ANDRÉA CARLA BASTOS ANDRADE

LINKAGE DISEQUILIBRIUM AND HAPLOTYPE BLOCK PATTERNS IN POPCORN POPULATIONS

Thesis submitted to the Breeding and Genetics Graduate Program of the Universidade Federal de Viçosa, in partial fulfillment of the requirements to obtain the degree of *Doctor Scientiae*.

Advisor: José Marcelo Soriano Viana

VIÇOSA – MINAS GERAIS

2019

Ficha catalográfica preparada pela Biblioteca Central da Universidade Federal de Viçosa - Câmpus Viçosa

T
A553L
2019

Andrade, Andréa Carla Bastos, 1989-
Linkage disequilibrium and haplotype block patterns in popocorn populations / Andréa Carla Bastos Andrade. – Viçosa, MG, 2019.
37 f. : il. (algumas color.) ; 29 cm.

Texto em inglês.
Orientador: Jose Marcelo Soriano Viana.
Tese (doutorado) - Universidade Federal de Viçosa.
Referências bibliográficas: f. 33-37.

1. Gametas. 2. Haplótipos. 3. Genética vegetal. 4. Milho de pipoca - Populações - Melhoramento genético. I. Universidade Federal de Viçosa. Departamento de Biologia Geral. Programa de Pós-Graduação em Genética e Melhoramento. II. Título.

CDD 22. ed. 576.58

ANDRÉA CARLA BASTOS ANDRADE

LINKAGE DISEQUILIBRIUM AND HAPLOTYPE BLOCK PATTERNS IN POPCORN POPULATIONS

Thesis submitted to the Breeding and Genetics Graduate Program of Universidade Federal de Viçosa, in partial fulfillment of the requirements to obtain the degree of *Doctor Scientiae*.

APPROVED: May 17, 2019.

Andréa Carla Bastos Andrade
Author

José Marcelo Soriano Viana
Advisor

À minha mãe Joseni,

À minha irmã Juliana,

À minha avó Carmen (in memoriam),

À minha madrinha Darcy (in memoriam),

DEDICO

ACKNOWLEDGEMENTS

Agradeço a Deus pela vida, pelos dons concedidos, por ser fortaleza ao longo dessa caminhada.

À Universidade Federal de Viçosa (UFV) e ao Programa de Pós-Graduação em Genética e Melhoramento, pela oportunidade de conduzir meus estudos de doutorado.

Ao meu orientador, Professor José Marcelo Soriano Viana, por me acolher sob sua orientação e pelos ensinamentos compartilhados.

À minha mãe Joseni e minha irmã Juliana, pelo amor incondicional que me tornou a mulher que sou, por em nenhum momento desacreditar de mim. Agradeço pelos conselhos e cuidado constantes.

Aos demais familiares, que mesmo distantes mantêm um elo de carinho, empatia e respeito.

Ao meu namorado Nilo, por ser carinho, atenção, incentivo e compreensão ao longo desses anos.

Aos amigos Perla, Jonathan, Filipe e Alysson, amizades que cultivei nos períodos de Graduação e Mestrado, agradeço pelo o que sonhamos juntos e estamos realizando a cada dia.

Aos primeiros amigos da Pós-Graduação em Genética e Melhoramento, Thaís e Téo, pelo companheirismo que permanece nesses quatro anos.

Aos amigos que ganhei nas andanças por Viçosa, com os quais convivi desde à Miltão, até à República Safe Zone: gratidão à Dreicita, Déa Souza, Gabi, Fred, Letícia Let's, Jonatas e Miller, por serem família.

Ao Coral da UFV e Madrigal da UFV, por me proporcionar a liberdade musical e me "tornar expressão musical" em todos os momentos, sobretudo àqueles momentos em que quase fui vencida pela exaustão. Um agradecimento especial ao Maestro Ciro Tabet e aos grandes amigos que conquistei nesse espaço, por me trazerem a alegria de viver a arte em diferentes nichos e por ser sorriso nos ensaios cotidianos.

Ao Professor Fabyano, por não medir esforços em auxiliar os estudantes que o procuram.

Aos companheiros do Programa Milho Pipoca, Vinicius, Geísa, Leonardo, Vitor, Aloísio e, de modo muito especial, ao Helcio, pela parceria nos trabalhos, atenção e paciência.

Aos Professores, funcionários e estudantes dos laboratórios de Informática do Melhoramento Animal, Biometria e do Setor de Meteorologia Agrícola, por abrirem as portas para a realização das análises dos estudos que conduzi.

Ao Grupo de Estudos em Genética e Melhoramento da UFV (GenMelhor), pelo ganho pessoal e profissional nos dois anos em que fui Coordenadora de Eventos e Coordenadora Geral.

Ao movimento *Uai Norma Lúcia!*, por aprofundar meu autoconhecimento, ajudar-me a enxergar além do caminho que escolhi percorrer, e todas as combinações possíveis do que há ao meu redor. Gratidão às "Norminhas" Gabi, Dreicita, Vanessa e Duanny.

Aos amigos e parceiros musicais do Déa Bastos Trio e banda Mvukáà (Alan, Lucas, Romário, Emerik e Murilo), por despertarem em mim um anseio que permanecia latente, mas que a cada dia que passa se faz presente.

Aos Professores e Pesquisadores Fabyano, Renata, Daniele e Antônio Carlos pela disponibilidade de participação na banca e pelas valiosas contribuições.

Ao Conselho Nacional de Desenvolvimento Cientifico e Tecnológico (CNPq), à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e à fundação de apoio à pesquisa do Estado de Minas Gerais (FAPEMIG), pelo apoio financeiro.

"Desde que me apaixonei por genética e melhoramento, tenho ouvido que tal ciência é uma arte… Hoje sou da música, da dança, da pintura, da poesia e do melhoramento genético: uma artista completa." (Déa Bastos)

ABSTRACT

ANDRADE, Andréa Carla Bastos, D.Sc., Universidade Federal de Viçosa, May, 2019. **Linkage disequilibrium and haplotype block patterns in popcorn populations.** Advisor: José Marcelo Soriano Viana.

Linkage disequilibrium (LD) and haplotype blocks have been a preliminary and powerful knowledge for genomic investigates in humans, animals and plants. We aimed to describe the LD measures, D' and r^2, and to characterize haplotype block patterns in popcorn populations (synthetic and breeding population) from tropical and temperate origins (biparental population). Three populations were genotyped by genotyping-by-sequencing (GBS), 190 or 192 individual from each population. After the data quality control the final marker set remaining for analysis ranged from 75,000 to 76,055 SNPs. D' and r^2 were assessed per chromosome, and haplotype blocks were characterized by number and size of block, and SNP number. LD and haplotype blocks were also evaluated within 12 genes containing polymorphisms related to the biosynthesis of zein, starch, cellulose and fatty acids. Our results showed that the synthetic is the population with higher LD, followed by the biparental population. The lower average D' value in the breeding population reflects its recombination history. We observed low average r^2 values in the populations. There is an initially higher LD decrease for SNPs separated by 51-100 kb (3 to 7% for D' and 28 to 66% for r^2, on average) and then a gradual decrease to the minimum LD value for SNPs separated by 451-500 kb. The number and length of the haplotype blocks and the number of SNPs per haplotype block were proportional to the average r^2. However, it is not expected a significant advantage of haplotype-based association study and along generations genomic selection, due to the reduced number of SNPs in the haplotype blocks (2 to 3). The LD intragenic approach revealed that tropical populations (synthetic and breeding population) developed by breeding program, based on expansion volume, have higher LD and presence of haplotype blocks. However, we cannot infer that the higher r^2 values observed in 11 of the 12 genes are due to selection for quality in populations.

Keywords: Gametic Phase Disequilibrium. Haplotypes. Biparental Population. Synthetic. Breeding Population.

RESUMO

ANDRADE, Andréa Carla Bastos, D.Sc., Universidade Federal de Viçosa, maio de 2019. **Padrões de desequilíbrio de ligação e blocos de haplótipos em populações de milho pipoca.** Orientador: José Marcelo Soriano Viana.

Desequilíbrio de ligação (DL) e blocos de haplótipos têm sido um conhecimento preliminar e poderoso para estudos genômicos em humanos, animais e plantas. Nós objetivamos descrever as medidas de DL, D' e r^2, e caracterizar os padrões de blocos de haplótipos em populações de milho-pipoca de origem tropical (sintético e população de melhoramento) e temperada (população biparental). As três populações foram genotipadas por genotipagem por sequenciamento (GBS), 190 ou 192 indivíduos de cada população. Após o controle de qualidade dos dados, o conjunto final de marcadores remanescentes para análise variou de 75.000 a 76.055 SNPs. D' e r^2 foram avaliados por cromossomo, e os blocos de haplótipos foram caracterizados por número e tamanho de blocos, e número de SNP. DL e blocos de haplótipos foram também avaliados em 12 genes contendo polimorfismos relacionados à biossíntese de zeína, amido, celulose e ácidos graxos. Nossos resultados mostraram que o sintético é a população com maior LD, seguida pela população biparental. Observamos valores médios baixos de r^2 nas populações. Há um decréscimo de LD inicialmente mais alto para os SNPs separados por 51-100 kb (3 a 7% para D' e 28 a 66% para r^2, em média) e então uma diminuição gradual para o valor mínimo de LD para SNPs separados por 451- 500 kb. O número e a extensão dos blocos de haplótipos e o número de SNPs por bloco de haplótipos foram proporcionais à média de r^2. No entanto, não se espera uma vantagem significativa do estudo de associação baseado em haplótipos e na seleção genômica ao longo das gerações, devido ao número reduzido de SNPs nos blocos de haplótipos (2 a 3). A abordagem intragênica de LD revelou que as populações tropicais (sintético e população de melhoramento) desenvolvidas pelo programa de melhoramento genético, com base no volume de expansão, têm maior LD e presença de blocos de haplótipos. No entanto, não podemos inferir que os maiores valores de r^2 observados em 11 dos 12 genes são devidos a seleção para qualidade em populações.

Palavras-chave: Desequilíbrio de Fase Gamética. Haplótipos. População Biparental. Sintético. População de Melhoramento.

FIGURE LIST

TABLE LIST

SUMMARY

1 INTRODUCTION

The knowledge about linkage disequilibrium (LD) was developed as a necessity for large-scale human studies, being subsequently applied in animal and plant genetic studies (LARSSON; LIPKA; BUCKLER, 2013). LD information increase power and precision in association mapping, define optimal marker spacing, and identify recombination hot-spots and regions influenced by natural selection (QANBARI et al., 2010). The feasibility and efficiency of genomic selection and association mapping depend strongly on the extent, distribution and structure of LD (KHATKAR et al., 2007).

