

RESEARCH ARTICLE

# Single Marker and Haplotype-Based Association Analysis of Semolina and Pasta Colour in Elite Durum Wheat Breeding Lines Using a High-Density Consensus Map

Amidou N'Diaye<sup>1\*</sup>, Jemanesh K. Haile<sup>1</sup>, Aron T. Cory<sup>1</sup>, Fran R. Clarke<sup>2</sup>, John M. Clarke<sup>1</sup>, Ron E. Knox<sup>2</sup>, Curtis J. Pozniak<sup>1</sup>

**1** Department of Plant Sciences and Crop Development Centre, University of Saskatchewan, Saskatoon, Saskatchewan, Canada, **2** Semiarid Prairie Agricultural Research Centre, Agriculture and Agri-Food Canada, Swift Current, Saskatchewan, Canada

\* [amidou.ndiaye@usask.ca](mailto:amidou.ndiaye@usask.ca)



**OPEN ACCESS**

**Citation:** N'Diaye A, Haile JK, Cory AT, Clarke FR, Clarke JM, Knox RE, et al. (2017) Single Marker and Haplotype-Based Association Analysis of Semolina and Pasta Colour in Elite Durum Wheat Breeding Lines Using a High-Density Consensus Map. PLoS ONE 12(1): e0170941. doi:10.1371/journal.pone.0170941

**Editor:** Pawan L. Kulwal, Mahatma Phule Krishi Vidyapeeth College of Agriculture, INDIA

**Received:** November 3, 2016

**Accepted:** January 12, 2017

**Published:** January 30, 2017

**Copyright:** © 2017 N'Diaye et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper and its Supporting Information files.

**Funding:** We gratefully acknowledge funding provided by Genome Canada to the CTAG project, as well as by the Western Grains Research Foundation, the Province of Saskatchewan, and Agriculture and Agri-Food Canada.

**Competing Interests:** The authors have declared that no competing interests exist.

## Abstract

Association mapping is usually performed by testing the correlation between a single marker and phenotypes. However, because patterns of variation within genomes are inherited as blocks, clustering markers into haplotypes for genome-wide scans could be a worthwhile approach to improve statistical power to detect associations. The availability of high-density molecular data allows the possibility to assess the potential of both approaches to identify marker-trait associations in durum wheat. In the present study, we used single marker- and haplotype-based approaches to identify loci associated with semolina and pasta colour in durum wheat, the main objective being to evaluate the potential benefits of haplotype-based analysis for identifying quantitative trait loci. One hundred sixty-nine durum lines were genotyped using the Illumina 90K Infinium iSelect assay, and 12,234 polymorphic single nucleotide polymorphism (SNP) markers were generated and used to assess the population structure and the linkage disequilibrium (LD) patterns. A total of 8,581 SNPs previously localized to a high-density consensus map were clustered into 406 haplotype blocks based on the average LD distance of 5.3 cM. Combining multiple SNPs into haplotype blocks increased the average polymorphism information content (PIC) from 0.27 per SNP to 0.50 per haplotype. The haplotype-based analysis identified 12 loci associated with grain pigment colour traits, including the five loci identified by the single marker-based analysis. Furthermore, the haplotype-based analysis resulted in an increase of the phenotypic variance explained (50.4% on average) and the allelic effect (33.7% on average) when compared to single marker analysis. The presence of multiple allelic combinations within each haplotype locus offers potential for screening the most favorable haplotype series and may facilitate marker-assisted selection of grain pigment colour in durum wheat. These results suggest a benefit of haplotype-based analysis over single marker analysis to detect loci associated with colour traits in durum wheat.

## Introduction

Marker-assisted selection (MAS) is increasing in use in plant breeding as a means to enrich selections from segregating populations for desirable alleles influencing economically important traits. In durum wheat (*Triticum turgidum* L. var *durum*), most MAS has focused on selection of traits controlled by single genes or large effect quantitative trait loci (QTL) [1]. Identification of robust markers is becoming easier because of the availability of high-density genetic maps (e.g., [2–4]). Although several QTL were reported in the literature, relatively few are practically used in breeding programs [5]. Reasons for their lack of practical use are mostly due to the difficulties with context dependencies caused by genotype-environmental interactions and/or epistasis, to the limitations of sampling bi-parental populations with multi-genic traits, and to lack of follow-through research to validate identified QTL [6–8]. Identification of marker-trait associations using association mapping techniques, could avoid some of these context dependencies.

Association mapping (AM) is a complementary strategy to QTL mapping to identify associations between genotype and phenotype [9] and is based on linkage disequilibrium (LD) in a collection of unrelated individuals. In contrast to bi-parental mapping, AM allows a broader population from which to sample multiple alleles and to map with higher resolution [9, 10]. Most AM studies test correlations between a single marker and phenotypes. However, because patterns of variation within genomes are inherited as linkage blocks [11–13], clustering markers into haplotypes is gaining acceptance in genome-wide association studies.

Advances in high-throughput genotyping technologies have made SNPs markers of choice for genome-wide association studies. SNPs are the most abundant class of sequence variability in the genome and thus have the potential to provide the highest map resolution (Jones et al. 2007). However, SNPs are usually bi-allelic so each provides less polymorphism information content (PIC) than markers such as SSRs (multi-allelic), thus marker density must be increased. This limitation can be overcome by merging SNPs into haplotypes (Lu et al. 2012). Haplotype-based analyses have been successfully carried out mostly in human genetics due to the availability of data from the HapMap project [14, 15]. Similar efforts are gaining ground in various crops such as maize [16–18], rice [19–22] and soybean [23–26]. In wheat, haplotype analyses were performed for QTL or marker-trait association studies [27–29], pattern of genetic variations [30–32] and gene diversity [33–35]. However, only a relatively low number of SNPs and/or SSR markers were used for marker-trait association studies.

Various arguments advocating for haplotype-based analysis rather than single marker analysis have been proposed. In particular, haplotype-based analysis could capture epistatic interactions between SNPs at a locus [36, 37]; provide more information to estimate whether two alleles are identical by descent [38]; elucidate the exact biological role played by neighbouring amino-acids on a protein structure [39]; reduce the number of tests and hence the type I error rate [40]; capture information from evolutionary history [41]; and provide more power than single marker when an allelic series exists at a locus [42–45]. The fundamental question that arises from all these rationales is to know whether the power and accuracy of association mapping can be improved by grouping SNPs into haplotype blocks (see [46] for a review). Intuitively, one could expect haplotypes to be more powerful due to the simultaneous use of multiple markers information [47–49]. Simulation studies have shown that clustering of markers into haplotypes can provide greater QTL detection power and mapping accuracy than single markers [43, 50–52], and this was supported in empirical studies [17, 18, 53–60]. Haplotype-based approach improves prediction accuracy compared with the individual SNP approach [61–63]. In contrast, a few studies found no apparent advantage of haplotype-based analysis over individual SNP analysis [64–66] for detecting QTL. The outcome of

the haplotype-based analysis could change under different models relating genotype to phenotype or under different demographic scenarios [67]. Indeed statistical adjustments for population structure and inclusion of kinship relationships is critical to reduce type I error rates of association mapping studies regardless of a haplotype or single marker approach is used [68–70].

There are various criteria for defining haplotype blocks [12, 46, 57, 71, 72]. In particular, haplotype blocks can be defined using a sliding window [28, 57, 73–75] or combining SNPs within a specific window size [22, 58, 76]. Studies in barley provided good support for the use of simple overlapping sliding windows of three SNPs [57]. Other studies proposed different numbers of SNPs for sliding windows, ranging from 2 to 10 SNPs [28, 77–80]. Although this approach is easy to implement, it could potentially lead to large degrees of freedom in the test statistic due to the large number of haplotypes.

A key factor in the success of whole-genome association mapping remains adequate marker coverage across the genome because sparse coverage reduces the power for marker identification [81]. However, the extent of genotyping required increases with rapid LD decay. Linkage disequilibrium is higher in autogamous species due to lower effective recombination [82]. In durum wheat, LD is limited to distances of 2 to 5 cM but is not uniform along chromosomes [83]. Advances in sequencing and genotyping technology allow generation of large amounts of SNP data and the Illumina 90,000 iSelect SNP chip [4] allows development of several robust high-density genetic maps of tetraploid wheat (see [84] for review). We published the first high-density SNP consensus map which anchored over 35,000 SNP markers to all 14 durum wheat chromosomes [3]. The average marker density was 0.079 cM/marker for the B genome and 0.101 cM/marker for the A genome, which provides a framework for association mapping. Because the majority of mapped SNPs are gene-derived markers, this map provides valuable anchor points for post-mapping genetic analysis of the loci and QTL [3].

Improvement of yellow pigment (YP) concentration in durum grain is targeted globally by breeding programs due to increased market demand for bright yellow colour of semolina and pasta products (see [85] for review). The genetics of YP is complex [86], and is due to carotenoid pigment content in the endosperm. Quantitative trait loci were detected on all chromosomes of the durum genome, and genomic regions housing known YP QTL were confirmed on groups 1, 2 and 3 chromosomes [87]. Quantitative trait loci analyses for YP was performed in both hexaploid [88–91] and durum wheat [86, 92–98]. In particular, a major QTL of YP was detected on chromosome 7AL by Parker et al. [90], explaining 60% of the genetic variation and supported by other studies [89, 91, 94, 95, 97, 98]. By contrast, Elouafi et al. [93] detected a major QTL of YP on 7B accounting for 53% of the total variation, and also reported by Kuchel et al. [88], Pozniak et al. [86], Zhang et al. [97] and Zhang and Dubcovsky [98]. Several minor QTL for YP were detected on chromosomes 3A [90], 4A and 5A [99], 2A, 4B and 6B [86], 4B and 6B [97], 1A, 3B and 5B [95], 3B and 5B [94], 1A, 1B, 3B and 4A [91].

Reimer et al. [87] utilized a genetically diverse collection of cultivars and breeding lines collected from global breeding programs, and performed association mapping for grain YP concentration. Although AM was successful at identifying QTL, we have not applied these to MAS because validation experiments showed most QTL did not explain sufficient proportions of phenotypic variation in our locally-adapted breeding materials. In addition, several of the QTL that we discovered were specific to lines from the diverse collection but most were identical by state in our breeding material, despite large phenotypic differences in trait expression [100, 101]. One strategy to overcome such limitations is to perform association mapping in locally-adapted breeding material [5]. Phenotypic data collected during the course of testing of inbred lines within a breeding program, often with replication over environments, is a valuable resource for discovery of marker associations because these lines are expected to carry a high

proportion of relevant, desirable alleles. However such phenotypic data sets are usually unbalanced because breeders tend to cull materials throughout the breeding cycle, making exploitation of such data complicated [5]. Utilization of a common set of check cultivars over successive breeding cycles in combination with mixed models which incorporate correlations among environments can be used to estimate best linear unbiased estimates (BLUEs) for individual lines and these could then be used to evaluate marker-trait associations. The utility of this approach has been demonstrated for durum wheat [101, 102], bread wheat [103], barley [104–107], potato [108] and sugarcane [109].

Taken together, the recent advances in SNP marker detection in durum wheat and robust phenotypic data collected from our breeding programs [100, 101] provided the opportunity to further assess association mapping strategies of practical use in a breeding program. Also, the availability of a high-density SNP consensus map allows the opportunity to assess haplotype based approaches for AM in durum wheat. The main objective of this study was to compare the two mapping approaches to explore the potential of haplotype-based analysis in durum wheat and to identify genomic regions associated with pigment colour in semolina and pasta.

## Materials and Methods

### Plant material

One hundred and sixty-nine durum lines were selected for the study (S1 Table) from the official Canadian durum cultivar registration trial grown in Canada between 1999 and 2013. Phenotypic data and trials were described in previous reports [101, 110]. Candidate lines were tested for one to three years but only lines with at least two years of data were included in the present study. Each trial included check cultivars; AC Avonlea [111], AC Morse, AC Navigator [112] and Strongfield [113] since 1999, and Commander [114] added in 2001. The checks AC Morse and Commander were dropped in 2013 and the new check Brigade [115] was brought in. Trials were arranged in lattice designs with four replications, except in 2013 where most locations comprised three replications.

### Quality analyses

End-use quality traits were measured on composite grain samples of locations within years. The composites included locations with acceptable physical condition (commercial grade Canada Western Amber Durum #3 or better), and blended to give a target grain protein concentration of about 13%. Yellow pigment (parts per million) of semolina was measured using the AACC method 14–50 (AACC 2000). Colour of semolina and of pasta dried at 70°C was measured with a Minolta CR–200 Chroma Meter (Minolta, Japan) equipped with a 50 mm measuring head to assess CIELAB  $a^*$  and  $b^*$  colour space units. Semolina  $a^*$  measurement was discontinued after 2008. The colour loss during pasta manufacture was estimated by regressing pasta  $b^*$  on semolina  $b^*$  [101]. The residuals for each genotype, actual minus predicted values, were used as a measure of colour loss in the analysis. Positive residuals indicate less pigment loss than the population average, while negative residuals indicate greater than average loss. The data were analysed with SAS version 9.3 [116] Proc Mixed using lines (fixed) with years (random) as replication to generate lsmeans. The analyses included all genotypes tested in the registration trial (approximately 300), not just those genotyped, so as to provide a better estimate of random variances and covariances. Pearson's correlations were performed among the lsmeans of the traits.

