Genome-wide association study of agronomic traits in bread wheat reveals novel putative alleles for future breeding programs

Youdes Rahimi 1, 2, Mohammad Reza Djamal 2, Alireza Taleei 1, Hadi Alipour 2 and Pat K. Ingvarsson* 2

Abstract

Background: Identification of loci for agronomic traits and characterization of their genetic architecture are crucial in marker-assisted selection (MAS). Genome-wide association studies (GWAS) have increasingly been used as a potential tool in identifying marker-trait associations (MTAs). The introduction of new adaptive alleles in the diverse genetic backgrounds may help to improve grain yield of old or newly developed varieties of wheat to balance supply and demand throughout the world. Landraces collected from different climate zones can be an invaluable resource for such adaptive alleles.

Results: GWAS was performed using a collection of 298 Iranian bread wheat varieties and landraces to explore the genetic basis of agronomic traits during 2016–2018 cropping seasons under normal (well-watered) and stressed (rain-fed) conditions. A high-quality genotyping by sequencing (GBS) dataset was obtained using either all original single nucleotide polymorphism (SNP; 10,938 SNPs) or with additional imputation (66,862 SNPs) based on W7984 reference genome. The results confirm that the B genome carries the highest number of significant marker pairs in both varieties (48,880, 27.37%) and landraces (55,685, 28.99%). The strongest linkage disequilibrium (LD) between pairs of markers was observed on chromosome 2D (2296). LD decay was lower in the D genome, compared to the A and B genomes. Association mapping under two tested environments yielded a total of 314 and 394 significant (-log10 p > 3) MTAs for the original and imputed SNP data sets respectively. Gene ontology results showed that 27 and 31.3% of MTAs of SNPs in the original set were located in protein coding regions of well-watered and rain-fed conditions, respectively. While, for the imputed data set 22.6 and 16.0% of MTAs represented in protein coding genes for the well-watered and rain-fed conditions, respectively.

Conclusions: Our finding suggests that Iranian bread wheat landraces harbor valuable alleles that are adaptive under drought stress conditions. MTAs located within coding genes can be utilized in genome-based breeding of new wheat varieties. Although imputation of missing data increased the number of MTAs, the fraction of these MTAs located in coding genes were decreased across the different sub-genomes.

Keywords: Wheat, Agronomic traits, GWAS, MTAs, Gene annotation
Background

Bread wheat (*Triticum aestivum* L.) is a staple crop both in developing and developed countries and there is a constant need to balance supply and demand [1]. The projected increase of human population is expected to increase the demand of wheat thereby highlighting the need for plant breeders to utilize all accessible tools to find new ways to sustainably increase the production of bread wheat over the coming decades [1, 2]. Wheat production increased significantly after green revolution in the 1960s and 1970s by better farm management practices and introduction of well-adapted wheat varieties. However, the global consumption of wheat has also steadily been increasing due to population boom [1]. The comparison between wheat production and consumption in 1962 and 2012 shows a substantial increase in demand for wheat, with China, EU, India, and the USA being the major consumers. However, looking at the increase in demand over the 50-year period, it becomes clear that Asian countries, including Indonesia, Bangladesh, and Thailand are at the top of the list [1]. There are several factors drastically limiting wheat production throughout the world, such as water deficiency, salt and cold stress, resulting in significant losses in both grain and biomass [3, 4].

Understanding drought tolerance mechanisms and identifying loci responsible for mediating drought tolerance are key steps for any breeding approach aimed at increasing stress tolerance induced by water deficiency in bread wheat. Recent progress in sequencing technologies have become genome databases available for model plants, such as *Arabidopsis thaliana* [5], *Oryza sativa* [6], and a number of important crop species, including *Hordeum vulgare* [7] and *T. aestivum* [8]. The availability of large-scale genome resources provides an opportunity to discover genetic and molecular mechanisms behind plant responses to different environmental stresses. As most agronomically important traits are likely controlled by a large number of genes, quantitative trait loci (QTL) mapping has been widely used to dissect the genetic architecture of such traits [9–11]. However, QTL-mapping has several drawbacks, such as low resolution and a limited number of alleles that can be screened per study. The high-throughput genotyping technologies providing large number of single nucleotide polymorphism (SNP) data has drastically improved the resolution of QTL mapping by providing high-resolution linkage maps [12].

Furthermore, due to the increased availability of large-scale genomic resources, genome-wide association studies (GWAS) are now a viable alternative to QTL mapping for dissecting the genetic architecture of quantitative traits [13]. In comparison to QTL-mapping, GWAS help accelerates the assessment of a more representative set of individuals in both time and cost-effective way [14]. GWAS are based on establishing correlations between genotype and phenotype, with the idea that linkage disequilibrium (LD) has been formed in a population across generations so that regions harboring QTLs can be detected even if the causal mutations aren’t necessarily included among the set of available genetic markers. Access to high-density genotyping spanning the entire genome makes GWAS invaluable tools for identifying genomic regions underlying the observed phenotypes. Several recent studies have successfully applied GWAS to identify the genetic basis of important traits in a number of crop species, including rice [15], barley [16], corn [17] and wheat [18].

These studies have also provided information about MTAs, which can help breeders in marker-assisted selection. In particular, a number of recent studies have focused on wheat by identifying QTLs associated with grain yield and related traits [19–21].

Genotyping by sequencing (GBS) can provide access to a large number of SNPs in a cost-efficient manner but it is often plagued by a high fraction of missing data that can limit the accuracy of any genome-wide association study. One approach to deal with missing data is through imputation and this has successfully been implemented in many studies on human and plant genomes [22–26]. Imputation can increase the number of variants that are included in a GWAS by relying on linkage information derived from common haplotypes after considering SNPs which are not directly genotyped [27]. Low depth sequencing and library complexity may contribute to missing information in SNP data and genotype imputation can thus be utilized to partly compensate for such issues through available reference genomes without the need for additional expensive resequencing [28, 29]. The main objective of the current study was to perform a GWAS experiment using GBS-SNP data [30] and SNPs imputed based on the W7984 reference genome for bread wheat, which has previously been demonstrated to yield the highest imputation accuracy [31]. A set of three categories of agronomic traits were measured among Italian wheat landraces and varieties and employed in an association study to explore putative QTLs to identify genes, which may be involved in important developmental pathways providing drought tolerance. The second objective of the study was to determine if there are any differences in the results produced using the original SNP data compared to the imputed SNP data by assessing the influence of imputation on MTAs.

Results

Phenotypic evaluation

The datasets for well-watered and rain-fed conditions were analyzed separately. ANOVA identified significant differences (P < 0.01) among varieties and landraces for all studied traits under both environments for two years, except thousand kernel weight under rain-fed conditions