Exploiting induced variation to dissect quantitative traits in barley


*Scottish Crop Research Institute, Invergowrie, Dundee DD2 5DA, Scotland, U.K., †Queensland Primary Industries and Fisheries, Hermitage Research Station, 604 Yangan Road, Warwick, QLD 4370, Australia, ‡Nordic Genetic Resource Center, Smedjevagen 3, P.O. Box 41, S-230 53, Alnarp, Sweden, §Department of Genetics, University of Silesia, Jagiellonska 28, Katowice 40-032, Poland, ¶Department of Agricultural and Environmental Sciences, University of Udine, Via delle Scienze, 208, 33100 Udine, Italy, †Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Corrensstrasse 3, 06466 Gatersleben, Germany, ‡KWS UK Ltd., 56 Church Street, Thriplow, Royston, Herts. SG8 7RE, U.K., and ††Applied Genomics Institute, University of Udine, Via J. Linusso 51, Udine 33100, Italy

Abstract
The identification of genes underlying complex quantitative traits such as grain yield by means of conventional genetic analysis (positional cloning) requires the development of several large mapping populations. However, it is possible that phenotypically related, but more extreme, allelic variants generated by mutational studies could provide a means for more efficient cloning of QTLs (quantitative trait loci). In barley (Hordeum vulgare), with the development of high-throughput genome analysis tools, efficient genome-wide identification of genetic loci harbouring mutant alleles has recently become possible. Genotypic data from NILs (near-isogenic lines) that carry induced or natural variants of genes that control aspects of plant development can be compared with the location of QTLs to potentially identify candidate genes for development-related traits such as grain yield. As yield itself can be divided into a number of allometric component traits such as tillers per plant, kernels per spike and kernel size, mutant alleles that both affect these traits and are located within the confidence intervals for major yield QTLs may represent extreme variants of the underlying genes. In addition, the development of detailed comparative genomic models based on the alignment of a high-density barley gene map with the rice and sorghum physical maps, has enabled an informed prioritization of ‘known function’ genes as candidates for both QTLs and induced mutant maps.

Quantitative versus Mendelian traits
Variation observed between individuals within a species can be categorized as phenotypic traits. Genetic analysis of such traits usually involves crossing two individuals with contrasting phenotypes to generate a hybrid followed by analysis of their segregation in subsequent generations [e.g. DH (doubled haploid), F2 or F3]. If two or three discrete phenotypic classes can be distinguished in the progeny, the trait is classed as exhibiting qualitative inheritance and is usually called a binary or Mendelian trait. If no obvious segregation classes are observed, the inheritance pattern is classed as quantitative. For both types of trait, by associating phenotypic segregation data with mapped genome-wide DNA polymorphisms, the genetic locus or loci that confer the phenotypic variation can be identified. The advantage of quantitative analysis is that more information can be gathered from a single QTL (quantitative trait locus)-mapping experiment, as opposed to Mendelian trait mapping where variance is attributed to a single locus. If cloning the gene underlying a quantitative trait is the objective, then considerably larger populations in which the target QTL has been effectively transformed into a Mendelian character are required. Extensive genotyping and phenotyping are required to compensate for residual variance and to achieve a satisfactory confidence level for the identification of candidate genes. Given that the population size required to clone a Mendelian trait in barley is already large and poses handling challenges [1] (Figure 1), cloning a QTL with low heritability would be very costly. Dealing with quantitative traits also adds complexity to the experimental designs needed to achieve satisfactory mapping results [2]. Accurate quantitative measurements need to be recorded for each individual in the population as opposed to binary characters which can often be scored visually. Understandably then, transforming QTLs into Mendelian traits by making an appropriate series of crosses and employing marker-assisted selection, is currently the major strategy to enhance QTL cloning [3].

---

**Key words:** barley, induced mutant, near-isogenic line, positional cloning, quantitative trait locus (QTL).

**Abbreviations used:** BAC, bacterial artificial chromosome; DH, doubled haploid; NIL, near-isogenic line; QTL, quantitative trait locus; SNP, single nucleotide polymorphism.

1 To whom correspondence should be addressed (email Arnis.Druka@scri.ac.uk).

2 Present address: School of Biological Sciences, University of Aberdeen, Aberdeen AB2 4UW, Scotland, U.K.
Estimates of population size needed to clone a gene from the 5000 Mb barley genome

Figure 1 | Estimates of population size needed to clone a gene from the 5000 Mb barley genome

Plotting the probability (y-axis) of finding in n gametes (x-axis) a minimum of two crossovers (one of each side of the target gene) at a physical distance <1 [1] Two T values, 150 and 250 kb, were selected based on the average BAC insert size of 107 kb [41]. The grey box indicates the population size range for these two values to achieve a 90% probability.

