

T.C.
KAFKAS ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
BİYOMÜHENDİSLİK ANABİLİM DALI

POPÜLASYON GENETİĞİ VERİ FORMATLARININ WEB TABANLI
UYGULAMALARDA BİRBİRİNE DÖNÜŞTÜRÜLMESİ

Mahmut AYDIN
YÜKSEK LİSANS TEZİ

DANIŞMAN
Yrd. Doç. Dr. Igor KRYVORUCHKO

OCAK-2018
KARS



T.C.
KAFKAS ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
BİYOMÜHENDİSLİK ANABİLİM DALI



POPÜLASYON GENETİĞİ VERİ FORMATLARININ WEB TABANLI
UYGULAMALARDA BİRBİRİNE DÖNÜŞTÜRÜLMESİ




Mahmut AYDIN
YÜKSEK LİSANS TEZİ

DANIŞMAN
Yrd. Doç. Dr. Igor KRYVORUCHKO

OCAK-2018
KARS

T.C. Kafkas Üniversitesi Fen Bilimleri Enstitüsü Biyomühendislik Anabilim Dalı yüksek lisans öğrencisi Mahmut AYDIN'ın Yrd.Doç. Dr. Igor KRYVORUCHKO danışmanlığında yüksek lisans tezi olarak hazırladığı “**Popülasyon Genetiği Veri Formatlarının Web Tabanlı Uygulamalarda Birbirine Dönüştürülmesi**” adlı bu çalışma, yapılan tez savunması sınavı sonunda jüri tarafından Lisansüstü Eğitim Öğretim Yönetmeliği uyarınca değerlendirilerek oy .. *birliği* ile kabul edilmiştir.

.. / .. / 20..

	Adı ve Soyadı	İmza
Başkan	: Yrd. Doç. Dr. Doğan İLHAN	
Üye	: Yrd. Doç. Dr. Özkan ÖZDEN	
Üye	: Yrd. Doç. Dr. Aydın AKBUDAK	

Bu tezin kabulü, Fen Bilimleri Enstitüsü Yönetim Kurulu'nun .. / .. / 20.. gün ve ...
... / sayılı kararıyla onaylanmıştır.

Doç.Dr. Fikret AKDENİZ
Enstitü Müdürü

ETİK BEYAN

Kafkas Üniversitesi Fen Bilimleri Enstitüsü Tez Yazım Kurallarına uygun olarak hazırladığım bu tez çalışmada;

- Tez içinde sunduğum verileri, bilgileri ve dokümanları akademik ve etik kurallar çerçevesinde elde ettiğimi,
- Tüm bilgi, belge, değerlendirme ve sonuçları bilimsel etik ve ahlak kurallarına uygun olarak sunduğumu,
- Tez çalışmada yararlandığım eserlerin tümüne uygun atıfta bulunarak kaynak gösterdiğimi,
- Kullanılan verilerde herhangi bir değişiklik yapmadığımı,
- Bu tezde sunduğum çalışmanın özgün olduğunu,

bildirir, aksi bir durumda aleyhime doğabilecek tüm hak kayıplarını kabullendiğimi beyan ederim.

Mahmut AYDIN

03.01.2018

ÖZET

20. yüzyılın ikinci yarısında, biyolojik bilgilerin çok hızlı bir şekilde artış göstermesi sonucunda dünya genelinde yapılan araştırma ve uygulamalar ile çok büyük miktarda genomik bilgi üretilmeye başlanmıştır. Buna paralel olarak, moleküler biyoloji, popülasyon genetiği gibi alanlarda yapılan çalışmalardan elde edilen DNA dizi analizi, protein ve nükleotid dizilimleri gibi genomik bilgilerin açığa çıkması ile bu bilgilerin bir araya getirilmesi, kolay erişilebilir bir ortamda saklanması ve gerektiğinde araştırmacıların kullanımına sunmak amacıyla dünya genelinde çeşitli veri tabanlarının oluşturulması ihtiyacı ortaya çıkmıştır. Araştırmacıların çalışmalar sonucu elde ettikleri çeşitli genomik verileri karşılaştırmak ya da veri tabanlarında saklamak amacıyla bu verileri farklı formatlara çevirmeleri gerekir.

Birçok genomik veri analiz programı, eldeki moleküler verinin kendine ait özel formatlarda hazırlanmasını gerektirir. Genomik analiz programlarının kendilerine göre düzenledikleri formatlar olduğu gibi veri tabanlarının da kendilerine özgü formatları vardır. Tam bu noktada en çok karşılaşılan problem, farklı programlar ve veri tabanları tarafından farklı formatların kullanılıyor olmasıdır. Bu amaçla bazı programlama dilleri kullanılarak araştırmacılar ya da kurumlar tarafından farklı formatlar arasında dönüşüm gerçekleştirebilen masaüstü programlar geliştirilmiştir. İstedığınız format dönüşümünü sağlayacak programı arayıp bulma ve masaüstü yazılımların web üzerinden indirilip kullanılması aşamasında çeşitli zorluklar yaşanmaktadır. Başka bir zorluk ise bu programların kurulum aşamasında, araştırmacıların çeşitli komut satırları da kullanmak zorunda kalabilmeleridir. Bu programların yeni versiyonu çıkınca güncelleme problemleri de ek olarak araştırmacıların zamanlarını kaybettirmektedir. Bazı programların lisans şartları da dikkate alınırsa masaüstü yazılımların, araştırmacılara çeşitli zorluklar yaşatması kaçınılmazdır.

Web tabanlı yazılımlar masaüstü yazılımlar gibi indirme, kurulum, sistem gereksinimleri ya da konsol uygulamaları istemezler ancak genomik veri formatlarının birbirine çevrilmesi için henüz kapsamlı bir platform mevcut değildir. Bu çalışmada, web üzerinden geliştirilen yazılım aracılığı ile program indirme ve sistem gereksinimleri gibi ihtiyaçlara gerek kalmadan araştırmacıların online (çevrimiçi) olarak yaygın kullanılan

genomik veri formatlarını birbirine çevirebilmelerine olanak tanıyan bir platformun geliştirilerek hem ülkemizdeki hem de tüm dünyadaki arařtırmacıların kullanımına sunulması saęlanmıřtır.

Anahtar Kelimeler: DNA sekanslama, Format dnřtrme, Genomik veri, Genomik veri tabanları, Molekler biyoloji, Poplasyon genetięi, Web tabanlı

ABSTRACT

In the second half of the 20th century, in parallel to the rise of biological science, genomic information began to be produced in large quantities with research conducted worldwide. The massive amount of genomic information such as protein sequences and nucleotide sequences, necessitated to store information in an easily accessible environment, and to provide easy access to researchers when necessary, and thus the creation of databases has emerged.

Many genomic data analysis programs require that the molecular data be formatted in specific formats. Just as genomic analysis programs are formats that are organized individually, databases also have their own specific formats. Therefore, researchers need to frequently change data into different formats in order to analyze them in various programs or to store them in databases. For this purpose, some programming languages have been developed in the form of standalone software programs that can transform data into different formats. There are various difficulties in identifying the appropriate program that will provide the desired format conversion, and in the process of downloading and using the desktop software from the web. Another difficulty is that, during the installation phase of these programs, researchers may have to use a variety of command lines as well. The new version of these programs become available update problems also can cost the time of the researchers. It is inevitable that desktop software will have various difficulties for the researcher if the license conditions of some programs are also taken into consideration.

Web-based software does not require downloads, installations, system requirements, or console applications contrary to the standalone software programs, but a comprehensive platform for genomic data format conversion is not yet available. In this study, a web-based platform has been developed for researchers which enable researchers to convert widely used genomic data formats without the need for program downloading and system requirements.

Keywords: DNA sequencing, Format converting, Genomic data, Genomic databases, Molecular biology, Population genetics, Web based

ÖNSÖZ

Biyoteknoloji alanında yaşanan gelişmelerle çok büyük miktarda genomik veri elde edilmiştir. Bu verilerin saklanması için veri tabanları, analiz edilmesi için ise veri analiz programları kullanılmaktadır.

Bu çalışmanın konusu, genomik verileri içeren dosya formatlarını, biyolojik veri tabanlarında saklanması ya da veri analiz programlarında işlenmesi için farklı formatlara çeviren web tabanlı bir program üretmektir.

Öncelikle yüksek lisans yaptığım süre boyunca her türlü desteği esirgemedi bana yardım eden değerli danışman hocam Yrd. Doç. Dr. Igor KRYVORUCHKO'a, bilgi ve deneyimlerinden yararlanma imkânı sunarak rehberlik eden, Doç. Dr. Muhammet ŞAKİROĞLU'na ve tüm Biyomühendislik Bölümü hocalarına içtenlikle teşekkür ederim.

Son olarak, desteklerini her zaman yanımda hissettiğim kıymetli anneme, babama ve sevgili eşime şükranlarımı sunarım.

Mahmut AYDIN

İÇİNDEKİLER

	Sayfa
ETİK BEYAN	i
ÖZET	ii
ABSTRACT	iv
ÖNSÖZ	v
İÇİNDEKİLER	vi
ŞEKİLLER DİZİNİ	viii
TABLolar DİZİNİ	x
SEMBOLLER VE KISALTMALAR DİZİNİ	xi
1. GENEL BİLGİLER	1
1.1. Giriş.....	1
1.2. Genomik Veri Formatları, Programları ve Veri tabanları.....	2
1.3. Bu Çalışmada Ele Alınacak Dosya Formatları ve Tiplerinin Temel Özellikleri.....	9
1.3.1. Pearson FASTA fa.....	10
1.3.2. PHYLIP.....	13
1.3.3. PAUP NEXUS.....	17
1.3.4. ARLEQUIN.....	19
1.3.5. MEGA.....	22
1.3.6. STRUCTURE.....	24
1.3.7. GENEPOP.....	25
1.4. Depolanan Moleküler Veri Çeşitleri.....	30
1.4.1. DNA ve RNA Sekans Verileri.....	31
1.4.2. Protein.....	31
1.4.3. SSR MICROSAT (Basit Dizi Tekrarları).....	32
1.4.4. SNP (Tek Nükleotit Polimorfizmi).....	33
1.4.5. RFLP (Restriksiyon Parça Uzunluk Polimorfizmi).....	34
1.4.6. AFLP (Çoğaltılmış Parça Uzunluğu Polimorfizmi).....	36
1.4.7. Frequency (Allel Frekans Verisi).....	36
1.4.8. Distance (Genetik Farklılık).....	37
2. MATERYAL VE YÖNTEM	38

3. BULGULAR	42
4. SONUÇLAR VE TARTIŞMA	47
5. KAYNAKLAR	48
ÖZGEÇMİŞ	52

ŞEKİLLER DİZİNİ

Resim 1.1: Genel FASTA Dosya Formatı Görünümü	11
Resim 1.2: Diğer FASTA Dosya Formatı Görünümleri	11
Resim 1.3: FASTA Dosya Formatı Görünümü (Single Sequences)	12
Resim 1.4: FASTA Dosya Formatı Görünümü (Multi Sequences)	12
Resim 1.5: PHYLIP DNA Dosya Formatı Görünümü	13
Resim 1.6: PHYLIP Interleaved Dizi Formatı Görünümü	14
Resim 1.7: PHYLIP Sequential Dizi Formatı Görünümü	14
Resim 1.8: PHYLIP Kare Genetik Farklılık Matrisi (Distance Matrix) Formatı Görünümü	15
Resim 1.9: PHYLIP Üçgen Genetik Farklılık Matrisi (Distance Matrix) Formatı Görünümü	15
Resim 1.10: PHYLIP Protein Dosya Formatı Görünümü	16
Resim 1.11: NEXUS Dosya Formatı Görünümü	17
Resim 1.12: Interleaved NEXUS Dosya Formatı Görünümü	18
Resim 1.13: Sequential NEXUS Dosya Formatı Görünümü	18
Resim 1.14: ARLEQUIN Dosya Formatı Görünümü	20
Resim 1.15: Haplotipik RFLP Verisi Görünümü	21
Resim 1.16: Haplotipik STANDARD HLA Verisi Görünümü	21
Resim 1.17: Genotipik DNA Sekans Verisi Görünümü.....	21
Resim 1.18: Genotipik Mikrosatelit Verisi Görünümü	21
Resim 1.19: ARLEQUIN Yorum İfadesi Görünümü.....	21
Resim 1.20: MEGA Dosya Formatı Görünümü.....	22
Resim 1.21: MEGA Interleaved Dosya Formatı Görünümü.....	23
Resim 1.22: MEGA Sequential (Noninterleaved) Dosya Formatı Görünümü.....	23
Resim 1.23: MEGA Distance Veri Tipi Dosya Formatı Görünümü.....	23
Resim 1.24: STRUCTURE Dosya Formatı Görünümü	24
Resim 1.25: GENEPOP Dosya Formatı Görünümü-1	25
Resim 1.26: GENEPOP Dosya Formatı Görünümü-2	26
Resim 1.27: GENEPOP Dosya Formatı Görünümü-3	27
Resim 1.28: GENEPOP Dosya Formatı Görünümü (Boşluk)	28
Resim 1.29: GENEPOP Dosya Formatı (Mixed) Görünümü	29

Resim 1.30: GENEPOP Dosya Formatı Eksik Veri (Missing Data) Görünümü	30
Resim 1.31: Örnek DNA Veri Tipi	31
Resim 1.32: Örnek Protein Veri Tipi	31
Resim 1.33: Örnek SSR/MICROSAT Veri Tipi	32
Resim 1.34: Örnek SNP Veri Tipi.....	33
Resim 1.35: Örnek RFLP Veri Tipi	35
Resim 1.36: Örnek 4 SSR/MICROSAT, 5 AFLP Lokusuna Sahip AFLP Veri Tipi.....	36
Resim 1.37: Örnek Frequency Veri Tipi	36
Resim 1.38: Örnek Distance Veri Tipi	37
Resim 2.1: Örnek HTML Kodları	39
Resim 2.2: JQuery -Ajax Metodu ile Sunucuya Bilgilerin Gönderilmesi.....	40
Resim 2.3: PHP Girdi Dosyası Okuma Kodları	41
Resim 2.4: PHP Çıktı Dosyası Yazma Kodları	41
Resim 3.1: Kullanıcı Arayüzü Adım-1.....	42
Resim 3.2: Kullanıcı Arayüzü Adım-2.....	43
Resim 3.3: Kullanıcı Arayüzü Adım-3.....	44
Resim 3.4: Kullanıcı Arayüzü Adım-4.....	45
Resim 3.5: Kullanıcı Arayüzü Adım-5.....	45
Resim 3.6: Kullanıcı Arayüzü Dosya İndirme Adım-5.....	46

TABLolar DİZİNİ

Tablo 1.1: Genomik Analiz Programlarının Listesi ve Temel Özellikleri.....	3
Tablo 1.2: Genomik Analiz Programlarının Listesi ve Bağlantıları	6
Tablo 1.3: Bu Çalışmada Ele Alınacak Dosya Formatları ve Dosya Tipleri	9
Tablo 1.4: Bu Çalışmada Ele Alınacak Dosya Formatları ve Çevrilebileceği Formatlar.	9
Tablo 1.5: FASTA Formatı Standart Veri Tabanı Başlık Formatları	10
Tablo 1.6: FASTA Dosya Uzantıları ve Anlamları.....	11

SEMBOLLER VE KISALTMALAR DİZİNİ

AFLP	: Amplified Fragment Length Polymorphism
CSS	: Cascading Style Sheets
DDJB	: DNA Data Bank of Japan
DNA	: Deoksiribo Nükleik Asit
EMBL	: European Molecular Biology Laboratory
HTML	: HyperText Markup Language
INSDC	: International Sequence Database Collaboration
IUBMB	: International Union of Biochemistry and Molecular Biology
IUPAC	: International Union of Pure and Applied Chemistry
NCBI	: National Center for Biotechnology Information
PHP	: Hypertext Preprocessor
PIR	: Protein Identification Resource
RFLP	: Restriction Fragment Length Polymorphism
RNA	: Ribonükleik Asit
SNP	: Single Nucleotid Polymorphism
SSR	: Simple Sequence Repeats

1. GENEL BİLGİLER

1.1. Giriş

Son yıllarda biyolojik bilimlerin çok hızlı bir şekilde gelişmesi büyük miktarda DNA'nın hızla azalan bir maliyetle sekanslanmasına olanak sağlamıştır. Tüm dünyada, farklı ülkelerdeki laboratuvarlarda elde edilen tüm genomik bilgilerin, bir araya getirilmesi, kolay erişilebilir bir ortamda saklanması ve gerektiğinde araştırmacıların kullanımına sunulması için genomik sekansları destekleyen biyolojik veri tabanları oluşturulmuştur [1, 2, 10]. Ayrıca araştırmacıların, elde ettikleri genomik verileri hızlı ve güvenilir bir şekilde analiz etmeleri, diğer verilerle karşılaştırmaları için çeşitli analiz programları mevcuttur. Birçok veri analiz programı, elde edilen bu verilerin kendilerine ait formatlarda programa yüklenmesini isterler. Genetik analiz programlarının kendilerine ait formatları olduğu gibi veri tabanlarının da kendilerine ait formatları vardır. Bu farklılıkların getirdiği en büyük zorluk, veri tabanlarının ve veri analiz programlarının formatları arasında yapısal olarak farklılıklar bulunmasıdır. Bu amaçla bazı araştırmacılar ve kurumlar tarafından farklı formatlar arasında dönüşüm yapabilen çeşitli masaüstü programlar geliştirilmiştir [3]. Masaüstü yazılımlar için programlama dilleri olarak Perl, Java, C#, C, C++, R, Pascal, FORTRAN, Python gibi diller kullanılmaktadır. İstedığınız format dönüşümünü sağlayacak programı arayıp bulma ve masaüstü yazılımların web üzerinden indirilip kullanılması aşamasında çeşitli zorluklar yaşanmaktadır. Bu programlama dillerinden Java ile yazılan masaüstü programlar sistem gereksinimlerinden (bilgisayarda SUN Java 1.6 RE kurulu olmalıdır) dolayı bu işlemlerin tamamlanması, araştırmacılar için zaman kaybına neden olabilmektedir. Ayrıca bilgisayarda kurulu olan her işletim sistemi için ayrı bir uzantıya sahip programın indirilmesi istenebilir. (Windows için PGDSpider2.exe, Linux için ./PGDSpider2.sh, Mac diğerleri için java Xmx1024m Xms512m jar PGDSpider2.jar) [31]. Başka bir zorluk ise bu programların kurulum aşamasında veya arayüzünde çevirme işlemleri sırasında, araştırmacıların çeşitli komut satırlarını kullanmak zorunda kalabilmeleridir. Bu komut satırlarına yazılacak komutları öğrenme ve uygulama aşamasında, araştırmacılar büyük sıkıntı yaşamaktadırlar. Bu programların yeni versiyonu çıkınca güncelleme problemleri de ek olarak araştırmacıların zamanlarını kaybettirmektedir. Bazı programların lisans şartları

da dikkate alınırsa masaüstü yazılımların, araştırmacılara çeşitli zorluklar yaşatması kaçınılmazdır.

