The Fifth International Scientific Conference PlantGen2019

Haplotypes-based genetic analysis: benefits and challenges

D. Sehgal ✉, S. Dreisigacker

International Center for Maize and Wheat Improvement (CIMMYT), Texcoco, México
✉ e-mail: d.sehgal@cgiar.org

The increasing availability of Single Nucleotide Polymorphisms (SNPs) discovered by Next Generation Sequencing will enable a range of new genetic analyses in crops, which was not possible before. Concomitantly, researchers will face the challenge of handling large data sets at the whole-genome level. By grouping thousands of SNPs into a few hundred haplotype blocks, complexity of the data can be reduced with fewer statistical tests and a lower probability of spurious associations. Owing to the strong genome structure present in breeding lines of most crops, the deployment of haplotypes could be a powerful complement to improve efficiency of marker-assisted and genomic selection. This review describes in brief the commonly used approaches to construct haplotype blocks and some examples in animals and crops are cited where haplotype-based dissection of traits were proven beneficial. Some important considerations and facts while working with haplotypes in crops are reviewed at the end.
Key words: haplotype; GWAS; genomic selection; SNP.

Генетический анализ на основе гаплотипов: проблемы и перспективы

Д. Сегал ✉, С. Дрейзигакер

Международный центр улучшения кукурузы и пшеницы (CIMMYT), Тексоко, Мексика
✉ e-mail: d.sehgal@cgiar.org

Возросшая доступность однонуклеотидных полиморфизмов (SNP), разработанных с помощью технологий секвенирования нового поколения (NGS), позволяет проводить генетические исследования культурных растений, которые ранее были невозможны. Одновременно с этим ученые сталкиваются с необходимостью обработки больших массивов данных, полученных на основе полногеномного скрининга. Сложность обработки и интерпретации экспериментальных данных может быть уменьшена за счет сокращения числа статистических тестов и снижения вероятности ложных ассоциаций путем группировки тысяч SNP в несколько сотен блоков гаплотипов. Благодаря устойчивой структуре генома в селекционных линиях большинства культур, построение гаплотипов может стать мощным дополнением для повышения эффективности маркер-ориентированной и геномной селекции. В настоящем обзоре кратко перечислены подходы, традиционно применяемые для конструирования гаплотипных блоков, а также приведены успешные примеры исследований, проведенных на культурных растениях и животных, по диссекции хозяйственно важных признаков на основе гаплотипного анализа. Представлены выводы и важные заключения, сделанные по результатам изучения генома культурных растений с использованием подхода, основанного на анализе гаплотипов.
Key words: гаплотип; GWAS; геномная селекция; SNP.

Introduction

Advances in Next Generation Sequencing (NGS) technologies by whole genome (Berkman et al., 2012; Chia et al., 2012), transcriptome (Cavanagh et al., 2013), reduced-representation (Elshire et al., 2011; Poland et al., 2012) and/or exome sequencing (Winfield et al., 2016) have led to new levels of Single Nucleotide Polymorphisms (SNPs) discovery. Hence, a paradigm shift from marker-based to sequencing-based genotyping of breeding populations and diversified germplasm panels has been observed in the post-genome sequencing era. These developments have facilitated development of high-density maps, identification of Quantitative Trait Loci (QTL) and discovery of new genes in several crops, thus assisting the breeding process (Sehgal et al., 2016, 2017; Singh et al., 2016; Pandey et al., 2017; Su et al., 2017). Especially polyploid crops such as wheat have benefited from these advances, as marker number and density were major gaps in conducting in depth genetic analyses. Dense sets of SNPs now available from different marker platforms [90K Illumina iselect, Genotyping-by-Sequencing (GBS), Diversity Array Technology Sequencing (DArTseq), high-density Affymetrix Axiom® genotyping array] have significantly upgraded the genetic toolkit available in wheat. Therefore, rapidly growing numbers of breeding lines are being genotyped at low cost (Poland, 2015). In addition, whole genome sequence (>15 Gb) of wheat is now available, by combining next generation (short Illumina reads)

and third generation sequencing data (long Pacific Biosciences reads), which will make cloning of genes feasible (Shi, Ling, 2018).

