

Chapter 18

Genome-Wide Association Scans (GWAS)

**Robbie Waugh, Bill Thomas, Andrew Flavell, Luke Ramsay,
Jordi Comadran, and Joanne Russell**

18.1 Introduction

Genetic analysis in barley using molecular markers has been conducted extensively over the past 20 years. Based initially on the framework provided by the development of genome-wide linkage maps (Graner et al. 1990; Kleinhofs et al. 1993), important major genes and quantitative trait loci (QTL) have been located using a range of F₂, RIL and doubled haploid mapping populations. These studies have yielded genetic markers that have been used extensively for the indirect selection of traits that are difficult to assess in a breeding programme context [e.g. resistance to the soilborne pathogen barley yellow mosaic virus (BaYMV) (Graner et al. 1998) and epiheterodendrin content in barley for the whisky industry (Thomas 2003)] and that, if translated into financial value, have generated millions of € by increasing yield under adverse conditions or improving product quality. These same studies have led to the identification of causal genes and corresponding alleles that confer a variety of traits, generally through the well-established route of positional cloning

R. Waugh (✉)

The James Hutton Institute, Invergowrie, Dundee DD2 5DA, Scotland

Division of Plant Sciences, The University of Dundee, JHI Campus, Invergowrie,
Dundee DD2 5DA, Scotland

e-mail: Robbie.Waugh@hutton.ac.uk

B. Thomas • L. Ramsay • J. Russell

The James Hutton Institute, Invergowrie, Dundee DD2 5DA, Scotland

A. Flavell

Division of Plant Sciences, The University of Dundee, JHI Campus, Invergowrie,
Dundee DD2 5DA, Scotland

J. Comadran

LIMAGRAIN Europe, BP1, 63720 Chappes, France

[e.g. *mlo* (Büschges et al. 1997), *Mla* (Wei et al. 1999), *Rym4/Rym5* (Stein et al. 2005), *Vrn3* (Yan et al. 2006), *Ppd-H1* (Turner et al. 2005)].

The use of experimental mapping populations derived from parents that contrast for a target trait has however been of limited use to the more applied research sector because the parents used are frequently irrelevant to current breeding germplasm and the traits identified are already frequently fixed in the elite breeding gene pool. Consequently a move to assess traits that still segregate in such much more closely related germplasm has been promoted. Genome-wide association scans (GWAS) provide a mechanism to assess variation that segregates in a gene pool, rather than in a biparental population. Fashioned originally in human genetics where it was developed to take account of the types of populations available for genetic analysis, it has become popular in plant genetic research over the last decade (Waugh et al. 2009). GWAS is attractive for multiple reasons, the first of which is that it potentially provides an opportunity to exploit existing and extensive phenotypic data collected during the plant registration process, thus making it directly relevant to current breeding material. Second, it holds the promise of increasing genetic resolution because GWAS populations typically contain more genetic breakpoints and more alleles than are found in conventional mapping populations. However, GWAS approaches also raise issues in genetic analysis. These are largely caused by the origins and history of the population, which introduce a tendency to reveal significant false-positive associations due to factors other than genetic linkage. Here, we will attempt to summarise some of the progress and the problems that have been encountered in establishing effective GWAS in barley and the approaches that have been developed or applied to take account of them. Whilst several studies from various groups have shown that GWAS in barley can be an effective tool for QTL analysis, within our group we have focused on the potential of the approach for identifying the actual genes underlying specific plant phenotypes.

18.2 Linkage Disequilibrium in Different Barley Gene Pools

Determining the extent of linkage disequilibrium (LD) in a target gene pool allows us to estimate the number of molecular markers required to conduct a saturated GWAS and the mapping resolution it is likely to achieve. Studies in outbreeding (e.g. maize; Remington et al. 2001) and inbreeding (e.g. *Arabidopsis*; Nordborg et al. 2002) species revealed that the extent of LD is very different according to breeding habit and, as predicted theoretically, tends to be considerably less extensive in outbreeders. For inbreeders, the derived homozygosity reduces the effective recombination rate at each round of meiosis, and LD is more extensive. However, LD is also highly population dependent, as reported for several species including barley (Caldwell et al. 2006). This has led subsequently to significantly revised estimates of LD (Nordborg et al. 2005; Kim et al. 2007; Yan et al. 2009). Thus, in barley, whilst the initial studies of Kraakman et al. (2004) assayed a collection of 146 modern two-row spring barley cultivars using 236 AFLPs observed significant

LD between markers extending up to 10 cM, Morrell et al. (2005) concluded that intra-locus LD decayed at a rate similar to that observed in outbreeding maize from looking at intra- and inter-gene LD in 18 nuclear genes in a collection of 25 wild barley accessions sampled from across its natural geographic range.

Caldwell et al. (2006) illustrated this population dependency issue very clearly. By resequencing genes present on a small BAC contig across cultivars, landraces and wild barley isolates, they observed a sharp decline in the extent of LD with increasing wildness, consistent with the evolutionary time between the individuals within each sampled set. Although this study was based on a small region of the barley genome, its general conclusions have been confirmed several times since in both diverse and narrow barley collections, and more importantly at a genome-wide scale (Malysheva-Otto et al. 2006; Cuesta-Marcos et al. 2010; Zhang et al. 2009). Subsequent studies have also shown that LD based on physical distance measurements varies enormously according to genomic position (Hamblin et al. 2010; Comadran et al. 2011a). Thus, within the same elite-cultivated gene pool, LD may extend from hundreds of kilobases in recombinogenic portions of the genome to hundreds of megabases in the rarely recombining (but gene rich) centromere-proximal regions.

18.3 Genetic Markers

A knowledge and understanding of how LD is elaborated in different gene pools allows us to estimate the number of genetic markers required to best capture the diversity and recombination history of the population (Fig. 18.1). In the cultivated gene pool where LD is extensive, a relatively small number of markers are theoretically required to capture the majority of the recombination events present

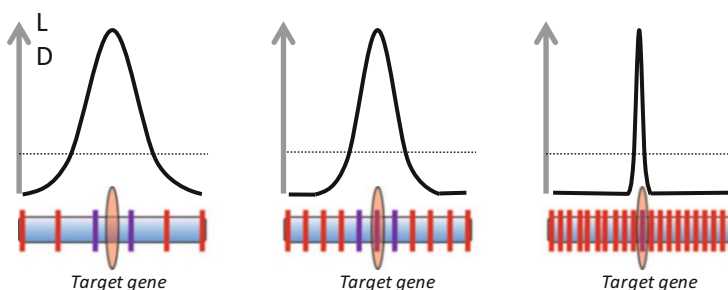


Fig. 18.1 A cartoon of linkage disequilibrium (LD) in three different barley gene pools (*black arc*). The three graphs from *left to right* symbolise the situation observed in the cultivated, landrace and wild gene pools of barley. They show the impact of changes in the extent in LD due to recombination on the number of genetic markers (*vertical red and purple lines* on a portion of a hypothetical chromosome) required to detect significant associations between a marker and a target gene. Thus, with more extensive LD in the cultivated gene pool, fewer markers are required to detect the target gene by GWAS when compared to the wild gene pool

in the population. Based on practical observations, this led Rostoks et al. (2006) to suggest that roughly 5×10^2 to 5×10^3 markers may be required to adequately survey the elite NW European barley gene pool. At the other end of the spectrum, the number required to capture the resolution afforded by thousands of years of effective recombination in wild species are likely to exceed this by orders of magnitude. In many respects the density of markers required has constrained the adoption of GWAS, and in many species there is still insufficient understanding of LD and available genetic markers and marker technologies that can be adequately applied for this purpose.

