410	0.61215	0.06340	0.55895	0.45543	0.45911	0.00000			
14	0.31795	0.31696	0.42205	0.43293	0.41417	0.43811	0.43449	0.58126	0.31471	0.58102	0.42786	0.43113	0.29285	0.00000		
15	0.28342	0.27918	0.43851	0.45230	0.45229	0.44904	0.46203	0.61254	0.25359	0.60762	0.44226	0.44913	0.23970	0.26806	0.00000	
16	0.50012	0.49061	0.06320	0.06408	0.07383	0.02776	0.28199	0.53660	0.47451	0.52778	0.30430	0.02550	0.46860	0.43510	0.44731	0.00000

Table 35. Gamma Distance

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	0.00000															
2	0.17189	0.00000														
3	1.35083	1.32449	0.00000													
4	1.39935	1.38243	0.18678	0.00000												
5	1.41794	1.37850	0.37825	0.34347	0.00000											
6	1.40027	1.38960	0.34247	0.33401	0.36088	0.00000										
7	1.36092	1.36367	0.95567	0.91304	0.90554	0.90807	0.00000									
8	1.93244	1.97517	1.57856	1.58068	1.60877	1.64863	1.51581	0.00000								
9	0.57328	0.55297	1.29938	1.32119	1.36823	1.38734	1.40948	1.89155	0.00000							
10	1.88158	1.82075	1.54868	1.56010	1.56733	1.54321	1.51897	1.04345	1.74519	0.00000						
11	1.41140	1.40245	0.96825	0.97076	0.98619	0.94996	0.43735	1.54262	1.39958	1.51100	0.00000					
12	1.40421	1.38151	0.32488	0.31663	0.34850	0.04585	0.88632	1.62318	1.36007	1.53237	0.94277	0.00000				
13	0.56593	0.57884	1.35002	1.33750	1.30214	1.36165	1.34451	1.88514	0.35797	1.62299	1.39359	1.33839	0.00000			
14	1.00267	1.01676	1.30709	1.28317	1.24184	1.30147	1.29074	1.74166	1.01156	1.74744	1.24969	1.28692	0.97044	0.00000		
15	0.91416	0.90534	1.31307	1.33650	1.32789	1.31363	1.36581	1.88546	0.84193	1.78672	1.32584	1.30963	0.80801	0.88492	0.00000	
16	1.43690	1.42201	0.36009	0.35139	0.38386	0.22533	0.93345	1.65006	1.39912	1.62009	0.96958	0.21347	1.36546	1.31321	1.30470	0.00000

Table 36. Jukes-Cantor Distance

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	0.00000															
2	0.16566	0.00000														
3	1.07546	1.05804	0.00000													
4	1.10730	1.09624	0.17946	0.00000												
5	1.11943	1.09366	0.35005	0.31996	0.00000											
6	1.10791	1.10093	0.31909	0.31171	0.33507	0.00000										
7	1.08210	1.08392	0.80323	0.77229	0.76681	0.76866	0.00000									
8	1.43995	1.46547	1.22244	1.22378	1.24149	1.26647	1.18255	0.00000								
9	0.51222	0.49581	1.04134	1.05585	1.08691	1.09945	1.11392	1.41539	0.00000							
10	1.40938	1.37254	1.20350	1.21075	1.21534	1.20002	1.18457	0.86588	1.32632	0.00000						
11	1.11517	1.10933	0.81229	0.81409	0.82516	0.79910	0.40032	1.19965	1.10746	1.17948	0.00000					
12	1.11048	1.09563	0.30372	0.29648	0.32433	0.04539	0.75272	1.25053	1.08155	1.19312	0.79390	0.00000				
13	0.50629	0.51669	1.07492	1.06665	1.04318	1.08259	1.07128	1.41153	0.33254	1.25042	1.10354	1.06724	0.00000			
14	0.83694	0.84697	1.04648	1.03052	1.00276	1.04273	1.03558	1.32415	0.84328	1.32771	1.00805	1.03303	0.81386	0.00000		
15	0.77310	0.76667	1.05045	1.06599	1.06029	1.05083	1.08532	1.41172	0.71991	1.35179	1.05893	1.04817	0.69457	0.75169	0.00000	
16	1.13175	1.12208	0.33438	0.32684	0.35486	0.21482	0.78714	1.26736	1.10715	1.24860	0.81324	0.20400	1.08509	1.05055	1.04489	0.00000