Web tabanlı yazılımlar; masaüstü yazılımlar gibi indirme, kurulum, sistem gereksinimleri ya da konsol uygulamaları istemezler ancak genomik veri formatlarının birbirine çevrilmesi için henüz kapsamlı bir platform veya çevrimiçi uygulama mevcut değildir. Bu çalışmada, web üzerinden geliştirilen yazılım aracılığı ile program indirme ve sistem gereksinimleri gibi ihtiyaçlara gerek kalmadan, belirtilen sorunların üstesinden gelerek araştırmacıların online (çevrimiçi) olarak yaygın kullanılan genomik veri formatlarını birbirine çevirebilmelerine olanak tanıyan bir platformun geliştirilerek hem ülkemizdeki hem de tüm dünyadaki araştırmacıların kullanımına sunulması sağlanacaktır.

1.2. Genomik Veri Formatları, Programları ve Veri tabanları

Dünyada çeşitli canlı ve organizmalara ait nükleotid dizi bilgilerinin organizasyon ve depolama işlemini üstlenen başlıca üç veri tabanı vardır. Bunlar, Amerika Birleşik Devletleri (ABD) merkezli National Center for Biotechnology Information (NCBI-GenBank), Japonya'nın DNA veri bankası olan DDJB (Japonya veri tabanı; Japonya-Mishima) veya Avrupa Moleküler Biyoloji Laboratuvarı olan EMBL/EBI'dır (Avrupa Moleküler Biyoloji Laboratuvarı; İngiltere-Hinxton) [1]. Araştırmacıların ürettikleri çeşitli moleküler genomik verilerini, depolamak ve karşılaştırmak amacıyla bu veri tabanlarından birine yüklemeleri teşvik edilir. Bu 3 veri tabanı, International Sequence Database Colloboration (INSDC) ile beraber günlük olarak verilerini güncelleyip depolayarak bu verileri erişime açık halde tutmaktadır. Ayrıca protein verileri için hizmet veren başlıca veri tabanları ise: GenBank, EMBL, PIR International (Protein Identification Resource) ve Swiss-Prot'tur [1].

GenBank'a ve diğer veri tabanlarına kayıt edilen moleküler veri bilgileri veri tabanlarında iki şekilde bulunur. İlk bölümü; moleküler/genomik verisini veri tabanına kaydeden araştırmacıya ait bilgiler, laboratuvar adresi, veri dizinlemesi yapılan organizmaya ve kaynağına ait bilgiler (Strain-izolat adı, ülke, izolasyon kaynağı, izole edilen canlı türü, tarih...), veri dizisinin başlangıç ve bitiş kodonları, ait olduğu gen ve kodladığı protein ismi gibi bilgileri bulunduran sayfa oluşturur. İkinci bölümü; gene ait nükleotid ve aminoasit popülasyon genetiği verilerini FASTA formatında içeren kısım oluşturur [1].

Buna ek olarak her bir araştırma grubu kendi laboratuvarında farklı markör(belirteç) sistemleri kullanarak bu sistemlere ait moleküler veriler oluştururlar. Bu veriler markör sistemlerinden markör sistemlerine farklılık gösterir. Ayrıca canlının biyolojisi (haploit, diploit ya da poliploit olması da) veri kodlama türünü etkilediği için yüksek sayıda format ortaya çıkmaktadır. Üstelik genomik analizler için her biri birbirinden farklı amaçlar için üretilmiş çok sayıda program bulunmaktadır ve her bir programında kendine özgü bir formatı vardır. Genomik veriler çeşitli formatlarda okunabilir ve yazılabilir. Eğer üretilen veri yaygın kullanılan bir formatta depolanmamışsa, başka formatlara çevirip analiz etmede çeşitli sıkıntılarla karşılaşılabilir. Yüksek format çeşitliliğinden dolayı da bu formatların birbirine kolayca çevrilmesini sağlayacak bir programa ihtiyaç vardır.

Aşağıda; internette ücretsiz bir şekilde indirilebilecek genomik alanında veri işleme ve analiz becerisine sahip programların listesi verilmiştir.

Tablo 1.1: Genomik Analiz Programlarının Listesi ve Temel Özellikleri [11]

Ad	Versiyon	Platform	Grafik arayüzü	Kabul edilen veri tipleri	İşlenen veri formatları
Çok Amaçlı Paketler					
ARLEQUIN	3.01	Win	Evet	DNA, SNP, STR, MULT, FREQ	Specific, GENEPOP
DnaSP	4.10	Win	Evet	DNA, SNP	In — MEGA, NEXUS, FASTA, PHYLIP; out —MEGA, NEXUS, FASTA, PHYLIP,
FSTAT	2.93	Win	Evet	STR, MULT	Specific, GENEPOP
GDA	1.1	Win	Evet	AFLP, MULT	In — NEXUS, BIOSYS, GeneStrut; out — NEXUS, BIOSYS, GeneStrut, GENESTAT-PC, SAS
Genepop	3.4	DOS	Hayır	STR, MULT	Specific

Tablo 1.1: (devam) Genomik Analiz Programlarının Listesi ve Temel Özellikleri [11]

GENETIX	4.05	Win	Evet	MULT	In — specific, FSTAT, Genepop; out — specific, FSTAT, Genepop, BIOSYS, ARLEQUIN
MEGA	3.1	Win	Evet	DNA, DIST	In — specific, CLUSTAL, NEXUS, PHYLIP, GCG, FASTA, NBRF/PIR, MSF, IG; out — specific, PHYLIP, NEXUS
MSA	4.0	DOS, MacOS, Linux	Hayır	STR, MULT	In — EXCEL; out — Genepop, MSVAR, STRUCTURE, ARLEQUIN, Migrate
SPAGeDi	1.2	DOS	Hayır	STR, MULT	Specific, FSTAT, Genepop
Bireysel Merkezli Programlar					
BayesAss+	1.3	Win, MacOS, Linux	Evet	MULT	Specific, IMMANC
BAPS	3.2	Win	Evet	MULT	Specific, Genepop
GeneClass	2.0g	Win	Evet	MULT	Genepop, FSTAT, GENETIX
Geneland	1.05	R		MULT	Specific
NewHybrids	1.1b3	Win, Linux	Hayır	MULT	Specific
STRUCTURE	2.1	Java	Evet	MULT	Specific

Tablo 1.1: (devam) Genomik Analiz Programlarının Listesi ve Temel Özellikleri [11]

Özel Programlar					
BATWING	–	DOS, MacOS, Linux	Hayır	MULT	Specific
COLONISE	1.0	Win	Evet		Specific
FDIST2	2.0	DOS, Linux	Hayır	DNA, STR, MULT	Specific
Hickory	1.0	Win, Linux	Evet	AFLP, RAPD, MULT	NEXUS
IM	–	DOS, MacOS	Hayır	DNA, STR, hapSTR	Specific
LAMARC	2.0.2	DOS, MacOS, Linux	Hayır	DNA, SNP, STR	Specific, PHYLIP, Migrate
Migrate	2.1.3	DOS, MacOS, Linux	Hayır	DNA, SNP, STR, MULT	Specific, PHYLIP
MSVAR	0.4.1.b	DOS, Linux	Hayır	STR	Specific
Dönüştürme Programları					
Convert	1.3	Win	Evet		In — EXCEL, Genepop; out — GDA, Genepop, ARLEQUIN, Popgene, MICROSAT, PHYLIP, STRUCTURE
Formatomatic	0.2	Java	Evet		In — Genepop; out — Genepop, ARLEQUIN, IMMANC

Tablo 1.2: Genomik Analiz Programlarının Listesi ve Bağlantıları [11]

Çok Amaçlı Paketler	
ARLEQUIN	http://cmpg.unibe.ch/software/ARLEQUIN3/
DnaSP	http://www.ub.es/dnasp/
FSTAT	http://www2.unil.ch/popgen/softwares/fstat.htm
GDA	http://hydrodictyon.eeb.uconn.edu/people/plewis/software.php
Genepop	http://ftp.cefe.cnrs.fr/PC/MSDOS/GENEPOP
GENETIX	http://www.univ-montp2.fr/~genetix/genetix/genetix.htm
MEGA	http://www.MEGAsoftware.net/
MSA	http://i122server.vu-wien.ac.at/MSA/MSA_download.html
SPAGeDi	http://www.ulb.ac.be/sciences/ecoevol/spagedi.html
Bireysel Merkezli Programlar	
BayesAss+	http://www.rannala.org/labpages/software.html
BAPS	http://www.rni.helsinki.fi/~jic/bapspage.html
GeneClass	http://www.montpellier.inra.fr/URLB/index.html
Geneland	http://www.inapg.inra.fr/ens_rech/mathinfo/personnel/guillot/Geneland.html
NewHybrids	http://ib.berkeley.edu/labs/slatkin/eriq/software/software.htm
STRUCTURE	http://pritch.bsd.uchicago.edu/software/STRUCTURE_2_1.html
Özel Programlar	
BATWING	http://www.mas.ncl.ac.uk/~nijw/
COLONISE	http://www-leca.ujf-grenoble.fr/logiciels.htm

Tablo 1.2: (devam) Genomik Analiz Programlarının Listesi ve Bağlantıları [11]

FDIST2	http://www.rubic.rdg.ac.uk/~mab/software.html
Hickory	http://darwin.eeb.uconn.edu/hickory/hickory.html
IM	http://lifesci.rutgers.edu/~heylab/HeylabSoftware.htm#IM
LAMARC	http://evolution.gs.washington.edu/lamarc/lamarc_prog.html
Migrate	http://popgen.csit.fsu.edu/
MSVAR	http://www.rubic.rdg.ac.uk/~mab/software.html
Dönüştürme Programları	
Convert	http://www.agriculture.purdue.edu/fnr/html/faculty/Rhodes/Students%20and%20Staff/glaubitz/software.htm
Formatomatic	http://taylor0.biology.ucla.edu/~manoukis/Pub_programs/Formatomatic/
XML Uyumlu	
BioPAX	http://www.biopax.org/
MAGE-ML	http://www.mged.org/Workgroups/MAGE
SBML	http://sbml.org/index.psp
R Kaynaklı	
HIERFSTAT	http://www2.unil.ch/popgen/software/hierfstat.htm
R-project	http://www.r-project.org/
Statistical Genetics Resources	http://cran.au.r-project.org/src/contrib/Views/Genetics.html

Tablo 1.2: (devam) Genomik Analiz Programlarının Listesi ve Baęlantıları [11]

Arayüzünde Komut Satırı Kullanılan Programlar	
CBSU	http://cbsuapps.tc.cornell.edu/index.aspx
Genepop on the web	http://wbiomed.curtin.edu.au/genepop
SNAP	http://www.cals.ncsu.edu/plantpath/people/faculty/carbone/workbench.html
Bireysel Programlar	
IMMANC	http://www.rannala.org/labpages/software.html
MESQUITE	http://mesquiteproject.org/Mesquite_Folder/docs/mesquite/manual.html
MR BAYES 3.1	http://mrbayes.csit.fsu.edu/
PHYLIP	http://evolution.genetics.washington.edu/PHYLIP.html
STRUCTURAMA	http://www.structurama.org/
TFPGA	http://www.marksgeneticsoftware.net/tfpga.htm

1.3. Bu Çalışmada Ele Alınacak Dosya Formatları ve Tiplerinin Temel Özellikleri

Nükleotid (DNA veya RNA) sekansları genellikle IUBMB (International Union of Biochemistry and Molecular Biology) standart kodları ile saklanır. Protein sekansları ise aynı şekilde IUPAC (International Union of Pure and Applied Chemistry) standart tek harfli kodlar ile kaydedilir [8].

Tablo 1.3: Bu Çalışmada Ele Alınacak Dosya Formatları ve Dosya Tipleri

Format Veri Tipleri	DNA	RNA	SSR/MICROSAT	SNP	RFLP	AFLP	Frequency	Distance	Protein
ARLEQUIN	✓		✓	✓	✓		✓		
FASTA	✓	✓		✓					✓
GENEPOP			✓	✓		✓			
MEGA	✓	✓						✓	✓
NEXUS	✓	✓							✓
PHYLIP	✓	✓		✓				✓	✓
STRUCTURE			✓	✓	✓	✓			

Tablo 1.4: Bu Çalışmada Ele Alınacak Dosya Formatları ve Çevrilebileceği Formatlar

Formatlar	ARLEQUIN	FASTA	GENEPOP	MEGA	NEXUS	PHYLIP	STRUCTURE
ARLEQUIN	-	✓	✓	✓	✓	✓	✓
FASTA	✓	-		✓	✓	✓	
GENEPOP	✓		-				✓
MEGA	✓	✓		-	✓	✓	
NEXUS	✓	✓		✓	-	✓	
PHYLIP	✓	✓		✓	✓	-	
STRUCTURE	✓		✓				-

1.3.1. Pearson | FASTA | fa

FASTA formatı nükleotid dizileri (DNA, RNA) ve protein (aminoasit) dizileri için en yaygın kullanılan metin tabanlı dosya formatıdır. Bu format genellikle hızlı benzerlik araştırmaları için kullanılır, sadece dizi (sekans) içeren bu format bazen "Pearson" olarak da adlandırılır. FASTA formatlarında dizi (sekans) bilgilerinden önce tek satırlık tanımlama satırı ve yorum satırları gelir. Tanımlama satırı çoğu kez büyükdür işareti (>) ile başlar. Büyükdür (>) işaretinden sonra isteğe bağlı olarak "AB000263" gibi bir kimlik kodu veya sekans adı gelir. Kimlik kodu ya da sekans adından sonra tanımlama satırına, isteğe bağlı olarak çeşitli sekans bilgileri eklenebilir. Birçok farklı dizi (sekans) veri tabanı, tanımlama satırlarından otomatik bilgi elde edilmesi için kendilerine ait standart başlıklar kullanırlar. Bu başlıklar içinde dikey çubuk (|) ile ayrılan ve belirli bir sıraya konulmuş çeşitli bilgiler bulunmaktadır [1, 6, 9, 15, 31].