Furthermore, LD also allows the identification of haplotype blocks (GUPTA; RUSTGI; KULWAL, 2005). Previously, haplotype studies were reported as necessary in designing and interpreting association studies of genotype and phenotype (GABRIEL et al., 2002). Recent studies have presented that using haplotype blocks as predictors may improve genomic predictions, since haplotypes are in stronger linkage disequilibrium with the quantitative trait loci (QTL) (CUYABANO; SU; LUND, 2015).

In a brief history, the interest in LD grew rapidly in the 1980s once the usefulness of LD for gene mapping became evident and large-scale surveys of closely linked loci became feasible (SLATKIN, 2008). In the early 2000s, the interest in humans LD was mainly due to the prospect of large-scale association studies to identify genes controlling complex disease (PRITCHARD; PRZEWORSKI, 2001). In 2002, an investigation at an intragenic level with 50 single nucleotide polymorphisms (SNPs) and 14 microsatellites genotyped from five genes of two X-chromosomal regions showed the differences and similarities around LD extent and haplotype-frequency distributions in small human populations in northern Eurasia (KAESSMANN et al., 2002).

Bourgain, Genin and Clerget-Darpoux (2002) evaluated haplotype methods to detect the role of candidate genes using intragenic SNPs, in a simulation study, and concluded that haplotypic methods perform better in situations of heterogeneity comparing to single SNP testing strategy. Mueller et al. (2005) described the SNP allelic variation within candidate-gene regions for complex traits in European populations, with sample size ranging from 98 to 170 individuals. They observed population differences regarding haplotype frequencies and LD structure, and concluded that these differences may affect the power to detect phenotype-genotype associations.

In the last ten years, human LD patterns from 17 population samples were used to novel inference on the complex emergence of modern *Homo sapiens* out of the African evolutionary cradle and their subsequent colonization of the globe (MCEVOY et al., 2011). Recently, LD was applied to ordering, orienting and positioning linked sequences to establish reference genome sequences by de novo sequence assembly, presented with a most informative method (PENGELLY; COLLINS, 2018).

An increasing number of studies have also aimed to characterize LD in domestic animals, especially in cattle (QANBARI et al., 2010). McKay et al. (2007) observed the importance to use a high-density of marker loci to determine LD extent in cattle, since linkage disequilibrium persists over a limited distance. Khatkar et al. (2008) computed LD estimates for 1,566,890 pair-wise SNPs combination on each chromosome and a sample of 365,400 SNP pairs from different chromosomes. They authors reported an ideal number of SNPs for association mapping in Holstein-Friesian cattle.

Using haplotype approach, Villumsen, Janss and Lund (2009) demonstrated through a simulation study that genomic selection models with haplotype effects perform better than single SNP effects, and that the optimal length of haplotypes will depend on factors such as LD, marker density, and population. Cuyabano, Su and Lund (2014) built haplotype blocks, based on the LD between markers, and used them to successfully predict three economically important traits. One year later, it was verified that haplotype blocks used as predictors can improve the reliability of genomic prediction (CUYABANO; SU; LUND, 2015).

In plants, LD mapping has been proposed to dissect the genetic basis of quantitative traits (TRUNTZLER et al., 2012). A pioneer study in maize (*Zea mays* L.) from Tenaillon et al. (2001) evaluated 25 individuals genotyped in 21 loci distributed along chromosome 1, to compare genetic diversity between exotic landraces and inbred lines, they found intragenic LD declining within 100–200 bp on average. Yan et al. (2009) investigated LD using 1,229 SNPs and 632 maize lines genotyped, the authors concluded that LD distance increased with the increase of minor allelic frequency (MAF), and LD decay was higher in temperate than in tropical and subtropical lines. Lu et al. (2011) also observed a rapid LD decay (5–10 kb) in the tropical maize inbred lines, using SNPs and haplotypes from genic and intergenic regions. The authors suggested marker numbers needed for whole genome scan in maize, in GWAS. The use of haplotype information present greater predictive ability and selection coincidence than single

SNPs for grain yield in maize population, in a genomic selection approach (MATIAS et al., 2017).

In popcorn populations, to date, few or none LD and haplotype block studies have been reported. Here, we described the LD measures and their respective decay, and characterized haplotype block patterns in tropical and temperate popcorn populations genotyped by genotyping-by-sequencing (GBS). LD and haplotype blocks also were evaluated in intragenic regions, for polymorphic genes related to the biosynthesis of zein, starch, cellulose and fatty acids.

2 MATERIAL AND METHODS

2.1 Populations

We used a biparental population (F_2 generation) derived from crossing AP4502, developed by the Agricultural Alumni Seed Improvement Association, Romney, IN, USA, and a tropical synthetic (Synthetic UFV) and a tropical breeding population (Beija-Flor cycle 4) developed by the Federal University of Viçosa (UFV), Minas Gerais, Brazil. The synthetic was derived by random crossings involving 20 inbred lines from the tropical population Viçosa and 20 inbred lines from the tropical population Beija-Flor, all also developed by UFV based on expansion volume (a measure of quality). Beija-Flor cycle 4 (BFc4) was also developed by UFV after four cycles of half-sib selection based on expansion volume. Thus, we have one temperate and two tropical populations with distinct structures. Theoretically, a biparental population shows LD only for linked genes and molecular markers. In a synthetic there is LD for genes and molecular markers with independent assortment.

2.2 DNA extraction, genotyping-by-sequencing, SNP calling, data quality control, and imputation

Leaf samples of young plants were collected from each population for DNA extraction. The DNA extraction was performed using the CTAB protocol (DOYLE; DOYLE, 1990). After quantification, the DNA samples of 574 plants (190 or 192 from each population) were sent to Institute of Biotechnology at Cornell University (two plates of 95 samples from the biparental population) and Institut de Recherche en Immunologie et en Cancérologie/IRIC at University of Montreal (four plates of 96 samples from the tropical populations) for GBS services (based on HiSeq 2500 and NextSeq500, respectively). The SNP variant call services were provided by Institute of Biotechnology and Omega Bioservices, Norcross, GA, using B73 version 4 as the reference genome. After reading the data using the R package vcfR (KNAUS; GRÜNWALD, 2017), we filtered by missing allele and chromosomes. Then, we computed the SNP and genotype call rates and the MAF, employing the R package HapEstXXR (KNUEPPEL; ROHDE; KNUEPPEL, 2015), and filtered by MAF > 0.01. Finally, we imputed based on Beagle (BROWNING; BROWNING, 2009), using the R package synbreed (WIMMER et al., 2012). The number of SNPs after the data quality control and imputation for the biparental population, Synthetic UFV, and Beija-Flor c4 were 145,420, 74,773, and 76,055, respectively. To keep a similar number of SNPs for the populations, we finally performed a random sampling of 75,000 SNPs from the biparental population.

2.3 LD and haplotype block analyses

For the Hardy-Weinberg equilibrium analysis by population and chromosome it was adopted the Bonferroni criterion to keep a global level of significance of 1%. The LD analyses were performed considering the two measures of LD for bi-allelic markers, D' and r^2, derived from Lewontin's D (LEWONTIN, 1964). LD statistics were computed pairwise for markers within 500 kb of distance, using a two marker expectation-maximization (EM) (BARRETT et al., 2005). To study D' and r^2 decay and LD extent, the physical distances between markers were classified into six intervals: 0-50 kb, 51–100 kb, 101-150 kb, 151–200 kb, 201–250 kb, 251–300 kb, 301–350 kb, 351–400 kb, 401–450 kb, and 451–500 kb in each chromosome. In the study of haplotype blocks, the criteria employed for definition of haplotypes blocks was given by GABRIEL et al. (2002), which consider pairs in "strong LD" if the one-sided upper 95% confidence bound on D' is > 0.98. The haplotypes were estimated using an accelerated EM algorithm with a partition–ligation approach (QIN; NIU; LIU, 2002), to generate phased haplotypes for population frequency (BARRETT, 2009).

The LD and haplotype blocks analyses were performed on the intragenic SNPs level. The approximately 75,000 SNPs covered from 23,710 to 24,498 genes in the three populations. We chose a set of 12 genes from 49 genes related biologic process for synthesis of zein, starch, cellulose and fatty acids (Table 1). With the exception of 2 genes, the remaining chosen genes for the intragenic analysis had at least five SNPs for each population. For the intragenic LD decay and LD decay extent analyses we computed the average D' and r^2 values defining intervals of 1 kb (0-1 to 10.1-11 kb). The information about the selected SNPs was extracted from the public MaizeGDB database (https://www.maizegdb.org/).

All analyses were performed using Haploview software (BARRETT et al., 2005). For assessing haplotype blocks information, the haplotype files in each population and chromosome were summarized using *REALbreeding* program (*Haplotype blocks summary*) developed in REALbasic 2009.

Table 1 – Gene name, annotation, and chromosome localization, and the number of intragenic SNPs in each population

Gene	Annotation	Chr.	SNPs	Population
Zm00001d002654	nkd1; naked endosperm1: double mutants have multiple (2-5) layers of peripheral endosperm cells that lack starch granules or other features of starchy endosperm	2	12	Biparental
			7	Synthetic
			7	BFc4
Zm00001d004817	Fatty acid amide hydrolase	2	14	Biparental
			11	Synthetic
			3	BFc4
Zm00001d005451	Cellulose synthase A catalytic subunit 5 [UDP-forming]	2	5	Biparental
			8	Synthetic
			6	BFc4
Zm00001d041972	Cellulose synthase-like protein G3	3	14	Biparental
			9	Synthetic
			10	BFc4
Zm00001d052263	Starch synthase 2 chloroplastic/amyloplastic	4	6	Biparental
			10	Synthetic
			6	BFc4
Zm00001d018033	Starch synthase IIb-2	5	19	Biparental
			6	Synthetic
			5	BFc4
Zm00001d035760	zp15; zein protein, 15kDa15: high methionine; genomic blot indicates one or two copies	6	5	Biparental
			2	Synthetic
			2	BFc4
Zm00001d036900	Cellulose synthase A catalytic subunit 7 [UDP-forming]	6	19	Biparental
			15	Synthetic
			8	BFc4
Zm00001d021731	Cellulose synthase-like protein D3	7	7	Biparental
			9	Synthetic
			9	BFc4
Zm00001d023810	Putative cellulose synthase-like family protein	10	5	Biparental
			12	Synthetic
			16	BFc4
Zm00001d025201	fab1; fatty acid biosynthesis1: endosperm cDNA 2C01H08 (uaz99) similar to fatty acid biosynthesis enzyme	10	9	Biparental
			5	Synthetic
			21	BFc4
Zm00001d026113	nkd2; naked endosperm2: double mutants have multiple (2-5) layers of peripheral endosperm cells that lack starch granules or other features of starchy endosperm	10	20	Biparental
			8	Synthetic
			11	BFc4

Source: Elaborated by the author.

