Linkage disequilibrium in barley

L. Ramsay, K. Caldwell, J. Russell, M. Macaulay, W.T.B. Thomas, W. Powell & R. Waugh

Linkage Disequilibrium (LD) is the non-random association of alleles at different loci. Such a correlation of allelic states of loci in different parts of the genome is usually caused by the physical proximity of the loci, hence the use of the term linkage disequilibrium although LD can be caused by other factors such as population substructure and selection. LD is of interest as it relates directly to the underlying patterns of polymorphism and therefore affects at a basic level the reliability of diagnostics and the general application of molecular markers to the characterisation of germplasm. Due to availability of molecular markers there has been a renewed interest in LD studies, particularly in humans, that utilize the pattern of ancestral recombination to delineate candidate genomic regions associated with traits of interest such as predisposition to disease etc. In plants such studies have to date been limited and have mainly been carried out in outbreeders such as maize that show similar patterns of LD to those found in human and animal studies. In the Genome Dynamics programme we have been interested in whether such approaches are applicable to barley as such an inbreeding species would be expected to show higher levels of spurious associations due to the relative lack of recombination and the resultant population structure.

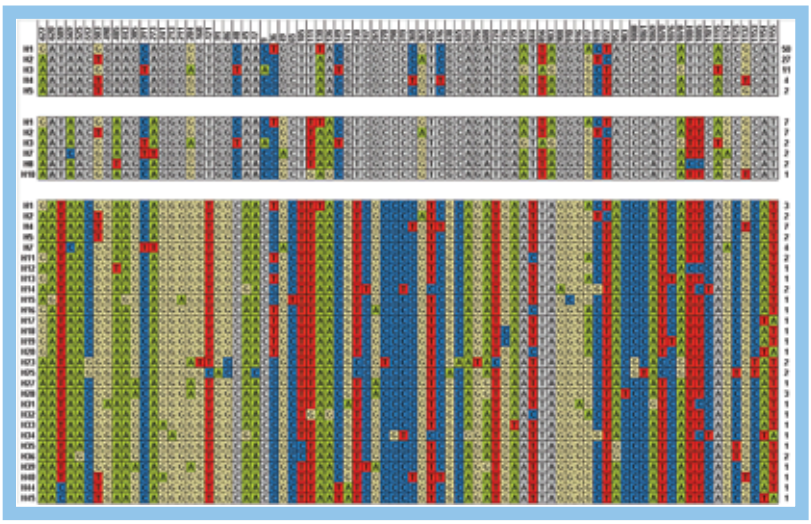

Figure 1 Sequence alignment showing polymorphism at the *Ha* locus

In germplasm surveys of barley cultivars using SSRs (simple sequence repeats) we found significant LD across the whole genome, but that closely linked loci showed significantly higher levels. This pattern was complicated by population structure that could be attributed to past breeding practice in maintaining the distinctness of differing cultivar types. By concentrating the analysis on a single cultivar type (elite two-row spring barleys) the severity of the effect of population structure was considerably reduced though not eliminated. In this elite germplasm the LD observed was maintained between loci up to 5cM apart, considerably further than that found in maize. An encouraging conclusion from these studies was that LD studies are feasible in cultivated barley although care is needed in choice of material and that the use of unlinked marker loci is necessary to estimate population substructure and give empirical thresholds for the appropriate levels of statistical significance.

A detailed investigation at the sequence level of the hardness locus (*Ha*) at the distal end of 5HS has also given considerable insights into LD in barley. Despite a substantial amount of diversity within wild barley, *Hordeum spontaneum*, where over 40 different combinations of polymorphisms or haplotypes were observed, a highly structured pattern of nucleotide diversity across the candidate genes revealed a limited number (4-5) for cultivated lines (Figure 1). Furthermore, the extent and magnitude of LD among these different sample sets revealed radical differences in the maintenance of LD between cultivated and wild barley at the sequence level. In cultivated barley, strong associations were found to extend across the entire 112 kb genic region; however, in wild barley these associations were relatively weak and rarely extended across a given gene. Such differences could prove to be invaluable as appropriate selection of germplasm could direct association studies towards both the location of genes involved in complex phenotypes and the discovery of causative mutations.