Barley genetic variation: implications for crop improvement

María Muñoz-Amatriain, Alfonso Cuesta-Marcos, Patrick M. Hayes and Gary J. Muehlbauer

Advance Access publication date 22 March 2014

Abstract

Genetic variation is crucial for successful barley improvement. Genomic technologies are improving dramatically and are providing access to the genetic diversity within this important crop species. Diverse collections of barley germplasm are being assembled and mined via genome-wide association studies and the identified variation can be linked to the barley sequence assembly. Introgression of favorable alleles via marker-assisted selection is now faster and more efficient due to the availability of single nucleotide polymorphism platforms. High-throughput genotyping is also making genomic selection an essential tool in modern barley breeding. Contemporary plant breeders now benefit from publicly available user-friendly databases providing genotypic and phenotypic information on large numbers of barley accessions. These resources facilitate access to allelic variation. In this review we explore how the most recent genomics and molecular breeding advances are changing breeding practices. The Coordinated Agricultural Projects (CAPs), Barley CAP and Triticeae CAP coupled with international collaborations, are discussed in detail as examples of a collaborative approach to exploit diverse germplasm resources for barley improvement.

Keywords: barley; breeding; genotyping; germplasm; genome-wide association studies

INTRODUCTION

Barley is unique among crop plants for being of tremendous importance to agriculture and to science. Advances on both fronts create a positive feedback loop, allowing barley to be in the forefront in meeting the great challenges of climate change and human population growth. In terms of agriculture, barley is the fourth most important cereal crop in the world (FAOSTAT: http://faostat.fao.org/site/339/default.aspx). Barley grain, in the form of malt, is the perfect nutritional source for yeast and is therefore the base of the brewing industry [1]. Barley, a truly ancient food grain, is on the rebound as a key component of healthy diets that can reduce the risk of obesity, Type II diabetes and heart disease [2]. The most common use of barley, worldwide, is as a feed grain and a robust barley feed market is an essential component of a viable barley industry, as it can absorb grain that does not meet malting or food specifications [3]. Despite this versatility in end-uses and justified reputation as a stress tolerant cereal crop, barley is in danger of descent to niche crop status.

Acreage in the USA is at an all-time low (USDA-NASS: http://www.agmrc.org/commodities__products/grains__oilseeds/barley/). In Europe, barley acreage is being lost to other more profitable crops [4]. The driving forces of the decline are price and productivity: in many areas of the world, current returns to farmers are simply not sufficient to ensure a viable barley industry. While barley yields are increasing [5], these increases are no match for maize (USDA-NASS: http://www.nass.usda.gov/...
Barley—a diploid \((2n = 2x = 14)\) with a genome size of 5.1 Gb—is the genetic model for the Triticeae. There is a tremendous history of barley genetics research (reviewed by [6]) and culminating most recently with a draft genome sequence [7]. Throughout the late 20th century, barley was a premier genetic model for agricultural research, with a tremendous outpouring of research reports, genetic stocks and data sets (GrainGenes: http://wheat.pw.usda.gov/GG2/ggdb.shtml). In the early days of genomics, barley briefly lost traction due to the high cost of DNA sequencing and the smaller genome size of Brachypodium distachyon. The availability of cost-effective sequencing and genotyping, however, has led to a renaissance in barley genetics, genomics and molecular breeding that should lead, in the relatively short term, to economically important increases in efficient and tailored barley production. A key anchor to all barley genetics efforts is the availability of a draft genome sequence [7] and the bioinformatics tools that facilitate nimble integration of information at the phenotypic level with differences in DNA sequence. A foundational area of endeavor in barley genetics is characterization of genetic resources, such as those present in the extensive germplasm collections housed at the Research Institute for Bioresources (RIB, Japan), the Leibniz Institute of Plant Genetics and Crop Plant Research (IPK, Germany) and the United States Department of Agriculture National Small Grains Collection (USDA-NSGC, USA). There have also been parallel efforts in Hordeum vulgare subsp. spontaneum, the wild ancestor of barley [8, 9]. As described in greater detail later in this article, there has been a significant investment, particularly in the USA, on characterizing and using genetic diversity in elite breeding materials under the auspices of the Coordinated Agricultural Projects (CAPs) funded by the United States Department of Agriculture (USDA). The CAP projects have focused primarily on identifying loci that control traits showing quantitative inheritance via genome-wide association studies (GWAS). These QTL discoveries are being channeled into genomic selection (GS) and marker-assisted selection (MAS) programs. At the same time, there has been an impressive stream of gene cloning and characterization work, including genes important in domestication [10], physiology [11], and resistance to biotic [12] and abiotic [13] stresses.