Tablo 1.5: FASTA Formatı Standart Veri Tabanı Başlık Formatları [31]

Veri Tabanı	Başlık Formatı
GenBank	"gi" gi- numarası "gb" erişim numarası lokus
EMBL Veri Kütüphanesi	"gi" gi- numarası "emb" erişim numarası lokus
DDBJ, DNA Japon Veri Tabanı	"gi" gi-numarası "dbj" erişim numarası lokus
Genel Veri Tabanı Tanımlayıcısı	"gnl" database tanımlayıcı
Basit Kullanım	tanımlayıcı

FASTA dosyası içindeki tek satırda sekans verisi 80 karakterden az ve her satırda aynı uzunlukta olmalıdır. İstenildiği takdirde yorum satırı kullanılabilir, yorum satırları noktalı virgül (;) işareti ile başlar. Yorum satırları dosya içinde satır sonu haricinde her yerde kullanılabilir ve kullanıldığı satırı tamamen yorum haline getirir [15].


```
>AB000263 |acc=AB000263|descr=Homo sapiens mRNA for prepro cortistatin like peptide, complete cds.|len=368
ACAAGATGCCATTGTCCCGGGCTCCTGCTGCTGCTCCTCCGGGGCCACGGCCACCGCTGCCCTGCC
CCTGGAGGGTGGCCACCAGCCGACAGACAGCATATGCAGGAAGCGGCAGGAATAAGGAAAGCAGC
CTCCTGACTTTCCTCGCTTGGTGGTTTGGTGGACCTCCAGGCCAGTGCCTGGGCCCTCATAGGAGAGG
AAGCTCGGGAGGTGGCCAGGGCAGGAAGGCGCACCCCCAGCAATCCGGCGCGGGACAGAAATGCC
CTGCAGGAACCTTCTTGGAAAGACCTTCTCCTCCTGCAATAAAACCTCACCCATGAATGCTACGCAAG
TTTAATTACAGACCTGAA
```

Sekans verisi

Tanımlama satırı

Resim 1.1: Genel FASTA Dosya Formatı Görünümü [6]

```
;LCBO - Prolactin precursor - Bovine
; a sample sequence in FASTA format
MDSKGSQKGSRL LLLL VVSNLL LCQGVVSTPVCNPGNQCQVSLRDLFDRAVMVSHYIHDLS
EMFNEFDKRYAQGGFITMALNSCHTSSLPTPEDKEQAQQTTHHEVLSLILGLLRSNWDPYHL
VTEVRGMKGPADAILSRATIEEENKRLLEGMEMIFGQVIPGAKETEPYVWSGLPSLQTKDED
ARYSAFYNLLHCLRRDSSKIDTYLKLNNCRIIYNHNC*

>MCHU - Calmodulin - Human, rabbit, bovine, rat, and chicken
ADQLTEEQIAEFKEAFSLFDKDGDTITTKELGTVMRSLGQNPTEAELQDMINEVDADGNGTID
PFEFLTMARKMKDTSDEEIEAFRVFDKDGNGYISAAELRHVMTNLGKELTDEEVDEMIREA
DIDGQVNYEEFVQMMTK*

>gi|5524211|gb|AAD44166.1| cytochrome b [Elephas maximus maximus]
LCLYTHIGRNIIYGSYLYSETWNTGIMLLITMATAFMGYVLPNGQMSFWGATVITNLSAIPYIGTNLV
EWIWGGFSVDKATLNRFFAFHFILPFTMVALAGVHLTFLHETGSNNPLGLTSDSDKIPFHPYYTIKDFLG
LILILLLLLLALLSPDMLGDPDNHMPADPLNTPHLIKPEHYFLFAVAILRSVPHKLGGLVLAFLSIVIL
GLMPFLHTSKHRSMMLRPLSQALFWTLTMDLLTLTWIGSQPVEYPYTIIGQMASILYFSIILAFPLIAG
IENY
```

Resim 1.2: Diğer FASTA Dosya Formatı Görünümleri [15]

FASTA dosya uzantılarında belli bir standart yoktur ancak, “.fasta”, “.fas”, “.fa”, “.seq”, “.fna”, “.fna”, “.ffn”, “.faa”, “.frn”, “.mpfa”, “.txt” dosya uzantılarına sahip olabilirler [15, 31]. Aşağıdaki tabloda dosya uzantıları ve bunların anlamları gösterilmektedir.

Tablo 1.6: FASTA Dosya Uzantıları ve Anlamları [15]

Uzanti	Anlamı	Açıklama
fasta	Genel FASTA formatı	Diğer uzantıları “.fas”, “.fa”, “.seq”, “.fna” olabilir.
fna	FASTA Nükleik Asit	Nükleik asitleri belirtmek için kullanılır.
ffn	Gen bölgelerinin FASTA nükleotidi	Bir genom için, kodlama bölgeleri içerir.
faa	FASTA Amino Asit	Amino asitler içerir. Çoklu protein FASTA dosyası, daha özel “.mpfa” uzantısına sahip
frn	FASTA Kodlanmayan RNA	Bir genom için kodlanmayan RNA bölgelerini içerir.

FASTA dosya formatları içerisinde DNA, RNA, Protein ve SNP dosya tiplerinden herhangi birisi bulunabilir [15, 31].

FASTA dosyası içerisinde kabul edilen karakterler büyük-küçük harf duyarlılığına sahip değildir. Baz çiftleri ve aminoasitler tek harfli kodlar kullanılarak temsil edilirler. Sekans verileri içinde nükleotidler, A, C, G, T karakterleri ile, boşluklar (gap) tire işareti (-) veya iki nokta üst üste (:) karakterleri ile, eksik veriler (missing data) soru işareti (?) veya (N) karakterleri ile temsil edilirler. Aminoasit sekanslarında U ve asteriks (*) karakterleri de kabul edilebilir. Dosya içinde “population:” ifadesinin geçtiği satırda, bu ifadeden sonraki boşluğa kadar olan bölüm, popülasyon adını içerir [15, 31].

FASTA dosyası içindeki verilerden, her birey için 2 sekansa sahip olanlara *diploit* veri, tek sekansa sahip olanlara ise *haploit* veri denir. Tek dosyada sadece bir dizi tutabileceği gibi (Single Sequences) birden fazla (Multi Sequences) dizide tutulabilir [15]. Bu çalışmada, FASTA dosya formatının çevrileceği dosya formatları MEGA, PHYLIP, ARLEQUIN ve NEXUS dosya formatlarıdır.

```
>AB000263 |acc=AB000263|descr=Homo sapiens mRNA for prepro cortistatin like peptide, complete cds.|len=368
ACAAGATGCCATTGTCCCCCGGCCTCCTGCTGCTGCTGCTCTCCGGGGCCACGGCCACCGCTGCCCTGCC
CCTGGAGGGTGGCCCCACCGGCCGAGACAGCGAGCATAATGCAGGAAGCGGCAGGAATAAGGAAAAGCAGC
CTCCTGACTTTCCTCCGCTTGGTGGTTTGGAGTGGACCTCCAGGCCAGTCCCGGGCCCTCATAGGAGAGG
AAGCTCGGGAGGTGGCCAGGGCCAGGAAGGGCCACCCCCAGCAATCCGCGCCGGGACAGAATGCC
CTGCAGGAACCTTCTTGGGAAGACCTTCTCCTCCTGCAAAATAAAACCTCACCCATGAATGCTCAGCGAAG
TTTAATTACAGACCTGAA
```

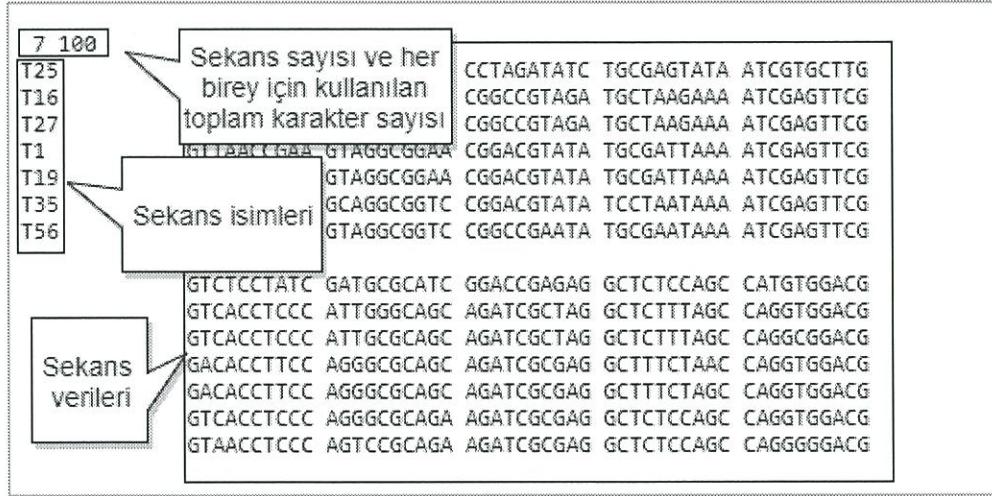
Resim 1.3: FASTA Dosya Formatı Görünümü (Single Sequences) [6]

```
>SEQUENCE_1
MTEITAAMVKELRESTGAGMMDCKNALSETNGDFDKAVQLLREKGLGKAAKKADRLAAEG
LVSVKVSDDFTIAAMRPSYLSYEDLDMTFVENEYKALVAELEKENEERRRLKDPNKPEHK
IPQFASRKQLSDAILKEAEEKIKEELKAQGKPEKIWDNIIPGKMNSFIADNSQLDSKLTLL
MGQFYVMDDKKTVEQVIAEKEKEFGGKIKIVEFICFEVGEGLKKTEDFAAEVAAQL
>SEQUENCE_2
SATVSEINSETDFVAKNDQFIALTKDTHAIQNSLSQSV EELHSSTINGVKFEEYKLSQI
ATIGENLVRRFATL KAGANGVVNGYIHTNGRVGVVIAAACDSAEVASKSRDLLRQICMH
```

Resim 1.4: FASTA Dosya Formatı Görünümü (Multi Sequences) [15]

1.3.2. PHYLIP

PHYLIP dosya formatı içindeki nükleotid sekans verileri farklı, genetik farklılık matris (distance matrix) verileri farklı yapılardadır. Nükleotid veri tipleri için, PHYLIP dosyasının ilk satırında dizi (sekans) sayısı ve bir boşluktan (space) sonra her birey için kullanılan karakter sayısı belirtilir. İlk satırda karakter sayısından sonra bir boşluk (space) bırakılarak interleaved (aralıklı) özelliğini belirtmek için “i” karakteri, sequential (sıralı) özelliğini belirtmek için “s” karakterleri kullanılabilir yani isteğe bağlıdır. İlk satırda boşluk (space) yerine tab (boşluk) kullanılamaz. Sonraki satır sekans adı ile başlar ve sekans verileri genellikle onlu bloklar halinde yazılır. PHYLIP dosya formatında birey isimleri içinde alt çizgi (_), boşluk (space), noktalama işaretleri ve rakam kullanılabilir [14, 18, 31].



Resim 1.5: PHYLIP DNA Dosya Formatı Görünümü [14]

PHYLIP dosya formatları, Interleaved (aralıklı) ve Sequential (sıralı) olmak üzere 2 çeşittir. Interleaved formatta genotip veriler, dosya içinde, 1 boş satır ile 2 parçaya ayrılır. İlk parçada birey sayısı kadar satır vardır. Her satır, maksimum 10 karakter kullanılacak şekilde birey ismi ile başlar ve 11. karakterden sonra genotip verilerin bir kısmı yazılır. Yazılan genotip veriler genellikle her 10 karakterde bir boşluk (space) kullanılarak 10’lu gruplar halinde yazılır. 2. parçada ise birey isimleri kullanılmadan yine her 10 karakterden sonra bir boşluk (space) kullanılarak genotip veriler yazılır. 2. parçada ki genotip verilerin

karakter sayısı ile ilk parçadaki karakter sayıları toplandığında ilk satırda belirtilen toplam karakter sayısı ile eşit olmalıdır [18].

```
2 250 i
CPZANT ATGGGAGCGGGGCGCTCTGTTT TGAGGGGAGAGAGCTAGATACATGGGA
U455 ATGGGTGCGAGCGCTCAGTATTAAGCGGGAAAAATAGATTCATGGGA
AACTATCAGGCTTCGGCCGGTGGCAAGAAAAGTACATGATAAAACATC
GAAAATTGCTTAAAGCCAGGGGAAACAAAAATATAGACTGAACATT
TGGTTTGGCCAGGATCGGAGCTGCAGCGT TTTGGGCTCAGCTCCTCCCTT
TAGTATGGCCAGCGAGGAGCTGAAAAATTCACACTTAACCCCTGGCCTT
CTAGAAACATCAGAAGGTTGTGAAAAGGCTATCCATCAATTGAGCCCTTC
TTAGAAACAGCAGAAGGATGTCAGCAATACTGGGCAATTAACAACCGC
CATAGAAATAGATCCCTGAAATAATATCTTGTTTAACCCATTTGTG
TCTCCAGCGGGAAACAGAACTTAGATCATATATAATACAGTAGCG
```

Resim 1.6: PHYLIP Interleaved Dizi Formatı Görünümü [18]

Sequential formatta ise, genotip veriler kullanılırken, yine maksimum 10 karakter olacak şekilde birey isimleri kullanılır, 11. karakterden sonra her bireyin tüm genom bilgisi yazılır ve diğer bireyin bilgilerini aynı şekilde yazmak için bir sonraki satırdan devam edilir [18].

```
2 247 =
CPZANT ATGGGAGCGGGGCGCTCTGTTT TGAGGGGAGAGAGCTAGATACATGGGA
AACTATCAGGCTTCGGCCGGTGGCAAGAAAAGTACATGATAAAACATC
TGGTTTGGCCAGGATCGGAGCTGCAGCGT TTTGGGCTCAGCTCCTCCCTT
CTAGAAACATCAGAAGGTTGTGAAAAGGCTATCCATCAATTGAGCCCTTC
CATAGAAATAGATCCCTGAAATAATATCTTGTTTAACCCATTT
U455 ATGGGTGCGAGCGCTCAGTATTAAGCGGGAAAAATAGATTCATGGGA
GAAAATTGCTTAAAGCCAGGGGAAACAAAAATATAGACTGAACATT
TAGTATGGCCAGCGAGGAGCTGAAAAATTCACACTTAACCCCTGGCCTT
TTAGAAACAGCAGAAGGATGTCAGCAATACTGGGCAATTAACAACCGC
TCTCCAGCGGGAAACAGAACTTAGATCATATATAATACAGTAG
```

Resim 1.7: PHYLIP Sequential Dizi Formatı Görünümü [18]

PHYLIP dosyaları içinde Distance dosya tipleri için ayrı bir yazım şekli vardır. Bu dosyasının ilk satırında sadece birey sayısı yazılır. Daha sonra genetik farklılık matrisi (distance matrix) yazılır. Matris içinde birey isimleri maksimum 10 karakter uzunluğunda olabilir. Bunlara boşluklar ve noktalama işaretleri dahildir [31].

```

5
Alpha      0.000 1.000 2.000 3.000 3.000
Beta       1.000 0.000 2.000 3.000 3.000
Gamma      2.000 2.000 0.000 3.000 3.000
Delta      3.000 3.000 3.000 0.000 1.000
Epsilon    3.000 3.000 3.000 1.000 0.000

```

Resim 1.8: PHYLIP Kare Genetik Farklılık Matrisi (Distance Matrix) Formatı Görünümü [31]

Genetik farklılık matrisi (distance matrix) üçgen formatında ise gereksinimlere göre veriler, yeni bir satırda devam edebilir [31].

```

14
Mouse
Bovine      1.7043
Lemur       2.0235  1.1901
Tarsier     2.1378  1.3287  1.2905
Squir Monk  1.5232  1.2423  1.3199  1.7878
Jpn Macaq   1.8261  1.2508  1.3887  1.3137  1.0642
Rhesus Mac  1.9182  1.2536  1.4658  1.3788  1.1124  0.1022
Crab-E.Mac  2.0039  1.3066  1.4826  1.3826  0.9832  0.2061  0.2681
BarbMacaq   1.9431  1.2827  1.4502  1.4543  1.0629  0.3895  0.3930  0.3665
Gibbon      1.9663  1.3296  1.8708  1.6683  0.9228  0.8035  0.7109  0.8132
0.7858
Orang       2.0593  1.2005  1.5356  1.6606  1.0681  0.7239  0.7290  0.7894
0.7140  0.7095
Gorilla     1.6664  1.3460  1.4577  1.5935  0.9127  0.7278  0.7412  0.8763
0.7966  0.5959  0.4604
Chimp       1.7320  1.3757  1.7803  1.7119  1.0635  0.7899  0.8742  0.8868
0.8288  0.6213  0.5065  0.3502
Human       1.7101  1.3956  1.6661  1.7599  1.0557  0.6933  0.7118  0.7589
0.8542  0.5612  0.4700  0.3097  0.2712

```

Resim 1.9: PHYLIP Üçgen Genetik Farklılık Matrisi (Distance Matrix) Formatı Görünümü [12]

