

T.C.
KASTAMONU ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
GENETİK VE BIYOMÜHENDİSLİK ANA BİLİM DALI



BAĞLANTI DENGESİZLİĞİ KATSAYISININ ÖRNEKLEME DAĞILIMI
ÜZERİNE BİR SİMÜLASYON ÇALIŞMASI

SELMA KÖKÇÜ

YÜKSEK LİSANS TEZİ

PROF. DR. ORHAN KAVUNCU

OCAK - 2022
KASTAMONU

TAAHHÜTNAME

Bu tezin tasarımı, hazırlanması, yürütülmesi, arařtırmalarının yapılması ve bulgularının analizlerinde bütün bilgilerin etik davranıř ve akademik kurallar çerçevesinde elde edilerek sunulduđunu; ayrıca tez yazım kurallarına uygun olarak hazırlanan bu çalışmada bana ait olmayan her türlü ifade ve bilginin kaynađına eksiksiz atıf yapıldıđını, bilimsel etiđe uygun olarak kaynak gösterildiđini bildirir ve taahhüt ederim.

Selma KÖKÇÜ

İmza

ÖZET

YÜKSEK LISANS TEZİ

BAĞLANTI DENGESİZLİĞİ KATSAYISININ ÖRNEKLEME DAĞILIMI ÜZERİNE BİR SİMÜLASYON ÇALIŞMASI

SELMA KÖKÇÜ

KASTAMONU ÜNİVERSİTESİ FEN BİLİMLERİ ENSTİTÜSÜ
GENETİK VE BIYOMÜHENDİSLİK ANA BİLİM DALI
DANIŞMAN: PROF. DR. ORHAN KAVUNCU

Bu çalışmanın amacı çeşitli bağlantı dengesizliği katsayılarına ve iki lokusta iki veya daha fazla allel sayılarına sahip sanal popülasyonlardan simülasyon yöntemiyle çekilen farklı genişlikteki örneklerden hesaplanan bağlantı dengesizliğiyle ilgili kuadratik istatistiklerinin nasıl bir dağılım göstereceğinin belirlenmesidir. Bağlantı dengesinde olan popülasyonlardan çekilen örneklerden hesaplanan bağlantı dengesizliği katsayısının (D) ortalamasının sıfır olduğu ($\mu_D=0$) ve kuadratik bir form olarak metinde 1.2 numara ile gösterilen istatistiğin bir Khi-Kare dağılımı gösterdiği bilinmektedir. $\mu_D=0$ ve $\mu_D=0,15$ bağlantı dengesizliği katsayılarına, iki lokus iki allel sayılarına sahip popülasyonlardan simülasyon yöntemiyle çekilen $n=10, 20, 50$ genişlikteki örneklerden hesaplanan birisi 1.2 numaralı kuadratik bir form olan istatistik olmak üzere üç ayrı kuadratik istatistiğin dağılımı çalışılmıştır. Sonuçta bağlantı dengesizliği katsayısının $0,15$ olduğu popülasyondan çekilen örneklerden hesaplanan 1.2 numaralı kuadratik formun 3 s.d.'li (noncentral) merkezi olmayan Khi-Kare dağılımı gösterdiği bulunmuştur. Paydaki D'nin, ortalamadan sapmasının karesi olarak hesaplanan kuadratik form metinde 3.1 numara ile gösterilen istatistik ise beklendiği gibi bağlantı dengesizliği katsayısı $\mu_D=0,15$ olduğu zaman da 1 s.d.'li Khi-Kare dağılımı göstermektedir. Klasik bir Khi-Kare formülü olan ve metinde 3.2 numara ile verilen istatistik de bağlantı dengesizliği katsayısı $\mu_D=0,15$ olduğu zaman da yine beklendiği gibi 3 s.d.'li Khi-Kare dağılımı göstermektedir. $\mu_D=0$ ve $\mu_D \neq 0$ bağlantı dengesizliği katsayılarına, iki lokus birinde iki allel diğerinde üç allel sayılarına sahip popülasyonlardan simülasyon yöntemiyle çekilen $n=100$ genişlikteki örneklerden hesaplanan birisi metinde 3.8 numaralı kuadratik bir form olan istatistik olmak üzere üç ayrı kuadratik istatistiğin dağılımı çalışılmıştır. Sonuçta bağlantı dengesizliği katsayısının sıfır olduğu popülasyondan çekilen örneklerden hesaplanan bu (3.8 numaralı istatistik) kuadratik formun 2 s.d.'li (central) merkezi Khi-Kare dağılımı gösterdiği bulunmuştur. Dengede olmayan popülasyonlardan çekilen örneklerden hesaplanan istatistikler için denenenler arasında uygun bir dağılım bulunamamış olup başka dağılımlara uygunluğunun test edilmesi gerekmektedir.

ANAHTAR KELİMELER:Bağlantı Dengesizliği Katsayısı(LD), Khi-Kare Dağılımı, Örneklem Dağılımı

Ocak 2022, 28 Sayfa

ABSTRACT

MSC THESIS

A SIMULATION STUDY ON THE SAMPLING DISTRIBUTION OF THE COEFFICIENT OF LINKAGE DISEQUILIBRIUM

SELMA KOKCU

KASTAMONU UNIVERSITY INSTITUTE OF SCIENCE
DEPARTMENT OF GENETICS AND BIOENGINEERIN
SUPERVISOR:PROF. DR. ORHAN KAVUNCU

The aim of this study is to determine the distribution of the quadratic forms related to the linkage disequilibrium coefficients calculated from the samples drawn from populations with non-zero linkage disequilibrium coefficient. It is known that the average of coefficient of linkage disequilibrium (D) calculated from samples drawn from populations in linkage equilibrium is zero ($\mu_D = 0$) and statistic with equation number 1.2 in the text, which is a quadratic form, shows a chi-square distribution. It is studied the distribution of three different quadratic statistics one is 1.2 calculated from samples of size $n = 10, 20, 50$ generated through simulation from populations with linkage disequilibrium coefficients of $\mu_D = 0$ and $\mu_D = 0,15$ for a model of two alleles, two loci. The results show that the first quadratic form (statistic number 1.2) which is much known statistics in literature, has found to be fitted noncentral chi square distribution with 3 df in the samples drawn from the populations of $\mu_D = 0,15$. The second quadratic form (equation number 3.1) in the methods was stable in being fitted to the central chi square distributions of 1 df in the existence of linkage disequilibrium. The third quadratic form (equation number 3. 2) has a distribution fitted to chi square with 3 df. The distribution of three different quadratic statistics , one of which is a quadratic form numbered 3.8 in the text, calculated from samples of size $n=100$ generated through simulation from populations with linkage disequilibrium coefficients of $\mu_D=0$ and $\mu_D\neq 0$ and two alleles in one locus and three alleles in the other were studied. As a result, it was found that this (3.8 statistic number) quadratic form, calculated from the samples drawn from the population with a linkage disequilibrium coefficient of zero, showed a central Chi-Square distribution with 2 df (central). For the statistics calculated from samples drawn from non-equilibrium populations, a suitable distribution could not be found among those tried, some other distributions should be tested.

KEYWORDS:The Coefficient of Linkage Disequilibrium(LD), Chi Square Distribution, Sampling Distribution.

January 2022, 28 Page

TEŐEKKÜR

Tez alıőmasında R yazılım sayesinde simülasyon alıőması kolaylıkla neticelendirilmiőtir. Bu konuda bana alıőma imkanı saėlayan ve baőtından sonuna kadar maddi ve manevi yardımını bir an olsun esirgemeyen tezin yürütülmesinde sonuçların deėerlendirilmesinde yol gösterici yardımlarından ve teőtviklerinden dolayı deėerli Prof. Dr. Orhan KAVUNCU hocama sonsuz Őükranlarımı sunarım. Eėitim-öėretim hayatımın baőtından bu yana beni destekleyen her zaman yanımda olan aileme sonsuz teőtekkür ederim.

SELMA KÖKÇÜ

Kastamonu, 2022

İÇİNDEKİLER

	<u>Sayfa</u>
TEZ ONAYI	ii
TAAHHÜTNAME	iii
ÖZET	iv
ABSTRACT	v
TEŞEKKÜR	vi
İÇİNDEKİLER	vii
ŞEKİLLER DİZİNİ	viii
TABLolar DİZİNİ	ix
SİMGELER VE KISALTMALAR DİZİNİ	x
1. GİRİŞ	1
2. LİTERATÜR ÖZETİ	3
3. MATERYAL ve YÖNTEM	10
3.1 Bağlantı Dengesizliği Katsayısından Hesaplanan Kuadratik İstatistikler	10
3.1.1 2*2 Lokus Modeli	10
3.1.2 3*2 Lokus Modeli	12
3.2 Simülasyonla Örnek Üretilen Popülasyonlar	14
4. SONUÇLAR	17
4.1 2*2 Lokus Modeli İçin Bulgular	17
4.2 3*2 Lokus Modeli İçin Bulgular	19
5. TARTIŞMA ve ÖNERİLER	23
KAYNAKLAR	24
ÖZGEÇMİŞ	28

ŞEKİLLER DİZİNİ

	<u>Sayfa</u>
Şekil 3.1.1.1 Üretilen örnekleme dağılımlarının uygunluğu araştırılan Khi-Kare dağılımları	12
Şekil 4.1.1 $\mu_D=0,15$ ve $n=20$ için üretilen 10.000 örnekten hesaplanan değerlerin grafikleri.....	17
Şekil 4.2.1 $\mu_D=0$ ve $n=100$ için üretilen 10.000 örnekten hesaplanan değerlerin grafikleri.....	22

TABLolar DİZİNİ

	<u>Sayfa</u>
Tablo 3.1.2.1 Gamet Frekansları.....	13
Tablo 3.1.2.2 Bağlantı Dengesinde Bir Popülasyonun Yapısı.....	13
Tablo 3.1.2.3 Bağlantı Dengesizliğindeki Bir Popülasyonun Yapısı.....	13
Tablo 3.2.1 Çalışmada Kullanılan R Programıyla Yazılan Algoritma (İki allelli iki lokus, $\mu_D=0,15$ ve $n=20$)	15
Tablo 3.2.2 Çalışmada Kullanılan R Programıyla Yazılan Algoritma (Birisini iki diğeri üç allelli iki lokus, $\mu_D=0$ ve $n=100$).....	16
Tablo 4.1.1 Altı Popülasyondan Elde Edilen Sonuçlar (1.2 numaralı istatistik için)	18
Tablo 4.1.2 Altı Popülasyondan Elde Edilen Sonuçlar (3.1 numaralı istatistik için)	19
Tablo 4.1.3 Altı Popülasyondan Elde Edilen Sonuçlar (3.2 numaralı istatistik için)	19
Tablo 4.2.1 Sekiz Popülasyondan Elde Edilen Sonuçlar (3.5 numaralı istatistik için)	20
Tablo 4.2.2 Sekiz Popülasyondan Elde Edilen Sonuçlar (3.7 numaralı istatistik için)	20
Tablo 4.2.3 Sekiz Popülasyondan Elde Edilen Sonuçlar (3.8 numaralı istatistik için)	21