Table 37. Poisson Distance

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	0.00000															
2	0.16490	0.00000														
3	1.03212	1.01634	0.00000													
4	1.06087	1.05090	0.17856	0.00000												
5	1.07178	1.04857	0.34644	0.31697	0.00000											
6	1.06142	1.05513	0.31612	0.30888	0.33178	0.00000										
7	1.03813	1.03977	0.78116	0.75210	0.74694	0.74868	0.00000									
8	1.35223	1.37384	1.16363	1.16481	1.18044	1.20241	1.12824	0.00000								
9	0.50407	0.48822	1.00117	1.01435	1.04248	1.05379	1.06682	1.33132	0.00000							
10	1.32619	1.29460	1.14686	1.15328	1.15734	1.14377	.	0.83969	1.25465	0.00000						
11	1.06795	1.06270	0.78965	0.79134	0.80170	0.77729	0.39553	1.14344	1.06101	1.12551	0.00000					
12	1.06373	1.05035	0.30104	0.29393	0.32126	0.04533	0.73367	1.18841	1.03763	1.13764	0.77241	0.00000				
13	0.49835	0.50838	1.03164	1.02415	1.00285	1.03857	1.02834	1.32803	0.32930	1.18830	1.05748	1.02468	0.00000			
14	0.81271	0.82207	1.00585	0.99133	0.96599	1.00244	0.99593	1.25276	0.81862	1.25585	0.97083	0.99361	0.79113	0.00000		
15	0.75287	0.74681	1.00946	1.02355	1.01838	1.00979	1.04104	1.32819	0.70267	1.27671	1.01715	1.00738	0.67865	0.73270	0.00000	
16	1.08284	1.07416	0.33111	0.32372	0.35115	0.21352	0.76606	1.20319	1.06074	1.18670	0.79055	0.20283	1.04083	1.00954	1.00440	0.00000

Table 38. P-Distance Distance

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	0.00000															
2	0.15202	0.00000														
3	0.64375	0.63808	0.00000													
4	0.65385	0.65038	0.16353	0.00000												
5	0.65760	0.64956	0.29280	0.27165	0.00000											
6	0.65403	0.65185	0.27103	0.26573	0.28235	0.00000										
7	0.64589	0.64646	0.54212	0.52862	0.52619	0.52701	0.00000									
8	0.74134	0.74687	0.68765	0.68802	0.69286	0.69953	0.67640	0.00000								
9	0.39593	0.38628	0.63255	0.63736	0.64742	0.65139	0.65590	0.73587	0.00000							
10	0.73451	0.72599	0.68237	0.68440	0.68568	0.68138	0.67698	0.56816	0.71482	0.00000						
11	0.65629	0.65448	0.54600	0.54676	0.55144	0.54035	0.32667	0.68128	0.65389	0.67551	0.00000					
12	0.65483	0.65019	0.25995	0.25467	0.27476	0.04432	0.51986	0.69529	0.64571	0.67943	0.53810	0.00000				
13	0.39247	0.39853	0.64358	0.64090	0.63317	0.64604	0.64240	0.73500	0.28058	0.69526	0.65267	0.64109	0.00000			
14	0.55635	0.56048	0.63426	0.62892	0.61939	0.63302	0.63062	0.71429	0.55896	0.71517	0.62123	0.62976	0.54667	0.00000		
15	0.52899	0.52612	0.63558	0.64068	0.63882	0.63571	0.64691	0.73504	0.50474	0.72105	0.63838	0.63482	0.49270	0.51939	0.00000	
16	0.66137	0.65842	0.28187	0.27655	0.29612	0.19226	0.53516	0.69977	0.65380	0.69477	0.54640	0.18358	0.64684	0.63561	0.63373	0.00000