PHYLIP dosyalarının uzantıları “.txt” , “.phy” veya “.py” uzantılıdır. Bu çalışmada PHYLIP dosya formatları içerisinde DNA, RNA, Protein, SNP veya Distance dosya tiplerinden herhangi birisi bulunabilir [31]. PHYLIP dosya formatının çevrileceği dosya formatları MEGA, NEXUS, ARLQUIN ve FASTA dosya formatlarıdır.

```

18 234
Cow      MAVPMQLGFQ  DAAAPVMEEL  LHFHDMALME  VFLESSLVLY  IISLMLTTKL
Carp     MAHPTQLGFQ  DAAAPVMEEL  LHFHDMALME  VLIISTLVLY  IITAMVSTKL
Chicken MAHNSQLGFQ  DAAAPVMEEL  VEFHDMALMV  ALAICSLVLY  LLTLMMEKL
Human   MAHAAQVGLQ  DAAAPVMEEL  IFFHDMALME  IFLICFLVLY  ALFLTITKL
Loach   MAHPTQLGFQ  DAAAPVMEEL  LHFHDMALME  VFLESSLVLY  VEITTVSTKL
Mouse   MAVPFQLGLQ  DAAAPVMEEL  MHFDMALME  VFLESSLVLY  IISLMLTTKL
Rat     MAVPFQLGLQ  DAAAPVMEEL  THFDMALME  VFLESSLVLY  IISLMLTTKL
SeaL    MAYPLQMLQ  DAAAPVMEEL  LHFHDMALME  VFLESSLVLY  IISLMLTTKL
Whale   MAVPFQLGFQ  DAAAPVMEEL  LHFHDMALME  VFLESSLVLY  IITLMLTTKL
Frog    MAHPSQLGFQ  DAAAPVMEEL  LHFHDMALMA  VLIISTLVLY  IITAMVTTKL

THTSTMDAQE  VETIWTILPA  IILILIALPS  LRILYHMEI  NNPSLTVKTM
TNKYILDSQE  IETVWTILPA  VILVLIALPS  LRILYHMEI  NDPHLTIKAM
S-SNTVDAQE  VELIWTILPA  IVLVLIALPS  LQILYHMEI  DEPDILKAI
TNTNISDAQE  METVWTILPA  IILVLIALPS  LRILYHMEI  NDPSLTIKSI
TNKYILDSQE  IETVWTILPA  IILILIALPS  LRILYHMEI  NDPHLTIKAM
THTSTMDAQE  VETIWTILPA  VILIMIALPS  LRILYHMEI  NNPSLTVKTM
THTSTMDAQE  VETIWTILPA  VILILIALPS  LRILYHMEI  NNPSLTVKTM
THTSTMDAQE  VETVWTILPA  IILILIALPS  LRILYHMEI  NNPSLTVKTM
THTSTMDAQE  VETVWTILPA  IELILIALPS  LRILYHMEI  NNPSLTVKTM
TNTNLMDAQE  IEMVWTIMPA  ISLIMIALPS  LRILYHMEI  NDPHLTIKAI

GHQWYNSVEY  TDYEDLSFDS  YMIPTSELKP  GELRLEVDN  RVVLRMENTI
GHQWYNSVEY  TDYENLGFDS  YMVPTQDLAP  GQFRLETDH  RMVVRMESPV
GHQWYNTVEY  TDFKDLSDS  YMTPTDLP  GHFRLEVDN  RIVIPMESPI
GHQWYNTVEY  TDYGGIFMS  YMLPPIFIEP  GDLRLEVDN  RVVLRTEAPI
GHQWYNSVEY  TDYENLGFDS  YMIPTQDLTP  GQFRLETDH  RMVVRMESPE
GHQWYNSVEY  TDYEDLCFDS  YMIPTNDLKP  GELRLEVDN  RVVLRMELPI
GHQWYNSVEY  TDYEDLCFDS  YMIPTNDLKP  GELRLEVDN  RVVLRMELPE
GHQWYNSVEY  TDYEDLNFD  YMIPTQELKP  GELRLEVDN  RVVLRMENTI
GHQWYNSVEY  TDYEDLSFDS  YMIPTSDLKP  GELRLEVDN  RVVLRMENTI
GHQWYNSVEY  TWYEDLSFDS  YMIPTNDLTP  GQFRLEVDN  RMVVRMESPT

RMLVSSDVL  HSNVAVSLGL  KTDATPGRIN  QTTLMSSRPG  LYGQCSEIC
RVLVAEDVL  HSNVAVSLGV  KMDAVPGRIN  QAAFIASRPG  VFYGCSEIC
RVITADDVL  HSNVAVPALGV  KTDATPGRIN  QTSFITTRPG  VFYGCSEIC
RMNITSQDVL  HSNVAVPTLGL  KTDATPGRIN  QTTFTATRPG  VFYGCSEIC
RILVAEDVL  HSNALPAMGV  KMDAVPGRIN  QTAFIASRPG  VFYGCSEIC
RALISSDVL  HSNVAVSLGL  KTDATPGRIN  QATVTSNRPG  LFYGCSEIC
RMLISSDVL  HSNATPSLGL  KTDATPGRIN  QATVTSNRPG  LFYGCSEIC
RMLISSDVL  HSNVAVSLGL  KTDATPGRIN  QTTLMSTRPG  LFYGCSEIC
RMLVSSDVL  HSNVAVSLGL  KTDATPGRIN  QTTLMSTRPG  LFYGCSEIC
RLLVTAEDVL  HSNVAVSLGV  KTDATPGRIN  QTSFIATRPG  VFYGCSEIC

GSNHSFMPIV  LELVPLKYPE  KHSASML--- ---
GANHSFMPIV  VEAVPLEHFE  NKSLSMLERA  SLGS
GANHSFMPIV  VESTPLKNFE  AKSSL-----  -LSS
GANHSFMPIV  LELIPLKIFE  M-----GP  VFTL
GANHSFMPIV  VEAVPLEHFE  NKSLSMLKDA  SLGS
GSNHSFMPIV  LEMVPLKYPE  NKSASMT--- ---
GSNHSFMPIV  LEMVPLKYPE  NKSASMT--- ---
GSNHSFMPIV  LELVPLEHFE  KHSASML--- ---
GSNHSFMPIV  LELVPLEHFE  KHSASML--- ---
GANHSFMPIV  VEAVPLTDFE  NKSLSML-EA  SL--

```

Resim 1.10: PHYLIP Protein Dosya Formatı Görünümü [17]

1.3.3. PAUP | NEXUS

NEXUS dosya formatı GDA, Paup*, Mesquite, ModelTest, MrBayes, MacClade ve Spectrum gibi çeşitli programlar tarafından kullanılmaktadır. NEXUS dosyası "#NEXUS" ifadesi ile başlar ve "END" ifadesi ile biter [13,14]. "matrix" ifadesinin geçtiği satıra kadar dosya içindeki veriler ile ilgili bilgiler yer alır, bu ifadeden sonraki satırlarda birey isimleri ve genotip verileri yer alır. Genotip bilgilerin bulunduğu satırlar, birey isimleri ile başlar ve bir boşluktan sonra genotip verileri gelir. Birey isimleri tekrarlanamaz, tamamen rakamlardan oluşamaz ve birey isimleri arasında boşluk (space) kullanılmaz. Boşluk yerine alt tire () karakteri kullanılmalıdır. Dosya içindeki karakterler büyük-küçük harf duyarlılığına sahip değildir ve köşeli parantez ([]) içindeki bilgiler yorum olarak kabul edilir. Yorum olarak kabul edilen satırlar çevirme programları tarafından okunabilirler fakat genetik verilerin içine dahil edilmezler ayrıca hesaplamalarda kullanılmazlar. Yorum satırları içinde kayıp veri (missing data), boşluk (gap) yerine kullanılan karakterler belirtilebilirler [31].

```
#NEXUS
[ Sample: Başlangıç ifadesi 638 -- data title]

[Name: DRONTGNC Len: 36 Check: 1ADB7C5]
[Name: Dros Len: 36 Check: C8051EB0]
[Name: YMU09206 Len: 36 Check: E84CA691]
[Name: LUCMPTIEA Len: 36 Check: 80147A4E]
[Name: MSQNCATR Len: 36 Check: 31F0A8]
[Name: Apis Len: 36 Check: 736F31AC]

begin data;
dimension n
format dat type=uno uncl=evr missing=. gap=-; ["missing=" changed to "missing=. gap=- ;"]
matrix
DRONTGNC TACTACCTGCTCTTTCTTT ATTATTAGTAAGAAGA
Dros ...TATCCTGCTCTTTCTTT ATTATTAGTAAGAAGA
YMU09206 TATTATCCATCCTTAACT ACTTAATTCTAGAAGA
LUCMPTIEA TTTTATCCTGCAT--ACTTT ACTATTAGTAAGTAGT
MSQNCATR TATTACCCCTCTTAACTCT TCTAA-TTCTAGAAGT
Apis TACTTTCCTCATTATTTAT ACTTTTATTAAGAAAT

end;
```

Yorum kullanımı

matrix kelimesinden sonra birey isimleri ve genetik veriler başlar

Birey isimleri ve genetik veriler

Bitiş ifadesi

Resim 1.11: NEXUS Dosya Formatı Görünümü [28]

Interleaved (aralıklı) ve Sequential (sıralı) olmak üzere 2 çeşit NEXUS formatı vardır [18].

```
#NEXUS

begin taxa;
dimensions ntax=2;
taxlabels
CPZANT
U455
;
end;

begin characters;
dimensions nchar=250;
format interleave datatype=DNA gap=-;
matrix
CPZANT ATGGGAGCGGGGGCGTCTGTTTTCAGGGGAGAGAGAGCTAGATACATGGGA
U455 ATGGGTCCGAGAGCGTCACTATTAAGCCGGCAAAAANTTAGATTCATGGGA

CPZANT AAGTATCAGGCTTCGGCCCGGTGCCAGGAAAAGTACATGATAAAACATC
U455 GAAAAATTGGTTAAGGCCAGGGGGAACAAAAAATATAGACTGAACATT

CPZANT TGGTTTGGGCAAGATCGGAGCTGCAGCGTTTTCGGCTCAGCTCCTCCCTT
U455 TAGTATGGGCAAGCAGGGGAGCTGAAAAAATTCACACTTAAACCCTGGCCCTT

CPZANT CTAGAAACATCAGGAGGTTGTGAAAAGGCTATCCATCAATTGAGCCCTTC
U455 TTGAAACAGCAAGAGGATGTCAGCAATACTGGGCAATTAACAACAGC

CPZANT CATAGAAATAGATCCCTGAAATATATCTTTGTTTAAACACCATTTGGG
U455 TCTCCAGACAGGAACAGAGAACTTAGATCATATATATATACAGTAGCAG

;
end;
```

Resim 1.12: Interleaved NEXUS Dosya Formatı Görünümü [18]

```
#NEXUS

Begin data;
Dimensions ntax=3 nchar=79;
Format datatype=dna gap=-;
Matrix

4axED43xco GGAAGDCCCTACCTCAAGTAGTGAAGCCCTACCTCCCGTTGGCTGTTTCCTCTTGGGTAGAACGCTACITTCGGGCAACC
2bxMD2b2x1 CGGTGTTGATCACCAATCGGAGGGCACCCTA-----GGACACAGCTCCTCATGGATCGAGGCTACTTTCIAACCSTGA
2bxMD2b3x1 CGGTGCCAATAACCGAGTCCGAAAGGCAATCAAGGTTGAGACACGGCTCCCCATGAACCGAGGCTATTCCTAACCTGG

;
End;
```

Resim 1.13: Sequential NEXUS Dosya Formatı Görünümü [19]

NEXUS formatının dosya uzantısı “.nex” ya da “.nxs” dir. NEXUS dosya formatları içerisinde DNA, RNA veya Protein dosya tiplerinden herhangi birisi bulunabilir [20]. Bu çalışmada NEXUS dosya formatının çevrileceği dosya formatları ARLEQUIN, MEGA, PHYLIP ve FASTA dosya formatlarıdır.

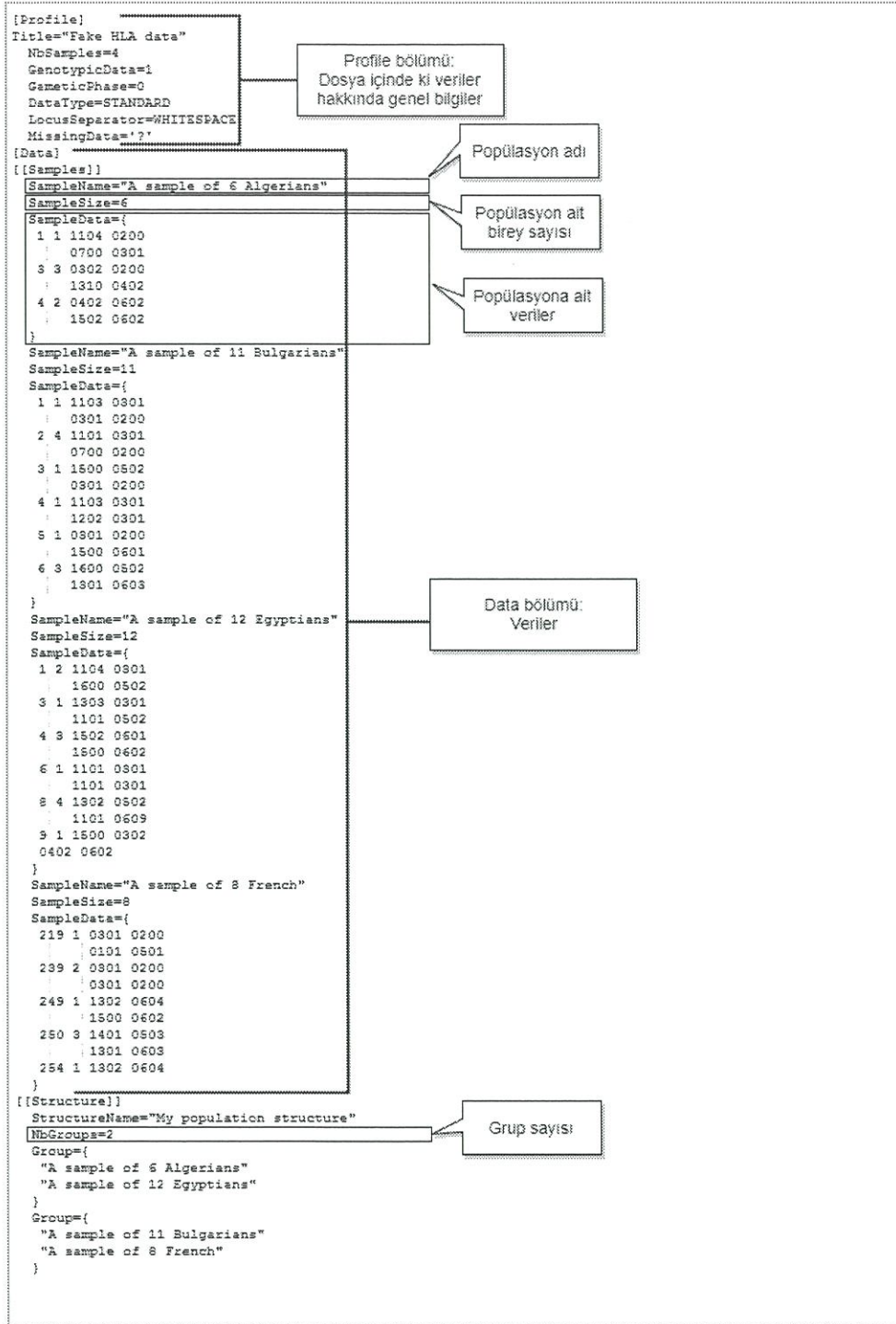
1.3.4. ARLEQUIN

ARLEQUIN dosyaları, *Profile* ve *Data* bölümleri olmak üzere 2 ana bölümden oluşmaktadır [21].

- *Profile* Bölümü
- *Data* Bölümü
 - Haplotip Listesi (Opsiyonel)
 - Genetik Veri Matrisleri (Opsiyonel)
 - Örnekler (Zorunlu)
 - Genetik Yapı (Opsiyonel)
 - Mantel Testleri (Opsiyonel)

Profile bölümünün başladığını belirtmek için satır başında “[Profile]” ifadesi kullanılır. ARLEQUIN dosyaları içindeki “Profile” bölümünde başlık, popülasyon sayısı, veri tipi, haplotip ya da genotip bilgisi, gametik safhası, resesif bilgisi, eksik veriler (missing data) için kullanılan karakter, lokus ayıraç karakteri gibi çeşitli bilgiler bulunmaktadır. “Data” bölümünde ise popülasyon adı, popülasyona ait birey sayısı, popülasyona ait veriler ve grup sayısı gibi bilgiler yer almaktadır [21].

“Data” bölümünün başladığını belirtmek için satır başında “[Data]” ifadesi kullanılır. ARLEQUIN dosyaları içindeki “Data” bölümünde, analiz edilmesi için çeşitli veriler vardır. “Data” bölümünün alt bölümü olan “Samples” içinde popülasyon sayısı, popülasyona ait birey sayısı bulunur. Ayrıca “SampleData” ifadesinden sonra gelen küme parantezleri ({}) içinde popülasyona ait veriler yazılır [21].