SİMGELER VE KISALTMALAR DİZİNİ

Simgeler

μ	: Ortalama
D	: Bağlantı Dengesizliği Katsayısı
LD	: Bağlantı Dengesizliği Katsayısı
μ_D	: Bağlantı Dengesizliği Katsayısının Ortalaması
χ	: Chi
p_1	: Büyük (A) geninin Frekansı
p_2	: Küçük (a) geninin Frekansı
q_1	: Büyük (B) geninin Frekansı
q_2	: Küçük (b) geninin Frekansı
Σ	: Toplam
π	: Pi
n	: Örnek Genişliği
k_1	: İki Yanlı Tablolar İçin Satır Sayısı
k_2	: İki Yanlı Tablolar İçin Sütun Sayısı
x_i	: Gametlerin Örnekteki Sayısı
p_i	: Gametlerin Popülasyondaki Beklenen Frekansları
λ	: Lamda (Noncentrallik Parametresi)

Kısaltmalar

ncpar	: Noncentrallik Parametresi
sd	: Serbestlik Derecesi
vd.	: Ve Diğerleri

1. GİRİŞ

Bağlantı dengesizliği üzerine çalışmalar 1950’li yıllarda başlamıştır (Kavuncu, 1987). İki allelli iki lokus için bağlantı dengesizliği katsayısı rastgele çiftleşen bir popülasyonda

$$\mu_D = P(AB) \cdot P(ab) - P(Ab) \cdot P(aB) \quad (1.1)$$

şeklinde hesaplanır. Burada bir lokustaki alleller A ve a, diğer lokustaki alleller ise B ve b harfleriyle gösterilmiştir. P(AB), oluşacak gametlerden rastgele birinin AB olma ihtimalidir. Popülasyon bağlantı dengesindeyse, bu gamet popülasyonundan çekilen n gametlik bir örnekte bağlantı dengesizliği katsayısı istatistiği D’nin sıfır olması beklenir ve böyle bir örnekte hesaplanan kuadratik formdaki

$$\chi_0^2 = \frac{n D^2}{p_1 \cdot p_2 \cdot q_1 \cdot q_2} \quad (1.2)$$

(1.2) numaralı istatistik 1 serbestlik dereceli Khi-Kare dağılımı gösterir (Gillespie, 2004; Hartl ve Clark, 2007; Nielsen ve Slatkin, 2013). Paydadaki p_1 A geninin, q_1 a geninin, p_2 B geninin ve q_2 ise b geninin frekanslarıdır. Lokuslarda allel sayısı ikiden fazla olduğunda da (1.1) numaralı istatistiğin çok allelli duruma uyarlanmasıyla, yine bağlantı dengesi varsayımı altında, $(k_1-1)(k_2-1)$ serbestlik dereceli bir Khi-Kare dağılımı gösteren bir istatistik elde edilir. Burada k_1 ve k_2 birinci ve ikinci lokustaki allellerin sayısıdır. (1.2) numaralı istatistiğin Khi-Kare dağılımı göstermesi, popülasyonun bağlantı dengesinde olduğu varsayımına, yani örneklerden hesaplanan D’lerin ortalamasının sıfır olduğu ($\mu_D=0$) varsayımına bağlıdır.

Ancak popülasyon bağlantı dengesinde değilse örneklerden hesaplanan bağlantı dengesizliği katsayısının nasıl bir dağılım göstereceğini bilmek gerekir. Ne var ki bağlantı dengesinde olmayan popülasyonlardan çekilen örneklerden hesaplanacak bağlantı dengesizliği katsayısı ve (1.2) numaralı kuadratik formdaki gibi istatistiklerin örnekleme dağılımı üzerine bir çalışmaya rastlanamamıştır. DNA

verilerinin istatistik analizinden bağlantı dengesizliğini ölçmeye yönelik teorik çalışmalara kadar taranan literatürde bu durum geçerlidir (Weir,1996).

Böyle bir çalışmanın teorik popülasyon genetiğinin ilgili alanlarında yararlı olması beklenmektedir. Ayrıca bu çalışmadan elde edilecek sonuçların, genetik varyans unsurlarının tahminindeki isabeti artıracığı, bu yüzden de kantitatif genetik ve bu tahminlere çokça ihtiyaç duyulan bitki ve hayvan ıslahı çalışmalarında kullanılma olanağı bulacağı düşünülmektedir. Çünkü eğer popülasyon bağlantı dengesinde değilse lokusların etkileri arasındaki kovaryans sıfır olmayacak, dolayısıyla bu kovaryans, genetik varyans unsurlarının tahmin edilmesinde sapmalara yol açacaktır (Kavuncu,1987). Bunun yanında nükleotid sitelerinin birlikte dağılımını dikkate alan moleküler ölçekli Crow ve Kimura (1970) çalışmalarda ve gerek polimorfizm ve gerekse protein varyasyonları gibi biyoinformatik alanlarda da katkı sağlaması ümit edilmektedir.

Bu tezde, lokuslarda allel sayısı ikiden fazla olduğunda da (1.1) numaralı istatistiğin çok allelli duruma uyarlanmasıyla ve çeşitli bağlantı dengesizliği katsayılarına sahip olan sanal popülasyonlardan simülasyon yöntemiyle çekilen farklı genişlikteki örneklerden hesaplanan bağlantı dengesizliği katsayısı ile ilgili istatistiklerin dağılımı belirlenmeye çalışılmıştır. Çünkü bağlantı dengesizliği katsayısı sıfır olmayan popülasyonlardan çekilen örneklerden hesaplanacak istatistiklerin örnekleme dağılımının bilinmesi, ilgili parametrelerin güven aralıkları tahminleri için gereklidir.

Böylece çalışmada bağlantı dengesizliği katsayıları, birinci lokustaki allel sayıları, ikinci lokustaki allel sayıları ve örnek genişlikleri olmak üzere dört faktörün seviyelerine göre sanal popülasyonlar belirlenmiştir. Bütün bu örneklerin simülasyonu, hesaplamalar ve örnekleme dağılımları R yazılım kullanılarak geliştirilmiş bilgisayar programları ile bulunmuştur.

2. LİTERATÜR ÖZETİ

Çeşitli bilim dallarında bağlantı dengesizliği katsayısı üzerine araştırmalar son yıllarda gittikçe artış göstermektedir. Bu çalışmaların çoğunluğunu simülasyon denemeleri oluşturmasının yanında uygulamalı alanlarda da çalışmalar mevcuttur. İstatistik analizinden bağlantı dengesizliğini ölçmeye yönelik teorik çalışmalarda dahil günümüze kadar taranan literatürde bağlantı dengesizliği katsayısının nasıl bir dağılım gösterdiği üzerine bir çalışmaya rastlanmamıştır. LD üzerine ilk çalışmalardan biri olan Asmussen ve Clegg (1981) tarafından iki lokuslu popülasyon genetik modellerinde, genetik danışma amacıyla kısıtlama parçası uzunluğu (RFLP) polimorfizmlerinin analiz edildiğini ve mutant polimorfik dengesine hareket ettikçe, nötr bir işaretleyici ile sıkı bir şekilde bağlanmış aşırı baskın mutant arasındaki bağlantı dengesizliğinin hızla artacağını göstermişlerdir. Bağlantı dengesizliğinin kantitatif bir özellik bakımından genetik varyans ve unsurları üzerine etkisi, çeşitli gen etki modelleri için teorik olarak çalışılmıştır (Kavuncu ve Kesici, 1982; Kavuncu ve Düzgüneş, 1983; Kavuncu 1983, 1984, 1987). Zapata vd. (1997, akt. Hamilton ve Cole, 2006) LD'nin standartlaştırılmış gametik ölçüsünün varyansını sunmuştur. Excoffier ve Slatkin (1998) çalışmalarında Haplotip frekanslarını çıkarmak ve bağlantı dengesinden önemli sapmaları test etmek için akrabaların genotiplerinin EM algoritmasına nasıl dahil edileceğini göstermişlerdir. Sabatti ve Risch (2002) İki veya daha fazla belirteç arasındaki dengesizlik seviyesini ölçmek için haplotiplerin homozigotluğunun nasıl kullanılabileceğini göstermişlerdir. Lee, (2003) Karmaşık insan hastalıklarının genetik çalışmalarının geleceğinin, giderek daha fazla epidemiyolojik ilişki paradigmasına dayanacağını ifade etmiş ve etkilenen bireylerin bir gen bankasında Hardy-Weinberg dengesinden sapmayı test ederek genomu hastalığa yatkınlık gen(ler)i için taramayı önermiştir. Gu vd. (2004) 148 hipertansif ailede LPL geni ile lipid profilleri, özellikle TG arasındaki ilişkiyi araştırmak için, yedi komşu mikro uydu işaretleyici ve LPL geninin dört dahili işaretleyici seçip ve SOLAR ve S.A.G.E. (genetik epidemiyoloji için istatistiksel analiz)/SIBPAL 2 programları ve QTDT (niceliksel aktarım/dengesizlik testi) ve GOLD (bağlantı dengesizliğine grafiksel genel bakış) ile bağlantı dengesizliği analizini yapmışlardır. Ailelerinde hipertansiyonu olmayan ve hipertansiyonu olan denekler arasında lipid

seviyelerinde istatistiksel olarak anlamlı farklılıklar olduğunu ortaya koymuşlardır. SOLAR tarafından D8S261 işaretçisinde TG ile 1,3'lük bir maksimum LOD puanı gözlemlenmiştir. S.A.G.E./SIBPAL 2 kullanarak, LPL geninin intron 6'sında yer alan 'ATTT' işaretçisinde TG ile bir bağlantı belirlenmiştir ($P=0,0095$). İki SNP (tek nükleotid polimorfizmi), HindIII ve HinfI, LDL kolesterol seviyeleri ile bağlantı dengesizliğinde bulunduğunu ifade etmişlerdir (sırasıyla $P=0,0178$ ve $P=0,0088$). İtron 8'deki HindIII ile ekson 9'daki HinfI arasında güçlü bir bağlantı dengesizliği gözlemlenmiştir ($P<0,00001$, $D =0,895$). İtron 6'daki 'ATTT' polimorfizmi ile iki SNP arasında bağlantı dengesizliğinin olduğunu tespit etmişlerdir (HinfI için $P=0,0021$ ve $D =0,611$; ve HinfI için $P=0,00004$, $D =0,459$). Hipertansiyonlu Çinli ailelerde yapılan bu çalışma, LPL geninin lipid seviyelerini, özellikle TG metabolizmasını etkileyebileceğini öne sürmüştür. Hem Çinlilerde hem de diğer popülasyonlarda yapılan replikasyon çalışmalarının, bu sonuçları doğrulamak için garantili olduğunu ifade etmişlerdir. Aynı yılda bir başka çalışmada Kim ve Nielsen (2004) aşırı bağlantı dengesizliğinin (LD) seçici bir taramanın güçlü bir imzası olduğu bilindiğinden, LD ile ilgili bilgilerin dahil edilmesiyle istatistiksel gücün ne kadar arttığını araştırmışlardır. Beklenen LD modelini, soykütüksel bir yaklaşımla tahmin etmişlerdir. Hem teori hem de simülasyon çalışmaları, faydalı mutasyonun bulunduğu yerin her iki tarafındaki dar bölgelerde güçlü LD'nin üretildiğini göstermektedir. Ancak, iki tarafta da LD eksikliğinin bekleniyor olduğunu Kim ve Nielsen ifade etmişlerdir. Bu seçici tarama imzasını istatistiksel testlerle tespit etmenin çeşitli yollarını araştırmışlardır. LD ile ilgili bilgileri birleştirmek için yeni bir bileşik olasılık yöntemi önermişlerdir. Bu yöntemin, seçici taramaları tespit etme ve seçim modelinin parametrelerini, LD'yi hesaba katmayan önceki bileşik olasılık yönteminden daha iyi tahmin ettiğini belirtmişlerdir. Hamilton ve Cole (2004) ve Zaykin (2004) ölçünün -1 ile 1 arasında olacak şekilde standartlaştırılmasına izin veren bileşik bir bağlantı dengesizliği (LD) ölçüsü için sınırlar önermişlerdir. Dekkers vd. (2005) varsayılan biallelik QTL ile çoklu allelik belirteçlerin kullanılabilir LD'sinin öngörücüleri olarak çok allelik belirteçler arasındaki gözlemlenebilir LD'nin alternatif ölçümlerini değerlendirmişlerdir. Drift ve bir dizi belirteç polimorfizmi tarafından oluşturulan bir LD aralığına sahip simüle edilmiş popülasyonlarda, standart bir Khi-Kare istatistiği ile ölçülen belirteç-işaretleyici LD'nin, bir grup için kullanılabilir belirteç-QTL LD'nin en iyi öngörücüsü olduğunu