In this review, we describe the most recent advances in genomic tools for assessing genetic diversity in barley and how they are being applied to characterize and exploit diverse germplasm resources. Current strategies for directing the continuous discoveries into barley breeding are also discussed.

TOOLS FOR ASSESSING GENETIC VARIATION

There are various types of genetic differences between individuals of the same species, including single nucleotide polymorphisms (SNPs), small insertion/deletion polymorphisms (INDELs) and copy-number variations (CNVs). Undoubtedly, SNPs have been the most studied type of intraspecific genetic variation. In barley, expressed sequence tag (EST) resources supported the first ‘large scale’ SNP identification studies [14–16] and the development of the first high-throughput SNP genotyping platform, based on the Illumina Oligo Pool Assays (OPAs) [17]. This platform, which allows the examination of 4596 total markers in sets of 1536 SNPs, has been extensively used by the barley community and made it possible to develop high-density consensus genetic maps [17, 18], and to assess genetic variation in germplasm collections (e.g. [19–22]) which can be associated with phenotypic variation via GWAS (see section below).

The rapid improvement and declining cost of next-generation sequencing (NGS) technologies is revolutionizing genotyping in barley. Thousands of SNPs discovered via resequencing of transcriptomes (RNAseq) have been used in the development of the new iSelect genotyping platform, which is based on the Illumina Infinium assay and allows the simultaneous testing of 7842 SNPs [11]. Higher density genetic maps have been generated (e.g. [11]) and integrated to build higher-resolution consensus maps that are better reference points for assessing and using genetic diversity [23]. Parallel SNP discovery and genotyping, via genotyping-by-sequencing (GBS) and population sequencing (POPSEQ), have been successfully applied to barley, resulting in
identifying and mapping thousands to millions of SNPs in two mapping populations [24–26]. The newly available barley genome assembly [7], which benefited from the high-density genetic maps for physical map anchoring, can likewise serve as a framework to construct genetic maps without the necessity of a de novo map construction [27]. Moreover, this reference assembly provides access to the majority of the barley genes, hence facilitating the exploitation of natural genetic diversity [7].

In the past year, the development of a barley comparative genomic hybridization (CGH) array using whole-genome shotgun (WGS) sequences of the reference genome has enabled exploration of the CNV landscape in the barley genome [28]. An extensive catalog of CNVs affecting both coding and non-coding regions has been identified among 14 accessions, revealing that CNV represents a major source of genetic variation in the barley genome [28]. Additionally, targeted sequencing of the barley exome has become an option due to the development of an exome capture platform covering over 60 Mbp of barley coding sequence [29]. This tool can be extremely useful for identifying genetic variation—SNPs, INDELs and CNVs—affecting the coding regions of the genome without the ascertainment bias associated with array-based platforms [30].

CONNECTING GENOTYPE TO PHENOTYPE

Genome-wide association studies
The development of the genomics tools described above has provided the resources to identify loci controlling important traits. Barley has been a model species for developing QTL mapping approaches and analyses based on biparental populations [6]. However, biparental populations present several disadvantages such as the time and resources needed for developing the specialized populations segregating for the trait(s) to be mapped, the limitation of evaluating effects of only two alleles per locus, and the low mapping resolution caused by the limited recombination in the population [31, 32]. The increase in marker density achieved with new genotyping technologies allowed the utilization of GWAS approaches in barley [33]. Association mapping (AM) relies on linkage disequilibrium to find marker-trait associations in unrelated sets of accessions (e.g. the non-random association of alleles), therefore allowing the assessment of multiple alleles per locus, and taking advantage of the historical recombination inherent in the population for high-resolution mapping (reviewed in [34]). While the diversity present in an association panel offers higher possibilities to identify beneficial alleles, population structure can cause false positive associations [35]. In barley, strong population structure is often associated with spike row number and growth habit, and accurate GWAS therefore requires proper correction for the stratification due to these traits [19, 20, 36, 37]. The barley genetics and breeding community has embraced AM as an approach to identify QTL and to provide the initial information for subsequent gene cloning.