Resim 1.14: ARLEQUIN Dosya Formatı Görünümü [21]

ARLEQUIN dosyaları içinde hem *haplotip*, hem de *genotip* çeşitlerindeki veriler kullanılabilir. *Haplotip* bir veya daha fazla lokusdaki allellerin birleşiminden oluşmaktadır [21].

```
100110100101001010
```

Resim 1.15: Haplotipik RFLP Verisi Görünümü [21]

```
DRB1*0101 DQB1*0102 DPB1*0201
```

Resim 1.16: Haplotipik STANDARD HLA Verisi Görünümü [21]

```
ACGGCATTTAAGCATGACATACGGATTGACA  
ACGGGATTTTAGCATGACATTCCGGATAGACA
```

Resim 1.17: Genotipik DNA Sekans Verisi Görünümü [21]

```
63    24    32  
62    24    30
```

Resim 1.18: Genotipik Mikrosatelit Verisi Görünümü [21]

ARLEQUIN dosyaları “.arp” uzantılıdır. ARLEQUIN dosya formatları içerisinde SSR/MICROSAT, DNA, RFLP, SNP ve Frequency dosya tiplerinden herhangi birisi *haploit* ve *diploit* olarak bulunabilir. ARLEQUIN interleaved (aralıklı) verileri desteklemez. Haplotipik, genotipik ya da Distance verilerin hepsi tek satırda yazılmalıdır. Her satırda maksimum 100.000 karakter kullanılabilir. ARLEQUIN dosya formatı içinde diyez (#) karakterinden sonra gelen her şey yorum olarak kabul edilir ve sadece bilgi amaçlıdır [21, 31]. Bu çalışmada ARLEQUIN dosya formatının çevrileceği dosya formatları GENEPOP, STRUCTURE, MEGA, PHYLIP, FASTA ve NEXUS dosya formatlarıdır.

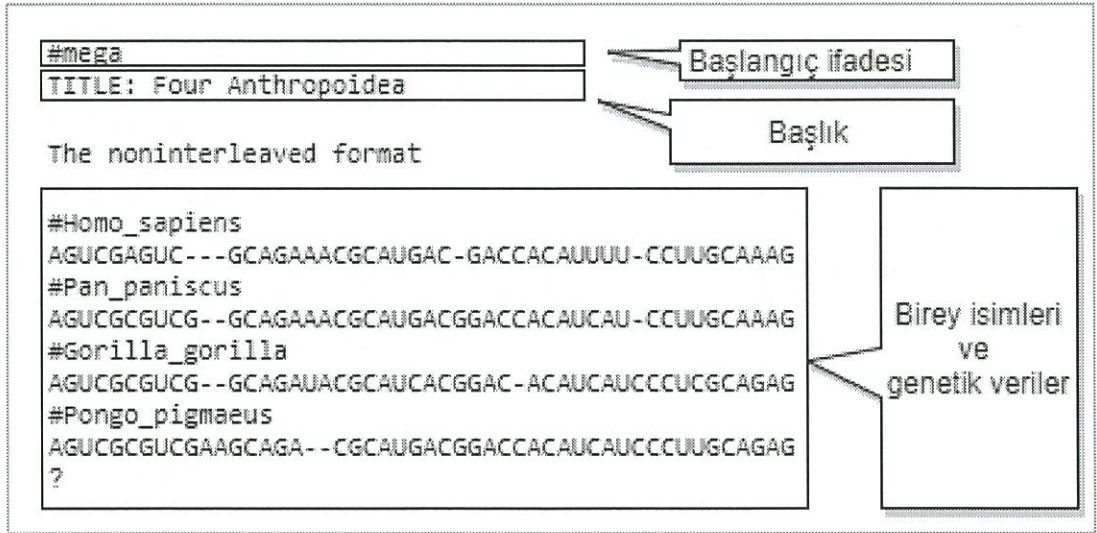
```
Id1 2  ACICGGGITCGCGCGC # the first pseudo-haplotype  
      ACICGGGCTCAGCGCGC # the second pseudo-haplotype
```

Diyez(#) karakterinden sonra gelen ifadeler yorum olarak kabul edilir.

Resim 1.19: ARLEQUIN Yorum İfadesi Görünümü [31]

1.3.5. MEGA

MEGA dosya formatı içindeki nükleotid sekans verileri farklı, genetik farklılık verileri farklı yapıdadır. Nükleotid veri tipleri için MEGA dosyasının ilk satırında, “#MEGA” ifadesi bulunurken sonraki satırlardaki “!Title” ifadesi ile başlık, “!Description” ifadesi ile açıklama, “!Gene” ifadesi ile gen kimliği ve özelliklerini gösteren açıklama yazılabilir. Genotip verilerin başladığı yer, diez (#) işaretinin olduğu satırdır, diez (#) işareti birey isimlerinin önüne bitişik yazılır. Birey isimlerinin uzunluklarında herhangi bir sınırlama yoktur. Birden fazla kelime kullanılan isimlendirmelerde boşluklar yerine alt çizgi () kullanılabilir. Sekans verileri genellikle IUPAC (International Union of Pure and Applied Chemistry) standart tek harfli kodlar ile kaydedilir ve büyük-küçük harf beraber kullanılabilir [29, 31].



Resim 1.20: MEGA Dosya Formatı Görünümü [30]

MEGA dosyaları “.txt”, “.meg” uzantısına sahiptirler ya da ASCII metin dosyaları olmak zorundadır. MEGA dosya formatları içerisinde DNA, RNA, Distance ve Protein dosya tiplerinden herhangi birisi bulunabilir. Eksik veriler (missing data) için soru işareti (?) ve boşluklar (gap) için ise tire işareti (-) kullanılır. Sekans verileri eşit uzunluklarda olmalıdır [23, 24, 31]. Bu çalışmada MEGA dosya formatının çevrileceği dosya formatları FASTA, PHYLIP, ARLEQUIN ve NEXUS dosya formatlarıdır.

Interleaved (aralıklı) ve Sequential (sıralı) olmak üzere 2 çeşit MEGA formatı vardır [18].

```
#MEGA
TITLE: Interleaved sequence data

#CPZANT ATGGGAGCGGGGCGCTCTGTTTTGAGGGGAGAGAGCTAGATACATGGGA
#U455 ATGGGTGCGAGAGCGTCAGTATTAAAGCGGAAALAAATTAGATTTCATGGGA

#CFZANT AAGTATCAGGCTTCGGCCCGGTGGCRAAGAAAAGTACATGATAAARCATC
#U455 GAAATTCGGTTAAGGCCAGGGGAAACAAAATATAGACTGAACATT

#CPZANT TGGTTTGGCAGAGATCGGAGCTGCAGCGTTTTGGGCTCAGCTCCTCCCTT
#U455 TAGTATGGCRAAGCGGAGCTGGAAATAATTCACACTTAACCCCTGGCCTT

#CPZANT CTAGAAACATCAGAGGTTGTGAAAGGCTATCCATCAATTGAGCCCTTC
#U455 TTAGAAACAGCAGAGGATGTGCAAAATCTGGGACAAATACACCCAGC

#CPZANT CATAGAAATAGATCCCTGARATAATATCTTTGTTTAAACCCATTTGTG
#U455 TCTCCAGCAGGAAACAGAGACTTAGATCATTATATAATACAGTAGCAG
```

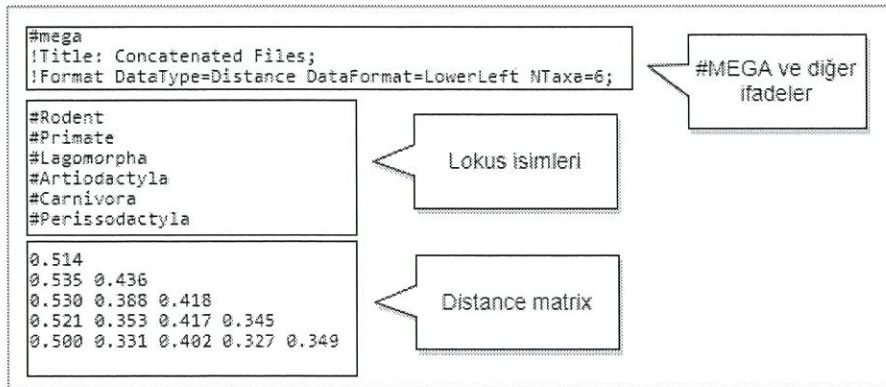
Resim 1.21: MEGA Interleaved Dosya Formatı Görünümü [18]

```
#MEGA
TITLE: Noninterleaved sequence data

#CPZANT
ATGGGAGCGGGGCGCTCTGTTTTGAGGGGAGAGAGCTAGATACATGGGA
AAGTATCAGGCTTCGGCCCGGTGGCRAAGAAAAGTACATGATAAARCATC
TGGTTTGGCAGAGATCGGAGCTGCAGCGTTTTGGGCTCAGCTCCTCCCTT
CTAGAAACATCAGAGGTTGTGAAAGGCTATCCATCAATTGAGCCCTTC
CATAGAAATAGATCCCTGARATAATATCTTTGTTTAAACCCATTTGTG

#U455
ATGGGTGCGAGAGCGTCAGTATTAAAGCGGAAALAAATTAGATTTCATGGGA
GAAATTCGGTTAAGGCCAGGGGAAACAAAATATAGACTGAACATT
TAGTATGGCRAAGCGGAGCTGGAAATAATTCACACTTAACCCCTGGCCTT
TTAGAAACAGCAGAGGATGTGCAAAATCTGGGACAAATACACCCAGC
TCTCCAGCAGGAAACAGAGACTTAGATCATTATATAATACAGTAGCAG
```

Resim 1.22: MEGA Sequential (Noninterleaved) Dosya Formatı Görünümü [18]



Resim 1.23: MEGA Distance Veri Tipi Dosya Formatı Görünümü [23]

MEGA dosyaları içinde Distance dosya tipleri için ayrı bir yazım şekli vardır. Bu dosyanın ilk satırında, “#MEGA” ifadesi bulunur ve sonraki satırlarda ki “!Title”, “!Description” ve “!Format” ifadeleri isteğe bağlı olarak kullanılabilir. Daha sonra lokus isimleri gelir. Lokus isimlerinden sonra genetik farklılık matrisi (distance matrix) yazılır [23, 31].

1.3.6. STRUCTURE

STRUCTURE dosya formatı matris gibi düzenlenmiştir. Satırlarda birey isimleri kolonlarda lokus isimleri yer alır. Dosyanın ilk satırında, isteğe bağlı olarak lokus isimleri yer alır. Lokus isimleri arasında 2 tab (boşluk) kullanılır. Lokus isimleri içinde boşluk veya tab karakterleri kullanılamaz. Sonraki satırlarda birey isimleri ve genotip veriler yer alır. Genotip veriler içinde sırasıyla birey adı, popülasyon numarası ve allel uzunlukları birer tab (boşluk) arayla yazılır. Birey isimleri ve popülasyon numaraları isteğe bağlı olarak kullanılabilir. Diploit organizmalar için veriler istenirse ardışık satırlar halinde ya da her bir lokus için iki ardışık kolon üzerine yazılabilir [31].

		loc_a	loc_b	loc_c	loc_d	loc_e
George	1	-9	145	66	0	92
George	1			64	0	94
Paula	1			68	1	92
Paula	1			64	0	94
Matthew	2	110	145	-9	0	92
Matthew	2	110	148	66	1	-9
Bob	2	108	142	64	1	94
Bob	2	-9	142	-9	0	94

Diagram annotations:

- Lokus isimleri:** loc_a, loc_b, loc_c, loc_d, loc_e
- Popülasyon numaraları:** 1, 2
- Eksik veri(Missing data):** -9
- Birey isimleri:** George, Paula, Matthew, Bob
- Birey verileri:** The second column (populations) and the data columns (loc_a to loc_e)

Resim 1.24: STRUCTURE Dosya Formatı Görünümü [31]

STRUCTURE veriler içindeki eksik veriler (missing data) için, genotip verilerin içinde bulunmayan bir sayı olan “-9” veya “-1” ile gösterilir. STRUCTURE içerisinde sadece *haploit* ya da sadece *diploit* veriler bulunur. STRUCTURE dosya uzantıları “.txt” ya da “.str” uzantılıdır. STRUCTURE dosya formatları içerisinde AFLP (Amplified Fragment Length Polymorphism), RFLP (Restriction Fragment Length Polymorphism), SSR/MICROSAT (Simple Sequence Repeats) ve SNP (Single Nucleotide Polymorphism)

dosya tiplerinden herhangi birisi bulunabilir [22, 25, 31]. Bu çalışmada STRUCTURE dosya formatının çevrileceği dosya formatları GENEPOP ve ARLEQUIN dosya formatlarıdır.

1.3.7. GENEPOP

GENEPOP dosyası, içerisindeki veri hakkında açıklayıcı bir bilgi satırı ile başlar. Açıklayıcı bilgi satırı içinde her karakter kullanılabilir. İkinci satır lokus isimlerinin başladığı satırdır. Lokus isimleri 2 farklı şekilde yazılabilir. Lokus isimleri tek satırda aralarında virgül ya da virgül+boşluk (space) kullanılarak yazılabilir [26, 42, 43].

```
Title line: "Grape populations in southern France"
ADH Locus 1, ADH #2, ADH three, ADH-4, ADH-5, mtDNA
Pop Grange_des_Peris
Grange des Peres , 02 003003 0102 0302 1011 01
Grange des Peres , 02 002001 0102 0303 1111 02
Grange d 102 1010 01
Grange d 202 1011 02
Grange d 102 1010 01
POP Tertre_Roteboeuf
Tertre Roteboeuf , 0102 002002 0201 0405 0807 01
Tertre Roteboeuf , 0102 002001 0201 0405 0307 01
Tertre Roteboeuf , 0201 002003 0101 0505 0402 01
Tertre Roteboeuf , 0201 003003 0301 0303 0603 01
Tertre Roteboeuf , 0101 002001 0301 0505 0807 01
pop
Bonneau 01 , 0101 002002 0304 0805 0304 01
Bonneau 02 , 0201 002002 0404 0505 0304 02
Bonneau 03 , 0101 002100 0304 0505 0101 01
Bonneau 04 , 0101 100100 0204 0805 0304 01
Bonneau 05 , 0101 100002 0104 0808 0304 03
Pop
, 0000 002001 0202 0402 0007 01
, 0200 002001 0202 0205 0707 01
, 0010 002001 0101 0105 0807 01
```

Resim 1.25: GENEPOP Dosya Formatı Görünümü-1 [43]

İkinci kullanımda ise 2. satırdan başlayarak, virgül (,) kullanmadan art arda satırlarda sadece lokus isimleri yazılarak devam edilir [26].

```
Title line: "Grape populations in southern France"
ADH Locus 1
ADH #2
ADH three
ADH-4
ADH-5
mtDNA
Pop
Grange des Peres , 0201 003003 0102 0302 1011 01
Grange des Peres , 0202 003001 0102 0303 1111 02
Grange des Peres , 0102 004001 0202 0102 1010 01
Grange des Peres , 0103 002002 0101 0202 1011 02
Grange des Peres , 0203 002004 0101 0102 1010 01
POP
Tertre Roteboeuf , 0102 002002 0201 0405 0807 01
Tertre Roteboeuf , 0102 002001 0201 0405 0307 01
Tertre Roteboeuf , 0201 002003 0101 0505 0402 01
Tertre Roteboeuf , 0201 003003 0301 0303 0603 01
Tertre Roteboeuf , 0101 002001 0301 0505 0807 01
pop
Bonneau 01 , 0101 002002 0304 0805 0304 01
Bonneau 02 , 0201 002002 0404 0505 0304 02
Bonneau 03 , 0101 002100 0304 0505 0101 01
Bonneau 04 , 0101 100100 0204 0805 0304 01
Bonneau 05 , 0101 100002 0104 0808 0304 03
Pop
, 0000 002001 0202 0402 0807 01
, 0200 002001 0202 0205 0707 01
, 0010 002001 0101 0105 0807 01
```

Lokus isimleri ayrı satırlarda, virgül (,) kullanılmadan yazılır.