bulmuşlardır. Yine aynı yılda Ioannidis vd. (2005) son genetik ilişkilendirme çalışmalarında Hardy-Weinberg dengesinin (HWE) test edilmesini ve raporlanmasını değerlendirip, HWE'nin ne sıklıkla ihlal edildiğini tespit edip ve bu literatürde HWE testinin gücünü tahmin etmişlerdir. Sonuç olarak mevcut genetik ilişkilendirme çalışmalarının çoğunda HWE testi için gücün sınırlı olduğunu belirtmişlerdir. Guo vd. (2006) rahatsız edici parametrelere genişletilen olabilirlik puanı teorisini kullanarak, birkaç bağımsız katmandaki dengesizlik katsayılarını karşılaştırmak için bir homojenlik puanı testi türetmişlerdir. Simülasyon sonuçları, homojenlik puanı testinin, ampirik boyutunun, küçük örnek boyutları için bile, önceden seçilen nominal seviyeyi nadiren %10'dan fazla aşması anlamında tatmin edici bir şekilde performans gösterdiğini belirtmişlerdir. Testi gerçek bir glioksalaz genotipi veri seti ile göstermişlerdir. Y. S. Song ve J. S. Song (2006) tarafından tekrarlayan mutasyon, genetik sürüklenme ve rekombinasyon içeren bir model için popülasyon çapında korelasyon katsayısının karesinin beklenen değerini hesaplama yöntemi tanıtılmıştır. Yaklaşımın cebirsel olduğunu ve difüzyon süreci yaklaşımına dayandığını belirtmişlerdir. Hamilton vd. (2006) genotip frekanslarına bağlı olan bileşik bir bağlantı dengesizliği ölçüsü için yaklaşık varyansı ve Simülasyonla, bu varyans yaklaşımının bir dizi allel olasılığı ve Hardy-Weinberg dengesinden sapmalar üzerinde geçerli olduğunu ifade etmişlerdir. Mano (2006) çalışmasında bağlantı dengesizliğinin bir ölçüsü olarak, momentlerin oranı ile tanımlanan standart bağlantı sapmasının karesini dikkate almıştır. Bu momentlerin uyduğu bir adi diferansiyel denklem sistemini elde etmiştir. Bu sistemin sayısal olarak çözülebilir olduğunu ifade ederek simülasyonlarla, karesi alınmış standart bağlantı sapmasının, gamet frekanslarının korelasyon katsayısının karesinin beklenen değerine iyi bir yaklaşım olduğunu göstermiştir. Aynı yılda bir başka çalışma da Hardy-Weinberg dengesinden (HWE) sapmayı yakalayan ve ölçen istatistiksel testler tasarlanmıştır, ancak çok allelik DNA lokuslarında sapma testi denendiğinde sınırlı kalmıştır. Bu noktada Pereira vd. (2006) tarafından HWE için tam bayes önemlilik testi (FBST) önerilmiştir. Bu testin, FBST, HWE hipotezi lehine bir kanıt indeksinin hesaplanmasına dayanmakta olduğunu belirtmişlerdir. FBST'yi rutin ebeveynlik testi sırasında birkaç multiallelik kısa tandem tekrar lokusunda elde edilen genotiplere uygulamışlar; D10S1214 ve D19S253 gibi ve Penta E konumu HWE'dekileri açık bir şekilde örneklediğini söylemişlerdir. Sabatti vd. (2006) Homozigotluğa dayalı

bağlantı dengesizliği ölçümünün nasıl korelasyon katsayısının karesi ile karşılaştırılabilir sonuçlara yol açtığını ve ayrıca D' (gözlenen allel frekansları için maksimum olası değer) değerlendirilirken küçük örnek varyasyonunun düzeltilmesinin önemini göstermişler ve NRH segmentlerini barındıran bölgelerin popülasyonlar arasında tutarlı olma eğiliminde olduğunu, gen açısından zengin olduğunu ve daha düşük rekombinasyon ile karakterize edildiğini gözlemlemişlerdir. Rosenberg ve Blum (2006) yaptıkları çalışmada bağlantı dengesindeki herhangi bir lokus çifti için, IIS fazlalık istatistiğinin dönüştürülmüş bir versiyonunun, numune boyutu sonsuza yaklaştıkça normal bir dağılıma yakınsamak gibi arzu edilen bir özelliğe sahip olduğunu göstermişlerdir. Bunun, bağlantı dengesizliği için asimptotik bir testin geliştirilmesini sağladığını: bağlantı dengesi sıfır hipotezi altında, istatistiğin standart normal dağılıma sahip olduğunu ifade etmişlerdir. Mahasirimongkol vd. (2006) Hapmap ve Thais'deki kuzey Doğu Asya popülasyonları arasındaki benzerliği kanıtlamak için Thais, Han Chinese ve Japanese arasında paylaşılan 166 ilaca bağlı gende 861 SNP, korelasyon istatistikleri için analiz edilip allel frekansı, Fst istatistikleri ve bağlantı dengesizliği istatistiklerinin (r^2) bu popülasyonlar arasında yüksek bir korelasyon göstermiş olduğunu ifade etmişlerdir. Tek nükleotid polimorfizmlerinin (SNP'ler), popülasyonların genetik heterojenliğini araştırmak ve bağlantı dengesizliği (LD) haritalaması yapmak için önemli belirteçler olduğunu ifade etmişlerdir. Yamada ve Matsuda (2007) çalışmalarında N_s di-allelik SNP'ler için 2^{N_x} haplotiplerinin sıklığını ifade etmek amacıyla yeni bir yöntem önermişlerdir. Haplotip frekansının yeni ifadesini kullanarak, sadece SNP çiftleri için değil, aynı zamanda çoklu işaretler için de yeni bir LD ölçüsü önermişlerdir. SNP çiftleri için geleneksel ikili LD endekslerinin, D_0 ve r^2 değerlerine benzer olduğu ortaya koymuş ve SNP çiftleri için yeni ölçünün geleneksel LD endeksleri tarafından ölçülmeyen LD bileşenlerini nicelediğini ifade etmişlerdir. Du vd. (2007) bağlantı mesafesine dayalı çok sayıda grup oluşturarak ve LD ölçümlerinin grup ortalamalarını kullanarak rekombinant oranların etkisini ayarlayarak LD ölçümlerinin minör allel frekansına (MAF) bağımlılığını ölçmeye çalışmışlardır. D' 'nin allel frekansına güçlü bir şekilde bağımlı olduğunu ve ortalama MAF arttıkça yani iki lokusun heterozigotluğu arttıkça bağımlılığın sürekli azaldığını

göstermişlerdir (çünkü heterozigotluk, biallelik lokuslar için MAF'ın artan bir fonksiyonudur). Zaykin vd. (2008) birden fazla allel içeren lokuslar için korelasyon katsayısının karesini (r^2) sunmuşlardır. r^2 'nin bağımsızlık sıfır hipotezine göre yaklaşık dağılımını incelemiş ve permütasyonla elde edilen kesin dağılımla yakın uyum içinde olduğunu belirtmişlerdir. r^2 kullanılarak yapılan bağımsızlık testinin, Pearson'ın ki karesi, Fisher'in kesin testi ve Cressie ile Read'in güç sapma istatistiğine dayalı bir test gibi yaklaşımlara güçlü bir rakip olduğunu belirtmişlerdir. Zhang vd. (2008) Contig.060226,1'in bir bölgesinde (200 kb) başlangıçta ortalama fiziksel mesafeye dayalı olarak GGA1 üzerinde toplam 36 SNP seçmişlerdir. Polimorfizm düzeylerini doğruladıktan sonra, son olarak bir vahşi ve iki evcil tavuk popülasyonunu genotiplemek için 21 SNP seçip işaretçi çiftleri arasında iki farklı bağlantı dengesizliği ölçümü, D' ve r^2 kullanmışlardır. Birçok işaretçi çiftinin tam LD gösterdiğini ifade etmişlerdir. Sonuç olarak, LD haritalama yöntemleri için çok yoğun bir işaretçi haritasının gerekli olacağını ifade etmişlerdir. Zintzaras (2008) Hardy-Weinberg denge sapması için allel tabanlı olasılık oranının varyansını hesaplamak için bir yaklaşım önermiştir. Yaklaşım, Woolf'un logit aralığı yöntemi ve dengesizlik katsayısı ile birlikte delta yöntemine dayanmaktadır. Önerilen varyans düzeltilmesi, önemli olasılık oranı tahminlerini tespit etmek için ayarlanmamış olandan daha iyi güç sağladığını ve varyans tahminini geliştirmiş olduğunu beyan etmiştir. Rogers ve Huff (2009) çalışmalarında, çiftleşmenin rastgele olması koşuluyla, D ve r 'nin diploid genotipler arasında kovaryanslar ve korelasyonlar olarak ifade edilebileceğini ifade etmişlerdir. Burada bu sonucun yaklaşık olarak çiftleşme rastgele olmadığında da geçerli olduğunu göstermişlerdir. Wang ve Zeng (2009) çalışmalarında F_{∞} ve Cockerham modelleri için modelleme şemalarını araştırmışlar. F_{∞} modellerinin, genotipleri farklı faktör seviyeleri olarak ele alarak temelde genotip tabanlı modeller olduğunu, Cockerham modellerinin ise bir doğrusallık problemini düzeltmek için özel bir modifikasyona sahip allel tabanlı modeller olduğunu göstermişlerdir. Zapata (2011) D (allel frekansları) ve r 'nin örnekleme özelliklerini, farklı aralıklar ile tahminlerin kesinliğini, doğruluğu ve verimliliğinin karşılaştırılmasına izin veren standartlaştırılmış istatistikler kullanılarak çeşitli gerçekçi popülasyon koşulları ve değişen örnek boyutları altında simülasyon ile değerlendirmiş olup simülasyonlar, standart istatistiklerle karşılaştırıldığında, r^2 'nin tahminlerinin D' (gözlenen allel frekansları için