TÜBİTAK
PROJE ÖZET BİLGİ FORMU

Proje Yürütücüsü:	Prof. Dr. AYŞE HÜMEYRA BİLGE
Proje No:	118F412
Proje Başlığı:	Sonlu Metrik Uzayların Gromov Çarpımları Ile İncelenmesi Ve Filogenetik Uygulamaları
Proje Türü:	1001 - Araştırma
Proje Süresi:	24
Araştırmacılar:	EBRU DEMET AKDOĞAN, MEHMET ŞAHİN KOÇAK, DERYA ÇELİK
Danışmanlar:	
Projenin Yürütüldüğü Kuruluş ve Adresi:	KADİR HAS Ü. MÜHENDİSLİK VE DOĞA BİLİMLERİ F. ENDÜSTRİ MÜHENDİSLİĞİ B.
Projenin Başlangıç ve Bitiş Tarihleri:	15/03/2019 - 15/03/2021
Onaylanan Bütçe:	350600.68
Harcanan Bütçe:	228879.63
Öz:	Sonlu metrik uzaylar, sonlu bir X kümesi üzerinde tanımlı bir d uzaklık fonksiyonu ile karakterize edilen uzaylardır. Projede, sonlu metrik uzayların sınıflandırılması ve çizge temsilleri problemlerine Gromov çarpımları vasıtasıyla yeni bir yaklaşım getirilmiş ve filogenetik analiz hesaplarına uygulanmıştır. Sonlu metrik uzaylar, üçgen eşitsizliklerini sağlayan ve negatif olmayan noktalar kümesi ile tanımlanır ve bir polihedral koni oluştururlar. Sonlu metrik uzayların sınıflandırılması, bu kümenin, hipersimpleks bölümlemesi olarak adlandırılan kanonik bir bölümlemesi ile yapılmaktadır. Literatürde, 4,5,6 elemanlı uzaylar için hipersimpleks bölünlemesi bilinmektedir ancak 6 dan büyük elemanlı uzaylar için herhangi bir sonuç bulunmamaktadır. Çalışmamızda, n elemanlı metrik uzayların sınıflandırma problemine, Gromov çarpımı ve dörtgen yapısı yöntemleri ile yaklaşılarak eleman sayısı 6dan küçük uzaylar için için dörtgen yapısı sınıflaması ve eleman sayısı 8 den küçük uzaylar için , için Gromov çarpımı sınıflamaları elde edilmiştir. Biyolojik sistemlerin incelenmesinde bir araç olarak kullanılan filogenetik ağaçlar, bir türün varyantları arasındaki mesafelerden hareketle, birbirleri arasındaki geçişleri ve oluşum aşamalarını temsil eden çizgelerdir. Literatürde, bu çizgelerin ağaç yapısında olduğundan hareketle çeşitli yöntemler geliştirilmiş, ancak ağaç yapısı varsayımının geçerli olmadığı durumlarda zorluklarla karşılaşılmıştır. Projede, ağaç varsayımı yapılmadan, verilen bir ailenin gen/protein yaısındaki dizilimlerden hareketle hesaplanan mesafe fonksiyonları hesaplanmış, sonlu bir metrik uzay olarak Gromov çarpımı ve dörtgen yapıları çıkarılmış, ağaç yapısına sahip olma/olmama ve döngülerin varlığına karar verme konusunda yaklaşıklık kriterleri getirilmiştir.
Anahtar Kelimeler:	Sonlu metrik uzaylar, Gromov çarpımları, Optimal temsiller, Filogenetik sınıflamalar
Fikri Ürün Bildirim Formu Sunuldu Mu?:	Evet
Projeden Yapılan Yayınlar:	1- Gromov product structures, quadrangle structures and splitmetric decompositions for finite metric spaces (Makale - Diğer Hakemli Makale), 2- Split Metric Decomposition of Finite Metric Spacesvia Gromov Products Structures (Bildiri - Uluslararası Bildiri - Sözlü Sunum),