Barley CAP and Triticeae CAP (TCAP)
A significant number of large-scale GWAS in barley have been conducted using tools provided by USDA-funded Barley CAP and TCAP projects. The Barley CAP was established with the goal of coordinated QTL mapping in breeding lines from 10 US breeding programs. Each breeding program submitted 96 lines per year for a total of 960 lines per year and 3840 lines for the 4 years of the project. Each line was genotyped with 3072 OPA SNPs and phenotyped for over 30 traits including those determining disease resistance, malting quality, agronomic characters and food quality. The Barley CAP was a very productive exercise in terms of generating new knowledge and results. Three representative papers from the Barley CAP addressing specific traits and gene/QTL discovery are highlighted in the following narrative: these reports cover resistance to an abiotic stress that will play an important role in climate change, a biotic stress that threatens the viability of barley as a crop in the Upper Midwest of the USA, and the genetic basis of malting quality—the suite of traits that add the greatest economic value to barley.

Intuitively, climate change is associated with global warming and therefore analysis of low-temperature tolerance (LTT) may appear to be an atavistic academic exercise. However, fall-sown cereals—in areas with winter precipitation patterns—have a significant yield advantage of spring-sown cereals. Capitalizing on this advantage requires the capacity to successfully overwinter, particularly as production moves steadily northward (USDA-NASS: http://www.nass.usda.gov/). Even though the average trend will be toward greater warming, there will an increasing likelihood of wide fluctuations in temperature [38]. VonZitzewitz et al. [39]
used a panel composed of winter and facultative growth habit elite lines to determine the utility of GWAS for discovery of genes determining LTT, simply defined as the ability to survive low temperatures. GWAS was successful in identifying the principal determinants of low-temperature survival in barley—the FRH-1 and FRH-2 loci. In one experiment, this GWAS revealed the same QTL that were painstakingly discovered in over 20 years of biparental QTL mapping. The analysis also generated novel and useful information on the non-additive interaction of favorable alleles at these two loci: notably, both positive alleles are required for maximum cold tolerance.

GWAS proved to be equally effective in validating and identifying genes associated with a biotic stress—resistance to Fusarium Head Blight (FHB). This devastating disease has always been a threat to barley: throughout the 20th century, the incidence of FHB steadily drove barley production westward. Relatively disease-free environments were found in Minnesota and North Dakota, which became major barley producing states in close geographic proximity to major urban centers, malt houses, breweries and consumers. More recently, however, climate change and advances in maize breeding and management have led to epidemics of FHB in the region [40]. In terms of climate change, altered precipitation and temperature patterns favor FHB disease development. maize crop residues favor inoculum buildup. The northward expansion of maize hybrids and no-till management system has, together with changes in climate, solidified two sides of the disease triangle (pathogen and environment). The third side of the triangle (susceptible host) has always existed. There are no sources of immunity to FHB in barley and the search for resistance to FHB has, of necessity, focused on unadapted germplasm. Massman et al. [41] used GWAS—based on multi-environment assessment of a large panel of barley CAP-derived breeding lines—to detect QTL for two FHB-related traits (severity and mycotoxin concentration in the grain). These authors found FHB resistance QTL that had been previously reported from biparental populations, as well as novel and useful information regarding markers that could be immediately applied to the development of resistant varieties.

Resistance to abiotic and biotic stresses are essential attributes for successful barley varieties, but these varieties must also have end-use properties that will command a premium in the marketplace. Barley—in the form of malt—is the base of beer and malting quality, therefore, is the most direct route to assuring economic value for barley grain. Malting quality is a complex trait with no simple consistent definition: different beer styles will require different malt profiles [1]. However, there are some baseline characters essential for all malts that lead to the necessary balance of carbohydrates and proteins essential for yeast nutrition. Extensive research over the past 100 years has identified genes, and more recently QTL, involved in the multiple metabolic pathways leading to malt quality (reviewed by [42]). Gutiérrez et al. [43] used the GWAS tools of the Barley CAP to identify QTL for five of the most important malting quality traits (malt extract, wort beta glucan, alpha amylase activity, diastatic power and grain protein content). Most of the QTL detected were close to previously mapped genes and QTL relevant for malt and beer quality. The analysis contributed to a deeper understanding of malting quality genetics via the unmasking of novel alleles and the quantification of inter-allelic interactions and interaction of alleles with environments. Cumulatively, these three studies typify the power and impact of GWAS and the tools developed by the Barley CAP.

