Breeding facilitates whole genome LD-mapping

Luke Ramsay, Nils Rostoks, David Marshall, Sharon Mudie, Timothy J. Close & Robbie Waugh

In plants, mapping traits and isolating the underlying genes has largely been based on following the inheritance of molecular markers in specific experimental populations derived from crosses between two parents that contrast for the trait under study. However, a more general population-based approach termed 'association genetics', or 'linkage disequilibrium (LD) mapping' (based on the non-random association of alleles in a population), is now being routinely used to map disease genes in humans. In crop plants, the potential of association mapping, with the objective of estimating the position of genes conferring a specific trait by exploiting LD between alleles of genetically mapped markers, has become a focus of considerable interest. A major attraction of LD-mapping is the potential to locate genes responsible for a wide range of traits in a sample population using pre-existing trait data collected during crop improvement and cultivar registration programmes. However, to be successful, the extent of LD must be known in the gene pool under study, and based on this, an appropriate number of molecular markers assembled into a format suitable for high throughput genotypic analysis.

Over the last several years we have directly assessed both the extent of LD, and the patterns of genetic diversity within elite cultivated barley (Fig. 1). Our observations allow us to predict that around one polymorphic marker per centi-Morgan (cM) will be necessary for whole genome LD-scans. Over the same period we also set out to develop a genotyping platform that would simplify genetic analyses in barley and facilitate the LD-mapping approach. Using expressed sequence tags (ESTs) as a template, we identified 1524 barley single nucleotide polymorphisms (SNPs) and used them to produce a pilot oligonucleotide pool array

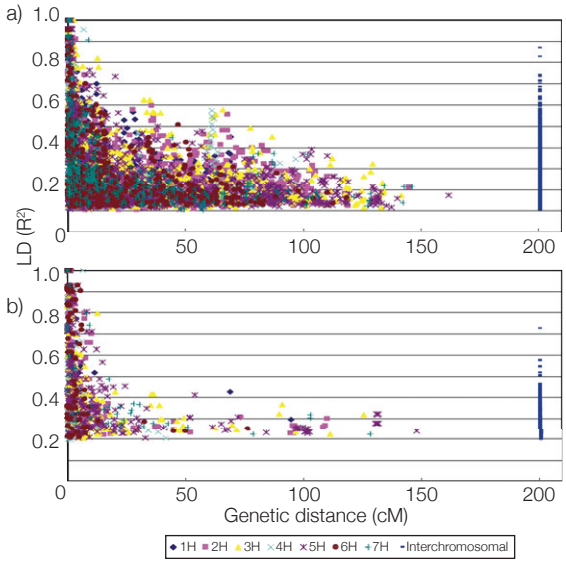
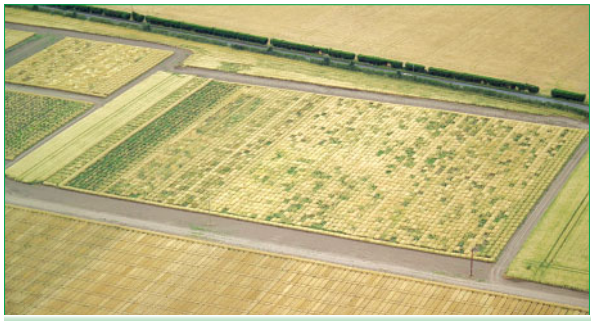


Figure 1 Plots of LD against genetic distance for all European barley lines (a) and spring varieties (b)

(OPA) for use with 'Illumina' GoldenGate Bead-Array technology (a gel-free parallel genotyping technology developed alongside the human genome sequencing and HapMap projects). We used the pilot OPA to genotype three bi-parental doubled haploid mapping populations and 102 barley cultivars. 91% of the SNP assays were successful, confirming the utility of the technology for high throughput genotypic analysis in barley. 1029 of the 1391 polymorphic SNPs segregated in at least one of the three populations with over 240 in any two cross comparisons and 146 in all three. We constructed linkage maps of each population and then developed a high quality integrated map comprising all 1029 SNPs.



We then used the SNP data from the 102 genotypes to study genetic diversity, population structure and the extent of LD. We observed three major subgroups within the germplasm, European spring and winter

material (n=91) and more exotic lines. Highly significant intra-chromosomal LD ($p>0.001$, $r^2>0.5$) extended over more than 60 cM (mean 3.9 cM, median 1.16 cM) in the combined European barley set. 20.4% of all significant ($p>0.001$) associations ($r^2>0.05$) were inter-chromosomal, demonstrating the impact of population sub-structure. In the spring 2-row subset (n=53), LD extended only up to 15 cM (mean 1.53 cM, median 0.8 cM) and the proportion of inter-chromosomal associations was reduced to 2%. To test whether we could use LD to locate genes we examined whether we could correctly position any of the 362 unmapped genes in our pilot OPA dataset via an LD-mapping approach. We calculated pair-wise LD (r^2) for each of the unmapped and mapped loci and assumed that strong LD indicated linkage. We were able to assign a putative map location to over 50% of the unmapped loci. Theoretical predictions suggest that LD in barley, an inbreeder, should persist over very long distances and generate many false positives in LD-mapping studies. However, we show that LD has broken down sufficiently in elite cultivated lines, probably due to repeated crossing by breeders, to allow LD-mapping to be carried out successfully.