Genetic stocks

Genetic studies in barley started over a century ago [4], with induced mutations reported over 80 years ago. Indeed, some of the earliest descriptions of induced mutational events were described in barley [5]. Since then, thousands of morphological and physiological mutants of barley have been identified worldwide [6]. A very small proportion have been cloned and those that were, in the pre-genomics era, were either inferred from homology with genes from other species or based on functional screens [7–9]. One such screen performed during the 1970s and 1980s at the Carlsberg Laboratory, Copenhagen, Denmark, and Washington State University, Pullman, WA, U.S.A., identified mutations that affected the formation of proanthocyanidins [10]. The screen involved phenotyping approx. 18.5 million M2 plants. Subsequently, using HPLC, 560 proanthocyanidin-free or ant mutants were identified. For example, the proanthocyanidin-free mutation, ant18, was found to lack leucocyanidin accumulation, whereas several other secondary metabolites upstream in the pathway were present [7] (Figure 2A). Based on this, dihydroflavonol 4-reductase was inferred and later confirmed as a causal gene of the ant18 phenotype [7,11]. Although clearly very powerful, this approach has limitations when considering mutations in regulatory genes or genes where the resulting molecular phenotypes are unknown.

To advance genetic mapping, and to compare mutations held in various different collections, one of us (J.F.), in parallel with several other individuals around the world, initiated a programme of back-crossing a wide spectrum of mutant alleles into a common genetic background. The mutants had a wide range of phenotypic variants (Figure 2B) and either were induced by different mutagens or occurred naturally. This effort generated a large number of back-cross-derived lines having the two-rowed spring cultivar ‘Bowman’ as the recurrent parent. In Figure 2(A), the ant18 mutant described above is an example of how these lines were developed. As a wide and representative sample of mutations was introgressed into the Bowman background, the resulting lines provide a powerful resource for both mutant characterization and gene identification, with each back-cross line retaining a segment of DNA from its donor parent that harbours the mutant allele.

Genomics tools for barley

Over the last 20 years, advances in molecular marker technology have made mapping quantitative traits more feasible [12]. One of the first populations generated for QTL mapping in barley was a DH recombinant line population from a cross of two distantly related six-rowed North American cultivars, Steptoe (high yield) and Morex (malt quality) [13]. The 150 DH lines of the Steptoe/Morex (St/Mx) population were used to construct a barley genetic linkage map and map QTLs for yield, malting quality and several biochemical, morphological and disease reaction genes [14]. Using several different molecular-marker technologies, other barley genetic linkage maps were also constructed to map QTLs and Mendelian loci. Despite this, to date, relatively few genes have been cloned by forward genetics approaches [15–23]. This is partially because genetic marker technologies used were neither of sufficiently high-throughput nor low cost and the mapping information generated from separate experiments was difficult to combine.

However, since 2001, several genomics tools and resources have been developed for barley and placed in the public domain, and these are enhancing projects focused on the isolation of barley genes. They include barley EST (expressed sequenced tag) libraries, comprising over 500,000 sequences [24], a 6.3 × genome equivalent BAC (bacterial artificial chromosome) library [25], a barley1 GeneChip (Affymetrix) for mRNA expression analysis [26], and a gene-based SNP (single nucleotide polymorphism) genotyping platform [27,28]. Although the barley1 GeneChip has been extensively used in classical expression-profiling experiments, it has also been used for gene mapping [29–33] and cloning [34,35]. The gene-based SNP genotyping platform was developed using Beadarray technology (from Illumina) and its use has helped to construct a barley consensus SNP-based map [24,26], which has subsequently been used for association mapping [27,28] and mapping mutants that had been back-crossed to generate NILs (near-isogenic lines) [36]. Raw datasets consisting of tens of millions of data points can be generated in a single experiment using either GeneChip or Beadarray technologies (Figure 3) that can be interrogated using various statistical algorithms.

Connecting quantitative and qualitative variation

As the genes used for the development of the genotyping and gene-expression platforms are complementary, but with a significant overlap, the data generated from analysing common genetic material using both platforms can easily be integrated [30,32,33] (Figure 3). For example, both gene-expression-based markers [31] and SNPs [26] have been used...
Figure 2 | The principle of the back-cross-based gene mapping

(A) Generation of the NIL BW018 that carries recessive mutation in the dihydroflavonol 4-reductase (Dfr) gene causing the ant18 phenotype, white stem base, and the absence of the purple vanillin HCl staining of the pericarp in developing kernels [10]. Crossing of the ant18 mutation, which was initially induced in the cultivar ‘Nordal’, seven times to Bowman resulted in a back-cross 6 (BC6) line, BW018, which carried on chromosome 3H a single, relatively small introgression from the donor parent that includes the original mutation. Black regions on chromosomes show the Bowman genotype, white ones show the Nordal genotype, the star indicates the approximate Dfr position. Note that Bowman was the male parent until the fourth back-cross. (B) The representation of different phenotype classes in the Bowman line population and some of the representative mutant groups from each class [36].