## SNP genotyping and genetic diversity analysis

Genomic DNA was extracted from fresh young leaf tissue using a modified CTAB method [117]. DNA was quantified using PicoGreen (Invitrogen) fluorescence assay, and diluted to 50 ng/ $\mu$ l. Genotyping was performed according to the method published previously [4]. The 90K iSelect assay chips were run on an Illumina HiScan for imaging and the resulting data were loaded into GenomeStudio v2011.1 software (Illumina) for SNP calling. After filtering those SNPs with ambiguous calls, having more than 25% missing values, or having MAF < 0.05, a total of 12,234 polymorphic SNP markers were used for analyses. PowerMarker V3.25 software [118] was used to calculate the summary statistics including allele number, allele frequency and PIC.

## Genotyping with *Rht-B1b* and *Lpx-B1* genes

Because the *Lpx-B1* deletion has been associated with reduced colour loss during processing [119, 120], the registration lines were genotyped with a *Lpx-B1* marker. The registration lines were also genotyped with *Rht-B1b*, an allele known to confer semi-dwarf growth habit in wheat [121] because the relatively few semi-dwarf lines in the panel were selected for very high pigment, presenting the possibility of spurious associations. In order to relate the association signals to *Lpx-B1* and *Rht-B1b*, pairwise LD ( $r^2$ ) was performed between all 4B association signals and these genes using MIDAS software [122].

## Population structure and linkage disequilibrium analysis

Population structure is one of several important factors that strongly influence LD. The presence of population stratification and an unequal distribution of alleles within groups can result in spurious associations [82]. Population structure was estimated using discriminant analysis of principal components (DAPC) as implemented in the Adegenet R package version 1.4 [123]. To avoid unstable results, the maximum number of principal components (PCs) should be  $\leq N/3$ ,  $N$  being the number of lines [123]. Therefore, 56 PCs were included in the model.

Single nucleotide polymorphism markers having MAF < 0.05 were filtered out prior to estimating the LD because the estimation of LD using  $r^2$  is dependent on allele frequency and rare alleles can inflate the  $r^2$  [124]. The LD was estimated as a correlation coefficient ( $r^2$ ) between all pairwise comparisons of loci both genome-wide and at the chromosome level, using the Genetics R package available at <http://cran.r-project.org/>. The  $r^2$  distribution of loci belonging to different chromosomes was used to calculate a threshold of  $r^2$  for LD which was taken from the parametric 95<sup>th</sup> percentile of that distribution [125]. The genetic distance corresponding to that  $r^2$  threshold was determined with nonlinear regression by plotting the genetic distance over which LD decayed, using R code written by F. Marroni that is available at <http://fabiomarroni.wordpress.com/>.

## Marker imputation and haplotype construction

Prior to haplotype construction, missing calls were imputed using the RF regression procedure [126] as implemented in the R package “randomForest” [127, 128]. The RF procedure has been described in detail for imputing missing genotypes for genomic selection [129] and has been successfully used for genetic diversity analysis [130] and genome-wide association studies [131–133].

For haplotype construction, redundant information known to introduce bias [134] was first filtered out using an in-house Ruby script. When two or more SNPs had the same genotype across all breeding lines along the same chromosome, they were represented by a single

genotype. Thus, a total of 8,581 SNPs were used for the analysis. The SNPs were sorted by position along each chromosome based on the durum high-density SNP-based consensus map [3]. Those SNPs spanned all 14 chromosomes of durum wheat with an average density of one marker per 0.3 cM (S2 Table). Then, SNPs within a window size of 5.3 cM (estimate of average LD decay) on the same chromosome were combined to form a haplotype block and assigned to the same locus. Loci for each chromosome were named as combination of the prefix 'hap', the chromosome and an index that is the incrementing number (1 to N, N being the total number of haplotypes) of the haplotype along the chromosome (e.g., *hap\_1A\_1* and *hap\_1B\_2* designate the first haplotype on chromosome 1A and the second haplotype on chromosome 1B, respectively). Only 17 haplotypes appeared to be rare (MAF < 0.05) and were excluded from further analyses.

## Association analysis

Marker-trait associations were carried out using the general linear model (GLM) and the mixed linear model (MLM) as implemented in TASSEL software version 3 [135]. In order to control spurious associations, population structure and/or relatedness between individuals were taken into account in both GLM and MLM procedures. The Q matrix was based on the four groups from the discriminant analysis of principal components and the kinship (K) matrix was calculated using TASSEL. To control for experiment-wise error, nominal *P*-values were adjusted according to Storey-Taylor-Siegmund's adaptive step-up procedure [136] as implemented in the Mutoss R package [137]. A false discovery rate (FDR) of 5% was used for computation and only SNPs and haplotypes having an adjusted *P*-value less than 0.05 were declared significant. The allelic effect of haplotypes and SNPs was estimated as the difference between the mean value of the lines carrying these haplotypes and SNPs, and the mean value of the entire population for each trait. Thus, only SNPs and haplotypes having relatively strong allelic effect were reported.

## Results

### Analysis of phenotypic data

Large phenotypic variation was observed among the breeding lines for all of the traits (Table 1). In particular, pasta *a*\* and semolina pigment values ranged from 1.66 to 5.79 and 6.0 to 12.05, respectively. Significant differences were observed between subpopulations (Table 1). The correlation among colour traits is presented in Table 2. Pasta *a*\* was significantly ( $P < 0.001$ ) correlated with all of the traits, and ranged from  $r = 0.40$  (pigment loss) to 0.69 (semolina pigment). Semolina *a*\* was correlated with only pasta *a*\*. However, semolina pigment exhibited strong correlation with semolina *b*\*, pasta *a*\* and pasta *b*\*. The highest correlation ( $r = 0.96$ ) was observed between semolina pigment and semolina *b*\*.

### Population structure and LD decay

Four subpopulations among the breeding lines were inferred using discriminant analysis of principal components (Fig 1). The accessions list with their subpopulations is shown in S1 Table. The total amount of genetic variation explained by the first 56 eigenvectors was 80%. Breeding lines were differentiated according to pedigree, source breeding program, and testing year. Subpopulation 1 is largely AC Avonlea [111] and/or Strongfield [113] heritage and comprised on average the most recent lines in the trial. Subpopulation 2 is based on Kyle [138] heritage, with the majority of the lines from the Agriculture and AgriFood Canada, Swift Current program and representing an earlier era of testing than subpopulation 1. Subpopulation 3

**Table 1. Average values (minimum, maximum) of color traits for the whole population (WP) and least squares means by sub-populations (SP).**

Traits	WP	SP1 (n = 46) <sup>1</sup>	SP2 (n = 38)	SP3 (n = 35)	SP4 (n = 50)	LSD <sub>.05</sub> <sup>2</sup>
Pasta a*	3.34 (1.66, 5.79)	3.26 <sup>a</sup>	3.95 <sup>a</sup>	3.28 <sup>b</sup>	2.98 <sup>c</sup>	0.28
Semolina a*	-3.02 (-3.24, -2.56)	-3.02 <sup>ab</sup>	-2.97 <sup>a</sup>	-3.03 <sup>b</sup>	-3.04 <sup>b</sup>	0.05
Pasta b*	64.85 (58.14–71.81)	64.51 <sup>b</sup>	66.15 <sup>a</sup>	64.71 <sup>b</sup>	64.26 <sup>b</sup>	1.08
Semolina b*	34.22 (29.39, 39.43)	34.77 <sup>a</sup>	34.58 <sup>a</sup>	34.06 <sup>ab</sup>	33.57 <sup>b</sup>	0.76
Pigment loss	-0.08 (-5.39, 6.21)	-0.92 <sup>c</sup>	0.89 <sup>a</sup>	-0.07 <sup>b</sup>	-0.06 <sup>b</sup>	0.73
Semolina pigment	8.82 (6.00, 12.05)	9.26 <sup>a</sup>	9.18 <sup>a</sup>	8.70 <sup>b</sup>	8.22 <sup>c</sup>	0.47

<sup>1</sup> Subpopulation size

<sup>2</sup> Average least significant difference, P<0.05

Values with the same appended letter are not significantly different according to the least significant difference test at p < 0.05 (for each trait). n: Subpopulation size.

\* Should be read as star (e.g., Pasta a\* is 'Pasta a star')

doi:10.1371/journal.pone.0170941.t001

contained lines with diverse ancestry from CIMMYT, University of North Dakota, Agriculture and AgriFood Canada, Winnipeg and Swift Current, and University of Saskatchewan. Subpopulation 4 was similar to subpopulation 3 but without the Swift Current component and represented the oldest era of testing of the four groups.

A total of 12,234 polymorphic SNPs were used to estimate the LD across all chromosomes. The critical  $r^2$  value from which the genome-wide LD decayed was estimated at 0.2 (Fig 2). The average genetic distance at which LD across all chromosomes decayed ( $r^2 < 0.2$ ) was 5.3 cM. Nonetheless, that distance varied among chromosomes, from 3.0 (chromosome 4A) to 9.4 cM (chromosome 5B). The LD pattern of all chromosomes is presented in S1 Fig. Only 4% of all pairs of SNPs showed very high LD ( $r^2 > 0.8$ ).

### Allele diversity as revealed by SNPs and haplotypes

After imputation, a total of 8,581 SNPs having a minor allele frequency greater than 5% and located on the high-density consensus map were used for analyses. Only 14.2% (1,222/8,581) of the SNPs showed almost equal allele frequencies between their two alternative alleles. The average PIC for these 8,581 SNPs was 0.27, ranging from 0.10 to 0.38 (Fig 3).

A total of 406 haplotype blocks containing 2 to 60 SNPs were generated. Of these haplotype blocks, 4.9% contained two SNPs, 47.5% contained three to nine SNPs and 47.6% had more than 10 SNPs. Haplotype blocks showed a higher level of allele diversity; the average PIC was 0.50, ranging from 0.10 to 0.93 (Fig 3). The number of allele combinations varied from 2 to 161 among haplotype blocks.

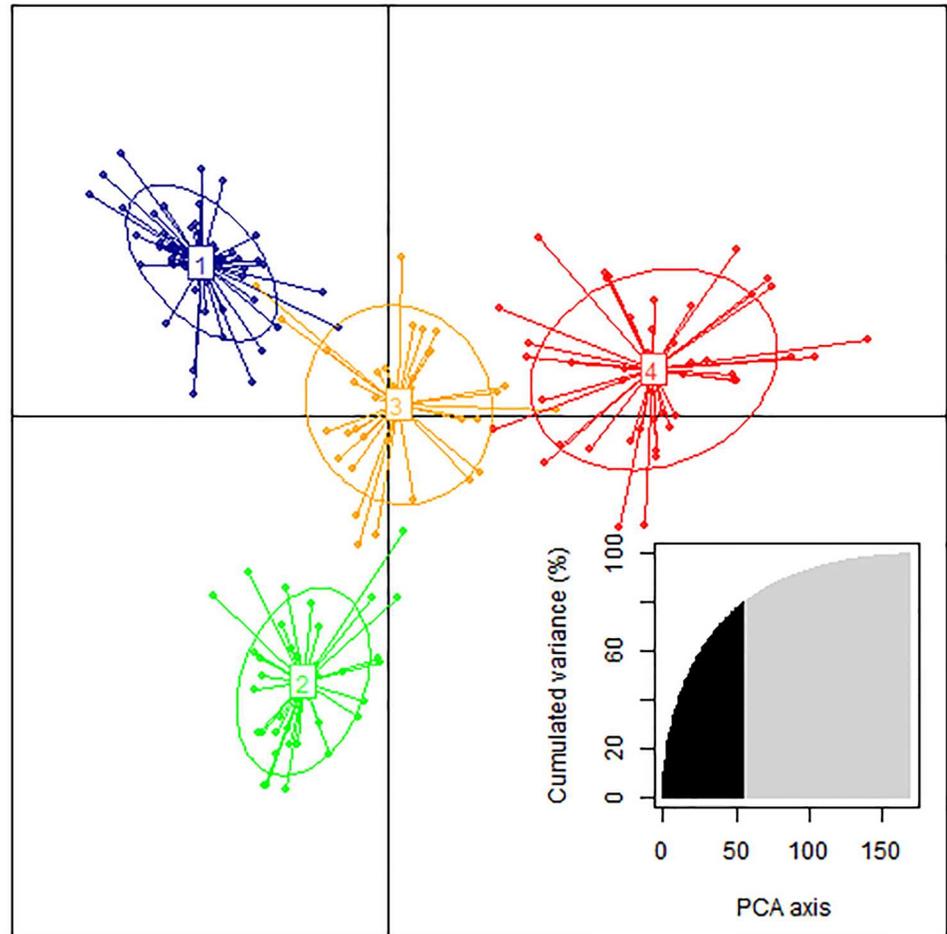
**Table 2. Correlation coefficients among colour traits for 169 durum lines.**

Traits	Semolina pigment	Pasta a*	Semolina a*	Pigment loss	Pasta b*
Pasta a*	0.69***				
Semolina a*	0.15	0.42***			
Pigment loss	0.13	0.40***	0.02		
Pasta b*	0.72***	0.65***	0.07	0.76***	
Semolina b*	0.96***	0.59***	0.09	0.1	0.72***

\*\*\*highly significant at P < 0.001

\* Should be read as star (e.g., Pasta a\* is 'Pasta a star')

doi:10.1371/journal.pone.0170941.t002



**Fig 1. Population structure of the breeding panel as revealed by discriminant analysis of principal components.** Each color represents a sub-population. The first 56 axes explained 80% of the total variance.