Genetic marker technologies have been evolving continuously for the last 25 years or more in barley, as in most major crops. Despite the early attempts by Kraakman et al. (2004, 2006) and Kraakman (2005) to apply AFLP technology to association mapping in barley, it only became realistic to attempt GWAS studies in large populations of related germplasm with the availability of high-throughput (HTP) single-nucleotide polymorphism (SNP) marker technologies such as Illumina's 'GoldenGate' oligo pool assays (OPAs) (Fan et al. 2003; Rostoks et al. 2005, 2006; Close et al. 2009). These technologies effectively eradicated unintentional error within genotypes introduced during serial marker assays and allowed the collection of massive marker datasets that were virtually inconceivable only a few years earlier. These markers also revealed much about legacy biparental mapping populations, highlighting genotypic errors and unintentional mix-ups, sometimes at frequencies of 10 % or higher, and by eradicating single-marker double recombinants, promoted map shrinkage to lengths broadly consistent with observed numbers of chiasmata during meiosis (Nilsson et al. 1993). HTP SNP marker sets were similarly informative in germplasm collections, revealing sample incongruence, heterogeneity and duplication at previously unprecedented resolution. Recently, SNP platforms containing many thousands of markers have been developed, such as Illumina's Barley-OPA1 (BOPA1), Barley-OPA2 (BOPA2) (Close et al. 2009) and iSELECT platforms (Comadran et al. 2012), and used widely to genotype thousands of samples in both the public and private sectors (e.g. AGOUEB, <http://www.agoueb.org>; BarleyCAP, <http://barleycap.cfans.umn.edu>; ExBarDiv: http://pgrc.ipk-gatersleben.de/barleynet/projects_exbardiv.php) (Waugh et al. 2010).

Despite their success, these SNP marker platforms are already coming under threat from methods that exploit the massive increase in data volumes and reduction in costs associated with next-generation sequencing technologies (NGS). Methods including the use of reduced-representation libraries (RRLs), complexity reduction of polymorphic sequences (CRoPSTM), restriction-site-associated DNA sequencing (RAD-seq) and low-coverage genotyping by sequencing (GbS) provide ultra-high density genotyping at extremely low cost per datapoint (reviewed in Davey et al. 2011). These sequence-based methods have no prior development requirements and can be used in species lacking reference genome sequences. In barley RAD-seq on the Oregon Wolfe Barley population generated 463 new RAD loci on all seven linkage groups (Chutimanitsakun et al. 2011) and GbS on the same population over 25,000 additional markers at exceedingly low cost (Elshire et al. 2011). However, at this point in time, the commercial propositions such as the iSELECT platform remain more accessible to the general user as the vendor

provides an ‘out-of-the-box’ informatics solution to capturing, analysing, recording and exporting defined genotypic data into a wide range of analytical software. At the time of writing, the sequence-based methods still require specialised bioinformatics support to collect and interrogate the genotypic data—a big disadvantage for many smaller groups. However, it is a logical development and a significant step forward. Not surprisingly, GbS has already been implemented in barley association mapping studies.

18.4 Marker Ascertainment Issues

Whilst the ‘marker constrained’ highly multiplex assays such as the OPA and iSELECT technologies from Illumina are tremendously effective and simple to use, they are not ideally suited to all applications. Because their development generally involves mining sequence data extracted from a limited number of individuals, the utility of the SNPs obtained is affected by this discovery protocol. Basically, SNPs are identified in a small panel of individuals selected from a much larger population. As they represent only a small subset of the individuals, only a fraction of total polymorphisms will be discovered. When these SNPs are then scored on a larger sample of individuals, an ‘ascertainment bias’ is introduced (Nielsen 2000). Because the SNP discovery panel is small, the probability that an SNP will be identified is a function of its frequency in the discovery population. Rare SNPs will go undiscovered more often than common SNPs, and SNPs not present in the discovery population will never be incorporated in the assay platform. When the platform is then used to screen a much broader set of germplasm, this ascertainment bias will compromise measures of relatedness and genetic diversity because statistical measures that rely on allele frequency, such as nucleotide diversity, population genetics parameters and linkage disequilibrium, will be affected (Nielsen 2000; Schlotterer and Harr 2002; Rosenblum and Novembre 2007; Storz and Kelly 2008).

BOPA1, BOPA2 and the 9K iSELECT platforms were developed from SNP data extracted from a limited number of barley accessions (Rostoks et al. 2005, 2006; Close et al. 2009; Comadran et al. 2012), and several large-scale projects have used them effectively to identify marker-trait associations in elite cultivars (AGOUEB, <http://www.agoueb.org>; Barley CAP, <http://barleycap.cfans.umn.edu>; ExBarDiv: http://pgrc.ipk-gatersleben.de/barleynet/projects_exbardiv.php) (Waugh et al. 2010) and in diversity panels comprising both elite cultivars and landraces (Pasam et al. 2012). Despite these apparent successes, we should be mindful that the extent and patterns of diversity observed have been affected by ascertainment issues and that results generated in these studies in most cases still need to be validated. This is particularly true when examining diverse genotypes. For example, understanding genetic diversity inherent within accessions that tolerate extreme conditions of temperature and water availability is likely to be particularly important in future breeding efforts that seek to respond to future environmental challenges. It is therefore important that issues such as

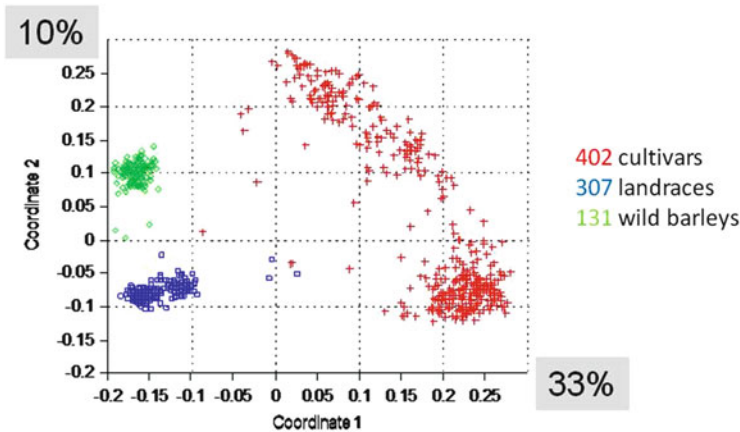


Fig. 18.2 Principle coordinates analysis illustrating the effect of ascertainment bias on estimations of genetic diversity in diverse barley gene pools. The SNPs were ascertained from the cultivated gene pool and were chosen based on high allele frequencies. The wild and landrace barley germplasm ‘looks’ as if it is much narrower than the cultivated germplasm—which we know from many other studies is completely the wrong way around

ascertainment bias are fully taken into account when using a marker platform derived from one gene pool to investigate another.

One example that highlights this issue and that has been examined in some detail is the use of SNPs sampled from the cultivated gene pool to examine diversity in collections of landrace barleys from Syria and Jordan (Fig. 18.2). Moragues et al. (2010) evaluated the effects of SNP number and selection strategy on estimates of germplasm diversity and population structure in different barley collections. Using the 1,536 BOPA1 SNP data and random or optimised subsets of 384 and 96 SNPs, they compared diversity statistics for 161 landraces from Jordan and Syria with 171 European cultivars that had previously been studied using SSRs (Russell et al. 2003). They observed differences in the patterns of SNP polymorphisms and, somewhat counter-intuitively, a lower estimate of diversity in the landraces, contradicting the SSR results. This bias could be at least partially nullified by selecting an appropriate subset of SNPs.

More recently Russell et al. (2011) described the first application of BOPA1 to assess the evolution of barley in a portion of the Fertile Crescent. Specifically, they were interested in examining diversity across the genome but in particular those regions that have been previously identified as playing a role in domestication. They genotyped geographically matched landrace and wild barleys (448 accessions) from Jordan and Syria. One consequence of ascertainment bias would be to skew the landrace-wild comparison by excluding rarely polymorphic markers in the wild barleys, resulting in an underestimate of their true genetic diversity. However, the experimental data showed higher levels of genetic variation in wild material, and furthermore, the differences were similar to those found in previous work (Russell et al. 2004). Also, if the effect of bias introduced by using SNPs sampled from elite

cultivars was problematic, the expectation would be a reduction of diversity in the wild compared to landraces around the domestication genes (because SNPs in the wild would not have been assayed). But they identified 141 cases where rolling diversity estimates were significantly different between wild and landrace barley genotypes, with diversity higher in wild material for 94 % of the cases, many in regions where domestication genes are known. As ascertainment bias would have pushed this comparison in the other direction, their observations become increasingly significant.