With upsurge in dense marker data sets coming from different genotyping platforms leading to more markers than observations, scientists will face the challenge of handling large data sets at the whole-genome level for both reliable gene discovery and genomic predictions. Therefore, new approaches will be required to deal with cumbersome data and to make analysis easier. Constructing haplotypes from SNPs is one of the options to deal with bulky datasets. Being multiallelic, haplotypes are more informative than SNPs and allow more powerful and less exhaustive genome-wide scan. In this review, we have first defined what haplotypes are and what approaches are available to make haplotypes. Many examples are cited in animals and crops where haplotypes-based analysis have yielded better results than using SNPs in Genome Wide Association studies (GWAS), Genomic Prediction (GP) and in candidate gene identification.

What are haplotype blocks?

A haplotype block defines a region in the genome that comprises a set of neighboring SNPs, whereby their phased alleles are likely inherited together with little chance of contemporary recombination (Fig. 1). Mainly, three approaches are used to construct haplotype blocks: (1) user-defined length, (2) sliding-window, and (3) linkage disequilibrium (LD). Any of these three methods can be used depending on the skills of the user and/or on the objective of the research. The user-defined fixed length of haplotype blocks (2 to 15 bp) is the easiest approach; however, generated haplotypes do not reflect any biological phenomenon such as LD (Gabriel et al., 2002) or shared evolutionary history (Templeton et al., 2005). The sliding-window approach is the most widely used, and has been used intensively for building haplotypes in GWAS for quantitative or qualitative traits. In this method, a genomic region under study is divided into windows, either of uniform-size or variable-size (Tang et al., 2009), and a multiple-marker association test is performed for each window. This approach is easy to use and handle, however, when adjacent SNPs are in strong LD, it provides redundant information thus making the sliding-window approach no more informative than a SNP. Similarly, when LD pattern vary over large genomic regions, it is difficult to determine window-size for a genome-wide scan. The LD-based approaches are the most advantageous because they focus directly on the detection of historical recombination in a given population and LD coefficients are easy to visualize.

Today most genomic analyses such as GWAS or GP use bi-allelic SNP markers. However, SNPs can be combined into short, multi-allelic haplotypes to overcome bi-allelic problem and to perform a powerful and less exhaustive genome scan. By using haplotype blocks, information on multiple markers jointly can be used and hence local epistatic interactions can be naturally modelled, and the reduced number of parameters enables a range of genomic analyses including GWAS, GP, and/or detection of selection signatures. Further, haplotype blocks can be coded in a simple numeric (binary) form to be used in different R codes or Java-based programs. Figure 2 shows how a haplotype block composed of two adjoining SNPs and having four alleles (AC, GT, AT and GC) can be converted to a simple binary 1-0 format.
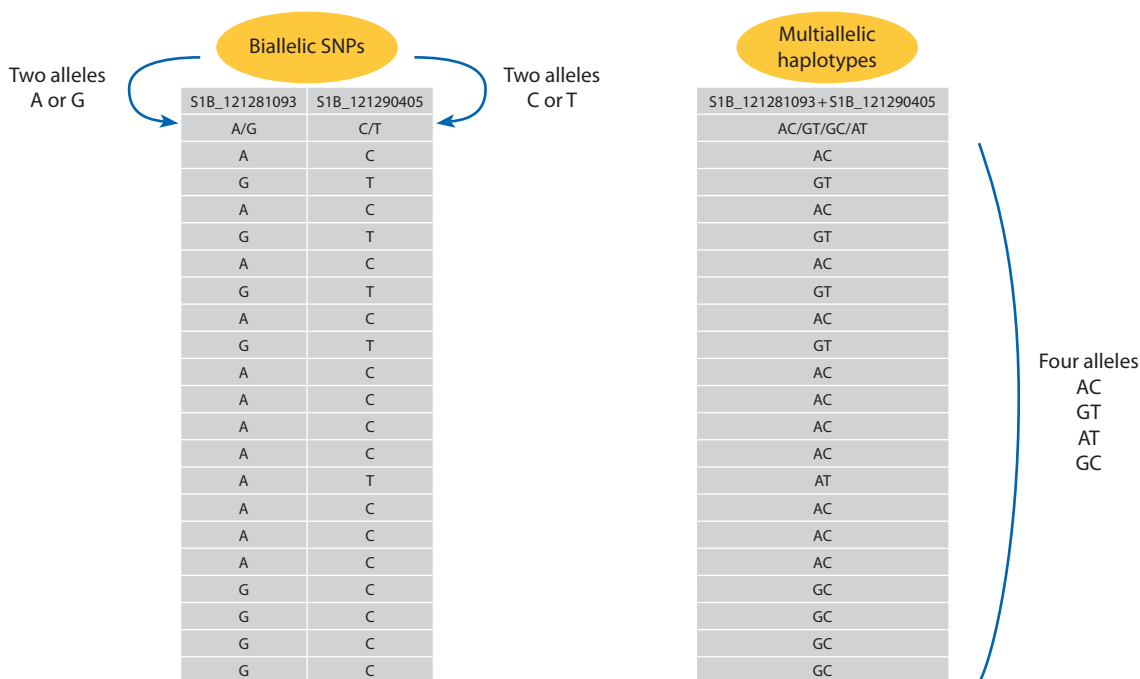
Case studies in animals and humans