Resim 1.26: GENEPOP Dosya Formatı Görünümü-2 [42]

Lokus isimleri içinde virgül (,) haricinde boşluk (space) dâhil herhangi bir karakter kullanılabilir. Çünkü lokus isimleri birbirlerinden ancak virgül(,) ile ayrılabilir. Lokus isimlerinden sonraki satırda, bir popülasyonun başladığını belirtmek için satır başında "POP" kelimesi kullanılır. Bu kelimedeki büyük-küçük harf duyarlılığı olmadığı için "Pop" veya "pop" kelimeleri de kullanılabilir. "POP" kelimesi ile başlayan satırda, isteğe bağlı olarak, bu kelimedeki boşluk bırakılıp popülasyon ismi yazılabilir. Eğer popülasyon ismi belirtilmemişse GENEPOP programının versiyonlarına ya da türlerine göre (desktop veya web), popülasyonun ilk bireyinin ya da son bireyinin ismi, otomatik olarak bu popülasyona isim olarak verilebilir. "POP" kelimesinin bulunduğu satırdan sonraki her satırda, bu popülasyona ait bir bireyin bilgileri verilir. Bu satırlar birey isimleri ile başlar


```

Title line: "Grape populations in southern France"
ADH Locus 1
ADH #2
ADH three
ADH-4
ADH-5
mtDNA
Pop
Grange des Peres , 0201 003003 0102 0302 1011 01
Grange des Peres , 0202 003001 0102 0303 1111 02
Grange des Peres , 0102 004001 0202 0102 1010 01
Grange des Peres , 0103 002002 0101 0202 1011 02
Grange des Peres , 0203 002004 0101 0102 1010 01
POP
Tertre Roteboeuf , 0102 002002 0201 0405 0807 01
Tertre Roteboeuf , 0102 002001 0201 0405 0307 01
Tertre Roteboeuf , 0201 002003 0101 0505 0402 01
Tertre Roteboeuf , 0201 003003 0301 0303 0603 01
Tertre Roteboeuf , 0101 002001 0301 0505 0807 01
pop
Bonneau 01 , 0101 002002 0304 0805 0304 01
Bonneau 02 , 0201 002002 0404 0505 0304 02
Bonneau 03 , 0101 002100 0304 0505 0101 01
Bonneau 04 , 0101 100100 0204 0805 0304 01
Bonneau 05 , 0101 100002 0104 0808 0304 03
Pop
, 0000 002001 0202 0402 0007 01
, 0200 002001 0202 0205 0707 01
, 0010 002001 0101 0105 0807 01

```

Resim 1.28: GENETPOP Dosya Formatı Görünümü (Boşluk) [42]

Genotip verilerde lokus içindeki "0102" verisi 4 tane rakama (digit) sahip olduğu için *diploitt* olarak kabul edilir. İlk allel "01", ikinci allel "02"dir. Eğer lokus içindeki bu veri 4 rakamdan küçük, "01" ya da "010" ise haploittir ve tek allel olarak kabul edilir. 2 rakam (digit) alleller, "01-99" arasında numaralandırılırken 3 rakam (digit) alleller, "001-999" arası numaralandırılırlar. Aynı GENETPOP dosyasında, genotip veriler içinde, SSR/MICROSAT dosya tiplerinde, haploitt ve diploitt veriler birlikte (mixed) kullanılabilir [26, 31].

```

Title line: "Grape populations in southern France"
ADH Locus 1
ADH #2
ADH three
ADH-4
ADH-5
mtDNA
Pop
Grange des Peres , 0201 003003 0102 0302 1011 01
Grange des Peres , 0202 003001 0102 0303 1111 02
Grange des Peres , 0102 004001 0202 0102 1010 01
Grange des Peres , 0103 002002 0101 0202 1011 02
Grange des Peres , 0203 002004 0101 0102 1010 01
POP
Tertre Roteboeuf , 0102 002002 0201 0405 0807 01
Tertre Roteboeuf , 0102 002001 0201 0405 0307 01
Tertre Roteboeuf , 0201 002003 0101 0505 0402 01
Tertre Roteboeuf , 0201 003003 0301 0303 0603 01
Tertre Roteboeuf , 0101 002001 0301 0505 0807 01
pop
Bonneau 01 , 0101 002002 0304 0805 0304 01
Bonneau 02 , 0201 002002 0404 0505 0304 02
Bonneau 03 , 0101 002100 0304 0505 0101 01
Bonneau 04 , 0101 100100 0204 0805 0304 01
Bonneau 05 , 0101 100002 0104 0808 0304 03
Pop
, 0000 002001 0202 0402 0007 01
, 0200 002001 0202 0205 0707 01
, 0010 002001 0101 0105 0807 01

```



Resim 1.29: GENEPOP Dosya Formatı (Mixed) Görünümü [42]

Genotip veriler içindeki eksik veriler (missing data), 2 rakam (digit) allellerde 00 ile gösterilirken 3 rakam (digit) allellerde 000 ile gösterilirler. GENEPOP dosyası içinde boş satır olamaz. Dosya sonunda ise birden fazla boş satır bulunmamalıdır [26, 27, 31].


```

Title line: "Grape populations in southern France"
ADH Locus 1
ADH #2
ADH three
ADH-4
ADH-5
mtDNA
Pop
Grange des Peres , 0201 003003 0102 0302 1011 01
Grange des Peres , 0202 003001 0102 0303 1111 02
Grange des Peres , 0102 004001 0202 0102 1010 01
Grange des Peres , 0103 002002 0101 0202 1011 02
Grange des Peres , 0203 002004 0101 0102 1010 01
POP
Tertre Roteboeuf , 0102 002002 0201 0405 0807 01
Tertre Roteboeuf , 0102 002001 0201 0405 0307 01
Tertre Roteboeuf , 0201 002003 0101 0505 0402 01
Tertre Roteboeuf , 0201 003003 0301 0303 0603 01
Tertre Roteboeuf , 0101 002001 0301 0505 0807 01
pop
Bon , 0304 0805 0304 01
Bon , 0404 0505 0304 02
Bon , 0304 0505 0101 01
Bonnet , 0204 0805 0304 01
Bonnet 05 , 0101 100002 0104 0808 0304 03
Pop
, 0000 002001 0202 0402 0007 01
, 0200 002001 0202 0205 0707 01
, 0010 002001 0101 0105 0807 01

```

Eksik veri
(Missing data) gösterimi

Resim 1.30: GENEPOP Dosya Formatı Eksik Veri (Missing Data) Görünümü [42]

GENEPOP dosya uzantıları ya “.txt” uzantılıdır ya da dosya uzantısına sahip değildir. Ancak dosya uzantısına sahip olmayanlar ASCII metin dosyaları olmak zorundadır. GENEPOP dosya formatları içerisinde DNA, AFLP, SSR/MICROSAT ve SNP dosya tiplerinden herhangi birisi bulunabilir [31]. Bu çalışmada GENEPOP dosya formatının çevrileceği dosya formatları STRUCTURE ve ARLEQUIN dosya formatlarıdır.

1.4. Depolanan Moleküler Veri Çeşitleri

Eldeki verinin çeşidi, kullanılan laboratuvar tekniğine bağlı olarak ve bu verinin dijital kodlama tercihine bağlı olarak değişkenlik gösterir. Dolayısıyla eldeki bir dosyanın içindeki verilerin (örneğin nükleotid, aminoasit, allel frekansı vb.) nasıl bir veri tipi olduğu da bu verilerin analizinde ve kodlamasında önemlidir. Bu yüzden veri çeşitlerini de tıpkı dosya türleri gibi burada sunmak önem arz etmektedir [31].

1.4.1. DNA ve RNA Sekans Verileri

Bu veriler herhangi bir DNA veya RNA parçasına ait dizilerin bilgilerinin bulunduğu verilerdir. Bunlardan DNA sekans verisi içeren dosya tiplerinde nükleotidleri belirtmek için A, C, G, T, R, Y, M, K, S, W, H, B, V, D, N harfleri kullanılır. RNA dosya tiplerinde ise nükleotidleri belirtmek için ise A, C, G, U, R, Y, M, K, S, W, H, B, V, D, N harfleri kullanılır. Genellikle soru işareti (?) belirsiz nükleotidler, tire işareti (-) ise INDEL (insersiyon/delesyon) nükleotidleri belirtmek için kullanılır [21, 34].

```
#NEXUS
BEGIN TAXA;
  Dimensions NTax=4;
  TaxLabels fish frog snake mouse;
END;

BEGIN CHARACTERS;
  Dimensions NChar=20;
  Format DataType=DNA;
  Matrix
    fish  ACATA GAGGG TACCT CTAAG
    frog  ACATA GAGGG TACCT CTAAG
    snake ACATA GAGGG TACCT CTAAG
    mouse ACATA GAGGG TACCT CTAAG
END;

BEGIN TREES;
  Tree best=(fish, (frog, (snake, mouse)));
END;
```

Resim 1.31: Örnek DNA Veri Tipi [33]

1.4.2. Protein

Protein dosya tiplerinde, bir proteini oluşturan aminoasitlerin dizi bilgisi bulunur ve bu aminoasitleri belirtmek için uluslararası standartlar olan A, R, N, D, C, Q, E, G, H, I, L, K, M, F, P, S, T, W, Y, V harflerinin kombinasyonu kullanılır [34].

```
>gi|5524211|gb|AAD44166.1| cytochrome b [Elephas maximus maximus]
LCLYTHIGRNIYYGSYLYSETWNTGIMLLITMATAFMGYVLPWGQMSFWGATVITNLFSAIPYIGTNLV
EWIWGGFSDVKATLNRFFAFHFILPFTMVALAGVHLTFLHETGSNNPLGLTSDSDKIPFHPPYTIKDFLG
LLILLLLLLLLALLSPDMLGDPDNHMPADPLNTPLHIKPEWYFLFAYAILRSVPNKLGGVLALFLSIVIL
GLMPFLHTSKHRSMMLRPLSQALFWTLTMDLLTLTWIGSQPVEYPYTIIGQMASILYFSIILAFPLIAGX
IENY
```

Resim 1.32: Örnek Protein Veri Tipi [31]

1.4.3. SSR | MICROSAT (Basit Dizi Tekrarları)

SSR (Simple Sequence Repeats) canlı DNA'sında bulunan tekrar bölgelerinin dizilerindeki tekrar sayısına dayanan bir moleküler markör sistemidir. Tür içi çeşitlilik bakımından zengin olması, laboratuvarında uygulama kolaylığı ve genom içerisinde bolca bulunmasından dolayı geçen 20 yılda en yaygın kullanılan moleküler verilerden birisi olmuştur. Yerini son zamanlarda yaygınlaşan sekans temelli SNP markörlerine bırakma eğiliminde olsa da halen bu veri çeşidi ile çok sayıda çalışma yapılmaktadır. Bu veri çeşidi ko-dominant (diploit) olup genellikle tekrar sayısı ya da içerdiği DNA bant uzunluğu olarak kodlanır ve analizlerde kullanılır [37, 38].

```
[Profile]
Title="A small example of microsatellite data"
NbSamples=4
GenotypicData=1
#Unknown gametic phase between the 2 loci
GameticPhase=0
DataType=MICROSAT
LocusSeparator=WHITESPACE
[Data]
[[Samples]]
SampleName="MICR1"
SampleSize=28
SampleData= {
  Genot1    27    12 23 17
             13 22 16
  Genot2     1    15 22 16
             13 22 16
}
SampleName="MICR2"
SampleSize=59
SampleData= {
  Genot3    37    12 24 18
             12 22 16
  Genot4     1    15 20 18
             13 22 18
  Genot5    21    14 22 16
             14 23 16
}
SampleName="MICR3"
SampleSize=30
SampleData= {
  Genot6    17    12 21 16
             13 22 15
  Genot7     1    12 20 16
             13 23 16
  Genot8    12    10 22 15
             12 22 13
}
SampleName="MICR4"
SampleSize=16
SampleData= {
  Genot9    15    13 24 16
             13 23 17
  Genot10   1     12 24 16
             13 23 16
}
[[Structure]]
StructureName="Test microsat structure"
NbGroups=2
#The first group is made up of the first 2 samples
Group={
  "MICR1"
  "MICR2"
}
#The last 2 samples will be put into the second group
Group={
  "MICR3"
  "MICR4"
}
}
```

Resim 1.33: Örnek SSR/MICROSAT Veri Tipi [21]

1.4.4. SNP (Tek Nükleotit Polimorfizmi)

SNP (Single Nucleotid Polymorphism) markörleri tek nükleotid polimorfizmine dayalı bir markör olup, genomdaki belirli bir konumda meydana gelen tek bir nükleotid değişimidir. Tüm popülasyon göz önüne alındığında genom dizisi boyunca yüksek oranda aynı şifreyi taşır. Sadece genomun <%1 daha az kısmına tekabül eder. İnsanlarda hastalık genlerinin bulunması ya da zirai açıdan önemli karakterleri kontrol eden genlerin bulunması amacıyla DNA üzerindeki bu polimorfizm özel kodlar haline getirilerek kullanılır. Tek nükleotid varyantı (SNV), herhangi bir frekans sınırlaması olmaksızın tek bir nükleotide bir varyasyon olup, tek nükleotidli bir değişiklik olarak da bilinir [35, 36].

rs#	alleles	chrom	pos	strand	assembly#	center	protLSID	assayLSID	panelLSID	QCcode	103A	103C	103D	103K	105A	105B	105D																							
TF134	G/A	9	134	+	NA	NA	NA	NA	NA	N	H	H	H	G	A	N	R	G	G	H	R	H	R	R	G	G	M	N	R	G										
TF164	G/A	9	164	+	NA	NA	NA	NA	NA	G	S	N	G	N	M	N	R	A	H	N	G	G	M	N	N	N	G	M	G	G	G	M	N	R	G					
TF202	T/G	9	202	+	NA	NA	NA	NA	NA	H	H	H	T	H	N	M	I	M	N	M	T	H	H	H	N	N	H	H	M	M	M	M	T	H	N					
TF241	C/G	9	241	+	NA	NA	NA	NA	NA	C	C	C	C	C	C	C	C	S	H	H	N	C	C	N	C	C	C	C	C	C	C	C	C	H	C	H	H			
TF315	C/A	9	315	+	NA	NA	NA	NA	NA	N	C	N	H	A	H	C	N	C	N	C	N	N	N	N	C	H	H	C	C	C	C	C	C	N	M	N	A	C		
TF345	C/T	9	345	+	NA	NA	NA	NA	NA	N	N	Y	Y	C	C	C	C	C	C	N	C	N	H	H	H	C	C	N	N	C	C	C	C	C	C	N	R	C	C	
TF353	A/G	9	353	+	NA	NA	NA	NA	NA	H	A	A	A	A	A	A	A	A	A	N	M	N	M	H	H	H	N	H	N	N	N	N	N	N	N	N	A	A		
TF355	C/T	9	355	+	NA	NA	NA	NA	NA	N	M	T	N	M	N	C	C	C	Y	H	N	H	H	H	M	N	N	N	N	N	N	N	N	N	T	H	H	H		
TF375	G/C	9	375	+	NA	NA	NA	NA	NA	G	H	H	H	H	H	H	H	N	N	S	S	N	N	N	N	H	C	S	H	N	C	G	G	M	N	N	N			
TF381	A/C	9	381	+	NA	NA	NA	NA	NA	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A		
TF390	C/T	9	390	+	NA	NA	NA	NA	NA	N	H	H	H	H	Y	N	N	C	C	N	N	N	C	C	Y	H	H	H	N	N	C	C	C	C	C	C	C	C	N	
TF405	C/A	9	405	+	NA	NA	NA	NA	NA	N	M	M	M	A	N	C	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	
TF484	T/A	9	484	+	NA	NA	NA	NA	NA	N	T	T	T	T	H	H	N	H	T	T	M	T	A	N	T	N	N	N	N	N	N	N	N	N	N	N	N	N	N	
TF495	T/C	9	495	+	NA	NA	NA	NA	NA	H	T	N	H	T	N	T	N	T	N	T	N	T	H	T	H	N	H	N	Y	C	N	T	N	N	N	N	N	N	N	
TF550	T/A	9	550	+	NA	NA	NA	NA	NA	N	T	N	N	N	T	N	N	N	N	N	T	T	N	N	T	T	N	N	N	N	N	N	N	N	N	N	N	N	N	
TF573	A/T	9	573	+	NA	NA	NA	NA	NA	N	H	H	H	H	N	N	A	N	M	N	N	N	N	N	N	N	N	N	A	A	A	A	A	A	A	A	A	A	N	
TF574	G/A	9	574	+	NA	NA	NA	NA	NA	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	
TF605	T/C	9	605	+	NA	NA	NA	NA	NA	N	T	N	T	T	T	N	T	T	T	N	T	T	N	T	T	N	N	T	N	N	N	N	N	N	N	N	N	N	N	
TF626	G/A	9	626	+	NA	NA	NA	NA	NA	R	R	A	A	A	A	A	A	A	N	N	N	N	R	G	G	N	G	G	R	R	N	H	H	H	H	H	H	H	H	
TF653	A/T	9	653	+	NA	NA	NA	NA	NA	T	T	N	H	A	T	A	N	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
TF659	A/G	9	659	+	NA	NA	NA	NA	NA	N	A	M	M	M	A	A	N	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	
TF719	G/A	9	719	+	NA	NA	NA	NA	NA	H	H	H	G	G	H	N	H	G	N	M	M	M	M	G	H	H	H	G	H	G	M	M	G	G	G	G	G	G	G	
TF735	T/G	9	735	+	NA	NA	NA	NA	NA	T	T	G	N	N	T	N	N	N	N	N	N	T	H	T	H	T	N	N	N	N	N	N	N	N	N	N	N	N	N	
TF754	G/A	9	754	+	NA	NA	NA	NA	NA	N	A	N	R	H	H	N	A	N	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	
TF800	C/T	9	800	+	NA	NA	NA	NA	NA	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	
TF819	C/T	9	819	+	NA	NA	NA	NA	NA	N	C	N	N	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	
TF835	T/A	9	835	+	NA	NA	NA	NA	NA	N	M	N	N	H	N	T	I	T	T	N	M	I	T	T	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	
TF910	C/T	9	910	+	NA	NA	NA	NA	NA	N	M	N	C	T	N	I	N	H	T	Y	I	T	N	N	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	
TF921	G/T	9	921	+	NA	NA	NA	NA	NA	G	T	G	N	N	N	T	N	N	N	N	N	N	N	H	H	H	H	H	N	N	N	N	N	N	N	N	N	N	N	
TF945	A/G	9	945	+	NA	NA	NA	NA	NA	N	M	M	M	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	
TF950	C/G	9	950	+	NA	NA	NA	NA	NA	C	C	N	C	H	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C		
TF953	C/G	9	953	+	NA	NA	NA	NA	NA	N	M	M	C	N	C	N	C	A	H	H	H	H	N	C	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	
TF1026	A/G	9	1026	+	NA	NA	NA	NA	NA	N	H	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	
TF1037	A/G	9	1037	+	NA	NA	NA	NA	NA	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	
TF1090	A/G	9	1090	+	NA	NA	NA	NA	NA	H	H	H	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	