The success of the Barley CAP, as documented above, provided the momentum to develop the TCAP. The TCAP seeks to explore new germplasm sources to identify beneficial alleles for climate-change-related traits such as nitrogen and water use efficiency, LTT and fungal disease resistance (http://www.triticeaecap.org; Figure 1). The diverse germplasm targeted by the TCAP includes the NSGC Barley Core, AM panels of elite breeding lines, nested association mapping (NAM) populations and wild barley-introgression populations. Accessing the diversity within the NSGC Core is one of the main goals of the TCAP, as the locally adapted landraces and historical breeding materials stored in the collection can contain beneficial allelic variation which can be incorporated into breeding programs. Figure 2 illustrates the phenotypic diversity existing in the NSGC Core. Although most of the efforts are underway, a genetic characterization of the NSGC Core using the high-density iSelect genotyping platform has been completed. This information has been used to determine population structure and linkage disequilibrium within the Core, to conduct GWAS of ‘hull cover’, ‘spike row number’ and ‘heading date’, and to develop mini-core sets capturing the
majority of the allelic diversity present in the Core collection [23].

**Additional international efforts**

Internationally, GWAS has been effectively used for allele mining of agronomic traits [44] and salt tolerance [45] in materials from the Barley Core Collection [46] and the IPK Genebank. UK germplasm has also been used for GWAS of morphological [47, 48] and agronomic [48, 49] traits.

*Hordeum vulgare* subsp. *spontaneum* collections are also being established to exploit the genetic diversity present in the wild relative of barley. For example, the Wild Barley Diversity Collection, comprising over 300 wild barley accessions [8], was established with the main goal of identifying new resistance loci to important barley diseases. New spot blotch resistance loci have been already identified using this collection [50].

GWAS in germplasm collections can also provide the initial information to isolate genes that control phenotypes. A germplasm collection composed of North American and European barley varieties was used in the identification of one of the genes involved in spike row morphology on chromosome 4H. The QTL was coincident with the *Intermediate-C* (*Int-C*) locus. Rice–barley synteny was subsequently used to identify a *Teosinte branched1* (*Tb1*) ortholog as a candidate gene [51]. Similarly, in a collection of UK barley cultivars, a combination of GWAS and comparative genomics identified a candidate gene for the anthocyanin pigmentation locus *ANT2* [47]. Another approach to GWAS for gene isolation was used by Comadran et al. [11] in the identification of a barley flowering time gene. Signatures of divergent selection revealed by FST analysis in a selected panel of spring- and winter-habit cultivars were used to identify a barley homolog of *Antirrhinum CENTRORADIALIS* (*HvCEN*) as a contributor to the spring growth habit and environmental adaptation [11].

**Copy-number variation studies**

Recent studies have revealed that CNVs (i.e. deletions, insertions and duplications ranging from a few to several thousand base pairs in size) are a very common type of genetic variation in plant genomes (reviewed in [52]). CNVs can have important phenotypic consequences by changing gene dosage, interrupting coding sequences or altering the copy number. For example, a CNV in the *Tb1* locus was identified in a GWAS of spike row morphology and was associated with variation in the number of spikelets per spikelet (LTT) [53].