maksimum olası deęer) tahminlerinden önemli ölçüde daha kesin, doğru veya verimli olma eğiliminde olmadığını ortaya koyduğunu çalışmasında belirtmiştir. İlhan (2018) bu çalışmasında; 70 yonca popülasyonu için 31 SSR markörü kullanılarak bağlantı dengesizliği analizini yapmıştır. Bağlantı dengesizliği analizi için; 31 markörden 23 tanesinin yonca'nın 8 kromozomu üzerindeki fiziksel mesafeleri belirlemiştir. 8 kromozom üzerinde 103 lokustan oluşan mesafeler *Medicago truncatula* genomunu temel alınarak saptamıştır. Deęerlendirilen markörler için; bu popülasyonlarda önemli ($p < 0,0001$) bir bağlantı dengesizliği deęeri olduğunu gözlemlemiştir. Raya vd. (2018) çalıştıkları dört İber domuzu suşunun karışımıyla oluşturulan bir kompozit soydaki birkaç lokustaki allellerin birleşmesi ile ilgili olan üçüncü dereceden (üç lokus için) bağlantı dengesizliğinin genom genelindeki seviyelerini incelemiştir. Genom çapında analizler, üçüncü dereceden bağlantı dengesizliğinin bu İber domuz soyundaki bağlantılı lokuslar arasında oldukça yaygın olduğunu belirtmişlerdir. Jiang vd. (2019) çalışmalarında, altı keçi popülasyonunda LD, Ne ve minör allel frekans dağılımını deęerlendirmişlerdir. Buna göre, tek nükleotid polimorfizm (SNP) paneli (Caprine SNP 50K BeadChip, Lincoln, NE, ABD) kullanan LD ve Ne'nin sonuçlarının, Çin keçi popülasyonlarının sürdürülebilir korunması, uygun yönetimi ve kullanımı için faydalı olacağını belirtmişlerdir. Aynı yılda Cheng vd. (2019) Mutasyon varlığında veya yokluęunda dengede r^2 'nin beklenen deęerini belirlemek için herhangi bir t generasyonunda sonlu bir popülasyonda iki lokustaki allel frekanslarının kesin dağılımını, t - 1 generasyonundaki dağılıma baęlı olarak hesaplamak amacıyla olasılıęa dayalı bir yaklaşımlar önermişlerdir. Etkin popülasyon büyüklüğünün bir fonksiyonu olarak, t - 1 neslindeki dağılımlar göz önüne alındığında, t neslindeki iki bağlantılı lokustaki allel frekanslarının kesin dağılımının nasıl hesaplanacağını göstermişlerdir. Hui ve Burt (2020) çalışmalarında, r^2 tahminini tüm LD düzeylerine ve hem aşamalı hem de aşamasız veriler için genelleştirmişlerdir. İlk olarak, sonlu örneklem büyüklüğünün gözlemlenen r^2 deęerleri üzerindeki etkisi için yeni formül önermişlerdir. İkinci olarak, gözlemlenen r^2 'nin varyansı için yeni bir ampirik formül bulmuşlardır. Üçüncüsü, Hardy-Weinberg Dengesi altında diploid genotiplerden haplotip frekanslarını ve r^2 'yi doğrudan tahmin etmek için olasılıęa dayalı bir yöntem olan kısıtlı ML adlı yeni bir rutin önermişlerdir. Yine farklı bir çalışmada amaç bağlantı dengesizliği katsayısı sıfır olmayan iki lokusta iki allel sayısına sahip popülasyonlardan çekilecek

örneklerden hesaplanan bağlantı dengesizliği katsayılarının kuadratik formlarının nasıl bir dağılım göstereceğinin belirlenmesi adlı çalışma yapılmıştır (Kökçü vd. 2020). Bağlantı dengesizliği katsayısının genetik varyans ve unsurları üzerine etkisini sayısal olarak somut bir şekilde göstermek için yapılan bir çalışmada biri dengede diğeri dengede olmayan 2 popülasyon tanımlanmıştır. Sonuç olarak parametrelerin bağlantı dengesizliğinden etkilendiği açıkça görülmüştür (Kökçü vd. 2021a, 2021b).

3. MATERYAL ve YÖNTEM

3.1 Bağlantı Dengesizliği Katsayısından Hesaplanan Kuadratik İstatistikler

3.1.1 2*2 Lokus Modeli

Bağlantı dengesizliği katsayıları gamet frekanslarına bağlıdır. Buna göre, birinci lokustaki A allelinin gittiği lokusa ikinci lokusta B veya b allelinin gitmesi tamamen şansa bağlı ise popülasyon bu iki lokusta bağlantı dengesindedir. İki lokus bağlı, yani aynı kromozom üzerinde olsa da iki lokustaki allellerin gametlere dağılımı birbirinden bağımsız olabilir. Bu durumda bağlantı dengesizliği katsayısı (μ_D) sıfır olacaktır. Örnekten hesaplanan bağlantı dengesizliği katsayısının sıfır olduğu hipotezini kontrol etmek için (1.2) numaralı istatistikle hesaplanan değer kullanılır. Ancak örneğin çekildiği popülasyonda bağlantı dengesizliği katsayısı (μ_D) sıfır değilse o zaman (1.2) numaralı istatistik iyi bir istatistik olmayabilir. Bunun yerine

$$\chi_1^2 = \frac{n (D - \mu_D)^2}{p_1 \cdot p_2 \cdot q_1 \cdot q_2} \quad (3.1)$$

(3.1) numaralı istatistik kullanılabilir. Burada p_1 A geninin frekansı, p_2 ise B geninin frekansıdır. a geninin frekansı $q_1=1-p_1$ ve b geninin frekansı $q_2=1-p_2$ şeklinde yazılabilir. Bunun yanında popülasyon bağlantı dengesizliği katsayısından bağımsız olarak

$$\chi_2^2 = \sum_{i=1}^4 \frac{(x_i - np_i)^2}{np_i} \quad (3.2)$$

(3.2) numaralı istatistik de bir çeşit kontrol grubu üretmek üzere denenebilir. Burada x_i 'ler, $i=1, 2, 3$ ve 4 için sırasıyla AB, Ab, aB ve ab gametlerinin örnekteki sayısı, p_i 'ler de aynı sırada gametlerin popülasyondaki beklenen frekanslarıdır. Bu durumda bu istatistiğin serbestlik derecesi 3 olacaktır.

Bu çalışmada bu üç istatistiğin dağılımının Khi-Kare veya merkezi olmayan (noncentral) Khi-Kare olabileceği varsayılmıştır. Merkezi olmayan Khi-Kare dağılımı için merkezi olmama parametresi, (3.1) numaralı istatistik için,

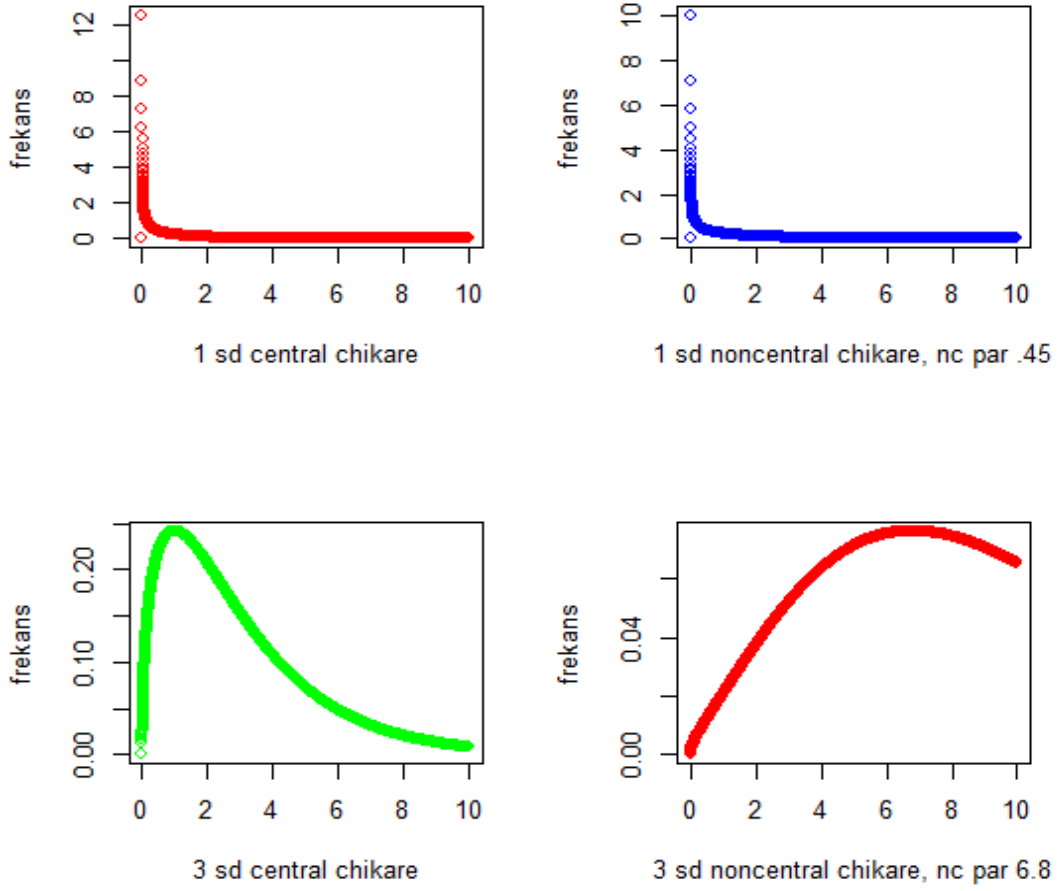
$$\lambda=n \cdot \mu_D^2 \quad (3.3)$$

(3.2) numaralı istatistik için ise

$$\lambda=n \sum_{i=1}^4 p_i^2 \quad (3.4)$$

alınmıştır. Merkezi olmayan Khi-Kare için Seber (1963)'e bakınız.