3 RESULTS

In the present study, the SNP spanned represented practically 100% of all the chromosome (Table 2). The average density of the SNPs ranged from 9.6 kb to 44.4 kb, with an overall average of 29.3 kb. Pengelly and Collins (2018) evaluated LD maps in humans in 216 kb region, with a average density of 0.87 kb. In the LD study on a bovine genome of 2.87 Gb size, the marker density was approximately 1000 kb (MCKAY et al., 2007). Employing GBS in maize inbred lines, Beckett et al.(2017) evaluated 77,314 SNPs distributed on all chromosomes, density similar to that applied in the present study. Although the populations had similar average MAFs of approximately 0.1. The MAF distribution reveals that biparental population is different from synthetic and breeding population, since it has the highest frequency of SNPs with MAF between 0.45 and 0.50, and between 0 and 0.05 (Figure 1). The MAF distribution in synthetic and breeding populations are similar.

Average D' was higher in biparental and synthetic populations, ranged from 0.74 (synthetic) until 0.81 (biparental), and minimum D' was higher in the breeding population per chromosome (Table 2). The r^2 analysis showed that minimum and average r^2 measures were higher in synthetic and breeding population. In these populations, the average r^2 values ranged from 0.021 to 0.054 (breeding population). The average r^2 value in the biparental population is approximately half of the corresponding average values observed in the other populations.

Table 2 – Number of SNPs, SNP coverage (kb), average SNP interval (bp), MAF, and minimum, average, and maximum LD measures by chromosome in each population

(to be continued)

Population	Chr.	SNPs	SNP coverage	SNP interval	MAF	D'			r^2		
						Min.	Av.	Max.	Min.	Av.	Max.
Biparental	1	11,816	307,039.27	25,982.75	0.09	0.0	0.78	1.0	0.00	0.023	1.00
	2	8,710	244,412.25	28,059.68	0.11	0.0	0.77	1.0	0.00	0.026	1.00
	3	8,205	235,520.19	28,693.18	0.11	0.0	0.75	1.0	0.00	0.032	1.00
	4	8,081	246,827.22	30,525.85	0.07	0.0	0.81	1.0	0.00	0.015	1.00
	5	8,697	223,657.67	25,708.94	0.09	0.0	0.79	1.0	0.00	0.019	1.00
	6	5,883	173,906.61	29,537.18	0.10	0.0	0.78	1.0	0.00	0.027	1.00
	7	6,401	182,200.48	28,440.64	0.11	0.0	0.77	1.0	0.00	0.025	1.00
	8	6,528	181,042.64	27,725.54	0.10	0.0	0.78	1.0	0.00	0.023	1.00
	9	5,625	159,429.26	28,336.11	0.11	0.0	0.76	1.0	0.00	0.027	1.00
	10	5,054	150,832.73	29,824.39	0.10	0.0	0.76	1.0	0.00	0.025	1.00

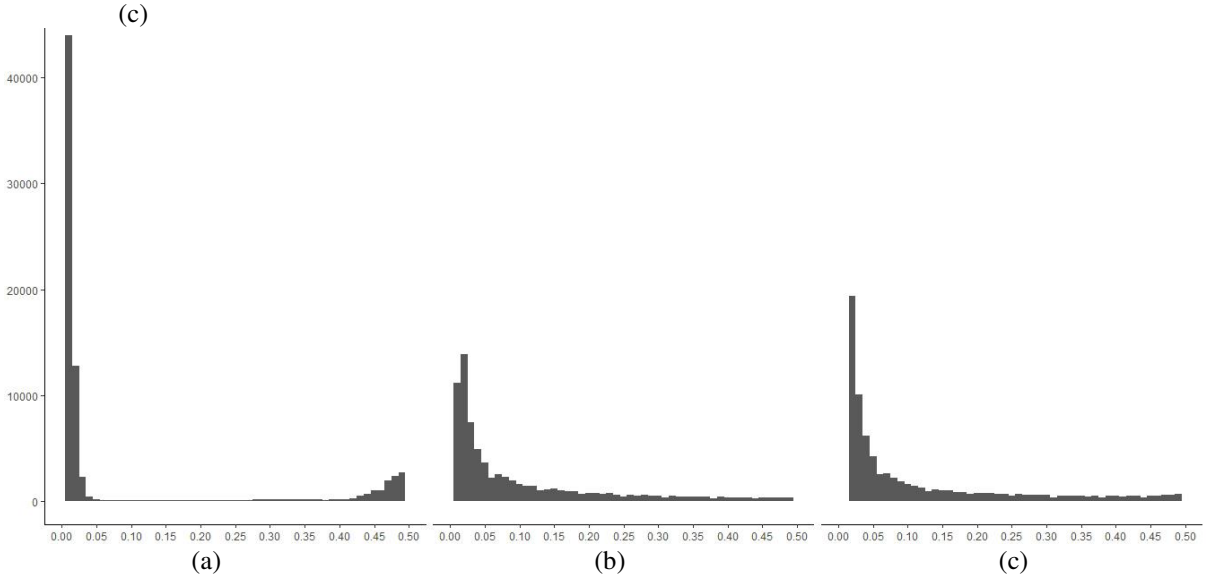
Table 2 – Number of SNPs, SNP coverage (kb), average SNP interval (bp), MAF, and minimum, average, and maximum LD measures by chromosome in each population

(conclusion)

Synthetic	1	11,224	306,909.66	27,341.76	0.10	0.0	0.75	1.0	0.02	0.046	1.00
	2	9,712	244,369.34	25,159.97	0.10	0.0	0.75	1.0	0.02	0.041	1.00
	3	9,374	235,478.72	25,083.00	0.10	0.0	0.76	1.0	0.02	0.042	1.00
	4	5,840	246,943.47	42,170.02	0.10	0.0	0.74	1.0	0.02	0.052	1.00
	5	9,460	223,706.51	23,589.54	0.10	0.0	0.74	1.0	0.02	0.040	1.00
	6	5,294	173,221.42	32,692.62	0.10	0.0	0.74	1.0	0.02	0.050	1.00
	7	6,299	182,159.80	28,857.92	0.11	0.0	0.74	1.0	0.02	0.042	1.00
	8	6,248	180,660.38	28,850.52	0.10	0.0	0.76	1.0	0.02	0.044	1.00
	9	5,161	159,553.33	30,909.31	0.11	0.0	0.75	1.0	0.02	0.045	1.00
	10	6,161	150,828.61	24,464.31	0.09	0.0	0.75	1.0	0.02	0.034	1.00
BFc4	1	10,182	306,774.01	30,126.80	0.11	0.20	0.71	1.0	0.02	0.047	1.00
	2	8,481	244,407.97	28,816.88	0.11	0.21	0.69	1.0	0.02	0.042	1.00
	3	8,005	235,478.74	29,373.18	0.11	0.20	0.70	1.0	0.02	0.040	1.00
	4	5,558	246,840.44	44,379.59	0.11	0.20	0.69	1.0	0.02	0.054	1.00
	5	7,674	223,706.51	29,080.32	0.11	0.20	0.70	1.0	0.02	0.039	1.00
	6	4,547	173,351.50	38,093.29	0.11	0.19	0.68	1.0	0.02	0.044	1.00
	7	5,602	182,155.19	32,448.24	0.11	0.20	0.69	1.0	0.02	0.040	1.00
	8	5,020	180,660.38	35,943.93	0.12	0.20	0.70	1.0	0.02	0.048	1.00
	9	5,353	159,489.87	29,788.60	0.11	0.20	0.69	1.0	0.02	0.042	1.00
	10	15,633	150,926.35	9,653.39	0.13	0.20	0.52	1.0	0.02	0.021	1.00

Source: Elaborated by the author.

Figure 1 – MAF distribution in the biparental population (a), in the synthetic (b), and in the breeding population (c)



(a) (b) (c)

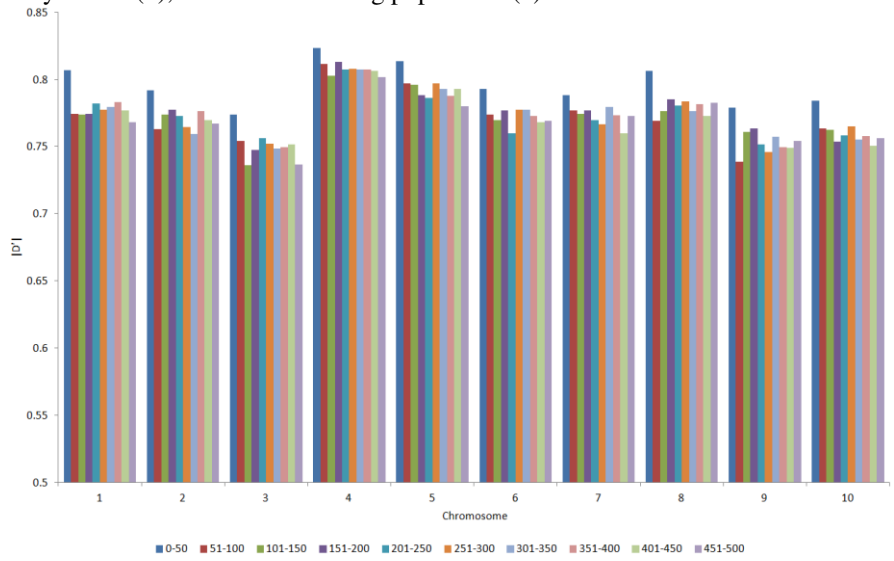
Source: Elaborated by the author.

Regardless of LD measure, in the populations it was observed the LD decay similar in the different chromosomes (Figures 2 and 3). The D' average values, on separated SNPs by 1 to 50 kb were from 0.83 to 0.86. There is an initially higher LD decrease for SNPs separated by 51-100 kb (3 to 7% for D' and 28 to 66% for r^2, on average) and then a gradual decrease to the minimum LD value for SNPs separated by 451-500 kb. Clearly, average LD is different on chromosome 10 of the breeding population.