18.5 Accounting for Population Structure

When mapping by association, underlying population structure can be a strong confounding factor that results in a high frequency of false-positive associations. (Rostoks et al. 2006; Mackay and Powell 2007). Considering a hypothetical trait, if this trait was frequently associated with any sub-population, then all corresponding background markers that identify alleles with a similar clustering distribution between populations would also be associated with the trait, regardless of whether they were physically linked to it. Minimising these false-positive effects has been the focus of considerable effort in the statistical genetics community, and a number of approaches have been developed in an attempt to nullify them whilst allowing true associations to be detected.

GWAS analysis that does not account for population substructure (a *naive* approach) is based on the same principles as those applied in biparental QTL mapping populations. Simply, it consists of regressing the phenotype against the alleles at each genetically mapped locus to detect QTLs and is successful because each marker allele in the genetic map has a given probability of being associated with the QTL of interest. The *naive* approach is not generally suitable for use in structured populations for the reasons given above. However, it is suitable for use in populations in which structure has been intentionally minimised. A popular example of this type is a multiparent advanced generation intercross (MAGIC) population (Cavanagh et al. 2008). Another possibility is to use substantially unstructured sub-populations identified by PCO or STRUCTURE analysis of the associated marker data (Waugh et al. 2010), although some would argue that even within these populations, a structure correction should always be applied.

The reality is that barley germplasm sampled across the world is strongly stratified into sub-populations, reflecting growth habit, ear morphology and geographical origin, and is linked to local adaptation and crop end use. As a naive approach is unsuitable in this case, several different statistical approaches that correct and/or account for the effects of population structure within such germplasm have been developed. Indeed, correcting for structure has guided most of the research on GWAS for the last few years (Pritchard et al. 2000; Mackay and Powell 2007). Issues arise when the application of different statistical approaches reveal an inconsistent number and/or identity significant associations or remove known

biological factors that are correlated at some level with population structure. This can result in uncertainty over what QTL to prioritise for further studies or to use as diagnostics in marker-assisted selection (MAS).

Structured association uses genome-wide molecular diversity data to compute statistics that define the genetic structure contained within the germplasm. The derived statistics can then be modelled within a mixed linear model (MLM) framework to account for the multiple levels of relatedness that result from historical stratification and kinship (Yu et al. 2006). Statistical softwares including Genstat (VSN International 2011), R (<http://www.R-project.org/>) and TASSEL (Bradbury et al. 2007, <http://www.maizegenetics.net>) can then provide (different) corrections for population structure. A variance covariance matrix containing coefficients of co-ancestry (kinship matrix) can be included in the mixed model to account for genetic relatedness between genotypes. Eigenanalysis (Patterson et al. 2006) uses the scores of the most significant PCA axes from the molecular marker matrix as co-variables in the mixed model, approximating the use of a kinship matrix. In barley Cockram et al. (2010) and Comadran et al. (2011b) found that a mixed linear regression model that accounts for relatedness due to kinship and historical population substructure to perform well. A significance threshold is usually estimated for each analysis using a Bonferroni-corrected p -value of 0.05. Importantly, with the observed increase in marker data volumes, methods that are able to cope with thousands to millions of computationally intensive analyses have emerged that provide a choice of both approximate [e.g. GRAMMAR (Aulchenko et al. 2007), implemented in GenABEL (<http://www.genabel.org/packages/GenABEL>); P3D (Zhang et al. 2010), implemented in TASSEL (<http://www.maizegenetics.net/tassel>); EMMAX (Kang et al. 2010) (<http://genetics.cs.ucla.edu/emmax/>)] and exact methods [e.g. FMM (W. Astle & D. Balding, <http://www.genabel.org/MixABEL/FastMixedModel.html>); FaST-LMM (Lippert et al. 2011) (<http://mscompbio.codeplex.com/>); GEMMA (M. Stephens lab, <http://stephenslab.uchicago.edu/software.html>)] to account for structure effects.

18.6 Data Management and Display

With the size of the datasets generated, both molecular and phenotypic, a key issue for longer-term value of an association mapping population surrounds data management, quality control and data visualisation, particularly if the dataset forms a reference for the wider research community and has been derived from multiple datasets generated by groups from remote locations. Whilst there may be local solutions to this issue, within our programme we have developed and implemented a GERMINATE data warehouse (Lee et al. 2005; http://bioinf.scri.ac.uk/public/?page_id=159) modified to hold high-density phenotypic and genotypic diversity data, Illumina iSELECT and GbS SNP metadata together with the results of our analyses. Working closely with the breeding community has prompted the development of a number of features in GERMINATE that assist data querying,

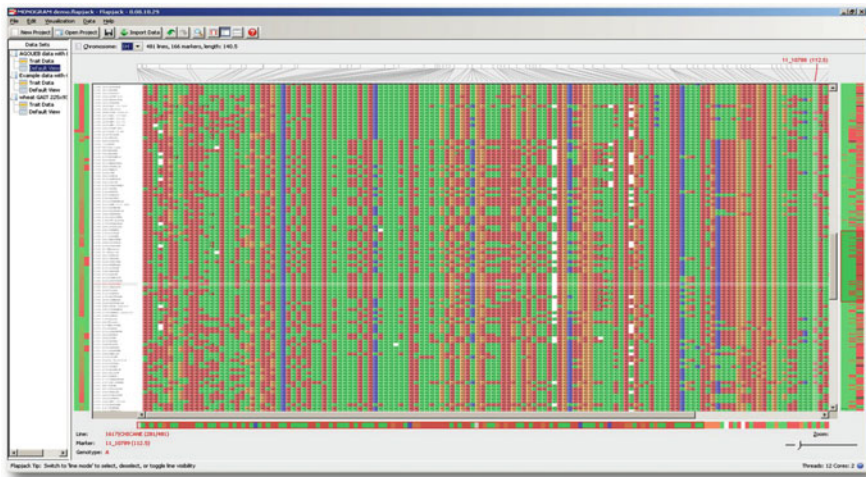


Fig. 18.3 A screenshot of the Flapjack graphical genotyping environment. The marker alleles are colour coded (A, C, G, T, white = missing data) and arranged in genetic marker order along each chromosome (horizontal axis). Individual accessions are shown in the vertical axis. Tracks for visualising trait data are available but not shown. The pattern of SNP alleles along a chromosome can be easily inspected visually (see <http://ics.hutton.ac.uk/flapjack/> for further details)

manipulation and visualisation. In particular, interfacing with the Flapjack graphical genotyping environment (Milne et al. 2010) has been of particular significance, with the Flapjack data model (Fig. 18.3) now being widely adopted by other plant breeding and germplasm diversity projects including the ‘SeeD’ programme at CIMMYT, the Triticeae CAP (T-CAP) project in the United States (<http://www.triticeaecap.org/?q=node/2>), Gates Foundation-funded GCP Integrated Breeding Platform (<http://wiki.cimmyt.org/confluence/display/MBP/Home>) and the Gramene Diversity project (http://www.gramene.org/db/diversity/diversity_view). Further developments in these latter projects will enable users to automatically load data and analysis results and provide enhanced tool integration with various genetic analysis platforms. Thus, efforts are underway to more intimately integrate Flapjack with data analysis software such as TASSEL, R, Genstat and genetic simulation tools like QuGene (Podlich and Cooper 1998).