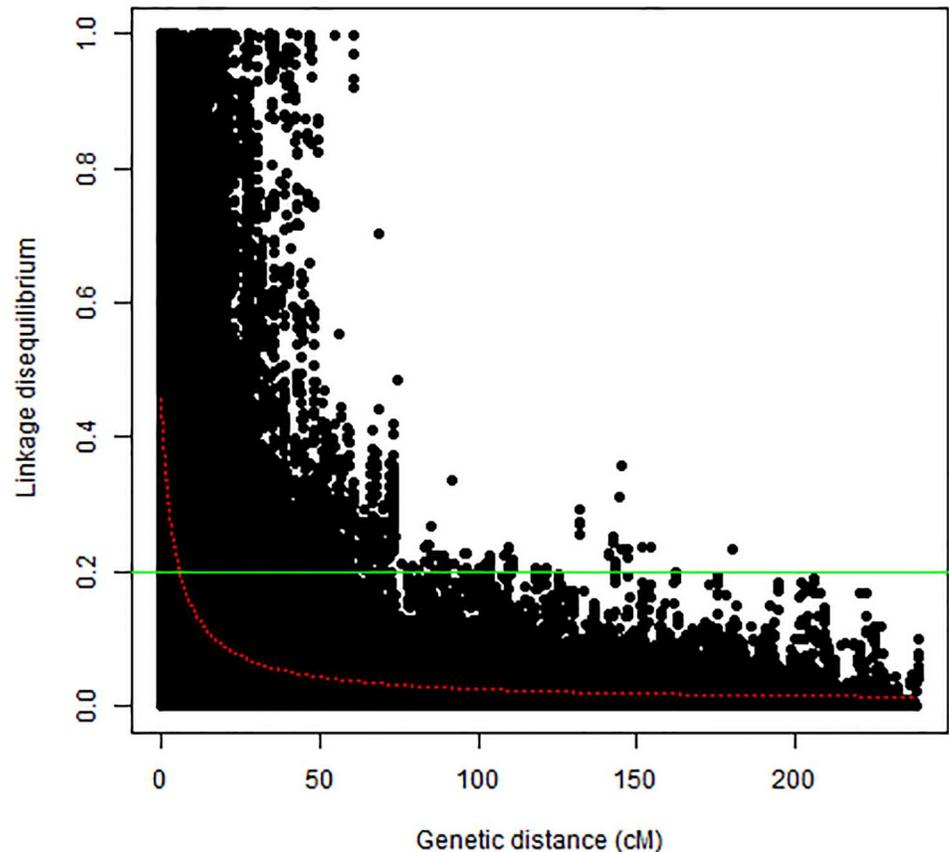
doi:10.1371/journal.pone.0170941.g001

### Loci associated with pigment colour

As shown by the quantile-quantile plots (S2 Fig), the MLM (K) and MLM (Q+K) models were significantly better than the GLM naïve and GLM (Q) models in reducing spurious associations. Only the MLM (Q+K) model was kept for the analyses because in general it performed a little better than the MLM (K) model.

Single marker-based analysis identified five loci associated with colour components (Table 3, Fig 4). The number of loci varied depending on the trait. Most of the loci revealed by the single marker-based analysis were associated with at least two traits, *Tdurum\_contig51688\_681* on 4B with pasta a\*, pasta b\* and pigment loss; *Tdurum\_contig54634-815* on 2A with pasta b\* and pigment loss; *BobWhite\_c41527\_201* on 2A and *Tdurum\_contig54832\_139* on 7A with semolina b\* and semolina pigment. Three loci associated with pigment loss were detected on chromosome 2A and 4B, explaining 11.9 to 26.2% of the phenotypic variation. A total of three loci, located on 2A and 4B, were associated with pasta b\*, explaining 9.5 to 26.2% of the variation.

Haplotype-based analysis identified a total of 12 loci associated with pigment colour components (Table 3, Fig 4). Detailed information (number and list of SNPs) on these haplotype



**Fig 2. Linkage disequilibrium (LD) scatterplot based on all pairwise comparisons between adjacent loci in the breeding panel.** The green horizontal line shows the critical  $r^2$  while the red curve displays the LD decay.

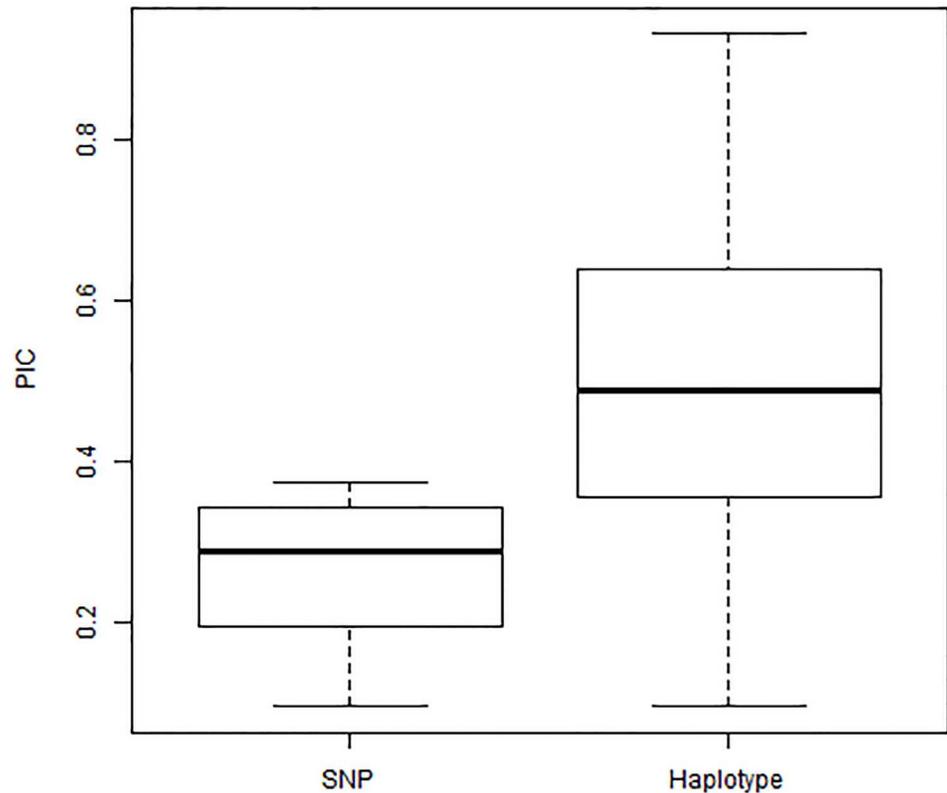
doi:10.1371/journal.pone.0170941.g002

loci are presented in [S3 Table](#). Most (8/12) of the loci were associated with at least two colour components. In particular, *hap\_4B\_6*, *hap\_4B\_7*, *hap\_4B\_12* and *hap\_5B\_25* were associated with pasta  $b^*$  and pigment loss while *hap\_2A\_18* and *hap\_7A\_32* were associated with semolina  $b^*$  and semolina pigment. For pasta  $a^*$ , a total of four loci were detected, located on chromosomes 2A, 3B and 4B. For pigment loss, six loci were detected on chromosome 2A, 3B, 4B and 5B. Six loci were detected for pasta  $b^*$ , located on 2A, 4B, 5B and 7B. Three haplotypes (*hap\_2A\_18*, *hap\_7A\_32* and *hap\_7B\_36*) were associated with both semolina  $b^*$  and semolina pigment. Over all pigment traits, the percentage of variance explained ranged from 8.5 to 40.2%.

Of the three loci on 4B associated with pigment loss, *Tdurum\_contig51688\_681* (*hap\_4B\_6*) showed strong LD ( $r^2 = 0.86$ ) with the lipoxygenase gene *Lxp-B1*, while *BS00023766\_51* (*hap\_4B\_7*) was strongly associated ( $r^2 = 0.92$ ) with the dwarfing gene *Rht-B1b* (Fig 4). These two loci appeared to be independent ( $r^2 = 0.31$ ).

### Comparison of loci identified by single marker- and haplotype-based analysis

The haplotype-based analysis identified a total of 12 loci associated with grain pigment colour traits, including all of the five loci identified by the single marker-based analysis. In particular,



**Fig 3. Boxplot of polymorphism information content (PIC) for individual SNP and haplotype.** The average PIC was 0.27 for individual SNP and 0.5 for haplotypes.

doi:10.1371/journal.pone.0170941.g003

the haplotype-based analysis detected at least one additional locus for each trait. The loci not detected by the single marker approach explained in general a relatively small amount of the phenotypic variation.

Haplotype-based analysis improved the amount of the phenotypic variance explained and the allelic effect (Table 3). Overall, there was substantial increase in the phenotypic variance explained (50.4% on average) and allelic effect (33.7% on average). For instance, the locus *hap\_4B\_6* showed an increase of 87.9% for the phenotypic variation of pasta  $a^*$ ; and the allelic effect of the locus *hap\_7A\_32* was 64.3% greater than that of the associated SNP for semolina pigment. The associated haplotype loci consisted of 2 to 10 SNPs although the number of SNPs ranged from 2 to 60 among the 406 haplotype blocks.

## Discussion

### Population structure and LD decay

In this study, the discriminant analysis of principal components [139] clustered the 169 breeding lines into four subpopulations. This population structure is in agreement with known differences in pedigree, breeding program source and era of testing in the trials.

The discriminant analysis of principal components successfully unraveled the population structure in germplasm such as cultivated sweet potato [140], rice [141], acacia [142] and sweet cherry [143]. The presence of genetic structure within a population can lead to spurious

**Table 3. Individual SNP and haplotype loci significantly associated with colour traits.**

Trait	Haplotype-based analysis					Single SNP-based analysis					Comparison		
	Haplotype	Position	nbM <sup>1</sup>	P-value	R <sup>2</sup> (%)	Effect	Marker	Position	P-value	R <sup>2</sup> (%)	Effect	IVE <sup>2</sup> (%)	IAE <sup>3</sup> (%)
Pasta a*	hap_2A_12	2A (74.6–78)	4	3.69E-04	15.1	1.6							
	hap_3B_32	3B (201.5–205.5)	3	1.04E-03	13.4	1.0							
	hap_4B_6	4B (28.5–30.8)	9	8.65E-04	35.7	1.6	Tdurum_contig51688_681	4B (28.8)	1.27E-07	19.0	1.6	87.9	0.0
	hap_4B_8	4B (37.4–41.7)	9	1.04E-03	27.6	1.7							
Pasta b*	hap_2A_5	2A (22.8–24.7)	4	8.08E-04	16.4	3.3	Tdurum_contig54634_815	2A (22.8)	1.37E-04	9.5	3.2	72.6	3.1
	hap_4B_6	4B (28.5–30.8)	9	1.62E-04	40.2	5.6	Tdurum_contig51688_681	4B (28.8)	9.66E-10	26.2	5.3	53.4	5.7
	hap_4B_7	4B (32.7–35.2)	7	1.51E-03	25.6	4.9							
	hap_4B_12	4B (58.7–60.4)	6	6.88E-04	14.1	4.5	Tdurum_contig37811_134	4B (60)	4.48E-05	10.9	3.4	29.4	32.4
	hap_5B_25	5B (129.7–131.2)	4	6.80E-04	14.1	3.1							
	hap_7B_36	7B (202.9–206.3)	2	2.48E-03	9.3	2.3							
Pigment loss	hap_2A_5	2A (22.8–24.7)	4	4.65E-04	17.6	3.5	Tdurum_contig54634_815	2A (22.8)	2.34E-05	11.9	2.4	47.9	79.2
	hap_3B_33	3B (209–209.6)	3	5.42E-04	14.6	5.4							
	hap_4B_6	4B (28.5–30.8)	9	2.03E-03	33.6	3.7	Tdurum_contig51688_681	4B (28.8)	1.17E-09	26.2	3.7	28.2	27.0
	hap_4B_7	4B (32.7–35.2)	7	4.01E-04	28.9	4.3							
	hap_4B_12	4B (58.7–60.4)	6	9.15E-05	17.5	3.3	Tdurum_contig37811_134	4B (60)	1.02E-05	13.0	2.6	34.6	26.9
	hap_5B_25	5B (129.7–131.2)	4	1.06E-03	13.5	2.1							
Semolina b*	hap_2A_18	2A (117.6–121.3)	9	8.63E-03	24.5	2.5	BobWhite_c41527_201	2A (117.7)	8.88E-08	19.2	1.9	27.6	31.6
	hap_7A_32	7A (180.2–181.8)	10	5.65E-03	34.6	3.3	Tdurum_contig54832_139	7A (181.4)	5.53E-08	19.8	2.0	74.7	65.0
	hap_7B_36	7B (202.9–206.3)	2	3.92E-03	8.5	1.3							
Semolina pigment	hap_2A_18	2A (117.6–121.3)	9	4.17E-03	27.5	1.9	BobWhite_c41527_201	2A (117.7)	2.00E-08	21.4	1.4	28.5	35.7
	hap_7A_32	7A (180.2–181.8)	10	4.42E-03	35.6	2.3	Tdurum_contig54832_139	7A (181.4)	2.67E-08	21.0	1.4	69.5	64.3
hap_7B_36	7B (202.9–206.3)	2	2.94E-03	8.9	1.0								