GWAS studies based on haplotypes are common in animals and humans (Grapes et al., 2004; Hayes et al., 2007; Calus et al., 2009; Shim et al., 2009; Khankhanian et al., 2015; Jónás et al., 2016; Sato et al., 2016). Studies have generated plethora of evidences to establish that multi-allelic haplotypes significantly improve the power and robustness of association as compared with individual SNPs. A common observation in SNP-based GWAS is the large gap between the variance explained by the identified SNP-associations and the total variance, termed as the 'missing heritability'. J. Yang et al. (2010) showed that a part of the 'missing heritability' could be attributed to a lack of LD between SNP markers and causative variants. Combining neighboring SNPs into haplotype blocks is a simple way to generate a more complete LD. It has been shown that the use of haplotype-based methods have reduced the heritability gap in many cases compared with SNP-based methods when both were applied to the same dataset. P. Khankhanian et al. (2015) investigated the genetic basis of Multiple Sclerosis (MS), a complex genetic disorder in humans controlled by a major histocompatibility complex (MHC) on the short arm of chromosome 6. Haplotypes of various lengths (from 1 up to 15 contiguous SNPs) were constructed at each of the 110 previously identified, MS-associated, genomic regions. The results based on haplotypes outperformed the results using individual SNPs by at least three orders of magnitude. Moreover, when 932 MS-associated haplotypes (identified from 102 genomic regions) were included as independent variables into a logistic linear model; the amount of MS heritability was 38 %, while with individual SNPs it was 29 %.

Simulations based on the LD and population history of livestock have shown that haplotypes can provide greater QTL detection power and mapping accuracy than single markers (Hayes et al., 2007; Calus et al., 2009). Use of haplotypes have also led to the discovery of new genetic regions of interest, which have not been identified by a SNP-based GWAS (Lu et al., 2003; Hagenblad et al., 2004; Shim et al., 2009). W. Barendse (2011) showed that haplotype analysis improved evidence for candidate genes for intramuscular fat (IMF) percentage in cattle as they explained around 80 % more of the phenotypic variance for the five genes that showed some evidence of association to IMF compared to individual SNP analyses. Further studies in animal breeding have also accumulated evidences that integration of haplotypes or haplotype-tagged QTL in genomic selection models can improve GP accuracies for complex traits (Boichard et al., 2012; Cuyabano et al., 2014, 2015a, b; Jónás et al., 2016; Hess et al., 2017; Jiang et al., 2018).

Case studies in crops

Although only a few case studies have been reported in crops, results have been encouraging towards haplotype-based analyses. A.J. Lorenz et al. (2010) used a sliding window approach and explored the utility of haplotype blocks over individual SNPs for GWAS in barley. They used heading date collected on a large set of barley germplasm from the Barley Coordi-