18.7 Phenotypic Analysis

One of the original attractions of association mapping was that it promised to be able to exploit rich phenotypic information that had already been collected either by prior academic studies or of the rigorous trialling and testing procedures that cultivars must go through as part of the official registration process. For example in the United Kingdom, up to 80 morphological-developmental traits are described

and available for use in assessing the distinctiveness, uniformity and stability (DUS) of prospective cultivars and up to 40 (including grain yield, quality and disease resistance) tested for value for cultivation and use (VCU) (<http://www.fera.defra.gov.uk/plants/plantVarieties/nationalListing/documents/protocolCereals10.pdf>). Work carried out in the AGOUEB population in the United Kingdom and cultivated barley collections at IPK in Germany have reported the use of such data (Cockram et al. 2010; Comadran et al. 2011a; Wang et al. 2012; Matthies et al. 2009, 2012). This may be because it can often be difficult to extract this type of data from archives or because it may be difficult to use as official testing protocols and ways of recording the phenotypic data have been modified over time and accessions may have undergone further selection between the point of testing for DUS/VCU and genotyping. However, where the data are clean, it remains a highly valuable asset that obviates the need for *de novo* phenotyping. Conducting the necessary quality control prior to analysis is however time consuming and may involve a considerable amount of retesting.

For certain phenotypes, like disease resistance, that are tested on relatively young leaf material using a common ‘treatment’ (e.g. a pathogen population), morphological-developmental differences between accessions can have limited impact on the collected data. However, the opposite can be true when attempting to collect equivalent data on diverse genotypes that may be confounded by significant developmental and morphological differentiation. For example, wild barley isolates and landraces from around the world have highly diverse heading dates and heights and using data such as ‘grain yield’ collected in a single environment across such a diverse population may be effectively meaningless. Because of these difficulties we have found it advantageous to ‘tune’ the accessions in our association mapping population by including only those with broadly similar developmental characteristics. Whilst this necessarily restricts the amount of variation that segregates in the population, we have found that this approach enables rather than restricts genetic dissection of the considerable genetic variation that remains in the population.

18.8 Association Mapping in Barley

Several individual groups and consortia have recently assembled collections of germplasm into association mapping panels and have phenotyped and genotyped them at varying depths with the objective of performing GWAS (e.g. Haseneyer et al. 2010). To date, none are artificially constructed populations such as nested association mapping (NAM; McMullen et al. 2009) or MAGIC (Cavanagh et al. 2008) that are promoted as exploiting the power of both linkage analysis and association mapping approaches and designed to avoid the population structure issues that inflate false-positive associations in natural populations. Such populations are currently under development (<http://triticeaecap.org/?q=node/1>). Examples of some of the populations already used for GWAS are as follows.

18.8.1 Wild Barley Populations

Steffenson et al. (2007) assembled a Wild Barley Diversity Collection (WDBC) comprising 318 accessions selected on the basis of eco-geographic parameters that included longitude/latitude, elevation, high/low temperature, rainfall and soil type. Most were from the Fertile Crescent, Central Asia, North Africa and the Caucasus region. Single plant selections were repeatedly selfed to near homozygosity and the resulting inbreds genotyped using 558 Diversity Array Technology (DArT[®]; Jaccoud et al. 2001) and 2,878 BOPA1 and BOPA2 SNPs. GWAS was conducted after correcting for structure, initially for leaf, stem and stripe rust (Steffenson et al. 2007) and latterly for spot blotch (Roy et al. 2010) resistance. 13–15 significant associations of small effect, some corresponding with the location of known resistance genes, were detected for each phenotype. Given the expected extent of LD in the WDBC (Caldwell et al. 2006; Morrell et al. 2005), these results are somewhat surprising and it will be interesting to see if any of the detected associations are subsequently validated. It is tempting to speculate that SNP ascertainment issues, combined with low levels of recombination in the genetic centromeres may have played some role in these findings.

18.8.2 Landraces

A European Union-funded project under the acronym EXBARDIV (http://pgrc.ipk-gatersleben.de/barleynet/projects_exbardiv.php) was founded on the hypothesis that stratified germplasm collections may allow genetic resolution to be manipulated in GWAS by shuttling between cultivated, landrace and wild association mapping populations. The Europe-wide team assembled a collection of 360 elite European barley cultivars (overlapping with the UK AGOUEB Project summarised below), 480 landraces from Jordan and Syria and known as the ICARDA Syrian-Jordanian Landrace Collection (SJLC; Ceccarelli et al. 1987) and two sets of wild barleys, including a subset of 131 individuals from the WDBC summarised above. These lines have been phenotyped for a wide range of characters at multiple sites across Europe and simultaneously genotyped with the barley 9K iSELECT SNP platform. Several manuscripts describing the analysis of the data associated with several of these phenotypes are currently in the pipeline (unpublished). In addition, Casas et al. (2011) surveyed the Spanish Core Collection of barley landraces (Igartua et al. 1998) to identify candidate genes affecting flowering time variation by GWAS. There are, however, few other GWAS studies specifically of barley landraces. Some include landraces as a subset of a wider germplasm collection, e.g. Comadran et al. (2011b), and others have used a limited number of SSR markers, e.g. Jones et al. (2011).

18.8.3 *Cultivars*

Several populations have been assembled specifically to exploit the potential power of GWAS in cultivated barley material starting with the relatively small population used in the original studies of Kraakman et al. (2004, 2006) and Kraakman (2005). We focus on two of these here. However, whilst we highlight these major efforts, other association mapping populations have been assembled and that have now exploited using the BOPA marker technology. These include MABDE (Comadran et al. 2009), EXBARDIV (see above) and GABI-Genobar (Rode et al. 2012), and results from these are now starting to emerge in the literature.

18.8.3.1 Barley CAP

In order to conduct association mapping (AM) studies of economically important traits in US barley breeding germplasm, a panel of 3,840 US barley breeding lines originating from 10 major breeding programmes was assembled and genotyped with 3,072 SNPs (BOPA1 and BOPA2). Population structure was examined using the programme STRUCTURE (Pritchard et al. 2000) and principle component analysis (PCA), revealing 7–9 sub-populations with some correspondence with the different breeding programmes (Hamblin et al. 2010; Zhou et al. 2012). The major population subdivisions were imposed by inflorescence morphology (two-row versus six-row) growth habit (spring vs. winter) and end use (malt vs. feed). Average LD within sub-populations was found to decay across a range of 20–30 cM in Hamblin et al. (2010) and between 4.0 and 19.8 cM in Zhou and Steffenson (2012) as determined by calculating r^2 . The authors estimated that quantitative trait loci (QTL) should be detected in their population with a 50 % probability within a genetic interval of 5 cM and with 95 % probability within 25 cM. These and other studies using subsets of the Barley CAP material (e.g. Cuesta-Marcos et al. 2010; von Zitzewitz et al. 2011; Wang et al. 2012; Massman et al. 2011) and phenotypic data from breeding programmes, were able to detect QTL previously detected in other studies, validating the investment in the association mapping approach. However, none so far have advanced as far as identifying the causal underlying genes. In each of these studies, the authors stress that careful consideration must be given to population diversity, size and experimental design.