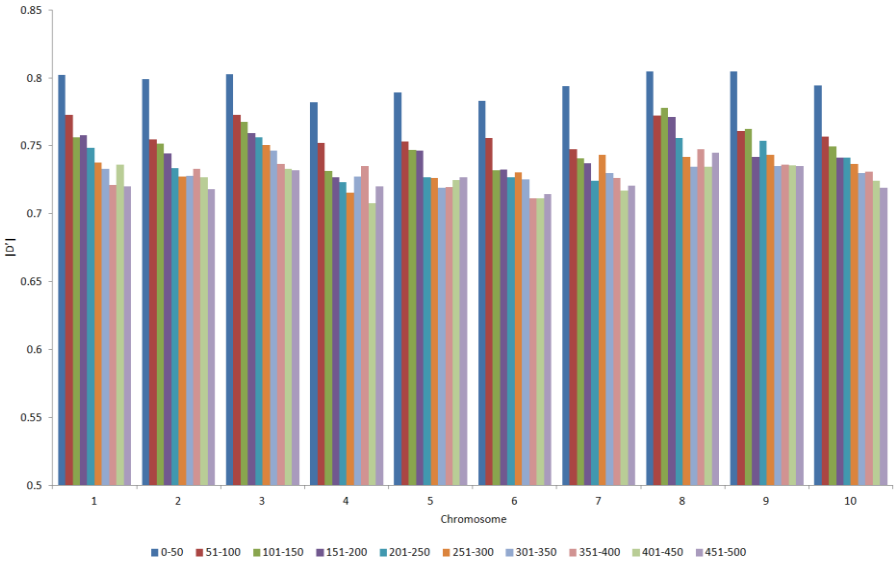
It is important to note that, in the three populations and regardless of the chromosome, the LD had higher for SNPs distant by up to 50 kb (Table 4). In the interval from 51 kb to 450 kb, the average D' values decreased from 0.69-0.77 to 0.64-0.77 in the three populations and the average r^2 values in the biparental population decreased from 0.025 to 0.020. In the other two populations the average r^2 value decreased in approximately 50%.

The haplotype blocks study revealed that biparental population has a smaller number of blocks per chromosome, ranging from 169 to 336, comparing to 486 to 1,126 blocks in synthetic population and 476 to 1,019 blocks in breeding population (Table 3). The average block size ranged from 6.89 to 15.31 kb in synthetic population, and from 5.36 to 18.89 kb in breeding population. In biparental population, it was observed an average block size eighteen times lower than others populations. In synthetic and breeding populations, the maximum number of SNPs ranged from 7 to 16, and in biparental population ranged from 4 to 6 SNPs. Despite the number of SNPs ranges from 2 to 16 per block, the most haplotype blocks are distributed with 2 SNPs in all populations (Figure 5).

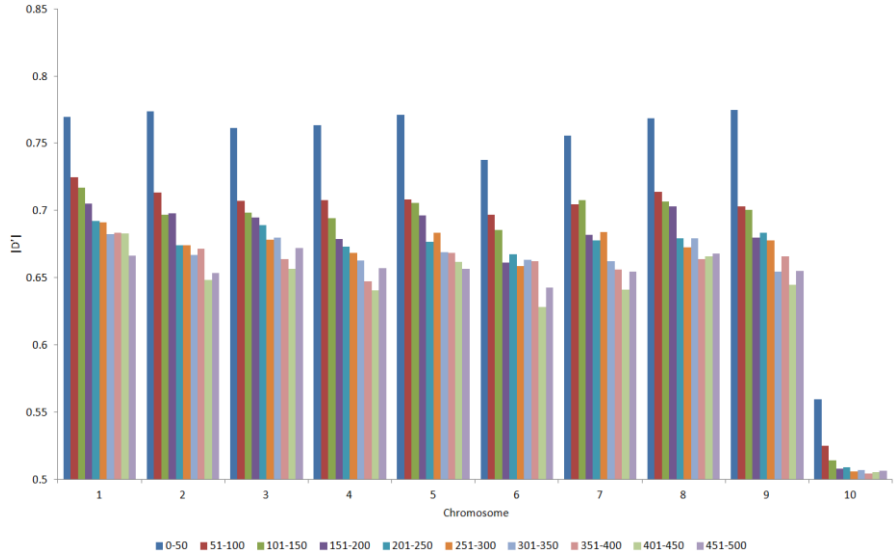
Figure 2 – Average D' values by chromosome and by distance interval (kb) in the biparental population (a), in the synthetic (b), and in the breeding population (c)



(a)



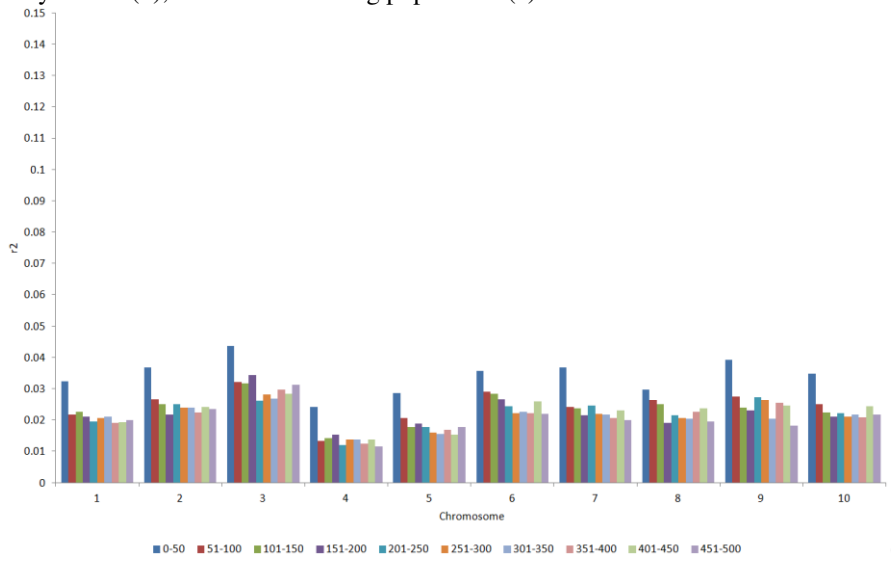
(b)



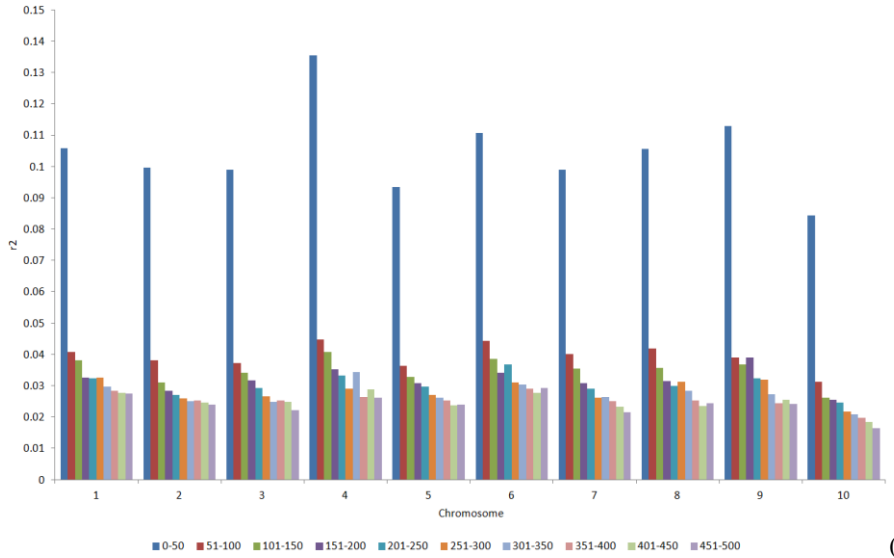
(c)

Source: Elaborated by the author.

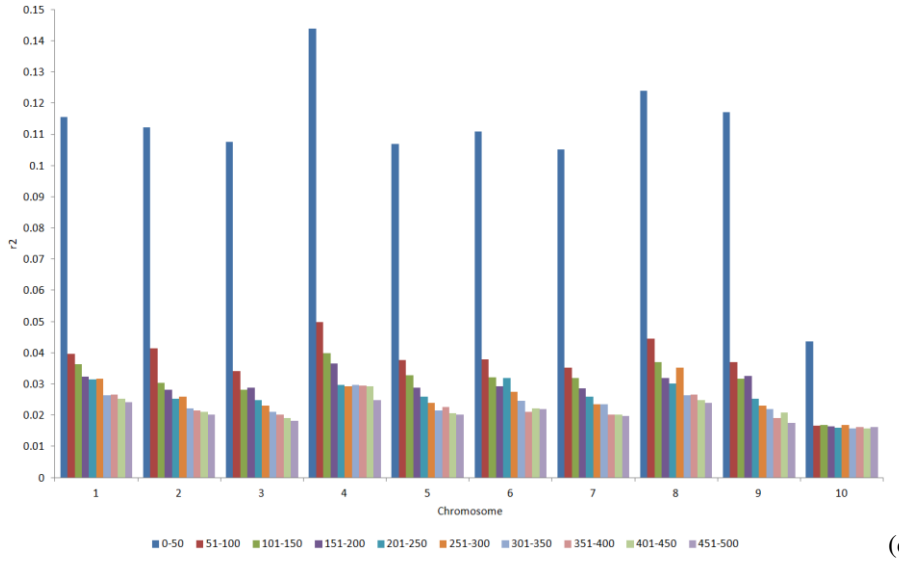
Figure 3 – Average r^2 values by chromosome and by distance interval (kb) in the biparental population (a), in the synthetic (b), and in the breeding population (c)



(a)



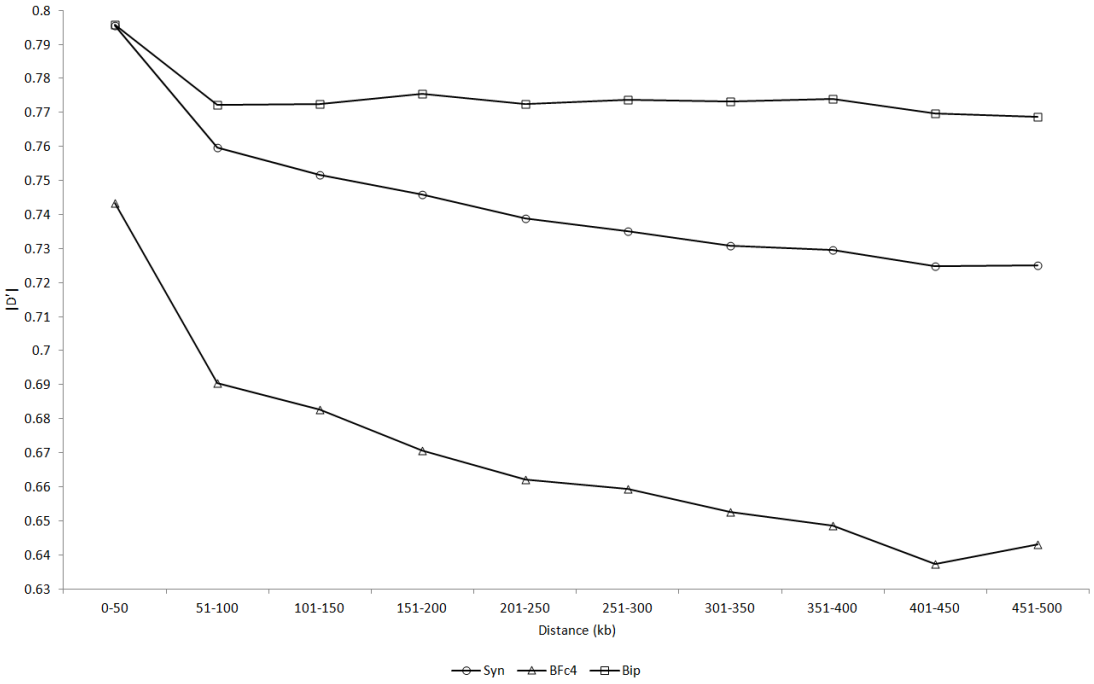
(b)