Д. Сегал
С. Дрейзигакер

Генетический анализ на основе гаплотипов:
проблемы и перспективы

2019
23•7

Fig. 1. Two close SNPs shown on the left side, each with two alleles. Haplotypes formed by combining these two SNPs are shown on the right side resulting in four alleles.
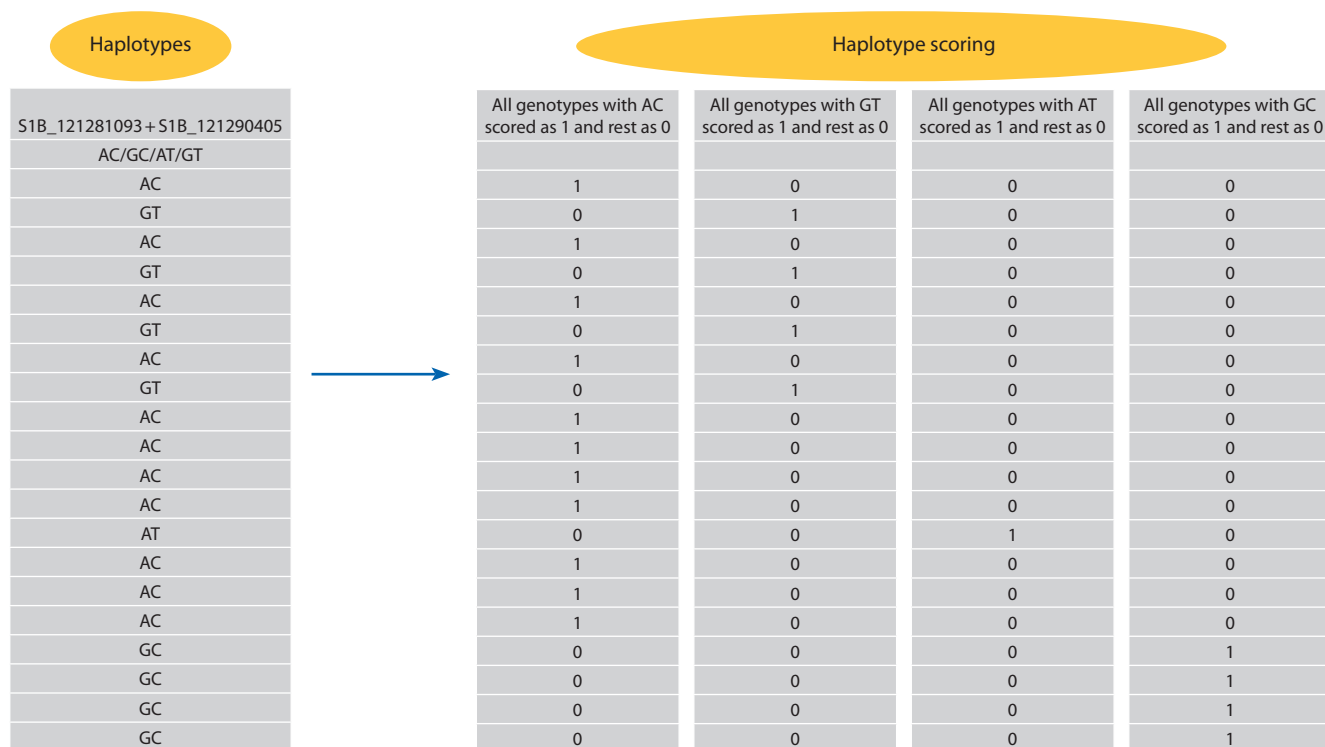
Fig. 2. Strategy to convert haplotypes to binary format to be used in well known R scripts and Java-based platforms for genetic analysis.

nated Agricultural Project. Three associations were found for heading date, two of which were detected by haplotype analyses only. Further, the authors determined the effect of three sets of QTL simulations. The power of individual SNP-based analysis was superior to that of haplotypes when the causal SNP was present in genotyping data. In the absence of causal SNP, haplotypes-based GWAS was more powerful to detect QTL than SNPs. In the latter case, however, the type of method used to construct haplotype blocks affected power of the GWAS. Y. Ma et al. (2016) studied the effect of marker preselection on the prediction accuracy in soybean on plant height and yield per plant. The three strategies tested were

(a) a random SNP sampling method (RSM), (b) a haplotype block analysis-based sampling (HBA), and (c) even SNP sampling method (ESM). They found that for grain yield, prediction accuracy increased by approximately 4 % based on HBA-based approach compared with RSM and ESM.

Y. Lu et al. (2012) conducted comparative LD mapping using SNPs and haplotype blocks to identify QTL for plant height and biomass under drought stress in maize. They used a 10 kb sliding-window approach accounting for the average length of LD to construct haplotype blocks. Using haplotype-based LD mapping, three and 12 significant haplotypes were identified for plant height and biomass, respectively, of which six haplotypes contained at least one SNP that was also significantly associated with the specific trait revealed by SNP-based LD mapping. The haplotype-based analysis explained higher phenotypic variation (on average 2.9 %) than SNPs for both traits.