Resim 1.34: Örnek SNP Veri Tipi

1.4.5. RFLP (Restriksiyon Para Uzunluk Polimorfizmi)

RFLP (Restriction Fragment Length Polymorphism) ilk geliřtirilip kullanılan moleküler markör tiplerindedir. DNA'nın enzimlerle kesilmesinden sonra enzim kesme bölgesinin bireyler arasında farklılık göstermesinden dolayı aynı türün ya da farklı türe ait bireylerin DNA'sının farklı uzunluklarda kesilmesi ve sonuç olarak jelde yürütüldüğünde farklı uzunluklarda DNA fragmentleri oluşturması esasına dayanır. Her restriksiyon alanı farklı bir lokus olarak düşünüldüğünde restriksiyon alanlarının varlıkları "1" olarak kodlanırken yoklukları "0" olarak kodlanır. Dolayısıyla dominant bir markördür. Nokta mutasyonun yokluğundan dolayı değil, restriksiyon alanının silindiğini belirtmek için tire işareti (-) kullanılmaktadır [21, 39, 40].


```

[Profile]
  Title="A small example of RFLP data: 3 populations"
  NbSamples=3
  GenotypicData=0
  DataType=RFLP
  LocusSeparator=WHITESPACE
  #We tell Arlequin to compute Euclidian square distances between
  #the haplotypes listed below
  MissingData='?'
[Data]
  [[HaplotypeDefinition]]
    HaplListName="A fictive list of RFLP haplotypes"
    HaplList= {
      1  000011100111010011011001001011001101110100101101100
      2  100011100111010011011001001011001101110100101100100
      6  000011100111010010011001001011001101110100101101100
      7  100011100111010011011001001011001101110100101101100
      8  000011100111010011011001001001001101110100101101100
      11 000001100111011011011001001011001101110100101111100
      12 000011100111010011011001101011001101110100101101100
      17 000011100111010011011001001011001100110100101101100
      22 000011100111011011011001001011001101110100101100100
      36 000011100111010011011001001010001100110100101101100
      37 000011100111011011011001001111001101110100101100100
      38 000111100111010011011001001011001101110100101101100
      40 000011100111000011011001001011001101110100101101100
      47 000011100111010011011001001011001101110100101100100
      139 000011100111010011011001001011001111110100101000110
      140 000011100111010011011001001011001101110100101100101
      141 000011100111010010011001000011001101110100101100100
    }
  [[Samples]]
    #1
    SampleName="pop 1"
    SampleSize=28
    SampleData= {
      1  27
      40  1
    }
    #2
    SampleName="pop 2"
    SampleSize=75
    SampleData= {
      1  37
      17  1
      6  21
      7  1
      2  1
      22  5
      11  2
      36  1
      139  1
      47  1
      140  1
      141  1
      37  1
      38  1
    }
    #3
    SampleName="pop 3"
    SampleSize=48
    SampleData= {
      1  46
      8  1
      12  1
    }
  [[Structure]]
    StructureName="A single group of 3 samples"
    NbGroups=1
    Group={
      "pop 1"
      "pop 2"
      "pop 3"
    }

```

Resim 1.35: Örnek RFLP Veri Tipi [21]

1.4.6. AFLP (Çoğaltılmış Parça Uzunluğu Polimorfizmi)

AFLP (Amplified Fragment Length Polymorphism) temelde RFLP markörleri ile benzer bir mantık ile üretilmektedir. Ancak bu sefer DNA bölgelerinin rastgele kesilmesinden ziyade primerler aracılığıyla rastgele çoğaltılması yani amplifikasyonu esasına dayanır. AFLP dosya tiplerinde, lokuslarda bant mevcutsa artı işareti (+), mevcut değilse eksi işareti (-) ile lokus belirtilmemişse sıfır rakamı (0) ile gösterilir [40, 41].

```
NumIndivs 60
NumLoci 9
Digits 1
Format Lumped

LocusNames      m1 m2 m3 m4 A1 A2 A3 A4 A5
1      11  12  13  11  +  +  +  -  +
2      22  33  11  22  -  -  0  -  -
3      12  13  13  11  +  -  -  -  +
```

Resim 1.36: Örnek 4 SSR/MICROSAT, 5 AFLP Lokusuna Sahip AFLP Veri Tipi [32]

1.4.7. Frequency (Allel Frekans Verisi)

Allel frekansı verisi genelde bireylerin heterojen ve heterozigot olduğu popülasyonlarda, her bir popülasyonun bir markör ya da bir lokusun allellerinin frekanslarının sayısal olarak kodlandığı bir formattır. Genomik analiz göz önüne alındığında kısmen işlenmiş bir moleküler veri olarak kabul edilebilir.

```
[Profile]
  Title="Frequency data"
  NbSamples=2
  GenotypicData=0
  DataType=FREQUENCY
[Data]
  [[Samples]]
    SampleName="Population 1"
    SampleSize=16
    SampleData= {
      000 1
      001 3
      002 1
      003 7
      004 4
    }
    SampleName="Population 2"
    SampleSize=23
    SampleData= {
      000 3
      001 6
      002 2
      003 8
      004 4
    }
  }
```

Resim 1.37: Örnek Frequency Veri Tipi [21]

1.4.8. Distance (Genetik Farklılık)

Moleküler markörlerin işlenmesi sonucu her bir bireyin bir diğer birey ile yakınlığının standartlaştırılması esasına dayanan bir veri türüdür. Genomik analizlerde sıkça kullanılmaktadır. Bu veri türü de tıpkı frekans verisi gibi moleküler markörlerin işlenmesi ile oluşmuş bir veridir. Bazen analiz sonucu ve son veri gibi raporlansa da özellikle ilişkilendirme haritalamaları gibi birçok genomik analizde popülasyon yapısının kontrolü amacıyla veri analizine yeniden dahil edilmektedir.

```
#mega
!Title: Concatenated Files;
!Format DataType=Distance DataFormat=LowerLeft NTaxa=6;

#Rodent
#Primate
#Lagomorpha
#Artiodactyla
#Carnivora
#Perissodactyla

0.514
0.535 0.436
0.530 0.388 0.418
0.521 0.353 0.417 0.345
0.500 0.331 0.402 0.327 0.349
```

Resim 1.38: Örnek Distance Veri Tipi [23]

2. MATERYAL VE YÖNTEM

Geliştirilen programın, platformlardan bağımsız ve donanım kısıtlı olmadan çalışabilmesi için program web tabanlı hazırlanmıştır. Web tabanlı çalışacak olan program, popülasyon genetiği veri dosyalarını, girilen doğru bilgilere göre istenilen yeni bir popülasyon genetiği veri formatına çevirebilmektedir. Aynı zamanda bu program, araştırmacılardan güncelleme istememekte, güncelleme olsa da araştırmacıya yansıtılmamaktadır. Web tabanlı bir program olduğu için, programın dağıtılması ya da kullanılması ile ilgili bir lisans problemi yaşanmayacaktır. Web tabanlı bir yazılımın hazırlanması için çeşitli alternatifler mevcutken öncelikle yazılacak programın çevirme işlemlerinin client (kullanıcı) ya da server (sunucu) tarafında mı çalışacağına karar vermek için her iki şekilde de performans ve güvenlik testleri yapılarak uygulamanın yazılacağı script ya da programlama diline karar verilmiştir. Sunucu tarafında çalışacak bir program için JAVA, PHP ya da ASP gibi serverside (sunucu tarafı) çalışan diller araştırılmıştır. Client(istemci-kullanıcı) tarafında çalışacak bir program için ise JQuery (javascript kütüphanesi) incelenmiştir. Tüm incelemelerden sonra programın sunucu tarafında PHP, client tarafında ise JQuery kütüphanesi tercih edilmiştir. Formatların birbiri arasında çevrilmelerini sağlamak için dosyaların içindeki verilerin yapılarını iyi bir şekilde analiz ederek aralarındaki farkları detaylı araştırılmıştır. Tüm bu kuralların, isteklerin ve verilerin sunucuya iletilmesi için JQuery kütüphanesi kullanılmıştır. Son olarak gelen verileri kullanarak sunucu tarafında programlama için seçtiğimiz uygun dil olan PHP'nin string (yazı) işleme, parçalama, düzenli ifadeler (regular expression) ve kendi özel fonksiyonlarımız yardımıyla bir formattaki veri başka bir formata rahat bir şekilde çevrilebilmektedir. Her dosya formatı için farklı dosya okuma kodları kullanılmaktadır.

Dosya formatlarının hızlı ve başarılı bir şekilde birbirine çevrilmesi aşamasında HTML, CSS, JQuery, PHP ve Linux tabanlı sınırsız trafik sağlayan bir sunucu tercih edilmiştir. Bu şekilde aynı anda binlerce araştırmacının formatlarını sorunsuz bir şekilde çevirmelerine imkân sağlanmaktadır.

Arayüz üzerinde girdi verinin haploit ya da diploit değerinin seçilebilmesi gibi birçok özellik için HTML ve JQuery kodları beraber kullanılmaktadır. Arayüz üzerinden girdi dosyasının ve çıktı dosyasının özellikleri çeşitli veri giriş ve seçim özellikleri ile alınıp veri çevirme fonksiyonlarının oldukları sayfalara bilgi olarak gönderilir.

```
<div class="col-sm-12">
  <!-- Start: ploidy_of_the_data <select> -->
  <div class="form-group">
    <label for="ploidy_of_the_data" class="p-label-required">What is the ploidy of the data?</label>
    <label class="input-group p-has-icon p-custom-arrow">
      <select id="ploidy_of_the_data" name="ploidy_of_the_data" required="required" class="form-control">
        <option class="p-select-default" value="haploid">haploid</option>
        <option value="diploid_on_one_row">diploid(on one row)</option>
        <option value="diploid_on_two_consecutive_row">diploid(on two consecutive row)</option>
      </select>
      <span class="input-group-state">
        <span class="p-position">
          <span class="p-text">
            <span class="p-required-text"><i class="fa fa-star"></i></span>
          </span>
        </span>
      </span>
      <span class="p-field-cb"></span>
      <span class="p-select-arrow"><i class="fa fa-caret-down"></i></span>
      <span class="input-group-icon"><i class="fa fa-file-text-o"></i></span>
    </label>
  </div>
  <!-- End: ploidy_of_the_data <select> -->
</div>
```

Resim 2.1: Örnek HTML Kodları

Jquery kodları kullanılarak arayüz üzerindeki formlardan alınan bilgiler sunucu üzerindeki php sayfasına gönderilerek dosya formatları içindeki bilgiler satır satır okunup sonuçları tekrar arayüz sayfasına bilgi olarak geri gönderilir. Eğer dosya yazdırma işlemi başarılı olursa çıktı dosyasının indirme linki sayfaya eklenerek kullanıcıların dosyayı indirmeleri sağlanır.

```
/**
 * id değeri {start_converting} olan butona tıklanınca - @start_converting click
 * Sunucuya tüm bilgiler ajax metodu ile gönderilir,
 * index.php sayfasında ki çevirme işlemlerinden sonra geriye response gönderilir,
 * gelen response değeri sayfaya eklenir.
 */
$('#start_converting').click(function(){
$.ajax
({
type : "POST",
url : 'php/index.php',
data : $('#formum').serialize()+ '&file_name=' + file_name,
beforeSend: function(){
$('#loader_step5').css({'display':'inline'});
},
success :function(response)
{
$('#section_step5').html(response);
$('#loader_step5') .css({'display':'none'});
},
error:function(ma,ydin)
{
if (ma.status == 0) {
alert('Not connected.\nPlease verify your network connection. [0]');
} else if (ma.status == 400) {
alert('Server understood the request, but request content was invalid. [400]');
} else if (ma.status == 401) {
alert('Unauthorized access. [401]');
} else if (ma.status == 403) {
alert('Forbidden resource can\'t be accessed. [403]');
} else if (ma.status == 404) {
alert('Requested page not found. [404]');
} else if (ma.status == 500) {
alert('Internal server error [500]');
} else if (ma.status == 503) {
alert('Service unavailable. ');
} else if (ydin == 'parsererror') {
alert('Error.\nParsing JSON Request failed. ');
} else if (ydin == 'timeout') {
alert('Request Time out. ');
} else if (ydin == 'abort') {
alert('Request was aborted by the server. ');
} else {
alert('Unknown Error \n' + ma.responseText);
}
}
}
});
```

Resim 2.2: JQuery -Ajax Metodu ile Sunucuya Bilgilerin Gönderilmesi

Arayüz üzerinden gelen çeşitli bilgiler doğrultusunda programa yüklenen girdi dosyası da dikkate alınarak PHP kodları ile bu dosya kendi kurallarına göre satır satır okunup, elde edilen bilgiler çeşitli fonksiyonlara parametre olarak gönderilir.

```
<?php
/*
 * Sunucu üzerinde, belirtilen dizine yüklenen input dosyasının belirtilmesi
 */
$file           = $_SERVER['DOCUMENT_ROOT'].'/converter/uploads/'.$file_name;
$read_file      = file($file);
/*
 * Dosyayı satır satır okumaya başlayacak
 */
foreach($read_file as $r => $line)
{
    if(preg_match('/^>/i',$line))
    {
        /*
         * Dosyanın okunan satırı ">" karakteri ile başlıyorsa
         * burada ki diğer kodlar çalışacaktır.
         */
    }
    else{
        /*
         * Dosyanın okunan satırı ">" karakteri ile başlamiyorsa
         * burada ki diğer kodlar çalışacaktır.
         */
    }
}
?>
```

Resim 2.3: PHP Girdi Dosyası Okuma Kodları

Dosya okunduktan sonra PHP dili kullanılarak popülasyon isimleri, popülasyon içindeki birey isimleri, bireylerin verileri ve dosya ile ilgili gerekli diğer bilgiler sanal olarak program içinde dizilere aktarılmaktadır. Daha sonra istenilen bir formatta çıktı almak için aşağıdaki gibi dosya yazdırma kodları kullanılmıştır. Her format için farklı dosya yazma kodları ve kuralları kullanılmıştır.

```
<?= "#mega" ?>
<?= "\r\n" ?>
<?= "!TITLE: ".$dizi['header']['Title']; ?>
<?= "\r\n" ?>
<?= "!Format DataType=".$dataType; ?>
<?= "\r\n" ?>
<?= "\r\n" ?>
<?php foreach ($dizi['Data'] as $key=>$value): ?>
<?php foreach ($value['PopData'] as $k=>$v): ?>
<?= "#".$v['individual'];?>
<?= "\r\n" ?>
<?= $v['data']['dataList'];?>
<?= "\r\n" ?>
<?php endforeach;?>
<?php endforeach;?>
```

Resim 2.4: PHP Çıktı Dosyası Yazma Kodları

3. BULGULAR

Programın kullanıcılar tarafından rahatlıkla kullanılabilmesi, dosyalarını sistem üzerine kolayca yükleyip ve sistemden anında indirebilmesi için yeni bir kullanıcı arayüzü tasarlanmıştır. Kullanıcı arayüzü üzerinde giriş verileri için girdi ve çıkış bölümleri için çıktı bölümü yer almaktadır. Platformun kullanıcı havuzu göz önüne alındığında programın dünyanın her yerinden araştırmacıların kullanımına sunulduğu açıktır. Bu sebeple program arayüzü İngilizce olarak hazırlanmıştır. Arayüzdeki bölümlerin içinde, dosya formatları, dosya tipleri seçilimi ve dosya içindeki verilerin özellikleri kullanıcı tarafından seçilebilmektedir. Yine kullanıcı arayüzü içerisinde ayrı ayrı bölümlerde girdi dosyaları için yükleme (upload) ve çıktı dosyaları için indirme (download) bölümleri yer almaktadır. Bunun için HTML, CSS, Javascript, JQuery gibi web dilleri, kütüphaneler ve çeşitli eklentiler kullanılmıştır. Kullanıcılar dosya formatlarını çevirmek istediklerinde aşağıdaki 5 adımdan sonra dosya indirme sayfasına ulaşabilirler.

İlk adımda kullanıcılar girdi dosya formatlarını, çıktı dosya formatlarını ve çevrilmek istenen veri tipini seçebilirler. İlk sayfada bütün alanlar zorunludur.

The screenshot displays a five-step process bar at the top, with Step 1 highlighted. Below the bar, a title bar reads "Choose Input, Output File Formats And File Type". The main content area contains three numbered dropdown menus:

- 1 Input File Format *: Structure
- 2 Output File Format *: Genpop
- 3 Specify which data type should be included in the Output File *: MICROSAT

A red "confirm" button with a checkmark icon is located at the bottom right of the form.