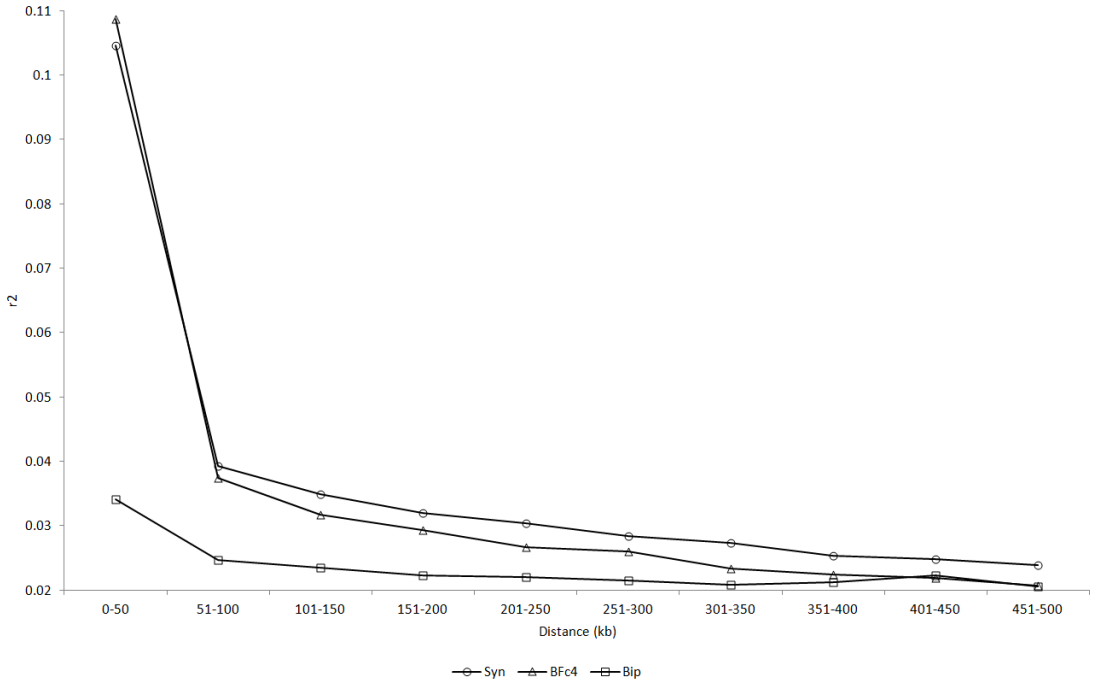
(c)

Source: Elaborated by the author.

Figure 4 – Overall average D' (a) and r^2 (b) values by distance interval (kb) in the biparental population (Bip), in the synthetic (Syn), and in the breeding population (BFc4)



(a)



(b)

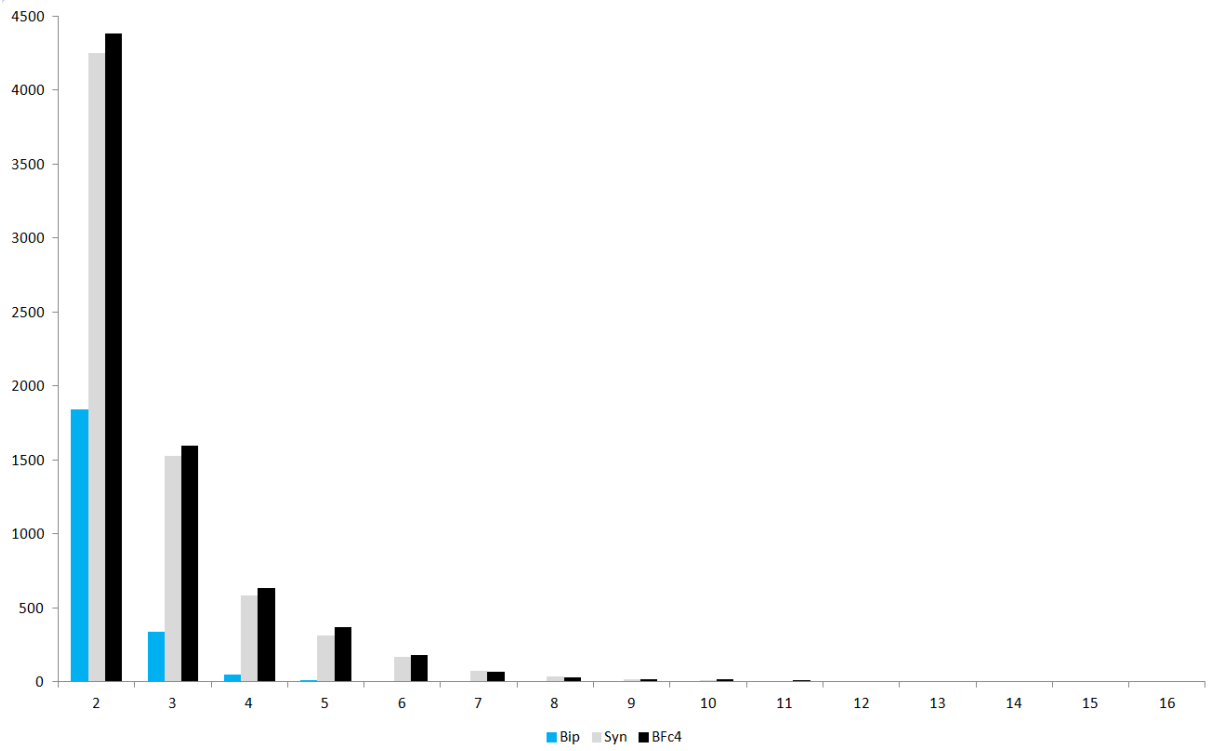
Source: Elaborated by the author.

Table 3 – Haplotype blocks structure of the populations

Population	Chr.	Blocks	Block size (kb)				SNPs			
			Total	Mean	Min.	Max.	Total	Mean	Min.	Max.
Biparental	1	336	58.60	0.17	0.001	10.30	727	2.2	2	5
	2	294	588.31	2.00	0.001	298.9	647	2.2	2	6
	3	273	307.66	1.13	0.001	101.90	622	2.3	2	5
	4	193	35.80	0.19	0.001	23.15	430	2.2	2	6
	5	218	47.49	0.22	0.001	20.39	484	2.2	2	4
	6	169	419.24	2.48	0.001	292.35	387	2.3	2	5
	7	215	45.60	0.21	0.001	11.68	479	2.2	2	5
	8	186	511.79	2.75	0.001	423.79	409	2.2	2	5
	9	195	58.19	0.29	0.001	15.58	432	2.2	2	5
	10	170	314.88	1.85	0.001	307.49	370	2.2	2	4
Synthetic	1	1126	11935.23	10.60	0.001	494.94	3093	2.7	2	10
	2	935	8501.15	9.09	0.001	451.74	2565	2.7	2	11
	3	810	9065.75	11.19	0.001	457.30	2257	2.8	2	11
	4	525	6615.63	12.60	0.001	423.71	1409	2.7	2	12
	5	933	6428.48	6.89	0.001	395.79	2527	2.7	2	11
	6	496	5051.01	10.18	0.001	492.95	1354	2.7	2	11
	7	569	5169.26	9.09	0.001	317.07	1594	2.8	2	15
	8	583	8927.76	15.31	0.001	476.37	1574	2.7	2	10
	9	486	6553.37	13.48	0.001	398.72	1375	2.8	2	9
	10	534	3905.24	7.31	0.001	434.32	1477	2.8	2	10
BFc4	1	1019	14352.62	14.09	0.001	499.04	2818	2.8	2	12
	2	861	7904.79	9.18	0.001	415.28	2432	2.8	2	11
	3	796	8682.69	10.91	0.001	418.18	2153	2.7	2	16
	4	539	6605.65	12.26	0.001	442.01	1492	2.8	2	12
	5	776	10870.44	14.01	0.001	479.50	2201	2.8	2	15
	6	476	5833.85	12.26	0.001	466.82	1278	2.7	2	7
	7	570	4471.35	7.84	0.001	479.70	1612	2.8	2	13
	8	491	9272.30	18.89	0.001	495.26	1390	2.8	2	12
	9	541	5188.65	9.59	0.001	449.77	1478	2.7	2	8
	10	1236	6619.87	5.36	0.001	471.30	3371	2.7	2	12

Source: Elaborated by the author.

Figure 5 – Distribution of the haplotype blocks based on the number of SNPs in the biparental population (Bip), in the synthetic (Syn), and in the breeding population (BFc4)



Source: Elaborated by the author.