A few genetic studies have attempted to model the effect of interactions between haplotypes (epistasis) on quantitative traits in crops. Some examples include the vernalization response in barley (Cockram et al., 2007) and chlorophyll content in rapeseed (Qian et al., 2016).

Studies in wheat (published and ongoing)
In wheat, studies are so far very few where haplotypes-based genetic analysis have been conducted. K. Voss-Fels et al. (2017) explored molecular interactions connecting root and shoot development and growth in European elite wheat germplasm to investigate plant's demand for water and nutrients along with its ability to access them. They mapped two highly significant haplotypes for root biomass in close proximity to a major locus known to affect spike development. It was concluded that possibly, strong selection for a haplotype variant controlling heading date, has eliminated a specific combination of two flanking, highly conserved haplotype variants whose interaction confers increased root biomass. Breeders could reverse this consequence of selection to recover root diversity that may be useful under stress environments.

N'Diaye et al. (2017) conducted a SNP- and haplotype-based GWAS of semolina and pasta color in elite durum wheat lines. They combined SNPs within a window size of 5.3 cM (based on average LD decay) on the same chromosome to form haplotype blocks. Haplotype-based GWAS resulted in an increase of the phenotypic variance explained (50.4 % on average) and the allelic effect (33.7 % on average) compared to SNP-based GWAS.

In the past decade, various high-throughput genotyping platforms have been adopted by CIMMYT including the 20K and 90K Illumina iselect SNP arrays, the Breeders' 35K Axiom® array (Affymetrix), DArTseq GBS. As a result, large data sets have been generated on different sets of germplasm. Several SNP-based GWAS studies have been performed (reviewed in Dreisigacker et al., 2019) and haplotype-based GWAS has been initially tested. A latest example include haplotype-based quantification of exotic (landrace, synthetics, etc.) genome imprints in pre-breeding germplasm (Singh et al., 2018). A set of 984 pre-breeding lines (PBLs) generated by a three-way cross (exotic/elite1//elite2) were genotyped with DArTseq and phenotyped for a range of agronomic traits under stress environments. Haplotype blocks, generated using

the LD approach, identified 361 and 367 blocks in PBLs and exotics, respectively. Haplotype block-by-block comparison on each chromosome revealed that 58 (16 %) blocks identified in PBLs were exotic-specific. Further, a rare and favorable haplotype (GT) was identified on chromosome 6D that minimized grain yield (GY) loss under heat stress without penalty under irrigated conditions.

A large GWAS using haplotypes and individual SNPs was performed for GY and superiority index Pi (measure of GY stability) using a large set of advanced bread wheat lines (4,302), which were genotyped with GBS markers and phenotyped under contrasting (irrigated and stress) environments (unpublished work). The average R^2 explained by haplotypes and SNPs showed a 6.1 to 9.9 % higher variation with the haplotype-based GWAS as compared to the individual SNP-based GWAS for GY and Pi (Sehgal et al. personal communication). We further explored whether integrating haplotype-tagged QTL for GY as fixed variables in prediction models improved prediction accuracy. It was observed that the model accounting for the haplotype-based GWAS results as fixed effects led to up to 9 to 10 % increase in prediction accuracy, whereas it was only 4 to 5 % with SNP-tagged QTL. Similarly, haplotype-based GWAS conducted for thousand-grain weight identified four major loci in CIMMYT germplasm; all the four loci showed higher p values than the associated individual SNPs on chromosomes 4A and 6A.

Considerations and challenges
Due to the growing availability of SNP datasets in crops, haplotype-based approaches for genomic analyses is likely to increase markedly. However, the power of analyses using haplotypes vs. SNPs must be evaluated on a case-by-case basis, as risk factors are common for both approaches. For example, under certain disease models (simple Mandelian or complex multi-gene additive or epistatic inheritance) and certain LD patterns one method outperforms the other, so different architectures of QTL and LD patterns interact with marker characteristics to influence power in GWAS. Similarly, bottlenecks are known to increase LD and shift allele frequency spectra toward higher minor allele frequencies. Hence, after a bottleneck, SNPs are more likely to be in LD with QTL and haplotypes might provide little advantage. Marker ascertainment is another important criterion and is a characteristic of SNP chips. In the standard method of developing a SNP chip or an array, a small SNP discovery panel is used, which means that low frequency mutations often go undetected and SNPs occurring at intermediate to high frequencies dominate in such chips or arrays. This over-sampling of mutations at intermediate frequencies results in lower levels of LD than if SNPs were selected randomly. For GP, haplotype-based prediction approaches are favored only if alleles at QTL are more closely linked to the haplotype than to individual SNPs. Finally, map order errors can play a significant role in determining the safe and best approach for analysis. For example, SNP analysis is unlikely to be affected by ordering errors and hence is the best approach when map order is doubtful.