**Figure 1**: Schematic diagram of the TCAP approach to barley breeding. A diverse set of germplasm including the NSGC Barley Core, NAM populations, wild barley-introgression populations and elite AM panels are being genotyped and phenotyped. SNP genotyping tools include the Illumina OPA, iSelect and custom platforms, and GBS. Phenotyping is being focused on climate-change-related traits such as yield, disease resistances, nitrogen and water use efficiency (NUE and WUE), and LTT. All the genotypic and phenotypic data being generated are placed in the Triticeae Toolbox (T3) database, and can be accessed by all members of the project to conduct GWAS or other type of QTL mapping. Discoveries are being directed into MAS and GS programs, and are also being used for gene isolation. Cumulatively, these strategies will lead to the development of improved barley germplasm in a more efficient manner.
number of regulatory regions. In barley, Sutton et al. [53] demonstrated that an increased copy number of the boron transporter gene Bot1 confers boron-toxicity tolerance to a barley landrace from Algeria. Similarly, Knox et al. [54] found an association between variation in the copy number of CBF (C-Repeat Binding Factor) genes at the Frost Resistant-2 (FR-2) locus and tolerance to low temperatures. More recently, a CGH array developed for barley has provided the tool to estimate the extent of CNV in the barley genome [28]. This study revealed that 9.5% of the coding sequences represented on the CGH array exhibited CNVs. Genes affected by CNV are enriched for sequences annotated as disease-resistance proteins, predominantly nucleotide-binding site leucine-rich repeat (NBS-LRR) and other classes of R proteins. Interestingly, this large-scale study also identified CNV at the CBF3 gene between winter and spring cultivars, providing more evidence that CNV of CBF genes may contribute to cold tolerance. A substantial loss in CNV diversity as a consequence of domestication and breeding was found, as almost half of the exon-affecting variants were present only in wild barley.

**APPLICATION TO BARLEY IMPROVEMENT**

Barley improvement is not possible without genetic variation. Breeders usually cross elite lines within a defined germplasm pool expecting to assemble better allele combinations in the offspring. This is called advanced cycle breeding [55] and leads to an accumulation of favorable alleles. Elite lines, as a result, become more alike within a breeding program and more different from unimproved germplasm [56]. In most barley programs, the sources of genetic diversity can be traced back to a small number of foundational varieties. Advanced cycle breeding reduces this limited genetic diversity even further. Despite this diversity compression, breeders continue to see genetic gains in yield and many agronomic traits [57]. Without new germplasm infusion, however, there is an increased risk of genetic vulnerability as a consequence of pathogen population evolution, loss of resistance genes [58], as well as the appearance of new abiotic stresses and yield constraints [59]. The narrow genetic base of breeding programs was quantified by Martin et al. [60] using pedigree information: seven ancestors contributed 52% of the North American barley varieties.

Figure 2: Morphological variation existing in the NSGC Barley Core. (A) Diversity of grain types from Ethiopian accessions belonging to the NSGC Core. (B) Examples of variation in spike morphology present in the Core collection.
American 6-row germplasm pool and another seven ancestors contributed 67% of the 2-row pool. More recently, using genotypic data from the iSelect genotyping platform we found that 10 parents had contributed over 90% of the alleles present in elite breeding lines of Oregon State University winter 6-row barley breeding program (unpublished data). These examples illustrate that breeders access only a small fraction of the overall available diversity in the barley germplasm pool to release new varieties.

Continued gains from selection will require efficient access to, and introgression of, novel alleles in wild relatives, exotic germplasm and elite germplasm from other breeding programs. Advanced back-cross QTL studies have shown that wild ancestors of barley may contribute favorable alleles, likely lost after a domestication bottleneck, for several agronomic traits, including grain yield and agronomic performance under drought conditions [61], disease resistance [62, 63] and malt quality [64]. There are also examples of successful introgression of alleles from exotic germplasm to address specific needs. Ethiopian landraces have been a source of favorable alleles for Barley Yellow Dwarf virus [65]; most of the powdery mildew resistance genes used commercially have been introgressed from landraces originating in West Asia, Ethiopia and North Africa [66]; the boron-toxicity tolerance found in many Australian varieties traces to an Algerian landrace [53]; and Spanish landraces were the source of a mild vernalization sensitivity allele that improves adaptation to Mediterranean environments [67].