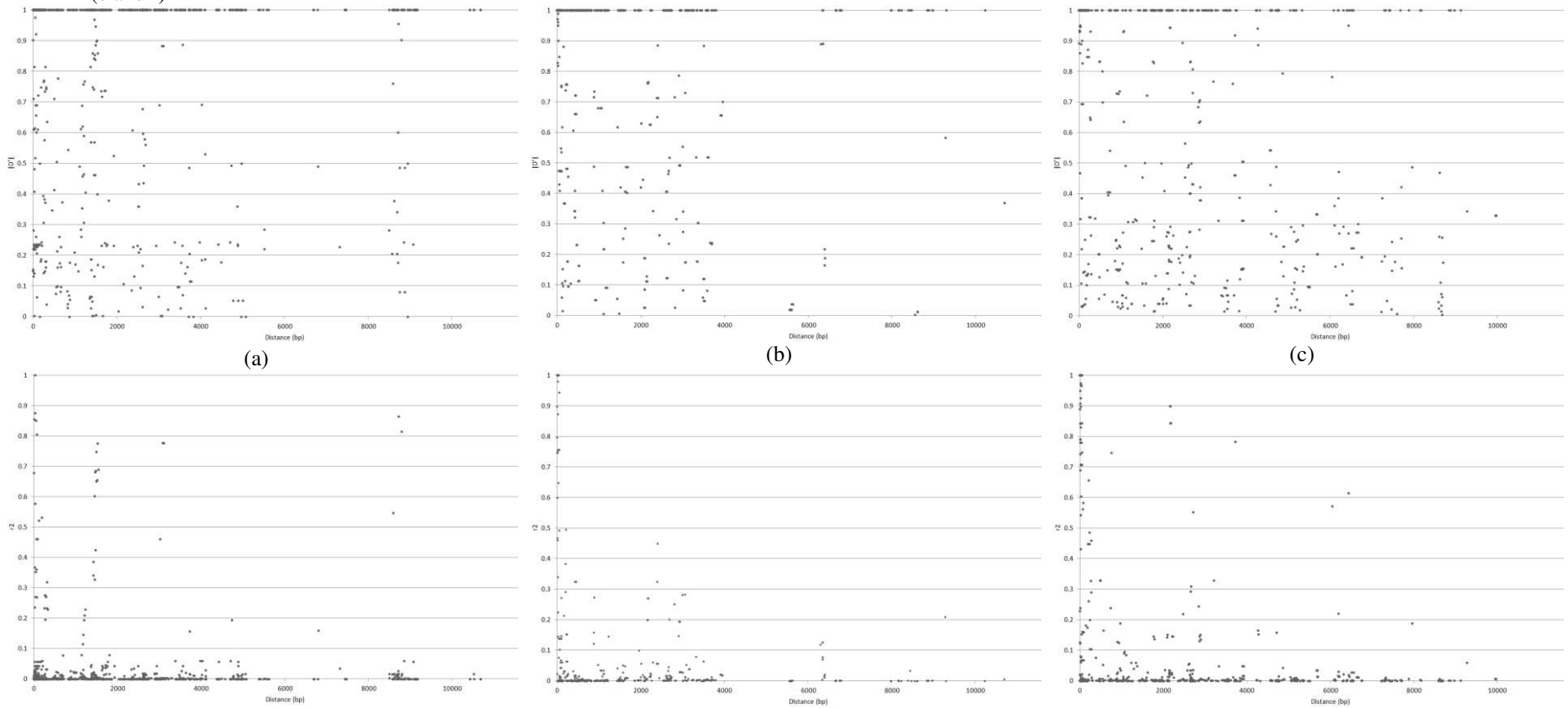
Twelve genes were chosen for the evaluation of LD intragenic (D' and r^2). The average D' was higher in the biparental population and the synthetic, compared to the value observed in the breeding population (0.74 and 0.88 vs. 0.67). The biparental population presented an r^2 average value much lower than the values observed in the other two populations (0.02 vs. 0.13 and 0.14) (Table 4). Regardless of the population, the maximum intragenic D' (1,0) was observed for SNPs separated by up to 10.6 kb while most of the higher intragenic r^2 values (0.7 or greater) were only observed for the closest SNPs (Figure 6). Haplotype blocks were found in 7 genes. The average block size ranged from 0.02 to 8.72 kb (Table 5). In general, the number of SNPs per haplotype was 2 SNPs, however, in the gene Zm00001d041972 in the breeding population, they were observed 6 SNPs in the one haplotype block.

Table 4 – Intragenic minimum, average, and maximum LD values in each population

Gene	Population	D'			r^2		
		Min.	Av.	Max.	Min.	Av.	Max.
Zm00001d002654	Biparental	0.176	0.96	1.0	0.000	0.005	0.19
	Synthetic	0.003	0.60	1.0	0.000	0.159	1.00
	BFc4	0.042	0.44	1.0	0.000	0.258	1.00
Zm00001d004817	Biparental	0.028	0.81	1.0	0.000	0.004	0.06
	Synthetic	0.059	0.62	1.0	0.000	0.089	1.00
	BFc4	1.000	1.00	1.0	0.002	0.310	0.93
Zm00001d005451	Biparental	0.148	0.91	1.0	0.000	0.003	0.01
	Synthetic	0.407	0.89	1.0	0.000	0.106	1.00
	BFc4	0.057	0.51	1.0	0.000	0.211	0.97
Zm00001d041972	Biparental	0.132	0.89	1.0	0.000	0.004	0.06
	Synthetic	0.263	0.79	1.0	0.000	0.191	1.00
	BFc4	0.193	0.88	1.0	0.000	0.280	1.00
Zm00001d052263	Biparental	0.236	0.85	1.0	0.000	0.011	0.06
	Synthetic	0.217	0.93	1.0	0.000	0.116	1.00
	BFc4	0.323	0.87	1.0	0.000	0.085	1.00
Zm00001d018033	Biparental	0.000	0.83	1.0	0.000	0.031	0.87
	Synthetic	0.488	0.97	1.0	0.000	0.025	0.21
	BFc4	0.137	0.77	1.0	0.001	0.070	0.46
Zm00001d035760	Biparental	0.187	0.84	1.0	0.000	0.007	0.06
	Synthetic	1.000	1.00	1.0	0.007	0.007	0.01
	BFc4	0.721	0.72	0.7	0.027	0.027	0.03
Zm00001d036900	Biparental	0.000	0.76	1.0	0.000	0.093	0.88
	Synthetic	0.005	0.77	1.0	0.000	0.026	1.00
	BFc4	0.031	0.60	1.0	0.000	0.019	0.24
Zm00001d021731	Biparental	0.094	0.59	1.0	0.000	0.037	0.68
	Synthetic	0.019	0.58	1.0	0.000	0.282	1.00
	BFc4	0.193	0.57	1.0	0.001	0.248	1.00
Zm00001d023810	Biparental	1.000	1.00	1.0	0.000	0.000	0.00
	Synthetic	0.026	0.76	1.0	0.000	0.093	1.00
	BFc4	0.004	0.48	1.0	0.000	0.066	0.97
Zm00001d025201	Biparental	0.097	0.84	1.0	0.000	0.004	0.06
	Synthetic	0.059	0.59	1.0	0.000	0.368	0.87
	BFc4	0.006	0.68	1.0	0.000	0.061	1.00
Zm00001d026113	Biparental	0.002	0.82	1.0	0.000	0.026	1.00
	Synthetic	0.105	0.81	1.0	0.000	0.057	0.90
	BFc4	0.015	0.52	1.0	0.000	0.073	1.00

Source: Elaborated by the author.

Figure 6 – Overall intragenic D' (a, b, c) and r^2 (d, e, f) by distance interval (bp) in the biparental population (a and d), in the synthetic (b and e), and in the breeding population (c and f)



(a) (b) (c)

(d) (e) (f)

Source: Elaborated by the author.

Table 5 – Intragenic haplotype blocks structure in each population

Population	Gene	Chr.	Blocks	Block size (kb)				SNPs			
				Total	Mean	Min.	Max.	Total	Mean	Min.	Max.
Biparental	Zm00001d018033	5	1	8.72	8.72	8.72	8.72	2	2	2	2
	Zm00001d026113	10	1	0.03	0.03	0.03	0.03	2	2	2	2
Synthetic	Zm00001d002654	2	1	0.05	0.05	0.05	0.05	3	3	3	3
	Zm00001d004817	2	2	0.22	0.11	0.02	0.21	4	2	2	2
	Zm00001d005451	2	1	0.03	0.03	0.03	0.03	2	2	2	2
	Zm00001d036900	3	1	0.06	0.06	0.06	0.06	2	2	2	2
	Zm00001d041972	6	1	0.02	0.02	0.02	0.02	2	2	2	2
BFc4	Zm00001d041972	3	1	2.22	2.22	2.22	2.22	6	6	6	6
	Zm00001d018033	5	1	0.26	0.26	0.26	0.26	2	2	2	2

Source: Elaborated by the author.

4 DISCUSSION

4.1 The LD degree and haplotype blocks in the popcorn populations

This is a descriptive study on LD and haplotype blocks in tropical and temperate popcorn populations using SNP genotypes from GBS. Our results showed that the synthetic is the population with higher LD, followed by the biparental population. The lower average D' value in the breeding population reflects its recombination history. The average D' was higher in biparental and synthetic because these populations have no recombination history. We observed low average r^2 values and the reduced frequency of SNPs with r^2 values greater than 0.25 (defined as useful LD in some studies) in the popcorn populations. Considering to a physic distance of up to 1 Mb in maize inbred lines from breeding programs, the average r^2 ranged from 0.02 to 0.23 (YAN et al., 2009).

When comparing populations that share a common origin, have similar effective population size, and did not have face an extreme reduction in size (population bottleneck), because similar allele frequencies the statistics D, D', and r^2 should provide a comparable characterization of the LD pattern. If the populations have distinct distributions of the allelic frequencies, D' can be used for analyzing the recombination history and r^2 should be the choice if recombination and mutation are important factors affecting the LD. The statistic r^2 is the most relevant for association mapping because it has a simple inverse relationship with the sample size required to detect association (WALL; PRITCHARD, 2003).

Our results showed that in each population, and regardless of the measure, the chromosomes had the same LD pattern. However, the LD pattern was notably different on chromosome 10 in breeding population compared to the other chromosomes. Among the studied populations, only this population, strictly speaking, has a history of recombination, since it was obtained for 4 cycles of selection. It cannot be guaranteed that the initial D' in this population was larger, because there were recombination and selection, and selection is a factor that can alter the disequilibrium. The difference observed in chromosome 10 may be related to the selection cycles that this population underwent. Concerning LD patterns along chromosomes, studies with maize inbred line panels have presented differences among chromosomes, and on different physic distances (VAN INGHELANDT et al., 2011; YAN et al., 2009).

In regard to haplotype blocks, synthetic and breeding populations had a larger number of blocks per chromosome, with larger blocks and with a greater number of SNPs (Table 3).

Overall, the number and length of the haplotype blocks and the number of SNPs per haplotype block were proportional to the average r^2. However, because the lower average r^2 values of the populations, the average number of SNPs and length of the haplotype blocks ranged from two to three and from one to 11 kb. It is noticed that there is a great variability between species and populations of the same species, and these differences may be related to the degree of LD in the population. Applying two different methods to build haplotype blocks for GWAS analysis in humans, Shim et al. (2009) detected 97,881 blocks with average block size of 7.33 kb, and 4 SNPs per block using the Gabriel et al. (2002) method. Matias et al. (2017) found 7,772 blocks in maize using Gabriel et al. (2002) method, with average block size of 35.16 kb, containing 2 to 10 SNPs per block.