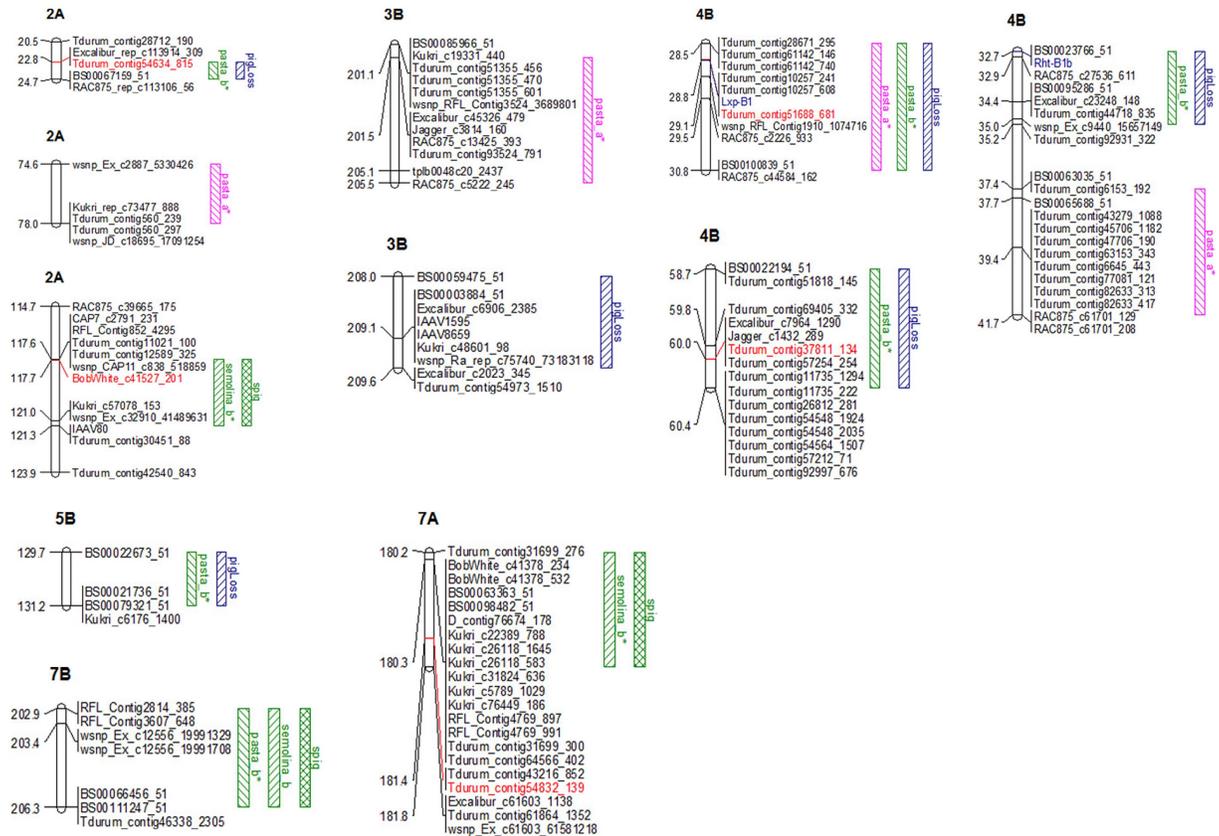
<sup>1</sup>nbM: Number of markers in the haplotype

<sup>2</sup>IVE: Increase in variance explained obtained from haplotype-based analysis compared to single SNP-based analysis

<sup>3</sup>IAE: Increase in allelic effect obtained from haplotype-based analysis compared to single SNP-based analysis

\* Should be read as star (e.g., 'Pasta a\*' is 'Pasta a star')

doi:10.1371/journal.pone.0170941.t003



**Fig 4. Genomic regions associated with semolina and pasta colour in durum wheat (based on the Maccaferri et al. (2014) map).** Markers highlighted in red are those detected by the individual SNP-based analysis.

doi:10.1371/journal.pone.0170941.g004

association signals [134, 144–148]. Understanding the actual population structure of the durum breeding panel was intended to limit the false discovery rate in the association analysis.

The average genetic distance results suggest that the LD mapping using our breeding panel can achieve a resolution of < 5 cM. Few (4%) markers showed very high LD ( $r^2 > 0.8$ ). Our results are congruent with those reported in bread wheat [149] and a geographically diverse durum wheat panel where the LD decayed within 5 cM on average [83]. However, a relatively higher (10 cM) LD decay distance was reported in a durum elite collection [150].

### Association mapping based on single marker and haplotypes

We used the 3-SNP sliding windows method and came up with a total of 8,537 haplotype blocks (data not shown) that is markedly greater than the 406 LD-based haplotype blocks we generated and used for analyses. A large number of haplotypes increases the degree of freedom for a test statistic [151]. Intuitively, the type I error rate would be higher for haplotypes derived from the 3-SNP sliding windows compared to the LD-based haplotypes. In addition, the sliding windows approach raises the question of the optimum number of markers to be included in the haplotype. A large window may include too many non-informative markers while a small window may ignore informative markers, both of which will lead to a reduction in testing power [152]. Alternatively, variable-sized sliding windows approaches have been proposed [73, 153–157]. However, most of the variable-sized methods require some computationally intensive phasing program to account for uncertain haplotype phases [158].

Because the optimal window size is always influenced by the underlying LD pattern [154, 159], we constructed haplotypes based on the average LD extent in our material. It is well known that LD patterns are variable across a large genomic region or the whole genome; therefore we also built haplotypes using chromosome-based LD. However, we found no substantial difference in size or number of haplotypes, using the chromosome-based LD distance rather than the average distance of LD decay (5.3 cM), suggesting that taking the average distance is reasonable for analysis. Similarly, the average LD distance has been used to build haplotypes in many studies when LD extent varied among chromosomes (e.g. [58, 75]). An advantage of using the LD-based method is that it avoids taking an arbitrary or suggestive number of markers to be included in the haplotype. This method is relatively easy to implement although it requires a pre-computation of the LD extent in the material under investigation. Haplotype blocks defined according to the LD usually reflect the variation patterns of the genome better than haplotype blocks artificially outlined by a fixed number of SNP [61].

The haplotype-based analysis was superior to the individual SNP analysis because it identified seven more loci associated with colour components. The same loci (*hap\_2A\_18*, *hap\_7A\_32* and *hap\_7B\_36*) detected for semolina pigment and semolina *b*\* were not surprising because these traits showed the highest correlation ( $r = 0.96$ ) amongst traits. Furthermore, the haplotype-based analysis resulted in a substantial increase (68.3% on average) in the phenotypic variance explained. The improvement ranged from an 87.9% increase of phenotypic variance explained for pasta *a*\* by haplotype *hap\_4B\_6* to 27.8% for pasta *b*\* by *hap\_2A\_18* compared to the associated single markers. Increases in the amount of phenotypic variance explained attributed to haplotype-based analysis were also reported in other crop species such as barley [57] and maize [75]. Similarly, haplotypes explained up to 80% more of the phenotypic variance for genes in cattle [53]. The increased allelic effect (e.g., 64.3% increased for semolina pigment attributed to *hap\_7A\_32*) from combining SNPs into haplotypes demonstrated an increase in power over the single marker method. However, no single allelic combination within any haplotype locus was able to select all of the lines having the desirable phenotype. Moreover, in general each haplotype carried more than one favorable allelic series. For example for pasta *a*\*, in addition to the most favorable allelic series (effect = 1.66) of *hap\_4B\_6*, two other allelic combinations showed good allelic effect on the trait, 1.41 and 1.37. Combinations of several allelic series within each haplotype, as well as the aggregation of the best haplotypes improved ability to select lines having the desirable phenotypes. These results confirm the complex genetic architecture of colour trait in durum.

Haplotype-based analysis was reported to increase the power of detecting QTL compared to single-marker analysis, based on simulated data [43]. Including more marker alleles in haplotypes leads to a higher proportion of the QTL variance being explained [52, 160] and provides additional power to the analysis [45, 161]. However, the haplotype loci detected in this study were not those having the highest number of SNPs. Thus, the power of haplotypes in increasing the variance explained could not be attributed mainly to the number of markers. The informativeness of markers within the haplotypes is more likely to be of greater importance. As functional nucleotide polymorphism (sequence variations responsible for alterations in gene function) databases are becoming available, including the most informative markers in haplotypes could enhance the potential utility of haplotype-based studies [21, 162]. In contrast, Zhao et al. [74] found no apparent advantage of haplotype-based analysis over individual SNP analysis in their simulation study that was designed to resemble the demography and population history of livestock. Lorenz et al. [163] reached similar conclusion but they noted that their conclusion may not be valid under different models relating genotype to phenotype or under different demographic scenarios. Despite of these contradictory results, haplotype-

based analysis could play a critical role in association mapping studies in crop plants as recently discussed by Gupta et al. [164].

### Comparison with QTL for pigment from previous reports

In durum wheat, many QTL for yellow pigment content have been reported on different chromosomes [86, 89, 92, 96], of which 4B. The locus *hap\_4B\_6* on 4B explained 33.6% and 40.2% of the variation of pigment loss and pasta  $b^*$ , respectively. The locus on 5B (*hap\_5B\_25*) explained 14.1% of the variance of pasta  $b^*$ , congruent with the results of Roncallo et al. [96] who reported a QTL associated to flour yellowness on 5B, explaining 12.2% of the phenotypic variance. Other studies reported QTL associated to yellow pigment on 4B in durum [86] and hexaploid wheat [89]. The locus *hap\_7A\_32* detected on 7A in our study explained only 35.6% of the phenotypic variance of semolina pigment. Similarly, a major QTL for yellow pigment concentration has been reported on 7A in both bread wheat [89, 90, 94] and durum wheat [95, 97, 98], and shown to be associated with the phytoene synthase *Psy-A1* locus. Other studies reported a major QTL for flour yellowness on chromosome 7B [88, 89, 98], supporting the existence of a second gene affecting yellow pigment concentration in the distal region of chromosome arm 7B. However, the locus *hap\_7B\_36* detected on 7B explained only 8.9% of the variation of semolina pigment in our material.

Our observations of semolina colour and marked by *hap\_7A\_32* on chromosome 7A and *hap\_7B\_36* on 7B for semolina  $b^*$  were similar to those of Roncallo et al. [96] whom recently reported QTL for flour yellow colour on 7A and 7B. The evidence is strong for involvement of these two chromosomes in controlling endosperm pigment with numerous reports of major QTL for yellow pigment on 7A [89, 90, 94, 95, 97, 98] and 7B [88, 89, 98].

The *Lxp-B1* gene has been mapped on chromosome 4B [86, 99, 165] as well as the *Rht-B1b* conferring semidwarfism in durum [166]. Therefore, we evaluated how these loci relate to *Lpx-B1.1* and *Rht-B1b* genes. Two of the three loci we identified on 4B associated with pigment loss, and explaining 28.9 to 33.6% of the phenotypic variation, were associated with *Rht-B1b* and *Lpx-B1*. The locus *hap\_4B\_6* showed strong LD ( $r^2 = 0.86$ ) with *Lpx-B1.1* gene with the locus *hap\_4B\_7* was strongly associated ( $r^2 = 0.92$ ) with the semidwarf height locus *Rht-B1b*. Both *Lpx-B1.1* and *Rht-B1b* are known to reside on chromosome 4B [101, 167]. Because *Lpx-B1* and *Rht-B1b* are both on 4BS, there could be undesirable linkage. However, these two loci showed relatively weak ( $r^2 = 0.31$ ) LD, suggesting an independent segregation in our material. Pozniak et al. [101] reached a similar conclusion based on DArT marker assessment of this breeding panel.

Carotenoid degradation (pigment loss) during pasta processing is controlled by lipoxygenases, polyphenol oxidases and peroxidases. The wheat genes isoforms *Lpx-1* and *Lpx-3* are located on chromosome 4, whereas the *Lpx-2* gene is located on chromosome 5 [97, 119, 120, 168–171]. In developing durum kernels, different transcript levels have been reported, with *Lpx-1* transcripts being the most abundant in mature grain [169]. This suggests that the *Lpx-1* gene might have a major role in oxidation of carotenoid pigments during pasta processing. In support to this hypothesis, a major QTL for total lipoxygenase activity, with three copies of the *Lpx-1* gene (*Lpx-B1.1*, *Lpx-B1.2* and *Lpx-B1.3*) has been mapped on chromosome 4BS [97, 99, 120, 168, 172]. Selection for and fixing this allele in all breeding lines could contribute to significantly reduced pigment loss during pasta processing and, consequently, to improve the aesthetic and nutritional qualities of the pasta products.

For pasta  $a^*$ , the four loci detected on chromosomes 2A, 3B and 4B suggest complex genetic control of pasta redness in durum wheat. To our knowledge, this is the first study of association mapping for pasta  $a^*$ . Half of the total number of loci associated with pasta  $a^*$  were located

on chromosome 4B. In particular, locus *hap\_4B\_6* explained 35.7% of the phenotypic variance. This locus also showed strong association with pasta  $b^*$  and pigment loss. Pasta  $a^*$  (redness) and pasta  $b^*$  (yellowness) being correlated ( $r = 0.65$ ), much effort should be put on breaking the LD between them to facilitate selecting against red colour pasta.

## Conclusion

Our results clearly showed that genome-wide association studies could benefit from haplotype-based analysis. The haplotype approach substantially increased the polymorphism information content and detected more loci associated with semolina and pasta pigment. The amount of phenotypic variance explained and the allelic effect were also improved over single marker analysis. In particular, the locus *hap\_4B\_6* on chromosome 4B was associated with pasta  $a^*$ , pasta  $b^*$  and pigment loss; and explained up to 40% of the phenotypic variation. This locus could be a good candidate for tagging the *Lpx-B1* gene. On the other hand, combinations of several allelic series within each haplotype locus, as well as the aggregation of the best haplotypes improved ability to select lines having the desirable phenotypes. The use of haplotype-based analysis in comparison with single marker analysis will provide more insight about the potential of combining SNPs into haplotypes in genome-wide association studies.