Resim 3.1: Kullanıcı Arayüzü Adım-1

İkinci adımda çevirmek istedikleri girdi dosyalarını programa yükleyebilirler. Maksimum yüklenebilir dosya boyutu, sunucunun maksimum dosya yükleme (upload) boyutuna göre değişmektedir. Kullanıcılar dosyalarını Browse File (dosya seç) butonuna tıklayarak bilgisayarlarından seçebildikleri gibi, sürükle bırak (drag and drop) yöntemi ile de programa yükleyebilirler. Sadece 1 adet girdi dosyası yüklenebilir, dosya değiştirilmek istenirse önce yüklenen dosya silinip yeniden dosya yüklenmelidir. Kullanıcının seçtiği girdi formatı hangi dosya uzantılarını destekliyorsa, program sadece bu dosya uzantılarının yüklenmesine izin vermektedir. Dosya yükleme işlemi sırasında confirm butonu pasif hale geçmektedir. Eğer dosya yükleme işlemi başarılı bir şekilde tamamlandı ise confirm butonuna tıklayarak sonraki adıma geçilebilir.

Step 1 Step 2 Step 3 Step 4 Step 5

Upload Input File

Input File: *

Drag&Drop input file here

or

Browse File

Structure_input_data.txt
size: 4.62 MB type: txt
Success

back confirm

Resim 3.2: Kullanıcı Arayüzü Adım-2

Üçüncü adımda kullanıcılar, girdi dosyasındaki verilerin özelliklerini seçebilmekte ya da el ile girebilmektedirler Bütün zorunlu alanlarda ki bilgiler doğru girildikten sonra confirm butonuna tıklayarak sonraki adıma geçilebilir.

Step 1 — Step 2 — Step 3 — Step 4 — Step 5

STRUCTURE Input File details

What is the ploidy of the data? *

*

Is the "Phase information" row present? *

*

What is the missing value code(-9,-999...)? *

*

Type of the data *

*

How are Microsat alleles coded? in case microsat data type *

*

Enter the size of the repeated motif(same for all loci:different comma separated list)(e.g.2,2,3,2). *

*

Are marker (locus) names included? *

*

Enter the number of markers (loci) listed in the input file.

*

Are individual name (labels) included in the input file? *

*

Is the "PopData" column (population identifier) present in the input file? *

*

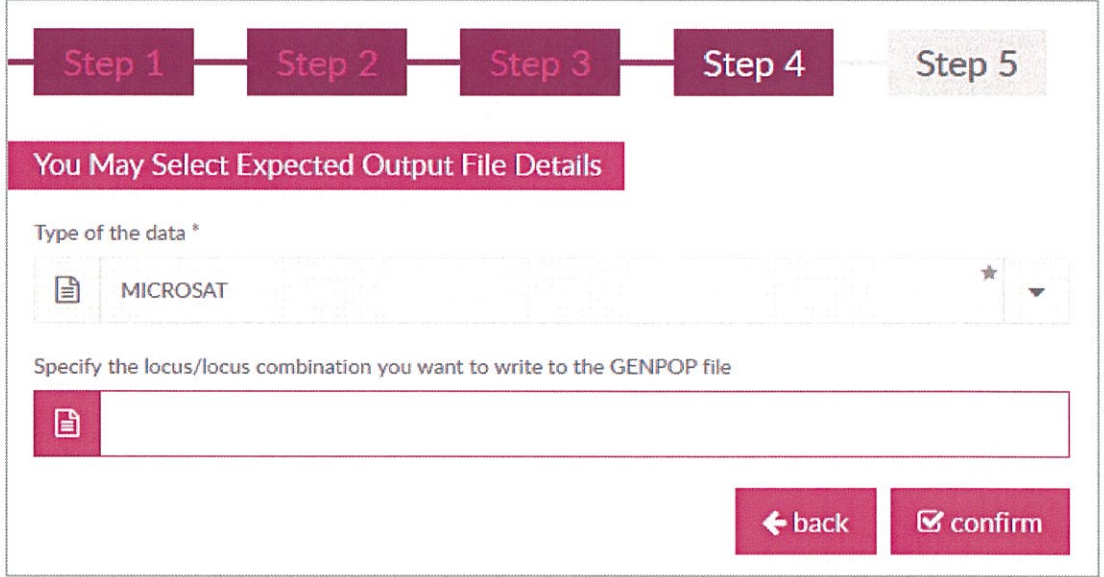
Are the "Recessive Alleles" row and/or the "Inter-Marker-Distance" row present in the input file? *

*

← back confirm

Resim 3.3: Kullanıcı Arayüzü Adım-3

Dördüncü adımda çıktı dosyasındaki verilerin nasıl yazılacağını belirlemek için gerekli bilgilerin doldurulması gerekmektedir. Bütün zorunlu alanlar doldurulduktan sonra confirm butonuna tıklayarak sonraki adıma geçilebilir.



Step 1 Step 2 Step 3 Step 4 Step 5

You May Select Expected Output File Details

Type of the data *

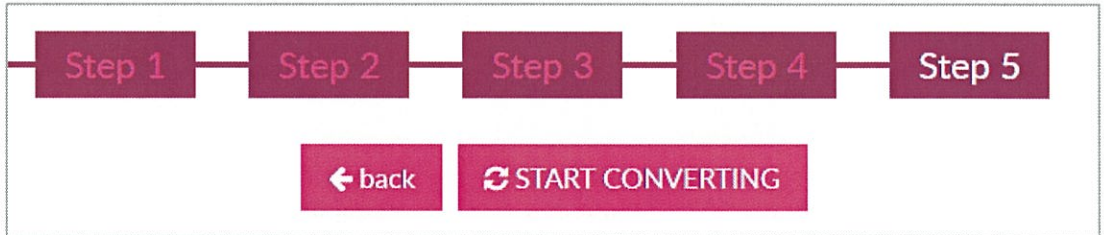
MICROSAT

Specify the locus/locus combination you want to write to the GENPOP file

back confirm

Resim 3.4: Kullanıcı Arayüzü Adım-4

Beşinci adımda girilen bütün bilgileri çıktı dosyasına yazdırmak için Start Converting (Çevirme işlemi başlat) butonuna tıklamak gerekmektedir. Bir süre sonra sayfa yenilenmeden çıktı dosyasının indirilebilmesi için Download (indir) butonu sayfaya eklenmektedir.

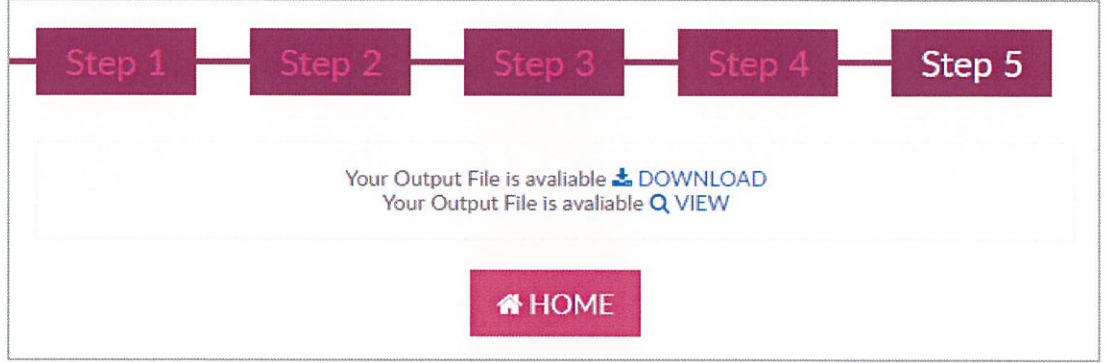


Step 1 Step 2 Step 3 Step 4 Step 5

back START CONVERTING

Resim 3.5: Kullanıcı Arayüzü Adım-5

Son olarak Download (indir) butonuna tıklanarak hazır hale gelen çıktı dosyası, kullanıcı bilgisayarına indirilebilmektedir. Eđer yeni bir çeviri işlemi yapılacaksa kullanıcı sayfayı yenileyebilir ya da Home butonuna basarak adım-1 sayfasına geri dönebilir.



Resim 3.6: Kullanıcı Arayüzü Dosya İndirme Adım-5

4. SONUÇLAR VE TARTIŞMA

Genomik bilgiler içeren veri dosyalarının yüksek format çeşitliliğine sahip olmasından dolayı araştırmacılar ya da çeşitli kurumlar tarafından veri formatlarını birbirine çevirebilecek çeşitli masaüstü yazılımlar geliştirilmesine karşın henüz kapsamlı ve çevrimiçi programlar mevcut değildir. Çevrimiçi hizmet veren programlar ise yeni çıkan diğer formatlara göre uyumlu olmadığı gibi önceki formatlar için gerekli güncel özellikleri eklememişlerdir.

Genomik verilerin çeşidi, kullanılan laboratuvar tekniğine göre ya da analiz edilen programların son çıktısına göre farklı şekillerde de kodlanmaktadır. Dolayısıyla genomik verilerin kodlanma şekilleri ya da veri tiplerine göre incelenmesi ve gruplandırılması yapılarak elde edilen tüm bilgilerin sonucunda dosya içindeki bu veriler belirli kurallar çerçevesinde çevrimiçi program tarafından önce okunup daha sonra yeni format kurallarına göre farklı bir formata çevrilmesi sağlanmıştır. Ayrıca en sık kullanılan dosya formatlarının çevrilmesi sağlanmış, yeni formatların kullanım sıklığı arttıkça bunlarında programa eklenmesi sağlanacaktır.

Tüm veri formatlarının web tabanlı tek bir sistem üzerinden birbirine çevrilmesi ile tüm dünyada araştırmacıların ihtiyaç duyduğu programı bulma, kurma ve çalıştırma süreçlerini bitirerek zaman kayıplarını önleyip araştırmalarında sonuca daha hızlı gitmeleri amaçlanmaktadır. Ayrıca sistem geliştirilerek araştırmacıların daha verimli çalışmaları sağlanıp ülkemiz için biyoinformatik alanında çevrimiçi (online-web tabanlı) yeni bir yazılım üretilmek istenmektedir.

5. KAYNAKLAR

- [1] Polat, M., ve Karahan, A. G. (2009). Multidisipliner yeni bir bilim dalı: biyoinformatik ve tıpta uygulamaları. *Medical Journal of Suleyman Demirel University*, 16(3), 41-50.
- [2] Kalaycıođlu A. T. (2013). Nükleotid Dizilerinin Aminoasit Formatına Dönüştürülmesi ve Dünya Veri Tabanlarındaki Verilerle Karşılaştırılması. *Kafkas Üniversitesi Veteriner Fakültesi Dergisi*, 19(2), 359-363.
- [3] http://yunus.hacettepe.edu.tr/~mergen/sunu/s_biyoinformatik.pdf, (27.11.2017).
- [4] Excoffier, L., and Heckel, G. (2006). Computer programs for population genetics data analysis: a survival guide. *Nature Reviews Genetics*, 7(10), 745-758.
- [5] <http://www.cmpg.unibe.ch/software/PGDSpider/>, (27.11.2017).
- [6] http://www.genomatix.de/online_help/help/sequence_formats.html, (27.11.2017).
- [7] http://www3.imperial.ac.uk/bioinfsupport/help/nuc_formats/, (27.11.2017).
- [8] <http://emboss.sourceforge.net/docs/themes/SequenceFormats.html>, (27.11.2017).
- [9] <https://zhanglab.ccmb.med.umich.edu/FASTA/>, (27.11.2017).
- [10] Lischer, H. E., and Excoffier, L. (2011). PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics*, 28(2), 298-299.
- [11] <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.192.3649&rep=rep1&type=pdf>, (27.11.2017).
- [12] <http://evolution.genetics.washington.edu/phylip/doc/distance.html>, (27.11.2017).
- [13] http://wiki.christophchamp.com/index.php/NEXUS_file_format, (27.11.2017).

- [14] http://phylogeny.lirmm.fr/phylo_cgi/documentation.cgi, (27.11.2017).
- [15] https://en.wikipedia.org/wiki/FASTA_format, (27.11.2017).
- [16] <http://bioinformatics.psb.ugent.be/webtools/ForCon/manual/>, (27.11.2017).
- [17] <http://evomics.org/resources/file-formats/PHYLIP/PHYLIP-protein/>, (27.11.2017).
- [18] <https://www.hiv.lanl.gov/content/sequence/HelpDocs/SEQsamples.html>, (27.11.2017).
- [19] https://www.hiv.lanl.gov/content/sequence/FORMAT_CONVERSION/FormatExplain.html, (27.11.2017).
- [20] http://wiki.christophchamp.com/index.php?title=NEXUS_file_format, (27.11.2017).
- [21] <http://cmpg.unibe.ch/software/arlequin35/man/Arlequin35.pdf>, (27.11.2017).
- [22] http://www.turkhataygen.gov.tr/doc/populasyon_genomigi_I_calistayi.pdf, (27.11.2017).
- [23] <http://heidi.chnebu.ch/doku.php?id=MEGA>, (27.11.2017).
- [24] <http://primerdigital.com/fastpcr/m2.html>, (27.11.2017).
- [25] http://www.ccg.unam.mx/~vinuesa/tlem09/docs/structure_doc.pdf, (27.11.2017).
- [26] http://genepop.curtin.edu.au/help_input.html, (27.11.2017).
- [27] <http://kimura.univ-montp2.fr/~rousset/Genepop4.7.pdf>, (27.11.2017).
- [28] <http://www.sfu.ca/~carmean/nexus.html>, (27.11.2017).
- [29] http://www.megasoftware.net/webhelp/walk_through_mega/mega_basics_hc.htm, (27.11.2017).

- [30] <http://bioinformatics.psb.ugent.be/webtools/ForCon/manual/>, (27.11.2017).
- [31] http://www.cmpg.unibe.ch/software/PGDSpider/PGDSpider%20manual_vers%202-0-0-3.pdf, (27.11.2017).
- [32] http://ib.berkeley.edu/labs/slatkin/eriq/software/new_hybs_doc1_1Beta3.pdf, (27.11.2017).
- [33] <http://heidi.chnebu.ch/doku.php?id=nexus>, (27.11.2017).
- [34] <http://mr bayes.sourceforge.net/Help/format.html>, (27.11.2017).
- [35] Liu, Z. (2007). Single nucleotide polymorphism (SNP). In *Aquaculture Genome Technologies* (pp. 59-72). Blackwell USA.
- [36] Seeb, J. E., Carvalho, G., Hauser, L., Naish, K., Roberts, S., and Seeb, L. W. (2011). Single-nucleotide polymorphism (SNP) discovery and applications of SNP genotyping in nonmodel organisms. *Molecular Ecology Resources*, 11(s1), 1-8.
- [37] Jones, C. J., Edwards, K. J., Castaglione, S., Winfield, M. O., Sala, F., Van de Wiel, C., and Brettschneider, R. (1997). Reproducibility testing of RAPD, AFLP and SSR markers in plants by a network of European laboratories. *Molecular Breeding*, 3(5), 381-390.
- [38] Şakiroğlu, M., Doyle, J. J., and Brummer, E. C. (2010). Inferring population structure and genetic diversity of broad range of wild diploid alfalfa (*Medicago sativa* L.) accessions using SSR markers. *Theoretical and Applied Genetics*, 121(3), 403-415.

- [39] Thormann, C. E., Ferreira, M. E., Camargo, L. E. A., Tivang, J. G., and Osborn, T. C. (1994). Comparison of RFLP and RAPD markers to estimating genetic relationships within and among cruciferous species. *Theoretical and Applied Genetics*, 88(8), 973-980.
- [40] Garcia-Mas, J., Oliver, M., Gomez-Paniagua, H., and De Vicente, M. C. (2000). Comparing AFLP, RAPD and RFLP markers for measuring genetic diversity in melon. *Theoretical and Applied Genetics*, 101(5), 860-864.
- [41] Roldàn-Ruiz, I., Dendauw, J., Van Bockstaele, E., Depicker, A., and De Loose, M. (2000). AFLP markers reveal high polymorphic rates in ryegrasses (*Lolium spp.*). *Molecular Breeding*, 6(2), 125-134.
- [42] http://www.med.kagawau.ac.jp/~genomelb/takezaki/poptree2/genepop_example1.html, (27.11.2017).
- [43] http://www.med.kagawau.ac.jp/~genomelb/takezaki/poptree2/genepop_example2.html, (27.11.2017).

ÖZGEÇMİŞ

Adı Soyadı : Mahmut AYDIN
Doğum Yeri ve Tarihi : OLTU, 14.07.1983
Yabancı Dili : İngilizce
İletişim (e-posta) : mahmut.aydin@kafkas.edu.tr

Eğitim Durumu (Kurum ve Yıl)

Lise : Nevzat Karabağ Anadolu Öğretmen Lisesi, 2001

Lisans : Atatürk Üniversitesi, Kazım Karabekir Eğitim Fakültesi,
Bilgisayar ve Öğretim Teknolojileri Eğitimi, 2012

Çalıştığı Kurum/Kurumlar ve Yıl : Kafkas Üniversitesi, Mühendislik Mimarlık
Fakültesi, Bilgisayar Mühendisliği, 2014