4.2 LD decay and LD extent

In this work, the LD extended for until 500 kb of distance interval in the three populations. Overall, studies with maize inbred line panels have showed LD extending up to 100 Mb. The LD decay distance for r^2 decreases to approximately 0.10 up to 100 Mb in maize temperate lines, and between 5 to 10 kb in tropical lines (LU et al., 2011; VAN INGHELANDT et al., 2011; YAN et al., 2009). In the present study, the tropical populations (synthetic and breeding population) also showed rapid LD decay compared to temperate population (biparental population). In study of Romay et al. (2013), the entire collection of inbred lines presented LD decaying very rapidly and r^2 reached an average of 0.2 within about 1 kb of distance.

4.3 Implications for GWAS and genomic selection studies

LD is the structural basis of genomic selection and GWAS contributing to figure out the genes responsible for variation of economically important traits (MEUWISSEN; HAYES; GODDARD, 2001; QANBARI et al., 2010). Although r^2 values are surprisingly low, this does not imply that these populations cannot be used for GWAS because there is a fraction of high r^2 values for SNPs separated by less than 5 kb. The populations are also not excluded for genomic selection because the most important factor affecting this selection process is the relatedness between individuals in the training and validation sets.

In addition to applications of LD knowledge, the haplotype block knowledge is desirable in selecting of a maximally informative set of SNPs in association studies, reducing cost and effort (QANBARI et al., 2010; ZHANG et al., 2005). Testing one SNP within each block for significant association with the trait might be sufficient to indicate association with every SNP in the block, thus reducing the number of SNPs that need to be tested (SLATKIN, 2008).

Furthermore, using haplotypes in association testing is expected to increase power relative to single-SNP approaches (SHIM et al., 2009). In genomic selection, the main benefit of using haplotypes for prediction is that haplotypes are expected to be in higher LD with the QTL than individual SNPs. Haplotypes used as predictors to estimate breeding values are expected to improve results (CUYABANO; SU; LUND, 2015). However, defining haplotype blocks for genomic predictions have followed different criterions (BOICHARD et al., 2012; SCHROOTEN et al., 2013; VILLUMSEN; JANSS; LUND, 2009). Cuyabano, Su and Lund (2014) proposed to define a minimum amount of LD between SNP markers, without a fixed length (number of SNPs), to reduce the number of explanatory variables required for the predictions. In this work, we do not expect to find a significant advantage in the haplotype-based GWAS and genomic selection along generations due to the reduced number of SNPs in the haplotype blocks (2 to 3).

4.4 Comparative analysis of intragenic LD and haplotype blocks

The LD intragenic approach revealed the presence of low intragenic LD and the minimum size of the haplotype blocks in the genes related to zein, starch, cellulose and fatty acids biosynthesis within three populations. Overall, the tropical populations developed by breeding program, based on expansion volume, have higher LD and presence of haplotype blocks. However, we cannot infer that the higher r^2 values observed in 11 of the 12 genes are due to selection for quality in populations, since there is no information on the LD and haplotype block patterns in the base populations Viçosa and Beija-Flor. Furthermore, we believe that the lower LD for the biparental population is due to crossing two similar high-quality inbred lines.

Compared with other studies, different genes in Eurasia populations had D' ranging from 0.67 to 0.94, with few molecular markers per gene (KAESSMANN et al., 2002). Eleven innate immune genes in *Bos taurus taurus* and *Bos taurus indicus* cattle presented r^2 ranging from 0.08 to 0.93, and SNPs per gene ranged from 4 to 50 (SEABURY et al., 2010). Wilson et al. (2004) evaluated six candidate gene and promoter regions for starch concentration and composition in maize, and they found LD declined rapidly less than 2,000 pb, with r^2 up to approximately 0.80. Here, the number of SNPs observed for haplotype block, in the gene *Zm00001d041972* (6 SNPs), was also comparable to related in Nilsen et al. (2008). They found 3 haplotype blocks, with SNPs ranging from 6 to 15 per block.

5 CONCLUSIONS

The findings from this study show that r^2 values are low the three populations – biparental population, synthetic, and breeding population. However, these populations can be used for GWAS and genomic selection. Because the reduced number of SNPs in the haplotype blocks, there is not a significant advantage in the haplotype-based GWAS and genomic selection. Furthermore, the results on LD decay (rapid decay after 5-10 kb) and LD decay extent (along up to 300 kb) are in the range observed with maize inbred line panels. The LD intragenic approach revealed that synthetic and breeding population, selected for quality (zein, starch, cellulose and fatty acids biosynthesis), have higher LD and presence of haplotype blocks. However, the higher r^2 values observed may not be limited to selection within populations.

REFERENCES

BARRETT, J. C. Haploview: Visualization and analysis of SNP genotype data. **Cold Spring Harbor Protocols**, v. 2009, n. 10, p. pdb. ip71, 2009.

BARRETT, J. C.; FRY, B.; MALLER, J.; DALY, M. J. Haploview: analysis and visualization of LD and haplotype maps. **Bioinformatics**, v. 21, n. 2, p. 263-265, 2005.

BECKETT, T. J.; MORALES, A. J.; KOEHLER, K. L.; ROCHEFORD, T. R. Genetic relatedness of previously Plant-Variety-Protected commercial maize inbreds. **PLOS ONE**, v. 12, n. 12, p. e0189277, 2017. Disponível em: < https://doi.org/10.1371/journal.pone.0189277 >.

BOICHARD, D.; GUILLAUME, F.; BAUR, A.; CROISEAU, P.; ROSSIGNOL, M.-N.; BOSCHER, M. Y.; DRUET, T.; GENESTOUT, L.; COLLEAU, J.-J.; JOURNAUX, L. Genomic selection in French dairy cattle. **Animal Production Science**, v. 52, n. 3, p. 115-120, 2012.

BOURGAIN, C.; GENIN, E.; CLERGET-DARPOUX, F. Comparison of family based haplotype methods using intragenic SNPs in candidate genes. **European Journal Of Human Genetics**, v. 10, p. 313, 06/18/online,2002. Disponível em: < https://doi.org/10.1038/sj.ejhg.5200808 >.

BROWNING, B. L.; BROWNING, S. R. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. **Am J Hum Genet**, v. 84, n. 2, p. 210-23, Feb,2009. Disponível em: < http://www.ncbi.nlm.nih.gov/pubmed/19200528 >.

CUYABANO, B. C.; SU, G.; LUND, M. S. Genomic prediction of genetic merit using LD-based haplotypes in the Nordic Holstein population. **BMC Genomics**, v. 15, n. 1, p. 1171, December 23,2014. Disponível em: < https://doi.org/10.1186/1471-2164-15-1171 >.

CUYABANO, B. C.; SU, G.; LUND, M. S. Selection of haplotype variables from a high-density marker map for genomic prediction. **Genetics Selection Evolution**, v. 47, n. 1, p. 61, 2015.

DOYLE, J. J.; DOYLE, J. L. Isolation ofplant DNA from fresh tissue. **Focus**, v. 12, n. 13, p. 39-40, 1990.

GABRIEL, S. B.; SCHAFFNER, S. F.; NGUYEN, H.; MOORE, J. M.; ROY, J.; BLUMENSTIEL, B.; HIGGINS, J.; DEFELICE, M.; LOCHNER, A.; FAGGART, M.; LIU-CORDERO, S. N.; ROTIMI, C.; ADEYEMO, A.; COOPER, R.; WARD, R.; LANDER, E. S.; DALY, M. J.; ALTSHULER, D. The Structure of Haplotype Blocks in the Human Genome. **Science**, v. 296, n. 5576, p. 2225, 2002. Disponível em: < http://science.sciencemag.org/content/296/5576/2225.abstract >.

GUPTA, P. K.; RUSTGI, S.; KULWAL, P. L. Linkage disequilibrium and association studies in higher plants: Present status and future prospects. **Plant Molecular Biology**, v. 57, n. 4, p. 461-485, 2005/03/01,2005. Disponível em: < https://doi.org/10.1007/s11103-005-0257-z >.

KAESSMANN, H.; ZÖLLNER, S.; GUSTAFSSON, A. C.; WIEBE, V.; LAAN, M.; LUNDEBERG, J.; UHLÉN, M.; PÄÄBO, S. Extensive Linkage Disequilibrium in Small Human Populations in Eurasia. **The American Journal of Human Genetics**, v. 70, n. 3, p. 673-685, 2002/03/01/,2002. Disponível em: < http://www.sciencedirect.com/science/article/pii/S0002929707602705 >.

KHATKAR, M. S.; NICHOLAS, F. W.; COLLINS, A. R.; ZENGER, K. R.; CAVANAGH, J. A.; BARRIS, W.; SCHNABEL, R. D.; TAYLOR, J. F.; RAADSMA, H. W. Extent of genome-wide linkage disequilibrium in Australian Holstein-Friesian cattle based on a high-density SNP panel. **BMC Genomics**, v. 9, n. 1, p. 187, April 24,2008. Disponível em: < https://doi.org/10.1186/1471-2164-9-187 >.

KHATKAR, M. S.; ZENGER, K. R.; HOBBS, M.; HAWKEN, R. J.; CAVANAGH, J. A. L.; BARRIS, W.; MCCLINTOCK, A. E.; MCCLINTOCK, S.; THOMSON, P. C.; TIER, B.; NICHOLAS, F. W.; RAADSMA, H. W. A primary assembly of a bovine haplotype block map based on a 15,036-single-nucleotide polymorphism panel genotyped in holstein-friesian cattle. **Genetics**, v. 176, n. 2, p. 763-772, 2007. Disponível em: < https://www.ncbi.nlm.nih.gov/pubmed/17435229 https://www.ncbi.nlm.nih.gov/pmc/PMC1894606/ >.

KNAUS, B. J.; GRÜNWALD, N. J. vcfr: a package to manipulate and visualize variant call format data in R. **Molecular Ecology Resources**, v. 17, n. 1, p. 44-53, 2017. Disponível em: < https://onlinelibrary.wiley.com/doi/abs/10.1111/1755-0998.12549 >.

KNUEPPEL, S.; ROHDE, K.; KNUEPPEL, M. S. Package 'HapEstXXR'. 2015.

LARSSON, S. J.; LIPKA, A. E.; BUCKLER, E. S. Lessons from Dwarf8 on the Strengths and Weaknesses of Structured Association Mapping. **PLOS Genetics**, v. 9, n. 2, p. e1003246, 2013. Disponível em: < https://doi.org/10.1371/journal.pgen.1003246 >.