References
Barendse W. Haplotype analysis improved evidence for candidate genes for intramuscular fat percentage from a genome wide association study of cattle. PLoS One. 2011;6(12):e29601.

Д. Сегал
С. Дрейзигакер

Генетический анализ на основе гаплотипов:
проблемы и перспективы

2019
23•7

Berkman P.J., Lai K., Lorenc M.T., Edwards D. Next-generation sequencing applications for wheat crop improvement. Am. J. Bot. 2012;99:365-371.

Boichard D., Guillaume F., Baur A., Croiseau P., Rossignol M.N., Boscher M.Y., Druet T., Genestout L., Colleau J.J., Journaux L., Ducrocq V., Fritz S. Genomic selection in French dairy cattle. Anim. Prod. Sci. 2012;52:115-120.

Calus M.P., Meuwissen T.H., Windig J.J., Knol E.F., Schrooten C., Vereijken A.L.J., Veerkamp R.F. 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.

Cavanagh C.R., Chao S., Wang S., Huang B.E., Stephen S., Kiani S., Forrest K., Saintenac C., Brown-Guedira G.L., Akhunova A., See D., Bai G., Pumphrey M., Tomar L., Wong D., Kong S., Reynolds M., da Silva M.L., Bockelman H., Talbert L., Anderson J.A., Dreisigacker S., Baenziger S., Carter A.K., Morrell P.L., Dubcovsky J., Morell M.K., Sorrells M.E., Hayden M.J., Akhunov E. Genome-wide comparative diversity uncovers multiple targets of selection for improvement in hexaploid wheat landraces and cultivars. Proc. Natl. Acad. Sci. USA. 2013;110:8057-8062.

Chia J.-M., Song C., Bradbury P.J., Costich D., de Leon N., Doebley J., Elshire R.J., Gaut B., Geller L., Glaubitz J.C., Gore M., Guill K.E., Holland J., Hufford M.B., Lai J., Li M., Liu X., Lu Y., McCombie R., Nelson R., Poland J., Prasanna B.M., Pyhäjärvi T., Rong T., Sekhon R.S., Sun Qi, Tenaillon M.I., Tian F., Wang J., Xu X., Zhang Z., Kaeppler S.M., Ross-Ibarra J., McMullen M.D., Buckler E.S., Zhang G., Xu Y., Ware D. Maize HapMap2 identifies extant variation from a genome in flux. Nat. Genet. 2012;44:803-807.

Cockram J., Chiapparino E., Taylor S.A., Stamati K., Donini P., Laurie D.A., O'Sullivan D.M. Haplotype analysis of vernalization loci in European barley germplasm reveals novel *VRN-H1* alleles and a predominant winter *VRN-H1/VRN-H2* multi-locus haplotype. Theor. Appl. Genet. 2007;115:993-1001.

Cuyabano B.C., Su G., Lund M.S. Genomic prediction of genetic merit using LD-based haplotypes in the Nordic Holstein population. BMC Genomics. 2014;15(1):1171.

Cuyabano B.C., Su G., Lund M.S. Selection of haplotype variables from a high-density marker map for genomic prediction. Genet. Sel. Evol. 2015a;47(1):61.

Cuyabano B.C., Su G., Rosa G.J.M., Lund M.S., Gianola D. Bootstrap study of genome-enabled prediction reliabilities using haplotype blocks across Nordic Red cattle breeds. J. Dairy Sci. 2015b; 98(10):7351-7363.

Dreisigacker S., Sehgal D., Singh R.P., Sansaloni C., Braun J.H. Application of genetic and genomic tools in wheat for developing countries. In: Miedaner T., Korzun V. (Eds.). Application of Genetic and Genomic Research in Cereals. Woodhead Publ., 2019;251-271.

Elshire R.J., Glaubitz J.C., Sun Q., Poland J.A., Kawamoto K., Buckler E.S., Mitchell S.E. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One. 2011;6: e19379.