18.8.3.2 AGOUEB

The AGOUEB (*pronounced Ag-web*) consortium was established as a public/private partnership in the United Kingdom and was set up to explore the diversity present in European plant breeding programmes using contemporary molecular marker technologies (BOPA1 and BOPA2). Using the same marker platform as

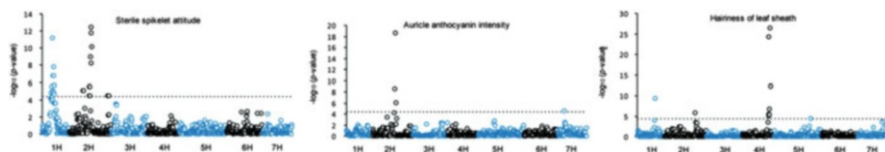


Fig. 18.4 GWAS for three morphological characters—sterile spikelet attitude, auricle anthocyanin intensity and hairiness of the leaf sheath using 1,536 SNPs on a collection of c. 500 mixed barley cultivars (adapted from Cockram et al. 2010). Resolution to single-gene level was achieved for anthocyanin pigmentation where a deletion in *HvbHLH1* was shown to be the causal polymorphism

Barley CAP, Cockram et al. (2010) genotyped a collection of c. 500 cultivars selected from UK registration trials over the past 20 years. As with Barley CAP significant population structure was detected generating high levels of false-positive associations between markers. Significant intrachromosomal LD was observed across the full length of chromosomes (mean distance between significant marker pairs = 40.2 cM, median = 30.7 cM, similar to that observed by Hamblin et al. (2010) in US germplasm). However, after adjustment using a mixed model to take account of population structure, this was reduced to <10 cM (mean = 1.2 cM, median = 0.6 cM), with the proportion of significant inter-chromosomal associations controlled to just 0.1 %. They examined historical phenotypic data for 32 different morphological traits, successfully identifying loci controlling 15 and attributing failure in the other 17 cases to low-quality or variably recorded phenotypic data (e.g. Fig. 18.4). Cockram et al. (2010) also modelled the power to detect 1, 2 and 10 independent loci distributed randomly across the genome, with heritabilities (h^2) of 0.5 and 0.9. Using a mixed model to correct for genetic substructure, simulations based on a trait controlled by one locus predicted that their experimental design had a high probability (≥ 0.92 for both values of h^2) of detecting significant (q value ≤ 0.1) associations within windows of ≤ 8 cM. However, for a ten-locus trait, they reported that the power to detect one or more loci after correction with the mixed model was low (0.25, $h^2 = 0.5$; 0.58, $h^2 = 0.9$). As with Barley CAP the issues associated with using highly structured populations in AGOUEB were therefore again highlighted as a potential impediment to successful GWAS.

18.9 GWAS to Single Gene Resolution

An advantage of GWAS over the use of biparental populations for trait dissection is that the amount of recombination that has occurred in the population should potentially afford single-gene resolution provided that the gene target does not reside in a genomic region with restricted recombination rate, such as the pericentromeric heterochromatin. Whilst the success of this depends on a large extent on the population assembled, several examples now exist in the literature where this

has indeed turned out to be the case. In *Arabidopsis*, Atwell et al. (2010) provide a number of examples where large-scale phenotyping combined with high-resolution genotyping and GWAS has identified a significant enrichment of a priori candidate genes for a wide range of traits. Thus, Todesco et al. (2010) demonstrated that allelic variation at *ACCELERATED CELL DEATH 6* was responsible for fitness benefits elaborated as resistance to microbial infection and herbivory. However, the same locus also had a marked impact on pleiotropic variation in vegetative growth. In the maize-nested association mapping population, Tian et al. (2011) recently showed that variation in leaf angle and size, parameters that have allowed maize planting density to be increased due to more efficient light capture, is partially controlled by allelic variation at the *LIGULELESS* genes. Similar successes have been achieved in a collection of c. 500 rice landraces (Huang et al. 2010).

In barley there are currently three examples in the literature of the successful use of GWAS to single-gene resolution (Fig. 18.3). In the first, Cockram et al. (2010) clearly demonstrated that this level of resolution was achievable in a germplasm collection comprised of winter and spring, two-rowed and six-rowed elite barley cultivars. By focusing on a robust single-gene phenotype, the presence or absence of anthocyanin pigmentation, they were able to show that variation in the anthocyanin pathway regulatory gene *HvbHLHI* was responsible for the observed phenotype. ‘White’ alleles contained a diagnostic deletion that resulted in a premature stop codon upstream of the basic helix-loop-helix domain. By assaying for the presence of the deletion in a collection of ‘red’ and ‘white’ alleles present in landrace germplasm originating from across Europe, they were able to infer the geographical origin of the white allele and map its subsequent spread throughout Europe.

In the second, Ramsay et al. (2011) were able to identify and prove that *SIX-ROWED SPIKE 5 (INTERMEDIUM-C)*, a gene that affects barley row type, was a functional orthologue of the maize domestication gene *TEOSINTE BRANCHED 1*. They achieved this despite the phenotype being a cause of major population subdivision in the germplasm used in the analysis. Although it is a simple two-state morphological character, GWAS identified four highly significant associations, suggestive of strong epistatic interactions. As would have been predicted, one association peak mapped to the *SIX-ROWED SPIKE 1 (Vrs1)* locus (Komatsuda et al. 2007), another with *SIX-ROWED SPIKE 5* and the remaining two with separate loci on chromosome 1H. One of these latter loci has subsequently been shown to correspond to the *SIX-ROWED SPIKE 3* locus (our unpublished results). Importantly, Ramsay et al. (2011) were able to validate their candidate gene using a legacy collection of independent spike mutants (Druka et al. 2011) that had previously been attributed to lesions in *SIX-ROWED SPIKE 5* by allelism tests.

Finally Comadran et al. (2012) used a modified analytical approach based on divergent selection between the winter and spring barley gene pools to identify regions of the barley genome where contrasting alleles had been selected in these different lifestyle types. They eventually focussed on one such region which from QTL studies had been called *EARLINESS PER SE 2* and mapped as the major determinant of earliness in a study examining adaptation of barley to droughted

environments. Using available mutant resources they were able to show that the gene responsible for the observed phenotype was the barley orthologue of the *Antirrhinum majus* gene *CENTRORADIALIS*, a paralogue of the *Arabidopsis* flowering repressor *TERMINAL FLOWER 1 (TFL1)*. Within our group we have now used GWAS to identify a number of additional genes and validated them using the same strategy, i.e. with independent barley mutants.

Conclusions

The successes in GWAS-associated identification of gene alleles encoding barley traits described above bode well for the future of this approach, especially since the potential power of the method is continuously increasing. It is not unreasonable to predict that in the next few years, hundreds of thousands of polymorphic sites that are mapped on a reliable physical framework for the barley genome will become available for GWAS in barley. Furthermore, the arrival of GWAS populations with lower substructure, more allelic variation and higher numbers of recombination breakpoints will increase the mapping resolution. In such circumstances single-gene resolution for GWAS will become commonplace.

Future directions of GWAS in barley will to some extent be driven by the falling cost of genotyping associated with next-generation sequencing technologies (NGS). Given the potential to saturate marker coverage of the genome, the discriminatory power of GWAS in barley will be determined by the size of the population studied and the patterns of LD and population structure within the population. The use of large more genetically balanced populations that are specifically developed for GWAS (McMullen et al. 2009; Cavanagh et al. 2008) will undoubtedly play an increasing role though recombination rates in this inbreeding crop will continue to be a limiting factor particularly in certain regions of the genome. In addition to the importance of choice of population, the potential discriminatory power of GWAS will certainly concentrate more attention onto experimental design and the opportunities offered by high-throughput phenotyping. Whilst it is now possible to conduct QTL x environment AM analyses using Genstat (VSN International 2011), current analytical methods are largely single-locus additive models. Future analytical developments will lead to multi-locus models with the potential to detect epistatic interactions, as now in biparental QTL mapping. Finally the discrimination of GWAS in barley down to the gene level will also necessitate the further development of validation strategies and the integration of future population studies with developments in functional genomics and systems analyses in the crop.

To avoid the majority of the potential issues with population substructuring, we have assembled a population of approaching 1,000 two-rowed spring barley varieties that exhibit low population substructure

(continued)

and show similar morpho-developmental characteristics (particularly flowering time). We are currently using this population extensively to investigate a range of simple and more complex traits, and our experience to date suggests that such populations do simplify underlying genetic complexity making it more amenable to statistical interpretation (Waugh et al. 2010). This population is a powerful resource for future genetic analysis in barley, and we welcome collaboration with groups who would like to exploit the power and resolution it affords.