## Supporting Information

**S1 Table. Lines pedigree and subpopulations they belong to, based on the discriminant analysis of principal components.**

(DOCX)

**S2 Table. Distribution of SNPs on the durum high-density SNP-based consensus map.**

(DOCX)

**S3 Table. Description of haplotypes associated with pigments colour traits.**

(DOCX)

**S1 Fig. Linkage disequilibrium (LD) scatterplot based on all pairwise comparisons between adjacent loci belonging to the same chromosome.**

(DOCX)

**S2 Fig. Quantile-quantile (Q-Q) plots comparing the distribution of observed versus expected  $P$ -values for association analyses of colour traits under different statistical models: GLM naïve (blue diamond), GLM\_Q (red square), MLM\_K (green triangle) and MLM\_QK (purple cross). The black dash line represents the null hypothesis of no true association.**

(DOCX)

## Acknowledgments

We gratefully acknowledge funding provided by Genome Canada to the CTAG project, as well as the by the Western Grains Research Foundation, the Province of Saskatchewan, and Agriculture and Agri-Food Canada. The technical assistance of J. Coulson, Lexie Martin, K. Wiebe and S. Yates is also gratefully acknowledged.

## Author Contributions

**Conceptualization:** AN CP FC JC.

**Data curation:** AN CP FC JC.

**Formal analysis:** AN CP FC JC.

**Funding acquisition:** CP.

**Investigation:** AN CP FC JC.

**Methodology:** AN CP.

**Project administration:** CP.

**Resources:** CP.

**Software:** AN.

**Supervision:** CP.

**Validation:** AN CP FC JC.

**Visualization:** AN CP FC JC RK JH.

**Writing – original draft:** AN.

**Writing – review & editing:** AN CP FC JC RK JH AC.

## References

1. Randhawa HS, Asif M, Pozniak C, Clarke JM, Graf RJ, Fox SL, et al. Application of molecular markers to wheat breeding in Canada. *Plant Breeding*. 2013; 132(5):458–71.
2. Li C, Bai G, Chao S, Wang Z. A High-Density SNP and SSR Consensus Map Reveals Segregation Distortion Regions in Wheat. *BioMed Research International*. 2015; 2015:10.
3. Maccaferri M, Ricci A, Salvi S, Milner SG, Noli E, Martelli PL, et al. A high-density, SNP-based consensus map of tetraploid wheat as a bridge to integrate durum and bread wheat genomics and breeding. *Plant Biotechnology Journal*. 2014; 13(5):648–63. doi: [10.1111/pbi.12288](https://doi.org/10.1111/pbi.12288) PMID: [25424506](https://pubmed.ncbi.nlm.nih.gov/25424506/)
4. Wang S, Wong D, Forrest K, Allen A, Chao S, Huang BE, et al. Characterization of polyploid wheat genomic diversity using a high-density 90 000 single nucleotide polymorphism array. *Plant Biotechnology Journal*. 2014; 12(6):787–96. doi: [10.1111/pbi.12183](https://doi.org/10.1111/pbi.12183) PMID: [24646323](https://pubmed.ncbi.nlm.nih.gov/24646323/)
5. Bernardo R. Molecular markers and selection for complex traits in plants: learning from the last 20 years. *Crop Sci*. 2008; 48:1649–64.
6. Holland JB. Genetic architecture of complex traits in plants. *Current Opinion in Plant Biology*. 2007; 10(2):156–61. doi: [10.1016/j.pbi.2007.01.003](https://doi.org/10.1016/j.pbi.2007.01.003) PMID: [17291822](https://pubmed.ncbi.nlm.nih.gov/17291822/)
7. Podlich DW, Winkler CR, Cooper M. Mapping As You Go: An Effective Approach for Marker-Assisted Selection of Complex Traits. *Crop Sci*. 2004; 44(5):1560–71.
8. Sneller CH, Mather DE, Crepieux S. Analytical Approaches and Population Types for Finding and Utilizing QTL in Complex Plant Populations. *Crop Sci*. 2009; 49(2):363–80.
9. Yu J, Buckler ES. Genetic association mapping and genome organization of maize. *Current Opinion in Biotechnology*. 2006; 17(2):155–60. doi: [10.1016/j.copbio.2006.02.003](https://doi.org/10.1016/j.copbio.2006.02.003) PMID: [16504497](https://pubmed.ncbi.nlm.nih.gov/16504497/)
10. Buckler ES, Thornsberry JM. Plant molecular diversity and applications to genomics. *Curr Opin Plant Biol*. 2002; 5:107–11. PMID: [11856604](https://pubmed.ncbi.nlm.nih.gov/11856604/)
11. Daly MJ, Rioux JD, Schaffner SF, Hudson TJ, Lander ES. High-resolution haplotype structure in the human genome. *Nature Genetics*. 2001; 29(2):229–32. doi: [10.1038/ng1001-229](https://doi.org/10.1038/ng1001-229) PMID: [11586305](https://pubmed.ncbi.nlm.nih.gov/11586305/)
12. Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, et al. The Structure of Haplotype Blocks in the Human Genome. *Science*. 2002; 296(5576):2225–9. doi: [10.1126/science.1069424](https://doi.org/10.1126/science.1069424) PMID: [12029063](https://pubmed.ncbi.nlm.nih.gov/12029063/)
13. Patil N, Berno AJ, Hinds DA, Barrett WA, Doshi JM, Hacker CR, et al. Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21. *Science*. 2001; 294(5547):1719–23. doi: [10.1126/science.1065573](https://doi.org/10.1126/science.1065573) PMID: [11721056](https://pubmed.ncbi.nlm.nih.gov/11721056/)
14. Li YC, Ding J, Abecasis GR. Mach 1.0: Rapid haplotype reconstruction and missing genotype inference. *American Journal of Human Genetics*. 2006; 79:S2290.

15. Tachmazidou I, Verzilli CJ, Iorio MD. Genetic Association Mapping via Evolution-Based Clustering of Haplotypes. *PLoS Genetics*. 2007; 3(7):e111. doi: [10.1371/journal.pgen.0030111](https://doi.org/10.1371/journal.pgen.0030111) PMID: [17616979](https://pubmed.ncbi.nlm.nih.gov/17616979/)
16. Inghelant VD, Melchinger A, Martinant J-P, Stich B. Genome-wide association mapping of flowering time and northern corn leaf blight (*Setosphaeria turcica*) resistance in a vast commercial maize germ-plasm set. *BMC Plant Biology*. 2012; 12(1):56.
17. Lipka AE, Gore MA, Magallanes-Lundback M, Mesberg A, Lin H, Tiede T, et al. Genome-Wide Association Study and Pathway-Level Analysis of Tocochromanol Levels in Maize Grain. *G3: Genes|Genomes|Genetics*. 2013; 3(8):1287–99. doi: [10.1534/g3.113.006148](https://doi.org/10.1534/g3.113.006148) PMID: [23733887](https://pubmed.ncbi.nlm.nih.gov/23733887/)
18. Weber AL, Zhao Q, McMullen MD, Doebley JF. Using Association Mapping in Teosinte to Investigate the Function of Maize Selection-Candidate Genes. *PLoS One*. 2009; 4(12):e8227. doi: [10.1371/journal.pone.0008227](https://doi.org/10.1371/journal.pone.0008227) PMID: [20011044](https://pubmed.ncbi.nlm.nih.gov/20011044/)
19. Lestari P, Lee G, Ham T-H, Reflinur, Woo M-O, Piao R, et al. Single Nucleotide Polymorphisms and Haplotype Diversity in Rice Sucrose Synthase 3. *Journal of Heredity*. 2011; 102(6):735–46. doi: [10.1093/jhered/esr094](https://doi.org/10.1093/jhered/esr094) PMID: [21914668](https://pubmed.ncbi.nlm.nih.gov/21914668/)
20. Shao G, Tang S, Chen M, Wei X, He J, Luo J, et al. Haplotype variation at *Badh2*, the gene determining fragrance in rice. *Genomics*. 2013; 101(2):157–62. doi: [10.1016/j.ygeno.2012.11.010](https://doi.org/10.1016/j.ygeno.2012.11.010) PMID: [23220350](https://pubmed.ncbi.nlm.nih.gov/23220350/)
21. Yonemaru J-i, Ebana K, Yano M. HapRice, an SNP Haplotype Database and a Web Tool for Rice. *Plant and Cell Physiology*. 2014; 55(1):e9. doi: [10.1093/pcp/pct188](https://doi.org/10.1093/pcp/pct188) PMID: [24334415](https://pubmed.ncbi.nlm.nih.gov/24334415/)
22. Yonemaru J-i, Yamamoto T, Ebana K, Yamamoto E, Nagasaki H, Shibaya T, et al. Genome-Wide Haplotype Changes Produced by Artificial Selection during Modern Rice Breeding in Japan. *PLoS One*. 2012; 7(3):e32982. doi: [10.1371/journal.pone.0032982](https://doi.org/10.1371/journal.pone.0032982) PMID: [22427922](https://pubmed.ncbi.nlm.nih.gov/22427922/)
23. Choi I-Y, Hyten DL, Matukumalli LK, Song Q, Chaky JM, Quigley CV, et al. A Soybean Transcript Map: Gene Distribution, Haplotype and Single-Nucleotide Polymorphism Analysis. *Genetics*. 2007; 176(1):685. doi: [10.1534/genetics.107.070821](https://doi.org/10.1534/genetics.107.070821) PMID: [17339218](https://pubmed.ncbi.nlm.nih.gov/17339218/)
24. Langewisch T, Zhang H, Vincent R, Joshi T, Xu D, Bilyeu K. Major Soybean Maturity Gene Haplotypes Revealed by SNPviz Analysis of 72 Sequenced Soybean Genomes. *PLoS One*. 2014; 9(4):e94150. doi: [10.1371/journal.pone.0094150](https://doi.org/10.1371/journal.pone.0094150) PMID: [24727730](https://pubmed.ncbi.nlm.nih.gov/24727730/)
25. Li Y-H, Zhang C, Gao Z-S, Smulders MJM, Ma Z, Liu Z-X, et al. Development of SNP markers and haplotype analysis of the candidate gene for *rhg1*, which confers resistance to soybean cyst nematode in soybean. *Molecular Breeding*. 2009; 24(1):63–76.
26. Patil G, Do T, Vuong TD, Valliyodan B, Lee J-D, Chaudhary J, et al. Genomic-assisted haplotype analysis and the development of high-throughput SNP markers for salinity tolerance in soybean. *Scientific Reports*. 2016; 6:19199. doi: [10.1038/srep19199](https://doi.org/10.1038/srep19199) PMID: [26781337](https://pubmed.ncbi.nlm.nih.gov/26781337/)
27. Haile JK, Hammer K, Badebo A, Singh RP, Roder MS. Haplotype analysis of molecular markers linked to stem rust resistance genes in Ethiopian improved durum wheat varieties and tetraploid wheat landraces. *Genetic Resources and Crop Evolution*. 2013; 60(3):853–64.
28. Hao C, Wang Y, Hou J, Feuillet C, Balfourier F, Zhang X. Association Mapping and Haplotype Analysis of a 3.1-Mb Genomic Region Involved in Fusarium Head Blight Resistance on Wheat Chromosome 3BS. *PLoS One*. 2012; 7(10):e46444. doi: [10.1371/journal.pone.0046444](https://doi.org/10.1371/journal.pone.0046444) PMID: [23071572](https://pubmed.ncbi.nlm.nih.gov/23071572/)
29. Sardouie-Nasab S, Mohammadi-Nejad G, Zebarjadi A. Haplotype analysis of QTLs attributed to salinity tolerance in wheat (*Triticum aestivum*). *Molecular Biology Reports*. 2013; 40(7):4661–71. doi: [10.1007/s11033-013-2561-z](https://doi.org/10.1007/s11033-013-2561-z) PMID: [23677711](https://pubmed.ncbi.nlm.nih.gov/23677711/)
30. Jordan K, Wang S, Lun Y, Gardiner L-J, MacLachlan R, Hucl P, et al. A haplotype map of allohexaploid wheat reveals distinct patterns of selection on homoeologous genomes. *Genome Biology*. 2015; 16(1):48.
31. Ma L, Li T, Hao C, Wang Y, Chen X, Zhang X. *TaGS5-3A*, a grain size gene selected during wheat improvement for larger kernel and yield. *Plant Biotechnology Journal*. 2016; 14(5):1269–80. doi: [10.1111/pbi.12492](https://doi.org/10.1111/pbi.12492) PMID: [26480952](https://pubmed.ncbi.nlm.nih.gov/26480952/)
32. Tsombalova J, Karafiatova M, Vrana J, Kubalaková M, Peusa H, Jakobson I, et al. A haplotype specific to North European wheat (*Triticum aestivum* L.). *Genetic Resources and Crop Evolution*. 2016:1–12.
33. Hou J, Jiang Q, Hao C, Wang Y, Zhang H, Zhang X. Global Selection on Sucrose Synthase Haplotypes during a Century of Wheat Breeding. *Plant Physiology*. 2014; 164(4):1918–29. doi: [10.1104/pp.113.232454](https://doi.org/10.1104/pp.113.232454) PMID: [24402050](https://pubmed.ncbi.nlm.nih.gov/24402050/)
34. Mago R, Tabe L, Vautrin S, Simkova H, Kubalaková M, Upadhyaya N, et al. Major haplotype divergence including multiple germin-like protein genes, at the wheat *Sr2* adult plant stem rust resistance locus. *BMC Plant Biology*. 2014; 14(1):379.