Bu (1.2), (3.1), (3.2) numaralı üç istatistik simülasyon yöntemi ile üretilen n genişliğindeki k örnek için hesaplanmış ve ilk iki istatistiğin (1.2 ve 3.1 numaralı istatistikler) merkezi (central) ve merkezi olmayan (noncentral) 1 s.d.'li Khi-Kare dağılımına uygunluğu, üçüncü istatistiğin (3.2 numaralı istatistik) de merkezi ve merkezi olmayan 3 s.d.'li Khi-Kare dağılımına uygunluğu, Kolmogorov-Smirnov testi ile kontrol edilmiştir. İşlem 20 kere tekrarlanmıştır. Merkezi olmama parametreleri 1 s.d.'li Khi-Kare için (3.3 numaralı istatistikte) 0,45, 3 s.d.'li Khi-Kare için (3.4 numaralı istatistikte) 6,8 olarak hesaplanmıştır. 1 ve 3 s.d.'li merkezi ve merkezi olmayan Khi-Kare dağılımlarının grafiği şekil 3.1.1.1'de gösterilmiştir.



Şekil 3.1.1.1 Üretilen örnekleme dağılımlarının uygunluğu araştırılan Khi-Kare dağılımları

3.1.2 3*2 Lokus Modeli

Bu lokus modelinde bir gamet popülasyonu tablo 3.1.2.1’de gösterilmiştir. Buna göre, birinde üç allel (A) diğesinde iki allel (B) olan iki lokus için, biri bağlantı dengesinde diğeri bağlantı dengesizliğinde olan iki popülasyonun yapısı tablo 3.1.2.2 ve tablo 3.1.2.3’deki gibi tanımlanmıştır.

Tablo 3.1.2.1 Gamet Frekansları

A Geni	B Geni	B1	B2	Toplam
A1		q11	q12	p11
A2		q21	q22	P12
A3		q31	q32	P13
Toplam		P21	P22	1

Tablo 3.1.2.2 Bağlantı Dengesinde Bir Popülasyonun Yapısı

A Geni	B Geni	B1	B2	Toplam
A1		0,24	0,16	0,4
A2		0,24	0,16	0,4
A3		0,12	0,08	0,2
Toplam		0,6	0,4	1,00

Tablo 3.1.2.3 Bağlantı Dengesizliğindeki Bir Popülasyonun Yapısı

A Geni	B Geni	B1	B2	Toplam
A1		0,30	0,25	0,55
A2		0,10	0,15	0,25
A3		0,15	0,05	0,20
Toplam		0,55	0,45	1,00

2*2 modelinde bağlantı dengesizliği katsayısı ($\mu_D=0$) ve bağlantı dengesizliği katsayısı ($\mu_D=0,15$) olan iki popülasyon için (1.2), (3.1), (3.2) numaralı üç istatistik ve merkezi olmama parametreleri (3.1.1 numaralı bölümde) tanıtılmıştır. Buna göre, birinde üç allel (A) diğerinde iki allel (B) olan iki lokus için, bağlantı dengesizliği katsayısı ($\mu_D=0$) ve bağlantı dengesizliği katsayısı ($\mu_D \neq 0$) olmak üzere iki popülasyondan simülasyon yöntemiyle çekilen $n=100$ genişlikteki örneklerden hesaplanan üç ayrı kuadratik istatistiğin formülleri aşağıdaki gibi yazılmıştır.

$$\chi_0^2 = n \cdot \sum_{i=1}^a \sum_{j=1}^b \frac{D_{ij}^2}{p_i \cdot (1 - p_i) \cdot p_j \cdot (1 - p_j)} \quad (3.5)$$

Burada a 1'inci Lokustaki allel sayısı, b 2'inci Lokustaki allel sayısıdır. D_{ij} , ilk lokusun i'inci alleli ile ikinci lokusun j'inci alleli arasındaki bağlantı dengesizliği

katsayısıdır. Meselâ A_1B_1 için bağlantı dengesizliği katsayısı $D_{11}=p_{11}-p_{1.}p_{.1}$ şeklinde yazılır ve genel olarak bağlantı dengesizliği katsayı

$$D_{ij}=p_{ij} - p_{i.}p_{.j} \quad (3.6)$$

Şeklinde yazılabilir.

$$\chi_1^2 = n \cdot \sum_{i=1}^a \sum_{j=1}^b \frac{(D_{ij} - \partial_{ij})^2}{p_{i.}(1-p_{i.})p_{.j}(1-p_{.j})} \quad (3.7)$$

$$\chi_2^2 = \sum_{i=1}^a \sum_{j=1}^b \frac{(x_{ij} - np_{ij})^2}{np_{ij}} \quad (3.8)$$

3.2 Simülasyonla Örnek Üretilen Popülasyonlar

İki allelli iki lokus için biri bağlantı dengesinde ($\mu_D=0$), diğeri bağlantı dengesizliğinde ($\mu_D=0,15$) olan iki popülasyondan $n=10, 20$ ve 50 genişliğinde on biner örnek çekilmiştir. Her simülasyon işlemi 20 defa tekrarlanmıştır. Buna göre araştırmada ele alınan faktörler iki bağlantı dengesizliği katsayısı (0 ve $0,15$) ve üç örnek genişliği ($10, 20$ ve 50) için olmak üzere $2*3=6$ popülasyonun her biri simülasyonla 10.000 defa üretilmiş ve bu altı simülasyonun her biri 20 defa tekrarlanmış; deneme tekniği tabiriyle simülasyon 20 tekerrürlü yapılmıştır. Çalışma için yazılan algoritma $\mu_D=0,15$ ve $n=20$ için Tablo 3.2.1.'de verilmiştir.

Tablo 3.2.1 Çalışmada Kullanılan R Programıyla Yazılan Algoritma (İki allelli iki lokus, $\mu_D=0,15$ ve $n=20$)

<pre>#parametreler: q1=P(AB), q2=P(Ab), #q3=P(aB), q4=P(ab) # nrep=deneme sayısı #ngen=örnek genişliği q1=.40 q2=.10 q3=.10 q4=.40 nrep=10.000 ngen=50 dort=q1*q4-q2*q3 ncpar=dort^2*ngen ncpar3=ngen*(q1^2+q2^2+q3^2+q4^2) dort ncpar ncpar3 d=c(rep(seq(0,0),nrep)) chi1=c(rep(seq(0,0),nrep)) chi2=c(rep(seq(0,0),nrep)) chi0=c(rep(seq(0,0),nrep)) for(j in 1:nrep){ x1=0 x2=0 x3=0 x4=0 y=runif(ngen) ncchi3=rchisq(ngen,3,ncpar3) ncchi1=rchisq(ngen,1,ncpar) cchi1=rchisq(ngen,1) cchi3=rchisq(ngen,3) par(mfrow=c(2,4)) for (i in 1:ngen){ ifelse(y[i]<=q1,(x1=x1+1),ifelse (y[i]<=(q1+q2),(x2=x2+1),ifelse (y[i]<=(q1+q2+q3),(x3=x3+1),(x4=x4+1)))) } d[j]=(x1*x4-x2*x3)/ngen^2 p1=(x1+x2)/ngen p2=(x1+x3)/ngen chi0[j]=ngen*d[j]^2/(p1*(1-p1)*p2*(1-p2))</pre>	<pre>chi1[j]=ngen*(d[j]-dort)^2/(p1*(1-p1)*p2*(1-p2)) chi2[j]=(x1-ngen*q1)^2/(ngn*q1)+(x2- ngn*q2)^2/(ngn*q2)+(x3-ngn*q3)^2/(ngn*q3)+(x4- ngn*q4)^2/(ngn*q4) } d=round(d,digits=3) d.table=table(d) d.table # d degerlerinin frekansini yazdir mean(d) # chi0 = nD^2/(p1*(1-p1)*p2*(1-p2)) # chi1 = n*(d-dort)^2/(p1*(1-p1)*p2*(1-p2)) # chi2 = tam chi-kare #histogramlar hist(d,ylab="frekans",xlab="D degerleri") hist(chi0,ylab="frekans",xlab="noncentral chikare") hist(chi1,ylab="frekans",xlab="central chikare") hist(chi2,ylab="frekans",xlab=" ??? chikare") hist(ncchi1,ylab="frekans",xlab="1 sd noncentral chi-kare") hist(ncchi3,ylab="frekans",xlab="3 sd noncentral chi-kare") hist(cchi1,ylab="frekans",xlab="1 sd central chi-kare") hist(cchi3,ylab="frekans",xlab="3 sd central chi-kare") ks.test(chi0,cchi1) ks.test(chi0,ncchi1) ks.test(chi0,cchi3) ks.test(chi0,ncchi3) ks.test(chi1,cchi1) ks.test(chi1,ncchi1) ks.test(chi1,cchi3) ks.test(chi1,ncchi3) ks.test(chi2,cchi1) ks.test(chi2,ncchi1) ks.test(chi2,cchi3) ks.test(chi2,ncchi3)</pre>
---	---

Birinde iki allel diğ erinde üç allele sahip iki lokus için biri bağlantı dengesinde, diğ eri bağlantı dengesizliğ inde olan iki popülasyondan $n=10, 20$ ve $50, 100$ genişliğ inde on biner örnek çekilmiştir. Her simülasyon işlemi yine 20 defa tekrarlanmıştır. Buna göre araştırmada ele alınan faktörler iki bağlantı dengesizliğ i

katsayısı ve dört örnek genişliği (10, 20 ve 50, 100) için olmak üzere $2*4=8$ popülasyonun her biri simülasyonla 10.000 defa üretilmiş ve bu sekiz simülasyonun her biri 20 defa tekrarlanmış; deneme tekniği tabiriyle simülasyon 20 tekerrürlü yapılmıştır. Çalışma için yazılan algoritma $\mu_D=0$ ve $n=100$ için Tablo 3.2.2.'de verilmiştir.

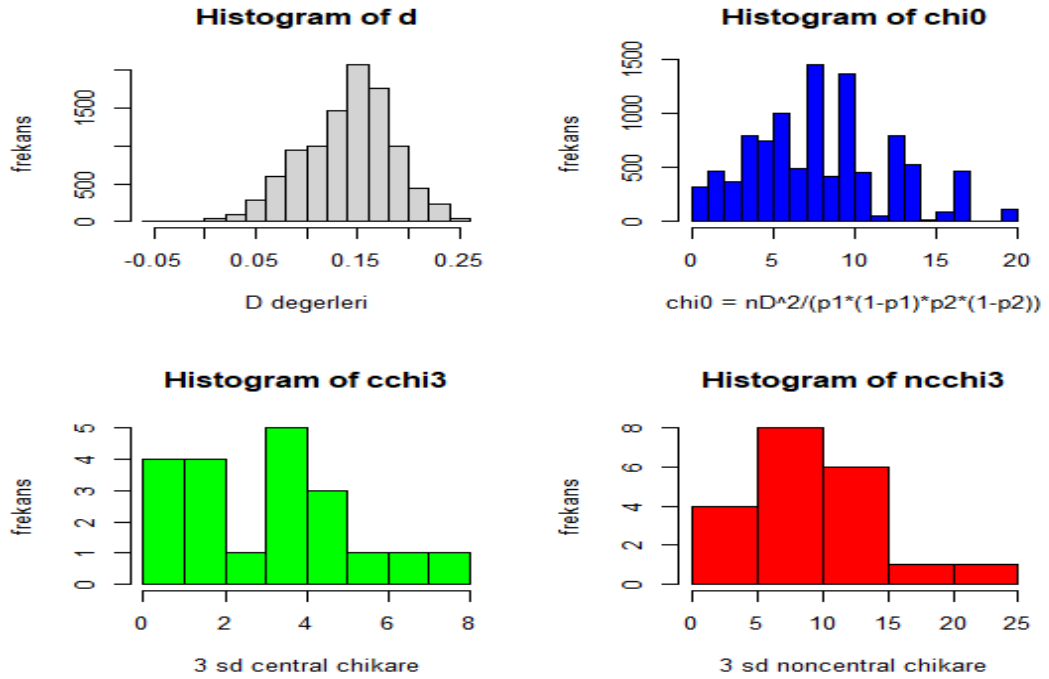
Tablo 3.2.2 Çalışmada Kullanılan R Programıyla Yazılan Algoritma (Birisini iki diğeri üç allelli iki lokus, $\mu_D=0$ ve $n=100$)