Gabriel S.B., Schaffner S.F., Nguyen H., Moore J.M., Roy J., Blumenstiel B., Higgins J., DeFelice M., Lochner A., Faggart M., Liu-Cordero S.N., Rotimi C., Adeyemo A., Cooper R., Ward R., Lander E.S., Daly M.J., Altshuler D. The structure of haplotype blocks in the human genome. Science. 2002;296:2225-2229.

Grapes L., Dekkers J.C.M., Rothschild M.F., Fernando R.L. Comparing linkage disequilibrium-based methods for fine mapping quantitative trait loci. Genetics. 2004;166:1561-1570.

Hagenblad J., Tang C.L., Molitor J., Werner J., Zhao K., Zheng H., Marjoram P., Weigel D., Nordborg M. Haplotype structure and phenotypic associations in the chromosomal regions surrounding two *Arabidopsis thaliana* flowering time loci. Genetics. 2004;168(3): 1627-1638.

Hayes B.J., Chamberlain A.J., McPartlan H., Macleod I., Sethuraman L., Goddard M.E. Accuracy of marker-assisted selection with single markers and marker haplotypes in cattle. Genet. Res. 2007; 89(4):215-220.

Hess M., Druet T., Hess A., Garrick D. Fixed-length haplotypes can improve genomic prediction accuracy in an admixed dairy cattle population. Genet. Sel. Evol. 2017;3:49(1):54.

Jiang Y., Schmidt R.H., Reif J.C. Haplotype-based genome-wide prediction models exploit local epistatic interactions among markers. G3: Genes Genomes Genetics. 2018;8(5):1687-1699.

Jónás D., Ducrocq V., Fouilloux M.-N., Croiseau P. Alternative haplotype construction methods for genomic evaluation J. Dairy Sci. 2016;994:4537-4546.

Khankhanian P., Gourraud P.A., Lizee A., Goodin D.S. Haplotype-based approach to known MS-associated regions increases the amount of explained risk. J. Med. Genet. 2015;52:587-594.

Lorenz A.J., Hamblin M.T., Jannink J.L. Performance of single nucleotide polymorphisms versus haplotypes for genome-wide association analysis in barley. PLoS One. 2010;5(11):e14079.

Lu X., Niu T.H., Liu J.S. Haplotype information and linkage disequilibrium mapping for single nucleotide polymorphisms. Genome Res. 2003;13(9):2112-2117.

Lu Y., Xu J., Yuan Z., Hao Z., Xie C., Li X., Shah T., Lan H., Zhang S., Rong T., Xu Y. Comparative LD mapping using single SNPs and haplotypes identifies QTL for plant height and biomass as secondary traits of drought tolerance in maize. Mol. Breed. 2012;30(1): 407-418.

Ma Y., Reif J.C., Jiang Y., Wen Z., Wang D., Liu Z., Guo Y., Wei S., Wang S., Yang C., Wang H., Yang C., Lu W., Xu R., Zhou R., Wang R., Sun Z., Chen H., Zhang W., Wu J., Hu G., Liu C., Luan X., Fu Y., Guo T., Han T., Zhang M., Sun B., Zhang L., Chen W., Wu C., Sun S., Yuan B., Zhou X., Han D., Yan H., Li W., Qiu L. Potential of marker selection to increase prediction accuracy of genomic selection in soybean (*Glycine max* L.). Mol. Breed. 2016;36:113.

N'Diaye A., Haile J.K., Cory A.T., Clarke F.R., Clarke J.M., Knox R.E., Pozniak C.J. 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. 2017;12(1): e0170941.

Pandey M.K., Agarwal G., Kale S.M., Clevenger J., Nayak S.N., Sriswathi M., Chitikineni A., Chavarro C., Chen X., Upadhyaya H.D., Vishwakarma M.K., Leal-Bertioli S., Liang X., Bertioli D.J., Guo B., Jackson S.A., Ozias-Akins P., Varshney R.K. Development and evaluation of a high density genotyping 'Axiom_Arachis' array with 58 K SNPs for accelerating genetics and breeding in groundnut. Sci. Rep. 2017;7:40577.

Poland J. Breeding-assisted genomics. Curr. Opin. Plant Biol. 2015; 24:119-124.

Poland J.A., Brown P.J., Sorrells M.E., Jannink J.-L. Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. PLoS One. 2012;7: e32253.