35. Prins R, Dreisigacker S, Pretorius Z, van Schalkwyk H, Wessels E, Smit C, et al. Stem Rust Resistance in a Geographically Diverse Collection of Spring Wheat Lines Collected from Across Africa. *Frontiers in Plant Science*. 2016; 7(973).
36. Bardel C, Danjean V, Hugot J-P, Darlu P, Genin E. On the use of haplotype phylogeny to detect disease susceptibility loci. *BMC Genetics*. 2005; 6(1):24.
37. Clark AG. The role of haplotypes in candidate gene studies. *Genetic Epidemiology*. 2004; 27(3):321–33.
38. Meuwissen TH, Goddard ME. Fine mapping of quantitative trait loci using linkage disequilibria with closely linked marker loci. *Genetics*. 2000; 155(1):421–30. PMID: [10790414](#)
39. Clark AG. The role of haplotypes in candidate gene studies. *Genetic Epidemiology*. 2004; 27(4):321–33. doi: [10.1002/gepi.20025](#) PMID: [15368617](#)
40. Zhao K, Aranzana MJ, Kim S, Lister C, Shindo C, Tang C, et al. An Arabidopsis Example of Association Mapping in Structured Samples. *PLoS Genetics*. 2007; 3(1):e4. doi: [10.1371/journal.pgen.0030004](#) PMID: [17238287](#)
41. Templeton AR, Boerwinkle E, Sing CF. A cladistic-analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping.I. basic theory and an analysis of alcohol-dehydrogenase activity in drosophila. *Genetics*. 1987; 117(2):343–51. PMID: [2822535](#)
42. Akey J, Jin L, Xiong M. Haplotypes vs single marker linkage disequilibrium tests: what do we gain? *European Journal of Human Genetics*. 2001; 9:291–300. doi: [10.1038/sj.ejhg.5200619](#) PMID: [11313774](#)
43. Hamblin MT, Jannink JL. Factors affecting the power of haplotype markers in association studies. *The Plant Genome*. 2011; 4:145–53.
44. Morris RW, Kaplan NL. On the advantage of haplotype analysis in the presence of multiple disease susceptibility alleles. *Genetic Epidemiology*. 2002; 23(3):221–33. doi: [10.1002/gepi.10200](#) PMID: [12384975](#)
45. Gawenda I, Thorwarth P, Günther T, Ordon F, Schmid KJ. Genome-wide association studies in elite varieties of German winter barley using single-marker and haplotype-based methods. *Plant Breeding*. 2015; 134(1):28–39.
46. Zhao H, Pfeiffer R, Gail MH. Haplotype analysis in population genetics and association studies. *Pharmacogenomics*. 2003; 4(2):171–8. doi: [10.1517/phgs.4.2.171.22636](#) PMID: [12605551](#)
47. Barton NH. Estimating multilocus linkage disequilibria. *Heredity (Edinb)*. 2000; 84 (Pt 3):373–89.
48. Service SK, Lang DW, Freimer NB, Sandkuil LA. Linkage-disequilibrium mapping of disease genes by reconstruction of ancestral haplotypes in founder populations. *American Journal of Human Genetics*. 1999; 64(6):1728–38. doi: [10.1086/302398](#) PMID: [10330361](#)
49. MacLean CJ, Martin RB, Sham PC, Wang H, Straub RE, Kendler KS. The trimmed-haplotype test for linkage disequilibrium. *American Journal of Human Genetics*. 2000; 66(3):1062–75. doi: [10.1086/302796](#) PMID: [10712218](#)
50. Calus M, Meuwissen T, Windig J, Knol E, Schrooten C, Vereijken A, et al. Effects of the number of markers per haplotype and clustering of haplotypes on the accuracy of QTL mapping and prediction of genomic breeding values. *Genet Sel Evol*. 2009; 41(1):11.
51. Grapes L, Dekkers JCM, Rothschild MF, Fernando RL. Comparing linkage disequilibrium-based methods for fine mapping quantitative trait loci. *Genetics*. 2004; 166(3):1561–70. PMID: [15082569](#)
52. Hayes BJ, Chamberlain AJ, McPartlan H, Macleod I, Sethuraman L, Goddard ME. Accuracy of marker-assisted selection with single markers and marker haplotypes in cattle. *Genetics Research*. 2007; 89(04):215–20.
53. Barendse W. Haplotype Analysis Improved Evidence for Candidate Genes for Intramuscular Fat Percentage from a Genome Wide Association Study of Cattle. *PLoS One*. 2010; 6(12):e29601.
54. Barrero R, Bellgard M, Zhang X. Diverse approaches to achieving grain yield in wheat. *Functional & Integrative Genomics*. 2011; 11(1):37–48.
55. Escamilla MA, McInnes LA, Spesny M, Reus VI, Service SK, Shimayoshi N, et al. Assessing the feasibility of linkage disequilibrium methods for mapping complex traits: an initial screen for bipolar disorder loci on chromosome 18. *American Journal of Human Genetics*. 1999; 64(6):1670–8. doi: [10.1086/302400](#) PMID: [10330354](#)
56. Hao D, Cheng H, Yin Z, Cui S, Zhang D, Wang H, et al. Identification of single nucleotide polymorphisms and haplotypes associated with yield and yield components in soybean (*Glycine max*) landraces across multiple environments. *Theoretical and Applied Genetics*. 2012; 124(3):447–58. doi: [10.1007/s00122-011-1719-0](#) PMID: [21997761](#)

57. Lorenz AJ, Hamblin MT, Jannink J-L. Performance of Single Nucleotide Polymorphisms versus Haplotypes for Genome-Wide Association Analysis in Barley. *PLoS One*. 2010; 5(11):e14079. doi: [10.1371/journal.pone.0014079](https://doi.org/10.1371/journal.pone.0014079) PMID: [21124933](https://pubmed.ncbi.nlm.nih.gov/21124933/)
58. Lu Y, Xu J, Yuan Z, Hao Z, Xie C, Li X, et al. Comparative LD mapping using single SNPs and haplotypes identifies QTL for plant height and biomass as secondary traits of drought tolerance in maize. *Molecular Breeding*. 2012; 30(1):407–18.
59. Martin ER, Lai EH, Gilbert JR, Rogala AR, Afshari AJ, Riley J, et al. SNPing Away at Complex Diseases: Analysis of Single-Nucleotide Polymorphisms around APOE in Alzheimer Disease. *The American Journal of Human Genetics*. 2000; 67(2):383–94. doi: [10.1086/303003](https://doi.org/10.1086/303003) PMID: [10869235](https://pubmed.ncbi.nlm.nih.gov/10869235/)
60. Van Inghelandt D, Melchinger A, Martinant J-P, Stich B. Genome-wide association mapping of flowering time and northern corn leaf blight (*Setosphaeria turcica*) resistance in a vast commercial maize germplasm set. *BMC Plant Biology*. 2012; 12(1):56.
61. Cuyabano BCD, Su G, Lund MS. Genomic prediction of genetic merit using LD-based haplotypes in the Nordic Holstein population. *BMC Genomics*. 2014; 15(1):1171.
62. Jonas D, Ducrocq V, Fouilloux M-N, Croiseau P. Alternative haplotype construction methods for genomic evaluation. *Journal of Dairy Science*. 2016; 99(6):4537–46. doi: [10.3168/jds.2015-10433](https://doi.org/10.3168/jds.2015-10433) PMID: [26995132](https://pubmed.ncbi.nlm.nih.gov/26995132/)
63. Ferdosi MH, Henshall J, Tier B. Study of the optimum haplotype length to build genomic relationship matrices. *Genet Sel Evol*. 2016; 48(1):75. doi: [10.1186/s12711-016-0253-6](https://doi.org/10.1186/s12711-016-0253-6) PMID: [27687320](https://pubmed.ncbi.nlm.nih.gov/27687320/)
64. Clark AG, Weiss KM, Nickerson DA, Taylor SL, Buchanan A, Stengard J, et al. Haplotype structure and population genetic inferences from nucleotide-sequence variation in human lipoprotein lipase. *American Journal of Human Genetics*. 1998; 63(2):595–612. doi: [10.1086/301977](https://doi.org/10.1086/301977) PMID: [9683608](https://pubmed.ncbi.nlm.nih.gov/9683608/)
65. Terwilliger JD, Weiss KM. Linkage disequilibrium mapping of complex disease: fantasy or reality? *Current Opinion in Biotechnology*. 1998; 9(6):578–94. PMID: [9889136](https://pubmed.ncbi.nlm.nih.gov/9889136/)
66. Zhao HH, Fernando RL, Dekkers JCM. Power and Precision of Alternate Methods for Linkage Disequilibrium Mapping of Quantitative Trait Loci. *Genetics*. 2007; 175(4):1975–86. doi: [10.1534/genetics.106.066480](https://doi.org/10.1534/genetics.106.066480) PMID: [17277369](https://pubmed.ncbi.nlm.nih.gov/17277369/)
67. Long AD, Langley CH. The power of association studies to detect the contribution of candidate genetic loci to variation in complex traits. *Genome Res*. 1999; 9:720–31. PMID: [10447507](https://pubmed.ncbi.nlm.nih.gov/10447507/)
68. Hoffman GE. Correcting for Population Structure and Kinship Using the Linear Mixed Model: Theory and Extensions. *PLoS One*. 2013; 8(10):e75707. doi: [10.1371/journal.pone.0075707](https://doi.org/10.1371/journal.pone.0075707) PMID: [24204578](https://pubmed.ncbi.nlm.nih.gov/24204578/)
69. Muller BU, Stich B, Piepho HP. A general method for controlling the genome-wide type I error rate in linkage and association mapping experiments in plants. *Heredity*. 2011; 106(5):825–31. doi: [10.1038/hdy.2010.125](https://doi.org/10.1038/hdy.2010.125) PMID: [20959861](https://pubmed.ncbi.nlm.nih.gov/20959861/)
70. Stich B, Mohring J, Piepho H-P, Heckenberger M, Buckler ES, Melchinger AE. Comparison of Mixed-Model Approaches for Association Mapping. *Genetics*. 2008; 178(3):1745–54. doi: [10.1534/genetics.107.079707](https://doi.org/10.1534/genetics.107.079707) PMID: [18245847](https://pubmed.ncbi.nlm.nih.gov/18245847/)
71. Templeton AR, Maxwell T, Posada D, Stengard JH, Boerwinkle E, Sing CF. Tree Scanning: A Method for Using Haplotype Trees in Phenotype/Genotype Association Studies. *Genetics*. 2005; 169(1):441–53. doi: [10.1534/genetics.104.030080](https://doi.org/10.1534/genetics.104.030080) PMID: [15371364](https://pubmed.ncbi.nlm.nih.gov/15371364/)
72. Zollner S, Pritchard JK. Coalescent-Based Association Mapping and Fine Mapping of Complex Trait Loci. *Genetics*. 2005; 169(2):1071–92. doi: [10.1534/genetics.104.031799](https://doi.org/10.1534/genetics.104.031799) PMID: [15489534](https://pubmed.ncbi.nlm.nih.gov/15489534/)
73. Guo Y, Li J, Bonham AJ, Wang Y, Deng H. Gains in power for exhaustive analyses of haplotypes using variable-sized sliding window strategy: a comparison of association-mapping strategies. *European Journal of Human Genetics*. 2009; 17:785–92. doi: [10.1038/ejhg.2008.244](https://doi.org/10.1038/ejhg.2008.244) PMID: [19092774](https://pubmed.ncbi.nlm.nih.gov/19092774/)
74. Zhao LP, Li SS, Shen F. A haplotype-linkage analysis method for estimating recombination rates using dense SNP trio data. *Genetic Epidemiology*. 2007; 31(2):154–72. doi: [10.1002/gepi.20198](https://doi.org/10.1002/gepi.20198) PMID: [17219374](https://pubmed.ncbi.nlm.nih.gov/17219374/)
75. Lu Y, Shah T, Hao Z, Taba S, Zhang S, Gao S, et al. Comparative SNP and Haplotype Analysis Reveals a Higher Genetic Diversity and Rapider LD Decay in Tropical than Temperate Germplasm in Maize. *PLoS One*. 2011; 6(9):e24861. doi: [10.1371/journal.pone.0024861](https://doi.org/10.1371/journal.pone.0024861) PMID: [21949770](https://pubmed.ncbi.nlm.nih.gov/21949770/)
76. Yan J, Shah T, Warburton ML, Buckler ES, McMullen MD, Crouch J. Genetic characterization and linkage disequilibrium estimation of a global maize collection using SNP markers. *PLoS One*. 2009; 4(12):e8451. doi: [10.1371/journal.pone.0008451](https://doi.org/10.1371/journal.pone.0008451) PMID: [20041112](https://pubmed.ncbi.nlm.nih.gov/20041112/)
77. Diaz A, Fergany M, Formisano G, Ziarsolo P, Blanca J, Fei Z, et al. A consensus linkage map for molecular markers and Quantitative Trait Loci associated with economically important traits in melon (*Cucumis melo* L.). *BMC Plant Biology*. 2011; 11(1):111.