<pre># Denegede popülasyon q11=.24 q12=.16 q21=.24 q22=.16 q31=.12 q32=.08 p11=q11+q12 p12=q21+q22 p13=q31+q32 p21=q11+q21+q31 p22=q12+q22+q32 ngen=100 nrep=10.000 ist1=c(rep(seq(0,0),nrep)) ist3=c(rep(seq(0,0),nrep)) chisqr=rchisq(nrep,2,0) ist1=NULL ist3=NULL say1=0 say2=0 for (jj in 1:nrep) { x11=0;x21=0;x31=0 x12=0;x22=0;x32=0 N11=0;N12=0;N13=0 N21=0;N22=0 y=runif(ngen) for (i in 1:ngen){ ifelse(y[i]<=q11,(x11=x11+1),ifelse (y[i]<=(q11+q12),(x12=x12+1),ifelse (y[i]<=(q11+q12+q21),(x21=x21+1), ifelse (y[i]<=(q11+q12+q21+q22),(x22=x22+1), ifelse (y[i]<=(q11+q12+q21+q22+q31),(x31=x3 1+1),(x32=x32+1)))))) } }</pre>	<pre>if(x11==0){x11=1} if(x12==0){x12=1} if(x21==0){x21=1} if(x22==0){x22=1} if(x31==0){x31=1} if(x32==0){x32=1} d11=(x11/ngen)-p11*p21 d12=(x12/ngen)-p11*p22 d21=(x21/ngen)-p12*p21 d22=(x22/ngen)-p12*p22 d31=(x31/ngen)-p13*p21 d32=(x32/ngen)-p13*p22 N11=x11+x12; N12=x21+x22;N13=x31+x32 N21=x11+x21+x31 N22=x12+x22+x32 N=N21+N22 # x11;x12;x31;x21;x22;x32 # N11;N12;N13 # N21;N22 # N ist1[jj]=d11^2/(N11/N*(1-N11/N)*N21/N*(1- N21/N))+d12^2/(N11/N*(1-N11/N)*N22/N*(1- N22/N))+d21^2/(N12/N*(1-N12/N)*N21/N*(1- N21/N))+d22^2/(N12/N*(1-N12/N)*N22/N*(1- N22/N))+d31^2/(N13/N*(1-N13/N)*N21/N*(1- N21/N))+d32^2/(N13/N*(1-N13/N)*N22/N*(1-N22/N)) ist3[jj]=(x11-N11*N21/ngen)^2/(N11*N21/ngen)+(x12- N11*N22/ngen)^2/(N11*N22/ngen)+(x21- N12*N21/ngen)^2/(N12*N21/ngen)+(x22- N12*N22/ngen)^2/(N12*N22/ngen)+(x31- N13*N21/ngen)^2/(N13*N21/ngen)+(x32- N13*N22/ngen)^2/(N13*N22/ngen) ist1[jj]=ngn*ist1[jj] if(ist1[jj]>qchisq(0.95,2)) say1=say1+1 if(ist3[jj]>qchisq(0.95,5)) say2=say2+1 } par(mfrow=c(2,2)) hist(ist1,ylab="frekans",xlab="dij^2") hist(ist3,ylab = "Frekans",xlab = "(fgoz-fbek)^2/fbek") cat(nrep, "denemedeki bağımlı sayısı",say1,"olarak elde edilmiştir.") cat(nrep, "denemedeki bağımlı sayısı",say2,"olarak elde edilmiştir.") ks.test(ist1,chisqr) ks.test(ist3,chisqr)</pre>
---	--

4. SONUÇLAR

4.1 2*2 Lokus Modeli İçin Bulgular

Simülasyondan önce, bağlantı dengesizliği katsayısı sıfır olan popülasyondan üretilen örneklerden hesaplanan (1.2) ve (3.1) numaralı kuadratik istatistiklerin örnekleme dağılımının 1 serbestlik dereceli Khi-Kare dağılımına uygunluğu öngörülürken (3.2) numaralı kuadratik istatistik için 3 s.d.'li Khi-Kare dağılımına uygunluk öngörülüyordu. Tablo 4.1.1, 4.1.2 ve 4.1.3'te bu üç kuadratik formun örnekleme dağılımının, denenen altı popülasyonda ele alınan Khi-Kare dağılımlarına uygunluk için yapılan yirmişer Kolmogorov-Smirnov testinden kaç tanesinde hipotezin kabul edildiği ($p>,05$) verilmiştir. Üretilen örneklerin histogramları da yapılmış ve Kolmogorov-Smirnov testi yanında gözle de gözlem yapmak niyetiyle 20 tekerrürün her birinde Şekil 3.1.1.1'de verilen teorik dağılımlarla karşılaştırılmıştır. Bu histogramlara örnek olarak $n=20$ ve $\mu_D=0,15$ için tekerrürlerin birisinde üretilen 10.000 örnekten elde edilen D değerleri ve (1.2) numaralı istatistiğe ait histogramlar şekil 4.1.1'de gösterilmiştir.



Şekil 4.1.1 $\mu_D=0,15$ ve $n=20$ için üretilen 10.000 örnekten hesaplanan değerlerin grafikleri

D değerlerinin histogram grafiği (gri), (1.2) numaralı istatistikle elde edilen değerlerin histogram grafiği (mavi), 3 s.d.'li Khi-Kare değerlerinin histogram grafiği (yeşil), 3 s.d.'li merkezi olmayan (noncentral) Khi-Kare değerlerinin histogram grafiği (kırmızı) olarak şekil 4.1.1' de ifade edilmiştir.

Tablo 4.1.1'de (1.2) numaralı kuadratik istatistik için $n=10, 20, 50$ örnek genişliğine sahip bağlantı dengesizliği katsayısı $\mu_D=0$ ve $\mu_D=0,15$ olan sanal popülasyonlardan hesaplanan dağılımlardan $\mu_D=0$ ve $n=10$ örnek genişliğine sahip bir popülasyonda 20 kez tekrarlanan simülasyon sonucunda 1 serbestlik dereceli merkezi Khi-Kare dağılımına uygunluk gösteren tekrarların sayısı 19, 1 serbestlik dereceli merkezi olmayan Khi-Kare 18, 3 serbestlik dereceli merkezi Khi-Kareye 1 ve 3 serbestlik dereceli merkezi olmayan Khi-Kareye 0 tekrar bulunmuştur. Buna karşılık $\mu_D=0,15$ ve $n=10$ için 3 s.d.'li merkezi ve merkezi olmayan Khi-Kare 20 tekrarın 15'inde uygun bulunmuştur. $n=20$ ve 50 için bulunan sonuçlar aynı bulguyu teyit etmektedir.

Tablo 4.1.1 Altı Popülasyondan Elde Edilen Sonuçlar (1.2 numaralı istatistik için)

Bağlantı Dengesizliği	Örnek Genişliği(n)	Khi Kare uygunluk			
		1 s.d central	1 s.d noncentral	3 s.d central	3 s.d noncentral
$\mu_D=0$	10	19	18	1	0
	20	19	19	0	0
	50	19	20	0	0
$\mu_D=0,15$	10	0	2	15	15
	20	0	0	0	17
	50	0	0	0	16

Tablo 4.1.2'de ikinci kuadratik form olan (3.1) numaralı istatistiğin sonuçları görülmektedir. Bu istatistiğin örnekleme dağılımı $\mu_D=0$ için 1 s.d.'li merkezi ve merkezi olmayan Khi-Kareye uyum göstermektedir. Merkezi olmama parametresi $\mu_D=0$ için sıfır, dolayısıyla merkezi olmayan Khi-Kare ile merkezi Khi-Kare'nin aslında aynı dağılım olduğu dikkate alınırsa bu sonuç olması beklenendir.

Tablo 4.1.2 Altı Popülasyondan Elde Edilen Sonuçlar (3.1 numaralı istatistik için)

Bağlantı Dengesizliği	Örnek Genişliği(n)	Khi Kare uygunluk			
		1 s.d central	1 s.d noncentral	3 s.d central	3 s.d noncentral
$\mu_D=0$	10	19	18	1	0
	20	20	19	0	0
	50	19	19	0	0
$\mu_D=0,15$	10	17	17	0	0
	20	16	9	0	0
	50	14	0	0	0

Tablo 4.1.3'te üçüncü kuadratik form (3.2) numaralı istatistik için bulunan sonuçlar verilmiştir. Bu kuadratik form üretilen gamet frekansları ile aynı varsayımlara göre beklenen gamet frekansları arasındaki farkın karesine dayandığından bulunan sonuçlar normaldir.

Tablo 4.1.3 Altı Popülasyondan Elde Edilen Sonuçlar (3.2 numaralı istatistik için)

Bağlantı Dengesizliği	Örnek Genişliği(n)	Khi Kare uygunluk			
		1 s.d central	1 s.d noncentral	3 s.d central	3 s.d noncentral
$\mu_D=0$	10	0	0	19	5
	20	0	0	19	0
	50	0	0	20	0
$\mu_D=0,15$	10	1	3	19	3
	20	0	2	20	0
	50	0	0	16	0

4.2 3*2 Lokus Modeli İçin Bulgular

Bağlantı dengesizliği katsayısı sıfır olan popülasyondan üretilen örneklerden hesaplanan (3.5) ve (3.7) numaralı kuadratik istatistiklerinin örnekleme dağılımı Khi-Kare dağılımına uygun çıkmamıştır. Aynı şekilde bağlantı dengesizliği katsayısı sıfır olmayan popülasyondan üretilen örneklerden hesaplanan (3.5), (3.7) ve (3.8) numaralı kuadratik istatistiklerinin örnekleme dağılımı yine Khi-Kare dağılımına uygun çıkmamıştır fakat bağlantı dengesizliği katsayısı sıfır olan popülasyondan üretilen örneklerden hesaplanan Tablo 4.2.3'te gösterildiği gibi (3.8) numaralı kuadratik istatistiğin örnekleme dağılımı $n=100$ örnek genişliğinde 2 s.d'li Merkezi Khi-Kare dağılımına uygun çıkmıştır. Denenen sekiz popülasyonda ele alınan Khi-

Kare dağılımlarına uygunluk için yapılan yirmişer Kolmogorov-Smirnov testinden kaç tanesinde hipotezin kabul edildiği ($p>,05$) verilmiştir. Üretilen örneklerin histogramları da yapılmış bu histogramlara örnek olarak $n=100$ ve $\mu_D=0$ için tekerrürlerin birisinde üretilen 10.000 örnekten elde edilen (3.5) ve (3.8) numaralı istatistiklere ait histogramlar şekil 4.2.1’de gösterilmiştir.