LEWONTIN, R. The interaction of selection and linkage. I. General considerations; heterotic models. **Genetics**, v. 49, n. 1, p. 49, 1964.

LU, Y.; SHAH, T.; HAO, Z.; TABA, S.; ZHANG, S.; GAO, S.; LIU, J.; CAO, M.; WANG, J.; PRAKASH, A. B.; RONG, T.; XU, Y. Comparative SNP and haplotype analysis reveals a higher genetic diversity and rapider LD decay in tropical than temperate germplasm in maize. **PloS one**, v. 6, n. 9, p. e24861-e24861, 2011. Disponível em: < https://www.ncbi.nlm.nih.gov/pubmed/21949770 https://www.ncbi.nlm.nih.gov/pmc/PMC3174237/ >.

MATIAS, F. I.; GALLI, G.; CORREIA GRANATO, I. S.; FRITSCHE-NETO, R. Genomic Prediction of Autogamous and Allogamous Plants by SNPs and Haplotypes. **Crop Science**, v. 57, p. 2951-2958, 2017. Disponível em: < http://dx.doi.org/10.2135/cropsci2017.01.0022 >.

MCEVOY, B. P.; POWELL, J. E.; GODDARD, M. E.; VISSCHER, P. M. Human population dispersal “Out of Africa” estimated from linkage disequilibrium and allele frequencies of SNPs. **Genome Research**, v. 21, n. 6, p. 821-829, June 1, 2011,2011. Disponível em: < http://genome.cshlp.org/content/21/6/821.abstract >.

MCKAY, S. D.; SCHNABEL, R. D.; MURDOCH, B. M.; MATUKUMALLI, L. K.; AERTS, J.; COPPIETERS, W.; CREWS, D.; NETO, E. D.; GILL, C. A.; GAO, C.; MANNEN, H.; STOTHARD, P.; WANG, Z.; VAN TASSELL, C. P.; WILLIAMS, J. L.; TAYLOR, J. F.; MOORE, S. S. Whole genome linkage disequilibrium maps in cattle. **BMC Genetics**, v. 8, n. 1, p. 74, October 25,2007. Disponível em: < https://doi.org/10.1186/1471-2156-8-74 >.

MEUWISSEN, T. H. E.; HAYES, B. J.; GODDARD, M. E. Prediction of total genetic value using genome-wide dense marker maps. **Genetics**, v. 157, 2001.

MUELLER, J. C.; LÕHMUSSAAR, E.; MÄGI, R.; REMM, M.; BETTECKEN, T.; LICHTNER, P.; BISKUP, S.; ILLIG, T.; PFEUFER, A.; LUEDEMANN, J.; SCHREIBER, S.; PRAMSTALLER, P.; PICHLER, I.; ROMEO, G.; GADDI, A.; TESTA, A.; WICHMANN, H.-E.; METSPALU, A.; MEITINGER, T. Linkage Disequilibrium Patterns and tagSNP Transferability among European Populations. **The American Journal of Human Genetics**, v. 76, n. 3, p. 387-398, 2005/03/01/,2005. Disponível em: < http://www.sciencedirect.com/science/article/pii/S0002929707633369 >.

NILSEN, H.; HAYES, B.; BERG, P. R.; ROSETH, A.; SUNDSAASEN, K. K.; NILSEN, K.; LIEN, S. Construction of a dense SNP map for bovine chromosome 6 to assist the assembly of the bovine genome sequence. **Anim Genet**, v. 39, 2008. Disponível em: < https://doi.org/10.1111/j.1365-2052.2007.01686.x >.

PENGELLY, R. J.; COLLINS, A. Linkage disequilibrium maps to guide contig ordering for genome assembly. **Bioinformatics**, v. 35, n. 4, p. 541-545, 2018. Disponível em: < https://dx.doi.org/10.1093/bioinformatics/bty687 >. Acesso em: 3/14/2019.

PRITCHARD, J. K.; PRZEWORSKI, M. Linkage disequilibrium in humans: models and data. **American journal of human genetics**, v. 69, n. 1, p. 1-14, 2001. Disponível em: < https://www.ncbi.nlm.nih.gov/pubmed/11410837 https://www.ncbi.nlm.nih.gov/pmc/PMC1226024/ >.

QANBARI, S.; PIMENTEL, E. C. G.; TETENS, J.; THALLER, G.; LICHTNER, P.; SHARIFI, A. R.; SIMIANER, H. The pattern of linkage disequilibrium in German Holstein cattle. **Animal Genetics**, v. 41, n. 4, p. 346-356, 2010/08/01,2010. Disponível em: < https://doi.org/10.1111/j.1365-2052.2009.02011.x >. Acesso em: 2018/09/25.

QIN, Z. S.; NIU, T.; LIU, J. S. Partition-ligation–expectation-maximization algorithm for haplotype inference with single-nucleotide polymorphisms. **The American Journal of Human Genetics**, v. 71, n. 5, p. 1242-1247, 2002.

ROMAY, M. C.; MILLARD, M. J.; GLAUBITZ, J. C.; PEIFFER, J. A.; SWARTS, K. L.; CASSTEVENS, T. M.; ELSHIRE, R. J.; ACHARYA, C. B.; MITCHELL, S. E.; FLINT-GARCIA, S. A.; MCMULLEN, M. D.; HOLLAND, J. B.; BUCKLER, E. S.; GARDNER, C.

A. Comprehensive genotyping of the USA national maize inbred seed bank. **Genome Biology**, v. 14, n. 6, p. R55, June 11,2013. Disponível em: < https://doi.org/10.1186/gb-2013-14-6-r55 >.

SCHROOTEN, C.; SCHOPEN, G.; PARKER, A.; MEDLEY, A.; BEATSON, P. Across-breed genomic evaluation based on bovine high density genotypes and phenotypes of bulls and cows. In: Proc Assoc Advmt Anim Breed Genet, 2013, p. 138-141.

SEABURY, C. M.; SEABURY, P. M.; DECKER, J. E.; SCHNABEL, R. D.; TAYLOR, J. F.; WOMACK, J. E. Diversity and evolution of 11 innate immune genes in Bos taurus taurus and Bos taurus indicus cattle. **Proceedings of the National Academy of Sciences**, v. 107, n. 1, p. 151-156, 2010. Disponível em: < https://www.pnas.org/content/pnas/107/1/151.full.pdf >.

SHIM, H.; CHUN, H.; ENGELMAN, C. D.; PAYSEUR, B. A. Genome-wide association studies using single-nucleotide polymorphisms versus haplotypes: an empirical comparison with data from the North American Rheumatoid Arthritis Consortium. **BMC Proceedings**, v. 3, n. 7, p. S35, December 15,2009. Disponível em: < https://doi.org/10.1186/1753-6561-3-S7-S35 >.

SLATKIN, M. Linkage disequilibrium — understanding the evolutionary past and mapping the medical future. **Nature Reviews Genetics**, v. 9, p. 477, 06/01/online,2008. Disponível em: < http://dx.doi.org/10.1038/nrg2361 >.

TENAILLON, M. I.; SAWKINS, M. C.; LONG, A. D.; GAUT, R. L.; DOEBLEY, J. F.; GAUT, B. S. Patterns of DNA sequence polymorphism along chromosome 1 of maize (Zea mays ssp. mays L.). **Proceedings of the National Academy of Sciences of the United States of America**, v. 98, n. 16, p. 9161-9166, 2001. Disponível em: < https://www.ncbi.nlm.nih.gov/pubmed/11470895 https://www.ncbi.nlm.nih.gov/pmc/PMC55390/ >.

TRUNTZLER, M.; RANC, N.; SAWKINS, M. C.; NICOLAS, S.; MANICACCI, D.; LESPINASSE, D.; RIBIÈRE, V.; GALAUP, P.; SERVANT, F.; MULLER, C.; MADUR, D.; BETRAN, J.; CHARCOSSET, A.; MOREAU, L. Diversity and linkage disequilibrium features in a composite public/private dent maize panel: consequences for association genetics as evaluated from a case study using flowering time. **Theoretical and Applied Genetics**, v. 125, n. 4, p. 731-747, 2012/08/01,2012. Disponível em: < https://doi.org/10.1007/s00122-012-1866-y >.

VAN INGHELANDT, D.; REIF, J. C.; DHILLON, B. S.; FLAMENT, P.; MELCHINGER, A. E. Extent and genome-wide distribution of linkage disequilibrium in commercial maize germplasm. **Theoretical and Applied Genetics**, v. 123, n. 1, p. 11-20, 2011/06/01,2011. Disponível em: < https://doi.org/10.1007/s00122-011-1562-3 >.

VILLUMSEN, T. M.; JANSS, L.; LUND, M. S. The importance of haplotype length and heritability using genomic selection in dairy cattle. **J Anim Breed Genet**, v. 126, 2009. Disponível em: < https://doi.org/10.1111/j.1439-0388.2008.00747.x >.

WALL, J. D.; PRITCHARD, J. K. Haplotype blocks and linkage disequilibrium in the human genome. **Nature Reviews Genetics**, v. 4, n. 8, p. 587-597, 2003/08/01,2003. Disponível em: < https://doi.org/10.1038/nrg1123 >.

WILSON, L. M.; WHITT, S. R.; IBÁÑEZ, A. M.; ROCHEFORD, T. R.; GOODMAN, M. M.; BUCKLER, E. S. Dissection of Maize Kernel Composition and Starch Production by Candidate Gene Association. **The Plant Cell**, v. 16, n. 10, p. 2719, 2004. Disponível em: < http://www.plantcell.org/content/16/10/2719.abstract >.

WIMMER, V.; ALBRECHT, T.; AUINGER, H.-J.; SCHÖN, C.-C. synbreed: a framework for the analysis of genomic prediction data using R. **Bioinformatics**, v. 28, n. 15, p. 2086-2087, 2012.

YAN, J.; SHAH, T.; WARBURTON, M. L.; BUCKLER, E. S.; MCMULLEN, M. D.; CROUCH, J. Genetic Characterization and Linkage Disequilibrium Estimation of a Global Maize Collection Using SNP Markers. **PLOS ONE**, v. 4, n. 12, p. e8451, 2009. Disponível em: < https://doi.org/10.1371/journal.pone.0008451 >.

ZHANG, K.; QIN, Z.; CHEN, T.; LIU, J. S.; WATERMAN, M. S.; SUN, F. HapBlock: haplotype block partitioning and tag SNP selection software using a set of dynamic programming algorithms. **Bioinformatics**, v. 21, n. 1, p. 131-134, 2005. Disponível em: < https://doi.org/10.1093/bioinformatics/bth482 >. Acesso em: 4/22/2019.