78. Durrant C, Zondervan KT, Cardon LR, Hunt S, Deloukas P, Morris AP. Linkage Disequilibrium Mapping via Cladistic Analysis of Single-Nucleotide Polymorphism Haplotypes. *American Journal of Human Genetics*. 2004; 75(1):35–43. doi: [10.1086/422174](https://doi.org/10.1086/422174) PMID: [15148658](https://pubmed.ncbi.nlm.nih.gov/15148658/)
79. Mathias RA, Gao P, Goldstein JL, Wilson AF, Pugh EW, Furbert-Harris P, et al. A graphical assessment of p-values from sliding window haplotype tests of association to identify asthma susceptibility loci on chromosome 11q. *BMC Genetics*. 2006; 7:38-. doi: [10.1186/1471-2156-7-38](https://doi.org/10.1186/1471-2156-7-38) PMID: [16774684](https://pubmed.ncbi.nlm.nih.gov/16774684/)
80. Pan Y, Chen J, Guo H, Ou J, Peng Y, Liu Q, et al. Association of genetic variants of GRIN2B with autism. *Sci Rep*. 2014; 5.
81. Arbelbide M, Bernardo R. Mixed-model QTL mapping for kernel hardness and dough strength in bread wheat. *Theoretical and Applied Genetics*. 2006; 112(5):885–90. doi: [10.1007/s00122-005-0190-1](https://doi.org/10.1007/s00122-005-0190-1) PMID: [16402188](https://pubmed.ncbi.nlm.nih.gov/16402188/)
82. Flint-Garcia SA. Maize association population: a high resolution platform for QTL dissection. *Plant J*. 2005; 44:1054–64. doi: [10.1111/j.1365-313X.2005.02591.x](https://doi.org/10.1111/j.1365-313X.2005.02591.x) PMID: [16359397](https://pubmed.ncbi.nlm.nih.gov/16359397/)
83. Somers DJ, Banks T, DePauw R, Fox S, Clarke J, Pozniak C, et al. Genome-wide linkage disequilibrium analysis in bread wheat and durum wheat. *Genome*. 2007; 50(6):557–67. doi: [10.1139/g07-031](https://doi.org/10.1139/g07-031) PMID: [17632577](https://pubmed.ncbi.nlm.nih.gov/17632577/)
84. Tuberosa R, Pozniak C. Durum wheat genomics comes of age. *Molecular Breeding*. 2014; 34(4):1527–30.
85. Ficco DBM, Mastrangelo AM, Trono D, Borrelli GM, De Vita P, Fares C, et al. The colours of durum wheat: a review. *Crop and Pasture Science*. 2014; 65(1):1–15. <http://dx.doi.org/10.1071/CP13293>.
86. Pozniak CJ, Knox RE, Clarke FR, Clarke JM. Identification of QTL and association of a phytoene synthase gene with endosperm colour in durum wheat. *Theoretical and Applied Genetics*. 2007; 114(3):525–37. doi: [10.1007/s00122-006-0453-5](https://doi.org/10.1007/s00122-006-0453-5) PMID: [17131106](https://pubmed.ncbi.nlm.nih.gov/17131106/)
87. Reimer S, Pozniak CJ, Clarke FR, Clarke JM, Somers DJ, Knox RE, et al. Association mapping of yellow pigment in an elite collection of durum wheat cultivars and breeding lines. 2008; 51(12):1016–25.
88. Kuchel H, Langridge P, Mosionek L, Williams K, Jefferies SP. The genetic control of milling yield, dough rheology and baking quality of wheat. *Theoretical and Applied Genetics*. 2006; 112(8):1487–95. doi: [10.1007/s00122-006-0252-z](https://doi.org/10.1007/s00122-006-0252-z) PMID: [16550398](https://pubmed.ncbi.nlm.nih.gov/16550398/)
89. Mares DJ, Campbell AW. Mapping components of flour and noodle colour in Australian wheat. *Australian Journal of Agricultural Research*. 2001; 52(12):1297–309. <http://dx.doi.org/10.1071/AR01048>.
90. Parker GD, Chalmers KJ, Rathjen AJ, Langridge P. Mapping loci associated with flour colour in wheat (*Triticum aestivum* L.). *Theoretical and Applied Genetics*. 1998; 97(1–2):238–45.
91. Zhang Y, Wu Y, Xiao Y, He Z, Zhang Y, Yan J, et al. QTL mapping for flour and noodle colour components and yellow pigment content in common wheat. *Euphytica*. 2009; 165(3):435–44.
92. Blanco A, Colasuonno P, Gadaleta A, Mangini G, Schiavulli A, Simeone R, et al. Quantitative trait loci for yellow pigment concentration and individual carotenoid compounds in durum wheat. *Journal of Cereal Science*. 2011; 54(2):255–64.
93. Elouafi I, Nachit MM, Martin LM. Identification of a microsatellite on chromosome 7B showing a strong linkage with yellow pigment in durum wheat (*Triticum turgidum* L. var. *durum*). *Hereditas*. 2001; 135(2–3):255–61. PMID: [12152344](https://pubmed.ncbi.nlm.nih.gov/12152344/)
94. Howitt C, Cavanagh C, Bowerman A, Cazzonelli C, Rampling L, Mimica J, et al. Alternative splicing, activation of cryptic exons and amino acid substitutions in carotenoid biosynthetic genes are associated with lutein accumulation in wheat endosperm. *Functional & Integrative Genomics*. 2009; 9(3):363–76.
95. Patil R, Oak M, Tamhankar S, Sourdille P, Rao V. Mapping and validation of a major QTL for yellow pigment content on 7AL in durum wheat (*Triticum turgidum* L. ssp. *durum*). *Molecular Breeding*. 2008; 21(4):485–96.
96. Roncallo P, Cervigni G, Jensen C, Miranda Rn, Carrera A, Helguera M, et al. QTL analysis of main and epistatic effects for flour color traits in durum wheat. *Euphytica*. 2012; 185(1):77–92.
97. Zhang W, Chao S, Manthey F, Chicaiza O, Brevis JC, Echenique V, et al. QTL analysis of pasta quality using a composite microsatellite and SNP map of durum wheat. *Theoretical and Applied Genetics*. 2008; 117(8):1361–77. Epub 2008/09/11. doi: [10.1007/s00122-008-0869-1](https://doi.org/10.1007/s00122-008-0869-1) PMID: [18781292](https://pubmed.ncbi.nlm.nih.gov/18781292/)
98. Zhang W, Dubcovsky J. Association between allelic variation at the Phytoene synthase 1 gene and yellow pigment content in the wheat grain. *Theoretical and Applied Genetics*. 2008; 116(5):635–45. doi: [10.1007/s00122-007-0697-8](https://doi.org/10.1007/s00122-007-0697-8) PMID: [18193186](https://pubmed.ncbi.nlm.nih.gov/18193186/)
99. Hessler TG, Thomson MJ, Benscher D, Nachit MM, Sorrells ME. Association of a Lipoxxygenase Locus, Lpx-B1, with Variation in Lipoxxygenase Activity in Durum Wheat Seeds. *Crop Sci*. 2002; 42(5):1695–700.

100. Clarke FR, Clarke JM, Ames NA, Knox RE, Ross RJ. Gluten index compared with SDS-sedimentation volume for early generation selection for gluten strength in durum wheat. *Canadian Journal of Plant Science*. 2010; 90(1):1–11.
101. Pozniak C, Clarke J, Clarke F. Potential for detection of marker—trait associations in durum wheat using unbalanced, historical phenotypic datasets. *Molecular Breeding*. 2012; 1–14.
102. Kollers S, Rodemann B, Ling J, Korzun V, Ebmeyer E, Argillier O, et al. Whole Genome Association Mapping of Fusarium Head Blight Resistance in European Winter Wheat (*Triticum aestivum* L.). *PLoS One*. 2013; 8(2):e57500. doi: [10.1371/journal.pone.0057500](https://doi.org/10.1371/journal.pone.0057500) PMID: [23451238](https://pubmed.ncbi.nlm.nih.gov/23451238/)
103. Zanke CD, Ling J, Plieske J, Kollers S, Ebmeyer E, Korzun V, et al. Whole Genome Association Mapping of Plant Height in Winter Wheat (*Triticum aestivum* L.). *PLoS One*. 2014; 9(11):e113287. doi: [10.1371/journal.pone.0113287](https://doi.org/10.1371/journal.pone.0113287) PMID: [25405621](https://pubmed.ncbi.nlm.nih.gov/25405621/)
104. Beattie AD, Edney MJ, Scoles GJ, Rossnagel BG. Association Mapping of Malting Quality Data from Western Canadian Two-row Barley Cooperative Trials. *Crop Science*. 2010; 50(5):1649–63.
105. Kraakman ATW, Niks RE, Van den Berg P, Stam P, Van Eeuwijk FA. Linkage Disequilibrium Mapping of Yield and Yield Stability in Modern Spring Barley Cultivars. *Genetics*. 2004; 168(1):435–46. doi: [10.1534/genetics.104.026831](https://doi.org/10.1534/genetics.104.026831) PMID: [15454555](https://pubmed.ncbi.nlm.nih.gov/15454555/)
106. Matthies IE, Malosetti M, Roder MS, van Eeuwijk F. Genome-Wide Association Mapping for Kernel and Malting Quality Traits Using Historical European Barley Records. *PLoS One*. 2014; 9(11):e110046. doi: [10.1371/journal.pone.0110046](https://doi.org/10.1371/journal.pone.0110046) PMID: [25372869](https://pubmed.ncbi.nlm.nih.gov/25372869/)
107. Saade S, Maurer A, Shahid M, Oakey H, Schmockel SM, Negrao S, et al. Yield-related salinity tolerance traits identified in a nested association mapping (NAM) population of wild barley. *Scientific Reports*. 2016; 6:32586. doi: [10.1038/srep32586](https://doi.org/10.1038/srep32586) PMID: [27585856](https://pubmed.ncbi.nlm.nih.gov/27585856/)
108. Malosetti M, van der Linden CG, Vosman B, van Eeuwijk FA. A mixed-model approach to association mapping using pedigree information with an illustration of resistance to *Phytophthora infestans* in potato. *Genetics*. 2007; 175:879–89. doi: [10.1534/genetics.105.054932](https://doi.org/10.1534/genetics.105.054932) PMID: [17151263](https://pubmed.ncbi.nlm.nih.gov/17151263/)
109. Racedo J, Gutierrez L, Perera MF, Ostengo S, Pardo EM, Cuenya MI, et al. Genome-wide association mapping of quantitative traits in a breeding population of sugarcane. *BMC Plant Biology*. 2016; 16:142. doi: [10.1186/s12870-016-0829-x](https://doi.org/10.1186/s12870-016-0829-x) PMID: [27342657](https://pubmed.ncbi.nlm.nih.gov/27342657/)
110. Clarke JM, Clarke FR, Pozniak CJ. Forty-six years of genetic improvement in Canadian durum wheat cultivars. *Canadian Journal of Plant Science*. 2010; 90(6):791–801.
111. Clarke JM, McLeod JG, McCaig TN, DePauw RM, Knox RE, Fernandez MR. AC Avonlea durum wheat. *Can J Plant Sci*. 1998; 78:621–3.
112. Clarke JM, McLeod JG, DePauw RM, Marchylo BA, McCaig TN, Knox RE, et al. AC Navigator durum wheat. *Can J Plant Sci*. 2000; 80:343–5.
113. Clarke JM, McCaig TN, DePauw RM, Knox RE, Clarke F, Fernandez MR, et al. Strongfield durum wheat. *Can J Plant Sci* 85:651–654. 2005.
114. Clarke JM, McCaig TN, DePauw RM, Knox RE, Clarke FR, Fernandez MR, et al. Commander durum wheat. *Can J Plant Sci*. 2005; 85:901–4.
115. Clarke JM, Knox RE, DePauw RM, Clarke FR, Fernandez MR, McCaig TN, et al. Brigade Durum wheat. *Canadian Journal of Plant Science*. 2009; 89:505–9.
116. Littell RC, Milliken GA, Stroup WW, Wolfinger RD. SAS<sup>®</sup> system for mixed models. SAS Institute, Cary, p 633. 1996.
117. Hoisington D, Khairallah M, Gonzalez-de-Leon D. Laboratory Protocols: CIMMYT Applied Molecular Genetics Laboratory. Mexico, DF. 1994.
118. Liu K, Muse SV. PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics*. 2005; 21(9):2128–9. doi: [10.1093/bioinformatics/bti282](https://doi.org/10.1093/bioinformatics/bti282) PMID: [15705655](https://pubmed.ncbi.nlm.nih.gov/15705655/)
119. Garbus I, Carrera AD, Dubcovsky J, Echenique V. Physical mapping of durum wheat lipoxigenase genes. *Journal of Cereal Science*. 2009; 50(1):67–73.
120. Verlotta A, De Simone V, Mastrangelo A, Cattivelli L, Papa R, Trono D. Insight into durum wheat Lpx-B1: a small gene family coding for the lipoxigenase responsible for carotenoid bleaching in mature grains. *BMC Plant Biology*. 2010; 10(1):263.
121. Ellis M, Spielmeyer W, Gale K, Rebetzke G, Richards R. "Perfect" markers for the Rht-B1b and Rht-D1b dwarfing genes in wheat. *Theoretical and Applied Genetics*. 2002; 105(6–7):1038–42. doi: [10.1007/s00122-002-1048-4](https://doi.org/10.1007/s00122-002-1048-4) PMID: [12582931](https://pubmed.ncbi.nlm.nih.gov/12582931/)
122. Gaunt T, Rodriguez S, Zapata C, Day I. MIDAS: software for analysis and visualisation of interallelic disequilibrium between multiallelic markers. *BMC Bioinformatics*. 2006; 7(1):227.
123. Jombart T, Ahmed I. adegenet 1.3–1: new tools for the analysis of genome-wide SNP data. *Bioinformatics*. 2011; 27(21):3070–1. doi: [10.1093/bioinformatics/btr521](https://doi.org/10.1093/bioinformatics/btr521) PMID: [21926124](https://pubmed.ncbi.nlm.nih.gov/21926124/)