Tablo 4.2.1 Sekiz Popülasyondan Elde Edilen Sonuçlar (3.5 numaralı istatistik için)

Bağlantı Dengesizliği	Örnek Genişliği(n)	Khi Kare uygunluk			
		2 s.d central	2 s.d noncentral	5 s.d central	5 s.d noncentral
$\mu_D=0$	10	0	0	0	0
	20	0	0	0	0
	50	0	0	0	0
	100	0	0	0	0
$\mu_D\neq 0$	10	0	0	0	0
	20	0	0	0	0
	50	0	0	0	0
	100	0	0	0	0

Tablo 4.2.2 Sekiz Popülasyondan Elde Edilen Sonuçlar (3.7 numaralı istatistik için)

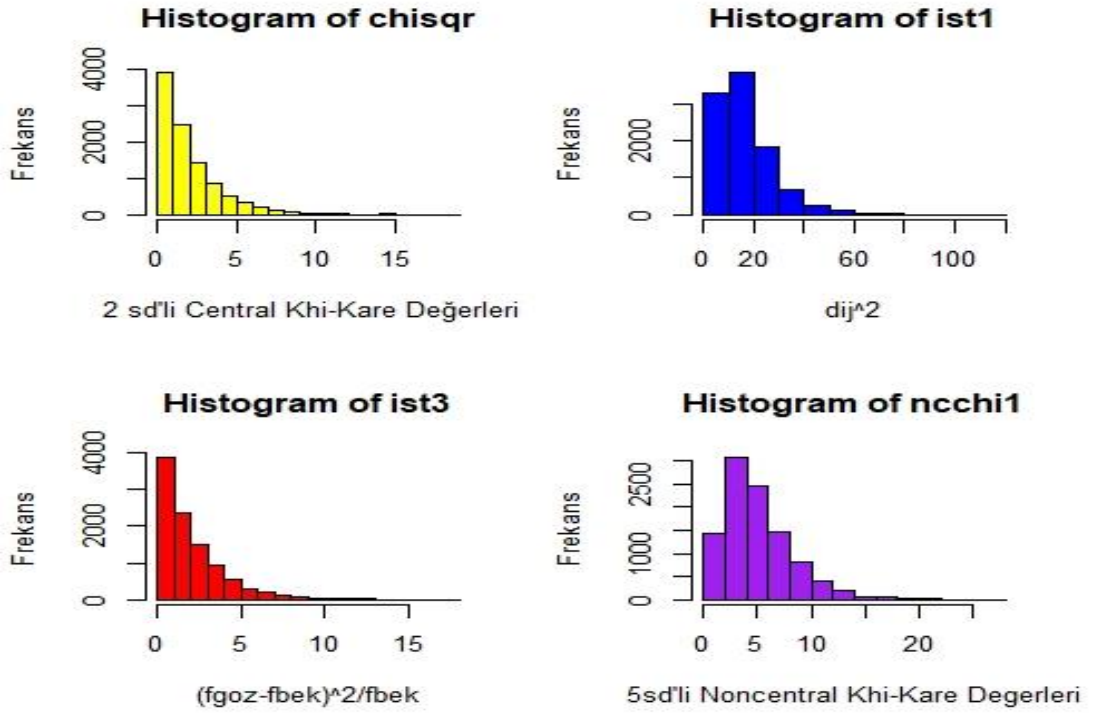
Bağlantı Dengesizliği	Örnek Genişliği(n)	Khi Kare uygunluk			
		2 s.d central	2 s.d noncentral	5 s.d central	5 s.d noncentral
$\mu_D=0$	10	0	0	0	0
	20	0	0	0	0
	50	0	0	0	0
	100	0	0	0	0
$\mu_D\neq 0$	10	0	0	0	0
	20	0	0	0	0
	50	0	0	0	0
	100	0	0	0	0

Denenen sekiz popülasyonda ele alınan Khi-Kare dağılımlarına uygunluk için yapılan yirmişer Kolmogorov-Smirnov testinden kaç tanesinde hipotezin kabul edildiği ($p>,05$) tablo 4.2.1 ve tablo 4.2.2’de verilmiştir. Tablolardan anlaşılacağı gibi (3.5) ve (3.7) numaralı kuadratik istatistikler hem dengede hem de dengede olmayan popülasyonlardan elde edilen sonuçlara göre denenen dağılımlara uygun bir dağılım göstermemişlerdir. Bu (3.5) ve (3.7) numaralı kuadratik istatistiklerin normal dağılıma da uygunlukları kontrol edilmiştir.

Tablo 4.2.3 Sekiz Popülasyondan Elde Edilen Sonuçlar (3.8 numaralı istatistik için)

Bağlantı Dengesizliği	Örnek Genişliği(n)	Khi Kare uygunluk			
		2 s.d central	2 s.d noncentral	5 s.d central	5 s.d noncentral
$\mu_D=0$	10	0	0	0	0
	20	0	0	0	0
	50	12	0	0	0
	100	18	0	0	0
$\mu_D \neq 0$	10	0	0	0	0
	20	0	0	0	0
	50	0	0	0	0
	100	0	0	0	0

Tablo 4.2.3'te üçüncü kuadratik form olan (3.8 numaralı istatistik) için bulunan sonuçlar verilmiştir. Bu kuadratik form üretilen gamet frekansları ile aynı varsayımlara göre beklenen gamet frekansları arasındaki farkın karesine dayanmaktadır. Tablo 4.2.3 incelenirse (3.8 numaralı istatistik) $n=100$ olduğunda 2 s.d.'li Khi-Kare dağılımına uygun olduğu görülmektedir. Daha az örnek genişliğine sahip popülasyonlarda (3.8 numaralı istatistik) 2 s.d.'li Khi-Kare dağılımına uygun olmadığı görülmektedir. $n=10, 20, 50$ için bulunan sonuçlar, bu örnek genişliklerinin Khi-Kare dağılımının ön şartlarını sağlamaya yetmemesinden kaynaklanıyor olabilir. Çünkü bu kadar küçük bir örnek genişliğinde, neredeyse bütün sınıflarda beklenen frekanslar 5'ten küçük çıkacaktır. Bu durumda popülasyon genetiği ile ilgili parametre tahmini veya hipotez kontrolü çalışmalarında örnek genişliğinin 50'den büyük olması gerektiği anlaşılmaktadır.



Şekil 4.2.1 $\mu_D=0$ ve $n=100$ için üretilen 10.000 örnekten hesaplanan değerlerin grafikleri

2 s.d.'li Khi-Kare değerlerinin histogram grafiği (sarı), (3.5) numaralı istatistikle elde edilen değerlerin histogram grafiği (mavi), (3.8) numaralı istatistikle elde edilen değerlerin histogram grafiği (kırmızı), 5 s.d.'li merkezi olmayan (noncentral) Khi-Kare değerlerinin histogram grafiği (mor) olarak şekil 4.2.1' te ifade edilmiştir.

5. TARTIŞMA ve ÖNERİLER

Çalışmada bulunan sonuçlara göre 2*2 lokus modelinde (Tablo 4.1.1) bağlantı dengesizliği katsayısı sıfır olmayan popülasyonlarda birinci kuadratik form (1.2 numaralı istatistik) merkezi olmayan Khi-Kare dağılımı göstermektedir. Sonuçları Tablo 4.1.2’de verilen ikinci kuadratik form (3.1 numaralı istatistik) beklendiği gibi, bağlantı dengesizliği parametresinden farkın karesi olarak hesaplandığından 1 s.d.’li Khi-Kare dağılımı göstermektedir. Tablo 4.1.3’te verilen üçüncü kuadratik form (3.2 numaralı istatistik) için sonuçlar bir çeşit kontrol grubu oluşturmak üzere elde edilmiş sonuçlardır. Beklendiği gibi burada sonuçlar her durumda 3 s.d.’li merkezi Khi-Kare dağılımına uygunluk göstermektedir.

3*2 lokus modelinde bağlantı dengesizliği katsayısı sıfır olan popülasyonda birinci kuadratik form (3.5 numaralı istatistik) tablo 4.2.1’de görüldüğü gibi denenen dağılımlara uygun bir dağılım göstermemiştir. Yine aynı şekilde tablo 4.2.2’de görüldüğü gibi ikinci kuadratik formda (3.7 numaralı istatistik) denenen dağılımlara uygun bir dağılım göstermemiştir. Bunlara ek olarak normal dağılıma uygunlukları denenen bu (3.5, 3.7 ve 3.8 numaralı istatistik) üç kuadratik istatistik için sonuçlar, daha başka dağılımları da çalışmak düşüncesiyle teze konulmamış olup bunun yanında G istatistiği de çalışılmış ve yine sonuçları teze konulmamıştır. Bağlantı dengesinde olan popülasyonda üçüncü kuadratik istatistik formu olan (3.8 numaralı istatistik) n=100 örnek genişliğinde 2 s.d’li merkezi Khi-Kare dağılımına uygun olduğu görülmüştür. Daha az örnek genişliğine sahip popülasyonlarda (3.8 numaralı istatistik) 2 s.d’li Khi-Kare dağılımına uygun olmadığı görülmektedir. n=10,20,50 için bulunan sonuçlar, bu örnek genişliğinin Khi-Kare dağılımının ön şartlarını sağlamaya yetmemesinden kaynaklanıyor olabilir. Bu durumda popülasyon genetiği ile ilgili parametre tahmini veya hipotez kontrolü çalışmalarında örnek genişliğinin 50’den büyük olması gerektiği anlaşılmaktadır.

Bağlantı dengesizliğindeki popülasyonda ise üç istatistik için de denenen dağılımlar arasında uygun bir dağılım bulunamamıştır. Sonraki süreçte üzerinde başka dağılımlara uygunluk üzerinde de çalışılması gerektiği kanaatine varılmıştır.