References

- Atwell S, Huang YS, Vilhjálmsson BJ, Willems G, Horton M, Li Y, Meng D, Platt A, Tarone AM, Hu TT, Jiang R, Muliyati NW, Zhang X, Amer MA, Baxter I, Brachi B, Chory J, Dean C, Debieu M, de Meaux J, Ecker JR, Faure N, Kniskern JM, Jones JD, Michael T, Nemri A, Roux F, Salt DE, Tang C, Todesco M, Traw MB, Weigel D, Marjoram P, Borevitz JO, Bergelson J, Nordborg M (2010) Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. *Nature* 465:627–631
- Aulchenko YS, de Koning D-J, Haley C (2007) Genomewide rapid association using mixed model and regression: a fast and simple method for genomewide pedigree-based quantitative trait loci association analysis. *Genetics* 177:577–585
- Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007) TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* 23:2633–2635
- Büschges R, Hollricher K, Panstruga R, Simons G, Wolter M, Frijters A, van Daelen R, van der Lee T, Diergaarde P, Groenendijk J, Töpsch S, Vos P, Salamini F, Schulze-Lefert P (1997) The barley *Mlo* gene: a novel control element of plant pathogen resistance. *Cell* 88:695–705
- Caldwell KS, Russell J, Langridge P, Powell W (2006) Extreme population-dependent linkage disequilibrium detected in an inbreeding plant species, *Hordeum vulgare*. *Genetics* 172:557–567
- Casas AM, Djemel A, Ciudad FJ, Yahiaoui S, Ponce LJ, Contreras-Moreira B, Gracia MP, Lasa JM, Igartua E (2011) HvFT1 (VrnH3) drives latitudinal adaptation in Spanish barleys. *Theor Appl Genet* 122:1293–1304
- Cavanagh C, Morell M, Mackay I, Powell W (2008) From mutations to MAGIC: resources for gene discovery, validation and delivery in crop plants. *Curr Opin Plant Biol* 11:215–221
- Ceccarelli S, Grandi S, Leur JAG (1987) Genetic diversity in barley landraces from Syria and Jordan. *Euphytica* 36:389–405
- Chutimanitsakun Y, Nipper RW, Cuesta-Marcos A, Cistué L, Corey A, Filichkina T, Johnson EA, Hayes PM (2011) Construction and application for QTL analysis of a Restriction Site Associated DNA (RAD) linkage map in barley. *BMC Genomics* 2011:4
- Close TJ, Bhat PR, Lonardi S, Wu Y, Rostoks N, Ramsay L, Druka A, Stein N, Svensson JT, Wanamaker S, Bozdog S, Roose ML, Moscou MJ, Chao S, Varshney RK, Szucs P, Sato K, Hayes PM, Matthews DE, Kleinhofs A, Muehlbauer GJ, DeYoung J, Marshall DF, Madishetty K, Fenton RD, Condamine P, Graner A, Waugh R (2009) Development and implementation of high-throughput SNP genotyping in barley. *BMC Genomics* 10:582
- Cockram J, White J, Zuluaga D, Smith D, Comadran J, Macaulay M, Luo ZW, Kearsley MJ, Werner P, Harrap D, Tapsell C, Liu H, Hedley PE, Stein N, Schulte D, Steuernagel B, Marshall DF, Thomas WTB, Ramsay L, Mackay I, Balding DJ, Waugh R, O’Sullivan D (2010)

- Genome-wide association mapping of morphological traits to candidate gene resolution in the un-sequenced barley genome. *Proc Natl Acad Sci U S A* 107:21611–21616
- Comadran J, Thomas WT, van Eeuwijk FA, Ceccarelli S, Grando S, Stanca AM, Pecchioni N, Akar T, Al-Yassin A, Benbelkacem A, Ouabbou H, Bort J, Romagosa I, Hackett CA, Russell JR (2009) Patterns of genetic diversity and linkage disequilibrium in a highly structured *Hordeum vulgare* association-mapping population for the Mediterranean basin. *Theor Appl Genet* 119:175–187
- Comadran J, Ramsay L, MacKenzie K, Hayes P, Close TJ, Muehlbauer G, Stein N, Waugh R (2011a) Patterns of polymorphism and linkage disequilibrium in cultivated barley. *Theor Appl Genet* 122:523–531
- Comadran J, Russell JR, Booth A, Psarayi A, Ceccarelli S, Grando S, Stanca AM, Pecchioni N, Akar T, Al-Yassin A, Benbelkacem A, Ouabbou H, Bort J, van Eeuwijk FA, Thomas WT, Romagosa I (2011b) Mixed model association scans of multi-environmental trial data reveal major loci controlling yield and yield related traits in *Hordeum vulgare* in Mediterranean environments. *Theor Appl Genet* 122:1363–1373
- Comadran J, Kilian B, Russell J, Ramsay L, Stein N, Ganai M, Shaw P, Bayer M, Thomas WTB, Marshall D, Hedley P, Tondelli A, Pecchioni N, Francia E, Korzun V, Walther A, Waugh R (2012) A homologue of *Antirrhinum CENTRORADIALIS* is a component of the quantitative photoperiod and vernalization independent *EARLINESS PER SE 2* locus in cultivated barley. *Nat Genet* 44:1388–1392
- Cuesta-Marcos A, Szucs P, Close TJ, Filichkin T, Muehlbauer GJ, Smith KP, Hayes PM (2010) Genome-wide SNPs and re-sequencing of growth habit and inflorescence genes in barley: implications for association mapping in germplasm arrays varying in size and structure. *BMC Genomics* 11:707
- Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nat Rev Genet* 12:499–510
- Druka A, Franckowiak J, Lundqvist U, Bonar N, Alexander J, Houston K, Radovic S, Shahinnia F, Vendramin V, Morgante M, Stein N, Waugh R (2011) Genetic dissection of barley morphology and development. *Plant Physiol* 155:617–627
- Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell Sharon E (2011) A robust, simple genotyping-by-sequencing (gbs) approach for high diversity species. *PLoS One* 6:e19379
- Fan JB, Oliphant A, Shen R, Kermani BG, Garcia F, Gunderson KL, Hansen M, Steemers F, Butler SL, Deloukas P, Galver L, Hunt S, McBride C, Bibikova M, Rubano T, Chen J, Wickham E, Doucet D, Chang W, Campbell D, Zhang B, Kruglyak S, Bentley D, Haas J, Rigault P, Zhou L, Stuelpnagel J, Chee MS (2003) Highly parallel SNP genotyping. *Cold Spring Harb Symp Quant Biol* LXVIII:69–78
- Graner A, Jahoor A, Schondelmaier J, Siedler H, Pillen K, Fischbeck G, Wenzel G, Herrmann RG (1990) Construction of an RFLP map of barley. *Theor Appl Genet* 83:250–256
- Graner A, Streng S, Kellermann A, Schiemann A, Bauer E, Waugh R, Pellio B, Ordon F (1998) Molecular mapping and genetic fine-structure of the *rym5* locus encoding resistance to different strains of the barley yellow mosaic virus complex. *Theor Appl Genet* 98:285–290
- Hamblin MT, Close TJ, Bhat PR, Chao S, Kling JG, Joseph Abraham K, Blake T, Brooks WS, Cooper B, Griffey CA, Hayes PM, Hole DJ, Horsley RD, Obert DE, Smith KP, Ullrich SE, Muehlbauer GJ, Jannink JL (2010) Population structure and linkage disequilibrium in U.S. barley germplasm: implications for association mapping. *Crop Sci* 50:556
- Haseneyer G, Stracke S, Paul C, Einfeldt C, Broda A, Piepho HP, Graner A, Geiger HH (2010) Population structure and phenotypic variation of a spring barley world collection set up for association studies. *Plant Breed* 129:271–279
- Huang X, Wei X, Sang T, Zhao Q, Feng Q, Zhao Y, Li C, Zhu C, Lu T, Zhang Z, Li M, Fan D, Guo Y, Wang A, Wang L, Deng L, Li W, Lu Y, Weng Q, Liu K, Huang T, Zhou T, Jing Y,