124. Weir BS. Genetic data analysis II: Methods for discrete population genetic data. Sinauer Associates Inc, Sunderland, Mass. 1996.
125. Breseghello F, Sorrells ME. Association Mapping of Kernel Size and Milling Quality in Wheat (*Triticum aestivum* L.). Cultivars. *Genetics*. 2006; 172(2):1165–77. doi: [10.1534/genetics.105.044586](https://doi.org/10.1534/genetics.105.044586) PMID: [16079235](https://pubmed.ncbi.nlm.nih.gov/16079235/)
126. Breiman L. Random Forests. *Machine Learning*. 2001; 45:5–32.
127. Liaw A, Wiener M. Classification and regression by random-Forest. *R News* 2. 2002; 2:18–22.
128. R Development Core Team. R: A Language and Environment for Statistical Computing. Austria. R Foundation for Statistical Computing, Vienna. 2013.
129. Rutkoski JE, Poland J, Jannink J-L, Sorrells ME. Imputation of unordered markers and the impact on genomic selection accuracy. *G3 (Bethesda)*. 2013; 3:427–39.
130. Fu YB. Genetic diversity analysis of highly incomplete SNP genotype data with imputations: an empirical assessment. *G3 (Bethesda)*. 2014; 4(5):891–900.
131. Botta V, Louppe G, Geurts P, Wehenkel L. Exploiting SNP Correlations within Random Forest for Genome-Wide Association Studies. *PLoS One*. 2014; 9(4):e93379. doi: [10.1371/journal.pone.0093379](https://doi.org/10.1371/journal.pone.0093379) PMID: [24695491](https://pubmed.ncbi.nlm.nih.gov/24695491/)
132. Minozzi G, Pedretti A, Biffani S, Nicolazzi EL, Stella A. Genome wide association analysis of the 16th QTL- MAS Workshop dataset using the Random Forest machine learning approach. *BMC Proceedings*. 2014; 8(Suppl 5):S4–S.
133. Wang Y, Goh W, Wong L, Montana G, the Alzheimer's Disease Neuroimaging I. Random forests on Hadoop for genome-wide association studies of multivariate neuroimaging phenotypes. *BMC Bioinformatics*. 2013; 14(Suppl 16):S6–S.
134. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics*. 2006; 38(8):904–9. [http://www.nature.com/ng/journal/v38/n8/suppinfo/ng1847\\_S1.html](http://www.nature.com/ng/journal/v38/n8/suppinfo/ng1847_S1.html). doi: [10.1038/ng1847](https://doi.org/10.1038/ng1847) PMID: [16862161](https://pubmed.ncbi.nlm.nih.gov/16862161/)
135. Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES. TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics*. 2007; 23:2633–5. doi: [10.1093/bioinformatics/btm308](https://doi.org/10.1093/bioinformatics/btm308) PMID: [17586829](https://pubmed.ncbi.nlm.nih.gov/17586829/)
136. Storey JD, Taylor JE, Siegmund D. Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach. *Journal of the Royal Statistical Society, B*. 2004; 66(1):187–205.
137. Blanchard G, Dickhaus T, Hack N, Konietschke F, Rohmeyer K, Rosenblatt J, et al. MuToss Multiple hypothesis testing in an open software system. *Journal of Machine Learning Research: Workshop and Conference Proceedings*. 2010; 11:12–9.
138. Townley-Smith TF, DePauw RM, Lendrum CW, McCrystal GE, Patterson LA. Kyle durum wheat. *Can J Plant Sci*. 1987; 67:225–7.
139. Jombart T, Devillard S, Balloux F. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics*. 2010; 11:94. doi: [10.1186/1471-2156-11-94](https://doi.org/10.1186/1471-2156-11-94) PMID: [20950446](https://pubmed.ncbi.nlm.nih.gov/20950446/)
140. Roullier C, Duputie A, Wennekes P, Benoit L, Fernandez Bringas VM, Rossel G, et al. Disentangling the Origins of Cultivated Sweet Potato (*Ipomoea batatas* (L.) Lam.). *PLoS One*. 2013; 8(5):e62707. doi: [10.1371/journal.pone.0062707](https://doi.org/10.1371/journal.pone.0062707) PMID: [23723970](https://pubmed.ncbi.nlm.nih.gov/23723970/)
141. Courtois B, Audebert A, Dardou A, Roques S, Ghneim-Herrera T, Droc G, et al. Genome-Wide Association Mapping of Root Traits in a Japonica Rice Panel. *PLoS One*. 2013; 8(11):e78037. doi: [10.1371/journal.pone.0078037](https://doi.org/10.1371/journal.pone.0078037) PMID: [24223758](https://pubmed.ncbi.nlm.nih.gov/24223758/)
142. Pometti CL, Bessega CF, Saidman BO, Vilardi JC. Analysis of genetic population structure in *Acacia caven* (Leguminosae, Mimosoideae), comparing one exploratory and two Bayesian-model-based methods. *Genetics and Molecular Biology*. 2014; 37:64–72. PMID: [24688293](https://pubmed.ncbi.nlm.nih.gov/24688293/)
143. Campoy JA, Lerigoleur-Balsemin E, Christmann H, Beauvieux R, Girollet N, Quero-Garcia J, et al. Genetic diversity, linkage disequilibrium, population structure and construction of a core collection of *Prunus avium* L. landraces and bred cultivars. *BMC Plant Biology*. 2016; 16:49. doi: [10.1186/s12870-016-0712-9](https://doi.org/10.1186/s12870-016-0712-9) PMID: [26912051](https://pubmed.ncbi.nlm.nih.gov/26912051/)
144. Cappa EP, El-Kassaby YA, Garcia MN, Acuna C, Borralho NMG, Grattapaglia D, et al. Impacts of Population Structure and Analytical Models in Genome-Wide Association Studies of Complex Traits in Forest Trees: A Case Study in *Eucalyptus globulus*. *PLoS One*. 2013; 8(11):e81267. doi: [10.1371/journal.pone.0081267](https://doi.org/10.1371/journal.pone.0081267) PMID: [24282578](https://pubmed.ncbi.nlm.nih.gov/24282578/)

145. Kang HM, Zaitlen NA, Wade CM, Kirby A, Heckerman D, Daly MJ, et al. Efficient Control of Population Structure in Model Organism Association Mapping. *Genetics*. 2008; 178(3):1709–23. doi: [10.1534/genetics.107.080101](https://doi.org/10.1534/genetics.107.080101) PMID: [18385116](https://pubmed.ncbi.nlm.nih.gov/18385116/)
146. Marchini J, Cardon LR, Phillips MS, Donnelly P. The effects of human population structure on large genetic association studies. *Nature Genetics*. 2004; 36:512–7. doi: [10.1038/ng1337](https://doi.org/10.1038/ng1337) PMID: [15052271](https://pubmed.ncbi.nlm.nih.gov/15052271/)
147. Mezouk S, Dubreuil P, Bosio M, Decousset L, Charcosset A, Praud S, et al. Effect of population structure corrections on the results of association mapping tests in complex maize diversity panels. *Theoretical and Applied Genetics*. 2011; 122(6):1149–60. doi: [10.1007/s00122-010-1519-y](https://doi.org/10.1007/s00122-010-1519-y) PMID: [21221527](https://pubmed.ncbi.nlm.nih.gov/21221527/)
148. Myles S, Peiffer J, Brown PJ, Ersoz ES, Zhang ZW, Costich DE, et al. Association Mapping: Critical Considerations Shift from Genotyping to Experimental Design. *Plant Cell*. 2009; 21(8):2194–202. doi: [10.1105/tpc.109.068437](https://doi.org/10.1105/tpc.109.068437) PMID: [19654263](https://pubmed.ncbi.nlm.nih.gov/19654263/)
149. Cavanagh CR, Chao S, Wang S, Huang BE, Stephen S, Kiani S, et al. Genome-wide comparative diversity uncovers multiple targets of selection for improvement in hexaploid wheat landraces and cultivars. *Proceedings of the National Academy of Sciences*. 2013; 110(20):8057–62.
150. Maccaferri M, Sanguineti MC, Noli E, Tuberosa R. Population structure and long-range linkage disequilibrium in a durum wheat elite collection. *Molecular Breeding*. 2005; 15(3):271–90.
151. Schaid DJ. Evaluating associations of haplotypes with traits. *Genetic Epidemiology*. 2004; 27(4):348–64. doi: [10.1002/gepi.20037](https://doi.org/10.1002/gepi.20037) PMID: [15543638](https://pubmed.ncbi.nlm.nih.gov/15543638/)
152. Yang H-C, Lin C-Y, Fann CSJ. A sliding-window weighted linkage disequilibrium test. *Genetic Epidemiology*. 2006; 30(6):531–45. doi: [10.1002/gepi.20165](https://doi.org/10.1002/gepi.20165) PMID: [16830340](https://pubmed.ncbi.nlm.nih.gov/16830340/)
153. Browning SR. Multilocus association mapping using variable-length Markov chains. *American Journal of Human Genetics*. 2006; 78:903–13. doi: [10.1086/503876](https://doi.org/10.1086/503876) PMID: [16685642](https://pubmed.ncbi.nlm.nih.gov/16685642/)
154. Gao Q, Yuan Z, He Y, Zhang JZX, Li F, Zhang B, et al. Exhaustive Sliding-Window Scan Strategy for Genome-Wide Association Study via Pca-Based Logistic Model. *Global Journal of Science Frontier Research Bio-Tech & Genetics*. 2012; 12(4):1–6.
155. Li Y, Sung WK, Liu JJ. Association mapping via regularized regression analysis of single-nucleotide-polymorphism haplotypes in variable-sized sliding windows. *Am J Hum Genet*. 2007; 80:705–15. doi: [10.1086/513205](https://doi.org/10.1086/513205) PMID: [17357076](https://pubmed.ncbi.nlm.nih.gov/17357076/)
156. Tang R, Feng T, Sha Q, Zhang S. A variable-sized sliding-window approach for genetic association studies via principal component analysis. *Ann Hum Genet*. 2009; 73:631–7. doi: [10.1111/j.1469-1809.2009.00543.x](https://doi.org/10.1111/j.1469-1809.2009.00543.x) PMID: [19735491](https://pubmed.ncbi.nlm.nih.gov/19735491/)
157. Yu Z, Schaid DJ. Sequential haplotype scan methods for association analysis. *Genetic Epidemiology*. 2007; 31(6):553–64. doi: [10.1002/gepi.20228](https://doi.org/10.1002/gepi.20228) PMID: [17487883](https://pubmed.ncbi.nlm.nih.gov/17487883/)
158. Sha Q, Tang R, Zhang S. Detecting susceptibility genes for rheumatoid arthritis based on a novel sliding-window approach. *BMC Proceedings*. 2009; 3(Suppl 7):S14–S.
159. Chen Y, Li X, Li J. A novel approach for haplotype-based association analysis using family data. *BMC Bioinformatics*. 2010; 11(1):S45.
160. Grapes L, Firat MZ, Dekkers JCM, Rothschild MF, Fernando RL. Optimal haplotype structure for linkage disequilibrium-based fine mapping of quantitative trait loci using identity by descent. *Genetics*. 2006; 172:1955–65. doi: [10.1534/genetics.105.048686](https://doi.org/10.1534/genetics.105.048686) PMID: [16322505](https://pubmed.ncbi.nlm.nih.gov/16322505/)
161. Liu N, Zhang K, Zhao H. Haplotype-Association Analysis. *Advances in genetics*. Volume 60: Academic Press; 2008. p. 335–405.
162. Kumagai M, Kim J, Itoh R, Itoh T. Tasuke: a web-based visualization program for large-scale resequencing data. *Bioinformatics*. 2013; 29(14):1806–8. doi: [10.1093/bioinformatics/btt295](https://doi.org/10.1093/bioinformatics/btt295) PMID: [23749962](https://pubmed.ncbi.nlm.nih.gov/23749962/)
163. Lorenz AJ, Hamblin MT, Jannink JL. Performance of single nucleotide polymorphisms versus haplotypes for genome-wide association analysis in barley. *PLoS One*. 2011; 5(11):e14079.
164. Gupta PK, Kulwal PL, Jaiswal V. Chapter Two—Association Mapping in Crop Plants: Opportunities and Challenges. *Advances in Genetics*. Volume 85: Academic Press; 2014. p. 109–47.
165. Cervigni G, Zhang W, Picca A, Carrera A, Helguera M, Manthey F, et al. QTL Mapping for LOX Activity and Quality Traits in Durum Wheat. In: *Proceedings 7th International Wheat Conference SAGPyA/INTA Mar del Plata, Argentina 27 November–2 December*. 2005.
166. Peng ZS, Su ZX, Cheng KC. Characterization of dwarf trait in the tetraploid wheat landrace, Aiganfanma. *Wheat Inf Serv*. 1999; 89:7–12.
167. Borner A, Roder M, Korzun V. Comparative molecular mapping of GA insensitive Rht loci on chromosomes 4B and 4D of common wheat (*Triticum aestivum* L.). *Theoretical and Applied Genetics*. 1997; 95(7):1133–7.

168. Carrera A, Echenique V, Zhang W, Helguera M, Manthey F, Schrager A, et al. A deletion at the Lpx-B1 locus is associated with low lipoxygenase activity and improved pasta color in durum wheat (*Triticum turgidum ssp. durum*). *Journal of Cereal Science*. 2007; 45(1):67–77.
169. De Simone V, Menzo V, De Leonardis AM, Maria Ficco DB, Trono D, Cattivelli L, et al. Different mechanisms control lipoxygenase activity in durum wheat kernels. *Journal of Cereal Science*. 2010; 52(2):121–8.
170. Feng B, Dong Z, Xu Z, An X, Qin H, Wu N, et al. Molecular analysis of lipoxygenase (LOX) genes in common wheat and phylogenetic investigation of LOX proteins from model and crop plants. *Journal of Cereal Science*. 2010; 52(3):387–94.
171. Manna F, Borrelli GM, Massardo DM, Wolf K, Alifano P, Del Giudice L, et al. Differential expression of lipoxygenase genes among durum wheat cultivars. *Cereal Research Communications*. 1998; 26:23–30.
172. Nachit MM, Elouafi I, Pagnotta A, El Saleh A, Iacono E, Labhili M, et al. Molecular linkage map for an intraspecific recombinant inbred population of durum wheat (*Triticum turgidum* L. var. *durum*). *Theoretical and Applied Genetics*. 2001; 102(2–3):177–86.