KAYNAKLAR

- Seber, G. A. (1963). The Non-Central Chi-Squared and Beta Distributions. *Biometrika*, 50 (3/4), 542-544.
- Crow, J. Kimura, M. (1970). *An Introduction to Population Genetics*. Newyork: Harper and Row Publisher Inc.
- Asmussen, M. A., Clegg, M. T. (1981). Use of Restriction Fragment Length Polymorphisms for Genetic Counseling: Population Genetic Considerations. *American Society of Human Genetics*, 34, 369-380.
- Kavuncu, O., Kesici, T. (1982). Calculation of The Genetic Covariance Between Relatives for Two-Locus Models with Dominance. *Ankara Üniversitesi Ziraat Fakültesi 1982 Yıllığı*, Sayı, 1,2,3,4.
- Kavuncu, O. (1983). The Total Genetic Variance and The Genetic Variance Between Full-Sib Lines in Self Fertilized Plants for Two-Locus Models with Dominance. *Ankara Üniversitesi Ziraat Fakültesi 1983 Yıllığı*, Sayı, 1,2,3,4.
- Kavuncu, O., Düzgüneş O. (1983). Rasgele Çiftleşen Popülasyonlarda Bağlantı Dengesizliğinin Genetik Varyansa Etkisi. *Ege Üniversitesi Mühendislik Bilimleri Dergisi*, Seri E Uygulamalı İstatistik.
- Kavuncu, O. (1984). Rastgele Çiftleşen Küçük Popülasyonlarda Bağlantı Dengesizliğinin Eklemeli Genetik Varyans Üzerine Etkileri. *Doğa Bilim Dergisi*, 8(1), 25-32.
- Kavuncu, O. (1987). Effect of Linkage Disequilibrium on the Additive and Dominance Variances in a Random Mating Small Population. *Communication Journal of Science Published by Ankara University*, Al, 36: 131–142.
- Weir, B. S. (1996). *Genetic Data Analysis II*. Sunderland, MA: Sinauer Assoc., Inc.
- Zapata, C., Alvarez, G., Carollo, C. (1997). Approximate Variance of The Standardized Measure of Gametic Disequilibrium D' . *American Journal of Human Genetics*, 61, 771–774.
- Excoffier, L., Slatkin, M. (1998). Incorporating Genotypes of Relatives into a Test of Linkage Disequilibrium. *American Journal of Human Genetics*, 62, 171–180
- Sabatti, C., Risch, N. (2002). Homozygosity and Linkage Disequilibrium. *Genetics* 160, 1707–1719.
- Lee, L. C. (2003). Searching for Disease-Susceptibility Loci by Testing for Hardy-Weinberg Disequilibrium in a Gene Bank of Affected Individuals. *American Journal Of Epidemiology*, 158,397–400.

- Gillespie, J. H. (2004). *Population Genetics*. Baltimore and London: The Johns Hopkins University Press.
- Hamilton, D.C., Cole, D. E.C. (2004). Standardizing a Composite Measure of Linkage Disequilibrium. *Annals of Human Genetics*, 68, 234–239.
- Zaykin, D. V. (2004). Bounds And Normalization of The Composite Linkage Disequilibrium Coefficient. *Genetic Epidemiology*, 27 (3), 252–257.
- Kim, Y., Nielsen, R. (2004). Linkage Disequilibrium as a Signature of Selective Sweeps. *Genetics*, 167, 1513–1524. DOI: 10.1534/103.025387.
- Yang, W., Huang, J., Yao, C., Su, S., Liu, D., Ge, D., Gu, D. (2004). Linkage and Linkage Disequilibrium Analysis of The Lipoprotein Lipase Gene with Lipid Profiles in Chinese Hypertensive Families. *Clinical Science* 108, 137–142.
- Ioannidis, J. P., Salanti, G., Amountza, G., Ntzani E. E. (2005). Hardy–Weinberg Equilibrium in Genetic Association Studies: An Empirical Evaluation of Reporting, Deviations, and Power. *European Journal of Human Genetics*, 13, 840–848.
- Dekkers, J. C. M., Zhao, H., Nettleton, D., Soller, M. (2005). Evaluation of Linkage Disequilibrium Measures between Multi-Allelic Markers as Predictors of Linkage Disequilibrium between Markers and QTL. *Genet. Res., Camb.*, 86, 77–87. doi:10.1017/S001667230500769X.
- Guo, J., Yin, X. L., Ma, W. Q., Tang, M. L. (2006). A Test of Homogeneity of Hardy Weinberg Disequilibrium Across Strata. *European Journal Of Human Genetics*, 14, 1223–1230. doi:10.1038/sj.ejhg.5201689.
- Song, Y. S., Song, J. S. (2006). Analytic Computation of The Expectation of The Linkage Disequilibrium Coefficient R^2 . *National Science Foundation*, 1-25.
- Hamilton, D. C., Liu, Q., Cole, D. E. C. (2006). Approximate Variance for A Standardized Composite Measure of Linkage Disequilibrium. *Annals of Human Genetics*, 70, 535–540. doi: 10.1111/j.1469-1809.2006.00254.x.
- Mano, S. (2006). Evolution Of Linkage Disequilibrium of The Founders in Exponentially Growing Populations. *Theoretical Population Biology*, 71, 95–108. doi:10.1016/j.tpb.2006.07.003.
- Pereira, C. A. B., Nakano, F., Stern, J. M., Whittle, M. R. (2006). Genuine Bayesian Multiallelic Significance Test for The Hardy-Weinberg Equilibrium Law. *Genetics and Molecular Research*, 5 (4), 619-631.
- Wang, H., Lin, C.H., Service, S., Chen, Y., Freimer, N., Sabatti, C. (2006). Linkage Disequilibrium and Haplotype Homozygosity in Population Samples Genotyped at a High Marker Density. *Human Heredity*, 62, 175–189. doi: 10.1159/000096599.

- Rosenberg, N. A., Blum, M. G. B. (2006). Sampling Properties of Homozygosity-Based Statistics for Linkage Disequilibrium. *Mathematical Biosciences*, 208, 33–47. doi:10.1016/j.mbs.2006.07.001.
- Mahasirimongkol, S., Chantratita, W., Promso, S., Pasomsab, E., Jinawath, N., Jongjaroenprasert, W., Lulitanond, V., Krittayapoositpot, P., Tongsimma, S., Sawanpanyalert, P., Kamatani, N., Nakamura, Y., Sura, T. (2006). Similarity of The Allele Frequency and Linkage Disequilibrium Pattern of Single Nucleotide Polymorphisms in Drug-Related Gene Loci between Thai and Northern East Asian Populations: Implications for Tagging SNP Selection in Thais. *Journal of Human Genetics*, 51, 896–904. doi 10.1007/s10038-006-0041-1.
- Hartl D. L., Clark A. G. (2007). *Principles of Population Genetics*. USA: Sinauer Assoc.
- Yamada, R., Matsuda F. (2007). A Novel Method to Express SNP-Based Genetic Heterogeneity, Ψ , and Its Use to Measure Linkage Disequilibrium for Multiple SNPs, Dg' and to Estimate Absolute Maximum of Haplotype Frequency. *Genetic Epidemiology*, 31, 709–726. doi: 10.1002/gepi.20235.
- Du, F. X., Clutter, A. C., Lohuis, M. M. (2007). Characterizing Linkage Disequilibrium in Pig Populations. *International Journal of Biological Sciences*, 3(3), 166-178.
- Hamilton, D. C., Cole, D. E. C. (2007). Testing for Equality of Standardized Composite Measures of Linkage Disequilibrium. *Annals of Human Genetics* 72, 292–296. doi: 10.1111/j.1469-1809.2007.00404.x.
- Zaykin, D.V. Pudovkin, A., Weir, B. S. (2008). Correlation-Based Inference for Linkage Disequilibrium with Multiple Alleles. *Genetics*, 180, 533–545. doi: 10.1534/genetics.108.089409.
- Zhang, X. Q., Rao, Y. S., Liang, Y., Xia, M. N., Shen, X., Du, Y. J., Luo, C. G., Nie, Q.H., Zeng, H. (2008). Extent of Linkage Disequilibrium in Wild and Domestic Chicken Populations. *Hereditas*, 145, 251-257. doi: 10.1111/j.1601 5223.2008.02043.x.
- Zintzaras, E. (2008). Variance Estimation of Allele-Based Odds Ratio in The Absence of Hardy–Weinberg Equilibrium. *Eur J Epidemiol*, 23,323–326. doi: 10.1007/s10654-008-9242-6.
- Rogers, A. R., Huff, C. (2009). Linkage Disequilibrium between Loci with Unknown Phase. *Genetics*, 182, 839–844. doi: 10.1534/genetics.108.093153.
- Wang, T., Zeng, Z. B. (2009). Contribution of Genetic Effects to Genetic Variance Components with Epistasis and Linkage Disequilibrium. *BMC Genetics*, 10, 52. doi:10.1186/1471-2156-10-52.
- Zapata, C. (2011). On The Uses and Applications of The Most Commonly Used Measures of Linkage Disequilibrium from The Comparative Analysis of Their Statistical Properties. *Human Heredity*, 71, 186–195. doi: 10.1159/000327732.

- Nielsen R., Slatkin, M. (2013). *An Introduction to Population Genetics*. Sunderland, Massachusetts U. S.A.: Sinauer Associates, Inc. Publishers.
- İlhan, D. (2018). Linkage Disequilibrium (LD) Analysis in Alfalfa (*Medicago sativa* L.) Populations Spreading in Different Geographies of The World. *KSU J. Agric Nat*, 21(6),846-853. doi:10.18016/ksutarimdog.vi.452852.
- Raya, L. G., Silio, L., Rauw, W. M., Cortés, L. A. G., Rodríguez, C. (2018). Extent of Third-Order Linkage Disequilibrium in a Composite Line of Iberian Pigs. *BMC Genetics*, 19,60. doi.org/10.1186/s12863-018-0661-4.
- Cheng, H., Qu, J., Kachman, S. D., Garrick, D., Fernando, R. L. (2019). Exact Distribution of Linkage Disequilibrium in The Presence of Mutation, Selection or Minor Allele Frequency Filtering. *Genetics*, doi: 10,1534/genetics.XXX.XXXXXX.
- Jiang, L., Berihulay, H., Islam, R., Ma, Y. (2019). Genome-Wide Linkage Disequilibrium and The Extent of Effective Population Sizes in Six Chinese Goat Populations Using a 50K Single Nucleotide Polymorphism Panel. *Animals*, 9, 350. doi:10.3390/ani9060350.
- Hui, T. Y. J., Burt, A. (2020). Estimating Linkage Disequilibrium from Genotypes under Hardy-Weinberg Equilibrium. *BMC Genetics*, 21, 21. doi.org/10.1186/s12863-020-0818-9.
- Kökçü, S., Kavuncu, O., Gedik, Y. (2020). Bağlantı Dengesizliği Katsayısının Örnekleme Dağılımı Üzerine Bir Simulasyon Çalışması. I. Uluslararası Uygulamalı İstatistik Konferansı, 142-150, Tokat.
- Kökçü, S., Kavuncu, O., Gedik, Y., Yılmaz, F., Akdoğan Y., Karakaya, K. (2021a). Bağlantı Dengesizliğinin Genotipik Varyans ve Unsurlarına Etkisi Üzerine bir Simulasyon Çalışması: I- Bir lokusta Dominans olan Modelde Popülasyon Parametreleri. II. Uluslararası Uygulamalı İstatistik Konferansı, 223, Tokat.
- Kökçü, S., Kavuncu, O., Gedik, Y. (2021b). Bağlantı Dengesizliğinin Genotipik Varyans Ve Unsurlarına Etkisi Üzerine Bir Simulasyon Çalışması: II- Bir Lokusta Dominans Diğer Lokusta Eklemeli Gen Etkisi Olan Modelde Popülasyon Parametrelerine Bağlantı Dengesizliğinin Etkisi. III. International Agricultural, Biological & Life Science Conference, 855-866, Edirne