Qian L., Qian W., Snowdon R.J. Haplotype hitchhiking promotes trait coselection in *Brassica napus*. Plant Biotechnol. J. 2016;14:1578-1588.

Sato S., Uemoto Y., Kikuchi T., Egawa S., Kohira K., Saito T., Sakuma H., Miyashita S., Arata S., Kojima T., Suzuki K. SNP- and haplotype-based genome-wide association studies for growth, carcass, and meat quality traits in a Duroc multigenerational population. BMC Genetics. 2016;17:60.

Sehgal D., Autrique E., Singh R.P., Singh S., Ellis M., Dreisigacker S. Identification of genomic regions for grain yield and yield stability and their epistatic interactions. Nat. Sci. Rep. 2017;7:41578.

Sehgal D., Dreisigacker S., Belen S., Küçüközdemir Ü., Mert Z., Özer E., Morgounov A. Mining centuries old *in situ* conserved Turkish wheat landraces for grain yield and stripe rust resistance genes. Front. Genet. 2016;18:7:201.

Shi X., Ling H.-Q. Current advances in genome sequencing of common wheat and its ancestral species. Crop J. 2018;6(1):15-21.

Shim H., Chun H., Engelman C.D., Payseur B.A. Genome-wide association studies using single-nucleotide polymorphisms versus haplotypes: an empirical comparison with data from the North American Rheumatoid Arthritis Consortium. BMC Proc. 2009;3(Suppl.7): S35.

Singh S., Vikram P., Sehgal D., Burgueño J., Sharma A., Singh S.K., Sansaloni C.P., Joynson R., Brabbs T., Ortiz C., Solís-Moya E., Govindan V., Gupta N., Sidhu H.S., Basandrai A.K., Basandrai D., Ledesma-Ramires L., Suaste-Franco M.P., Fuentes-Dávila G., Moreno J.I., Sonder K., Singh V.K., Singh S., Shokat S., Arif M.A.R., Laghari K.A., Srivastava P., Bhavani S., Kumar S., Pal D., Jaiswal J.P., Kumar U., Chaudhary H.K., Crossa J., Payne T.S., Imtiaz M., Sohu V.S., Singh G.P., Bains N.S., Hall A., Pixley K.V. Harnessing genetic potential of wheat germplasm banks through impact-oriented-prebreeding for future food and nutritional security. Nat. Sci. Rep. 2018;8:12527.

Singh V.K., Khan A.W., Saxena R.K., Kumar V., Kale S.M., Sinha P., Chitikineni A., Pazhamala L.T., Garg V., Sharma M., Sameer Kumar C.V., Parupalli S., Vechalapu S., Patil S., Muniswamy S., Ghanta A., Yamini K.N., Dharmaraj P.S., Varshney R.K. Next-generation sequencing for identification of candidate genes for *Fusarium* wilt and sterility mosaic disease in pigeonpea (*Cajanus cajan*). Plant Biotechnol. J. 2016;14:1183-1194.

Su C., Wang W., Gong S., Zuo J., Li S., Xu S. High density linkage map construction and mapping of yield trait QTLs in maize (*Zea mays*) using the genotyping-by-sequencing (GBS) technology. Front. Plant Sci. 2017;8:706.

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;73:631-637.

Templeton A.R., Maxwell T., Posada D., Stengard J.H., Boerwinkle E., Sing C.F. Tree scanning: a method for using haplotype trees in phenotype/genotype association studies. Genetics. 2005;169: 441-453.

Voss-Fels K.P., Qian L., Parra-Londono S., Uptmoor R., Frisch M., Keeble-Gagnère G. Linkage drag constrains the roots of modern wheat. Plant Cell Environ. 2017;40:717-725.

Winfield M., Allen A., Burridge A., Barker G., Benbow H., Wilkinson P., Coghill J., Waterfall C., Davassi A., Scopes G., Pirani A., Webster T., Brew F., Bloor C., King J., West C., Griffiths S., King I., Bentley A., Edwards K. High-density SNP genotyping array for hexaploid wheat and its secondary and tertiary gene pool. Plant Biotechnol. J. 2016;14(5):1195-1206.

Yang J., Benyamin B., McEvoy B.P., Gordon S., Henders A.K., Nyholt D.R. Common SNPs explain a large proportion of the heritability for human height. Nat. Genet. 2010;42:565-569.