<table>
<thead>
<tr>
<th>Phenotype class</th>
<th>Number of lines</th>
<th>Representative groups</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spike morphology</td>
<td>169</td>
<td>dense spike, laxatum</td>
</tr>
<tr>
<td>Plant pigmentation</td>
<td>142</td>
<td>proanthocyanidin-free, chlorina</td>
</tr>
<tr>
<td>Spikelet fertility</td>
<td>130</td>
<td>desynapsis, male sterile</td>
</tr>
<tr>
<td>Spikelet morphology</td>
<td>116</td>
<td>breviarsatum, brachytic</td>
</tr>
<tr>
<td>Disease reaction</td>
<td>97</td>
<td>reaction to Puccinia hordel</td>
</tr>
<tr>
<td>Surface wax</td>
<td>93</td>
<td>eceriferum, waxy spike</td>
</tr>
<tr>
<td>Plant height</td>
<td>83</td>
<td>semi-dwarf, gigas</td>
</tr>
<tr>
<td>Leaf morphology</td>
<td>49</td>
<td>curly, eligulum</td>
</tr>
<tr>
<td>Kernel shape</td>
<td>32</td>
<td>globosum, shrunken endosperm</td>
</tr>
<tr>
<td>Maturation</td>
<td>26</td>
<td>early maturity, lazy dwarf</td>
</tr>
<tr>
<td>Growth pattern</td>
<td>24</td>
<td>fragile stem, viviparoides</td>
</tr>
<tr>
<td>Bud development</td>
<td>17</td>
<td>uniculum, absant lower laterites</td>
</tr>
<tr>
<td>Waxy endosperm</td>
<td>1</td>
<td>waxy</td>
</tr>
</tbody>
</table>

to construct an integrated gene map of barley (Figure 3). The sequences of genes from this integrated map can then be used for similarity searches in genomic models such as rice, sorghum and Brachypodium to identify chromosomal blocks where conservation of synteny is high. Thus this integrated gene map is one of the key components connecting genotypic information from classical genetic stocks and QTL mapping experiments. In addition, it has provided the information to
Figure 3 | Barley genetic and genomic data integration framework
Gene-based platform technologies are central for connecting different classical genetic stocks to the sequenced genomes. Thus 4608 SNPs were used to genotype 335 DH lines representing three mapping populations to construct a consensus map [24]. A subset of 1536 SNPs was used to genotype the NIL population, enabling straightforward mapping of the introgressions. A subset of St/Mx SNPs together with expression data can be used to complement the consensus map with transcript-derived markers (TDMs) [33], resulting in the integrated barley gene map consisting of approx. 5000 genes. This set of genes then was used for a homologue search in the sequenced genomes and consecutively for identification of the synteny blocks.

build an accurate synteny model between the barley genetic map and available physical maps and genome sequences of related genomic model plants.
High-throughput genotyping of sufficient NILs, which are mutant-allele-containing, can result in the identification of chromosomal segments that are distributed across the barley genome [36]. Collectively, these NILs could represent a wide range of phenotypic variants (Figure 2B). Our hypothesis is that locational coincidence between mutant alleles and components of phenotypically equivalent quantitative traits may indicate that the mutant allele is an extreme variant of the gene underlying the QTL. Thus, for a particular QTL of interest, a subset of mutants mapping within the confidence limits of the QTL could be selected, and if a possible functional relationship is identified between mutant and trait phenotypes, the mutant allele adopted as a Mendelian trait for QTL gene isolation. As an example, yield can be divided into component traits such as number of tillers per plant, number of kernels per tiller and the size of individual kernels. Mutants that map within the vicinity of the QTL and that affect branching of the vegetative or inflorescence meristems, kernel size, plant height, spike length, spikelet density, flowering time, lodging or fertility are clear candidates that can be prioritized for gene identification by positional cloning by exploiting a simplified strategy afforded by Mendelian genetics. Furthermore, candidate genes can be predicted based on equivalent phenotypes and conservation of synteny with genes that have been functionally characterized in model species. Of course, once candidates have been identified, their barley orthologues still need to be cloned and sequenced. By aligning DNA sequences from mutant lines with those from the original accessions in which the
mutations were discovered, disruptive polymorphisms such as non-synonymous or nonsense mutations strongly suggest a link between gene and phenotype. This link can be strengthened to the point of proof if multiple independent phenotypic mutants are found to possess independent lesions in the same candidate gene. The DNA sequence of this gene now becomes the template for allele mining and characterization of the QTL.

Summary
Currently available barley genomics tools and resources, recently generated molecular marker datasets, comparative genetics and valuable biological materials in the form of genetically characterized mutants have great potential for simplifying the process of identifying the genes that underlie QTL. If the genomic region containing a mutant allele has already been genetically defined by marker analysis, PCR-based re-sequencing of the entire gene complement of both mutants and wild-types may alleviate some of the current bottlenecks in positional cloning. However, as barley genome sequencing is currently well underway [37–40], availability of the known barley gene content of a target region will further enable gene identification. Developing efficient solutions to provide experimental proof that a gene underlying a QTL is the same as that inferred from an extreme allelic variant remains a high priority.

Acknowledgements
The Bowman isolines referred to were developed at the Department of Plant Sciences, North Dakota State University, Fargo, ND 58108, U.S.A.

References
5 Stadler, L.J. (1928) Mutations in barley induced by X-rays and radium. Science 68, 186–187