Introgression of unadapted germplasm into a breeding program usually combines several cycles of backcrossing using an elite line as a recurrent parent and a landrace/wild barley as the donor of favorable alleles. The use of the new genotyping and mapping tools can make this process faster and more efficient, as SNPs used for MAS can be more closely associated with the desired gene. Recombinant individuals in a segregating population can be selected for a smaller introgression size and a larger percentage of the recurrent parent genome, thus reducing linkage drag [68]. However, introgression of exotic germplasm is always challenging. Very often, the favorable alleles in exotic germplasm are tightly linked with negative alleles determining deleterious or undesirable traits and then linkage drag is more difficult to solve, even with the increased marker density. Either because of linkage drag, or the epistatic effects of genetic background, favorable alleles may not have the same effect when they are introgressed into a different genetic background. In these cases, an intermediate step of validation of allele effects in different genetic backgrounds is recommended [69]. In contrast, if panels of elite materials are used for GWAS, the QTL detected could be more directly incorporated into breeding programs with no need of prebreeding.

The availability of high-density marker data has driven the emergence of a new breeding approach called genomic selection (GS). GS attempts to address some of the limitations of MAS by using all genome-wide SNP data to predict breeding values (genomic estimated breeding values, GEBVs [70]) for individual lines in a breeding program [71]. The genomic prediction for a trait of interest requires a ‘training population’ of breeding lines for which both phenotypic and genotypic data exist. Since individuals can be selected based on all marker information and without phenotypic tests, faster gains for quantitative traits can be achieved [71]. If GS is coupled with doubled haploid production, the breeding cycle can be as short as one generation. Indeed, GS has been implemented in barley breeding programs and simulations studies are showing promising results [72–74]. Encouragingly, in the first empirical study in barley, Lorenz et al. [75] show that GS approaches can be used to accelerate FHB resistance breeding.

Regardless of whether breeders are intending to access ancestral or elite germplasm, or the type of breeding strategy they are planning to apply, key considerations are access to germplasm, passport information, phenotypic data, genotypic data and connectivity to the barley gene sequence information. The development of databases that integrate and analyse the amount of data being generated is key to providing the accessibility and distribution of genetic materials to the barley community. In this context, a new and key resource integrating contemporary genotyping and sequencing data with phenotype data sets relevant to plant breeders is the Triticeae Toolbox (T3) (http://triticeaetoolbox.org/). T3 was previously developed by the Barley CAP and known as The Hordeum Toolbox [76]. T3 serves as the database for the data generated by the TCAP. T3 allows open access to SNP, phenotypic and pedigree data from barley (and wheat) germplasm from groups participating in the CAP and from the core collections of the USDA-NSGC. It also integrates the same information from the Barley CAP. T3 includes analysis and
visualization tools in a user-friendly environment, allowing breeders to select specific lines, traits and trials of interest and obtain marker–trait associations with GWAS or to make genomic predictions (GEBV predictions) for a set of lines. This resource connects the breeding programs from across the country, providing the opportunity for the barley breeding community to collaborate more effectively. T3 is a crucial component of the TCAP approach to achieve a sustainable nation-wide improvement of the barley crop (Figure 1).

New genomics and molecular breeding technologies are allowing us to better understand and mine genetic variation. Cumulatively, the new resources empower the current generation of barley breeders and geneticists to continue the tradition of barley making simultaneous contributions to agriculture and to science. The great challenges of climate change and human population growth will require innovative, insightful, and productive germplasm enhancement and variety development.

Key Points
- We are experiencing a revolution in genomics that is allowing access to barley genetic diversity (SNPs, INDELS and CNVs) on an unprecedented scale.
- High-density genome-wide SNP data are allowing GWAS in barley, providing information for MAS, GS and gene cloning.
- Useful genetic variation for barley breeding can be now connected to the assembled catalog of gene sequences, increasing the precision and speed of barley improvement.
- The Barley CAP and TCAP are successful examples of collaborative approaches to breeding that have resulted in a rapid exploitation of barley genetic resources for crop improvement.

Acknowledgements
We would like to thank Dr Harold Bockelman and Dr Mike Bonman (USDA-ARS Aberdeen, Idaho) for kindly providing the picture used in Figure 2A.

FUNDING
This work was supported by the Triticeae Coordinated Agricultural Project [2011-68002-30029], funded by the United States Department of Agriculture, National Institute of Food and Agriculture.

References


68. Hospital F. Size of donor chromosome segments around introgressed loci and reduction of linkage drag in marker-assisted backcross programs. Genetics 2001;158:1363–79.