- Li W, Lin Z, Buckler ES, Qian Q, Zhang QF, Li J, Han B (2010) Genome-wide association studies of 14 agronomic traits in rice landraces. *Nat Genet* 42:961–967
- Igartua E, Gracia MP, Lasa JM, Medina B, Molina-Cano JL, Montoya JL, Romagosa I (1998) The Spanish barley core collection. *Genet Resour Crop Evol* 45:475–481
- Jaccoud D, Peng K, Feinstein D, Kilian A (2001) Diversity arrays: a solid state technology for sequence information independent genotyping. *Nucleic Acids Res* 29:E25
- Jones H, Civián P, Cockram J, Leigh FJ, Smith LM, Jones MK, Charles MP, Molina-Cano JL, Powell W, Jones G, Brown TA (2011) Evolutionary history of barley cultivation in Europe revealed by genetic analysis of extant landraces. *BMC Evol Biol* 11:320
- Kang HM, Sul JH, Service SK, Zaitlen NA, Kong SY, Freimer NB, Sabatti C, Eskin E (2010) Variance component model to account for sample structure in genome-wide association studies. *Nat Genet* 42:348–354
- Kim S, Plagnol V, Hu TT, Toomajian C, Clark RM, Ossowski S, Ecker JR, Weigel D, Nordborg M (2007) Recombination and linkage disequilibrium in *Arabidopsis thaliana*. *Nat Genet* 39:1151–1155
- Kleinhofs A, Kilian A, Saghai Marrof MA, Biyashev RM, Hayes P, Chen FQ, Lapitan N, Fenwick A, Blake TK, Kanazin V, Ananiev E, Dahleen L, Kudrna D, Bollinger J, Knapp SJ, Liu B, Sorells M, Heun M, Franckowiak JD, Hoffman D, Skadsen R, Steffenson BJ (1993) A molecular, isozyme and morphological map of the barley (*Hordeum vulgare*) genome. *Theor Appl Genet* 86:705–712
- Komatsuda T, Pourkheirandish M, He C, Azhaguvel P, Kanamori H, Perovic D, Stein N, Graner A, Wicker T, Tagiri A, Lundqvist U, Fujimura T, Matsuoka M, Matsumoto T, Yano M (2007) Six-rowed barley originated from a mutation in a homeodomain-leucine zipper I-class homeobox gene. *Proc Natl Acad Sci U S A* 104:1424–1429
- Kraakman ATW (2005) Mapping of yield, yield stability, yield adaptability and other traits in barley using linkage disequilibrium mapping and linkage analysis. PhD Dissertation 3772, Wageningen University
- Kraakman ATW, Niks RE, Van den Berg PMMM, Stam P, Van Eeuwijk FA (2004) Linkage disequilibrium mapping of yield and yield stability in modern spring barley cultivars. *Genetics* 168:435–446
- Kraakman ATW, Martinez F, Mussiraliev B, van Eeuwijk FA, Niks RE (2006) Linkage disequilibrium mapping of morphological, resistance, and other agronomically relevant traits in modern spring barley cultivars. *Mol Breed* 17:41–58
- Lee JM, Davenport GF, Marshall D, Ellis TH, Ambrose MJ, Dicks J, van Hintum TJ, Flavell AJ (2005) GERMINATE: a generic database for integrating genotypic and phenotypic information for plant genetic resource collections. *Plant Physiol* 139:619–631
- Lippert C, Listgarten J, Liu Y, Kadie CM, Davidson RI, Heckerman D (2011) FaST linear mixed models for genome-wide association studies. *Nat Methods* 8:833–835
- Mackay I, Powell W (2007) Methods for linkage disequilibrium mapping in crops. *Trends Plant Sci* 12:57–63
- Malysheva-Otto LV, Ganal MW, Röder MS (2006) Analysis of molecular diversity, population structure and linkage disequilibrium in a worldwide survey of cultivated barley germplasm (*Hordeum vulgare* L). *BMC Genet* 7:6
- Massman J, Cooper B, Horsley R, Neate S, Dill-Macky R, Chao S, Dong Y, Schwarz P, Muehlbauer GJ, Smith KP (2011) Genome-wide association mapping of Fusarium head blight resistance in contemporary barley breeding germplasm. *Mol Breed* 27:439–454
- Matthies IE, Weise S, Roder MS (2009) Association of haplotype diversity in the alpha-amylase gene amy1 with malting quality parameters in barley. *Mol Breed* 23:139–152
- Matthies IE, Sharma S, Weise S, Roder MS (2012) Sequence variation in the barley genes encoding sucrose synthase I and sucrose phosphate synthase II, and its association with variation in grain traits and malting quality. *Euphytica* 184:73–83
- McMullen MD, Kresovich S, Villeda HS, Bradbury P, Li H, Sun Q, Flint-Garcia S, Thornsberry J, Acharya C, Bottoms C, Brown P, Browne C, Eller M, Guill K, Harjes K, Kroon D, Lepak N,

- Mitchell SE, Peterson B, Pressoir G, Romero S, Rosas OM, Salvo S, Yates H, Hanson M, Jones E, Smith S, Glaubitz JC, Goodman M, Ware D, Holland JB, Buckler ES (2009) Genetic properties of the maize nested association mapping population. *Science* 325:737–740
- Milne I, Shaw P, Stephen G, Bayer M, Cardle L, Thomas WTB, Flavell AJ, Marshall DF (2010) Flapjack – graphical genotype visualization. *Bioinformatics* 26:3133–3134. doi:10.1093/bioinformatics/btq580#_blank
- Moragues M, Comadran J, Waugh R, Milne I, Flavell AJ, Russell JR (2010) Effects of ascertainment bias and marker number on estimations of barley diversity from high throughput SNP genotype data. *Theor Appl Genet* 120:1525–1534
- Morrell PL, Toleno DM, Lundy KE, Clegg MT (2005) Low levels of linkage disequilibrium in wild barley (*Hordeum vulgare* ssp. *spontaneum*) despite high rates of self-fertilization. *Proc Natl Acad Sci U S A* 102:2442–2447
- Nielsen R (2000) Estimation of population parameters and recombination rates from single nucleotide polymorphisms. *Genetics* 154:931–942
- Nilsson NO, Säll T, Bengtsson BO (1993) Chiasma and recombination data in plants: are compatible? *Trends Genet* 9:344–348
- Nordborg M, Borevitz JO, Bergelson J, Berry CC, Chory J, Hagenblad J, Kreitman M, Maloof JN, Noyes T, Oefner PJ, Stahl EA, Weigel D (2002) The extent of linkage disequilibrium in *Arabidopsis thaliana*. *Nat Genet* 30:190–193
- Nordborg M, Hu TT, Ishino Y, Jhaveri J, Toomajian C, Zheng H, Bakker E, Calabrese P, Gladstone J, Goyal R, Jakobsson M, Kim S, Morozov Y, Padhukasahasram B, Plagnol V, Rosenberg NA, Shah C, Wall JD, Wang J, Zhao K, Kalbfleisch T, Schulz V, Kreitman M, Bergelson J (2005) The pattern of polymorphism in *Arabidopsis thaliana*. *PLoS Biol* 3:e196
- Pasam RK, Sharma R, Malosetti M, van Eeuwijk FA, Haseneyer G, Kilian B, Graner A (2012) Genome-wide association studies for agronomical traits in a world-wide spring barley collection. *BMC Plant Biol* 12:16
- Patterson N, Price AL, Reich D (2006) Population structure and eigenanalysis. *PLoS Genet* 2:e190
- Podlich DW, Cooper M (1998) QU-GENE: a platform for quantitative analysis of genetic models. *Bioinformatics* 14:632–653
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155:945–959
- Ramsay L, Comadran J, Druka A, Marshall DF, Thomas WTB, Macaulay M, MacKenzie K, Simpson CG, Fuller J, Bonar N, Hayes PM, Lundqvist U, Franckowiak JD, Close TJ, Muehlbauer G, Waugh R (2011) INTERMEDIUM-C, a modifier of lateral spikelet fertility in barley is an ortholog of the maize domestication gene TEOSINTE BRANCHED 1. *Nat Genet* 43:169–172
- Remington DL, Thornsberry JM, Matsuoka Y, Wilson LM, Whitt SR, Doebley J, Kresovich S, Goodman MM, Buckler ES IV (2001) Structure of linkage disequilibrium and phenotypic associations in the maize genome. *Proc Natl Acad Sci U S A* 98:11479–11484
- Rode J, Ahlemeyer J, Friedt W, Ordon F (2012) Identification of marker-trait associations in the German winter barley breeding gene pool (*Hordeum vulgare* L.). *Mol Breed* 30:831–843. doi:10.1007/s11032-011-9667-6
- Rosenblum EB, Novembre J (2007) Ascertainment bias in spatially structured populations: a case study in the eastern fence lizard. *J Hered* 98:331–336
- Rostoks N, Mudie S, Cardle L, Russell JR, Ramsay L, Booth A, Svensson JT, Wanamaker SI, Walia H, Rodriguez EM, Hedley PE, Liu H, Morris J, Close TJ, Marshall DF, Waugh R (2005) Genome wide SNP discovery and linkage analysis in barley based on genes responsive to abiotic stress. *Mol Genet Genomics* 274:515–527
- Rostoks N, Ramsay L, MacKenzie K, Cardle L, Bhat PR, Roose ML, Svensson JT, Stein N, Varshney RK, Marshall DF, Graner A, Close TJ, Waugh R (2006) Recent history of artificial outcrossing facilitates whole genome association mapping in elite crop varieties. *Proc Natl Acad Sci U S A* 103:18656–18661

- Roy JK, Smith KP, Muehlbauer GJ, Chao S, Close TJ, Steffenson BJ (2010) Association mapping of spot blotch resistance in wild barley. *Mol Breed* 26:243–256
- Russell J, Booth A, Fuller JD, Baum M, Ceccarelli S, Grando S, Powell W (2003) Patterns of polymorphism detected in the chloroplast and nuclear genomes of barley landraces sampled from Syria and Jordan. *Theor Appl Genet* 107:413–421
- Russell J, Booth A, Fuller F, Harrower B, Hedley P, Machray G, Powell W (2004) A comparison of sequence-based polymorphism and haplotype content in transcribed and anonymous regions of the barley genome. *Genome* 47:389–398
- Russell JR, Dawson IK, Flavell AJ, Steffenson B, Weltzien E, Booth A, Ceccarelli S, Grando S, Waugh R (2011) Analysis of more than 1,000 SNPs in geographically-matched samples of landrace and wild barley indicates secondary contact and chromosome-level differences in diversity around domestication genes. *New Phytol* 191:564–578
- Schlotterer C, Harr B (2002) Single nucleotide polymorphisms derived from ancestral populations show no evidence for biased diversity estimates in *Drosophila melanogaster*. *Mol Ecol* 11:947–950
- Steffenson BJ, Olivera P, Roy JK, Jin Y, Smith KP, Muehlbauer GJ (2007) A walk on the wild side: mining wild wheat and barley collections for rust resistance genes. *Aust J Agric Res* 2007:532–544
- Stein N, Perovic D, Kumlehn J, Pellio B, Stracke S, Streng S, Ordon F, Graner A (2005) The eukaryotic translation initiation factor 4E confers multiallelic recessive Bymovirus resistance in *Hordeum vulgare* (L.). *Plant J* 42:912–922
- Storz JF, Kelly JK (2008) Effects of spatially varying selection on nucleotide diversity and linkage disequilibrium: insights from deer mouse globin genes. *Genetics* 180:367–379
- Thomas WTB (2003) Prospects for molecular breeding of barley. *Ann Appl Biol* 142:1–12
- Tian F, Bradbury PJ, Brown PJ, Hung H, Sun Q, Flint-Garcia S, Rocheford TR, McMullen MD, Holland JB, Buckler ES (2011) Genome-wide association study of leaf architecture in the maize nested association mapping population. *Nat Genet* 43:159–162
- Todesco M, Balasubramanian S, Hu TT, Traw MB, Horton M, Epple P, Kuhns C, Sureshkumar S, Schwartz C, Lanz C, Laitinen RAE, Huang Y, Chory J, Lipka V, Borevitz JO, Dangel JL, Bergelson J, Nordborg M, Weigel D (2010) Natural allelic variation underlying a major fitness trade-off in *Arabidopsis thaliana*. *Nature* 465:632–636
- Turner A, Beales J, Faure S, Dunford RP, Laurie DA (2005) The pseudo-response regulator *Ppd-H1* provides adaptation to photoperiod in barley. *Science* 310:1031–1034
- von Zitzewitz J, Cuesta-Marcos A, Condon F, Castro AJ, Chao S, Corey A, Filichkin T, Fisk SP, Gutierrez L, Haggard K, Karsai I, Muehlbauer GJ, Smith KP, Veisz O, Hayes PM (2011) The genetics of winterhardiness in barley: perspectives from genome-wide association mapping. *Plant Genome* 4:76–91
- VSN International (2011) *GenStat for Windows* (14th edn). VSN International, Hemel Hempstead (Web page: GenStat.co.uk)
- Wang M, Jiang N, Jia TY, Leach L, Cockram J, Comadran J, Shaw P, Waugh R, Luo ZW (2012) Genome-wide association mapping of agronomic and morphologic traits in highly structured populations of barley cultivars. *Theor Appl Genet* 124:233–246
- Waugh R, Jannink JL, Muehlbauer GJ, Ramsay L (2009) The emergence of whole genome association scans in barley. *Curr Opin Plant Biol* 12:218–222
- Waugh R, Marshall DF, Thomas WTB, Comadran J, Russell JR, Close T, Stein N, Hayes P, Muehlbauer G, Cockram J, O'Sullivan D, Mackay I, Flavell AJ, AGOUEB, BarleyCAP, Ramsay L (2010) Whole-genome association mapping in elite inbred crop varieties. *Genome* 53:967–972
- Wei F, Gobelman-Werner K, Morroll SM, Kurth J, Mao L, Wing R, Leister D, Schulze-Lefert P, Wise RP (1999) The *Mla* (powdery mildew) resistance cluster is associated with three NBS-LRR gene families and suppressed recombination within a 240-kb DNA interval on chromosome 5S (1HS) of barley. *Genetics* 153:1929–1948

- Yan L, Fu D, Li C, Blechl A, Tranquilli G, Bonafede M, Sanchez A, Valarik M, Yasuda S, Dubcovsky J (2006) The wheat and barley vernalization gene *VRN3FT*. *Proc Natl Acad Sci U S A* 103:19581–19586
- Yan J, Shah T, Warburton ML, Buckler ES, McMullen MD, Crouch J (2009) Genetic characterization and linkage disequilibrium estimation of a global maize collection using SNP markers. *PLoS One* 4:e8451
- Yu J, Pressoir G, Briggs WH, Bi IV, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Holland JB, Kresovich S, Buckler ES (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat Genet* 38:203–208
- Zhang LY, Marchand S, Tinker NA, Belzile F (2009) Population structure and linkage disequilibrium in barley assessed by DArT markers. *Theor Appl Genet* 119:43–52
- Zhang Z, Eroz E, Lai C-Q, Todhunter HK, Gore MA, Bradbury PJ, Yu J, Arnett DK, Ordovas JM, Buckler ES (2010) Mixed linear model approach adapted for genome-wide association studies. *Nat Genet* 42:355–360
- Zhou H, Muehlbauer G, Steffenson B (2012) Population structure and linkage disequilibrium in elite barley breeding germplasm from the United States. *J Zhejiang Univ Sci B* 13:438–451