Unprecedented Genetic Diversity of Barley Epiheterodendrin

Joanne Russell, Allan Booth, Ingo Hein, Sandie Williamson, Jenny Morris, Wayne Powell and Pete Hedley

Scottish Crop Research Institute, Dundee and NIAB, Cambridge (WP)

Understanding genetic diversity

Over the last decade substantial effort has gone into the development of different technologies for the generation of extensive genomic resources: genome sequencing, vast EST collections, microarray systems, large-insert libraries, genetic and physical maps, and structured mutant populations are becoming increasingly available. Coupled with increased efforts in bioinformatics, these tools provide the necessary framework for investigating patterns and levels of genetic diversity present between and within different biological systems. For crop species, the challenge is to understand the genetic basis of variation and patterns of nucleotide diversity underlying traits important in crop improvement, and to identify polymorphisms responsible for these traits.

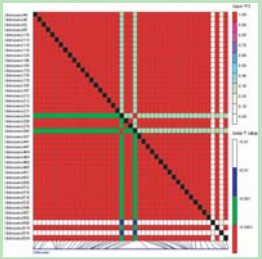
Linkage disequilibrium (LD) mapping of candidate gene associations is an emerging approach for identifying genes underlying phenotypic variation. By examining nucleotide polymorphism across many individuals, we can infer associations between phenotype and genotype, on the basis of the evolutionary process which has shaped the haplotype structure of the species in question. In a previous study, we have examined patterns of diversity and LD across the hardness locus at the distal end of chromosome 5H in barley (Caldwell et al., 2005) and, although this was an important step towards designing effective association mapping strategies, it did not allow us to identify the polymorphisms associated with the phenotypic trait. This current study aims to address this question of associating genotype with phenotype by detailed analysis of diversity at the epiheterodendrin (EPH) locus in barley.

Molecular diversity of EPH-associated P450 genes

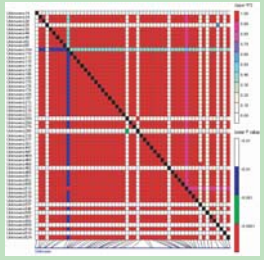
In order to determine the levels of diversity and whether there was any association with the EPH producer and non-producer phenotype, oligonucleotide primer sets were designed to re-amplify regions of the cyp79 gene in phenotypically characterised barley cultivars (25 producers and 10 non-producers). In a total of 860bp sequenced across cyp79, 45 SNPs were detected, which were organised into five different haplotypes, four representing producers (green) and only one haplotype representing non-producers (red). When the sample was increased to include landrace and wild barley (Hordeum spontaneum) accessions from across a geographical range, the number of SNPs detected increased to 60 and, despite the increase in the number of haplotypes, the organisation into producer and non-producer haplotypes was maintained. Of these 60 SNPs, 45 give perfect separation into producers and non-producers, and of the remaining 15 polymorphisms, all segregate exclusively within producers, or exclusively within non-producers. No site which is polymorphic is shared between phenotypes.

Across this region of the cyp79 gene the level of diversity is ten times greater than previously reported studies in both cultivated and wild accessions, with values more commonly associated with highly outbreeding species such as maize. By estimating Tajima's D statistic we can determine whether the diversity observed is solely influenced by mutation and drift, or by selection. The significant Tajima's D for this region (2.638 for cultivated and 2.526 for wild barley) is due entirely to the difference between producers and non-producers. Tajima's D within producers is -0.63 and within non-producers 0.79, both non-significant. This pattern is consistent with the divergence between producers and non-producers being very old, with the two groups evolving independently of each other.

The majority of SNPs are in linkage disequilibrium (LD), establishing the basis for an association between haplotype and phenotype across this region. Pairwise comparison between SNPs across cyp79 is represented in a LD matrix, where the upper triangle represents LD measured by r2, and the lower triangle those pairwise comparisons that are significant. Similar patterns of LD were also found in the wild barley accessions and this is in contrast to what we have observed in our previous studies (Caldwell et al., 2005), providing further evidence that these are ancient and have been maintained following domestication.

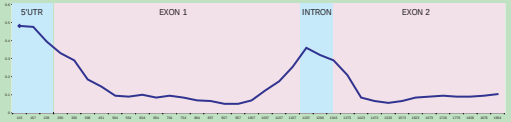


SWRI



H spontaneum

Due to the extraordinarily high levels of polymorphism, it was difficult to re-amplify across the full-length gene in non-producers and therefore we have employed a gene walking strategy to sequence the complete gene. The level of DNA divergence between producer and non-producer is unprecedented in barley. Within 280bp of 5' untranslated region (UTR) and the 122 bp intron, almost half of the nucleotide sequence has diverged and, although this value is somewhat lower in the two exons, it is still very high for barley.



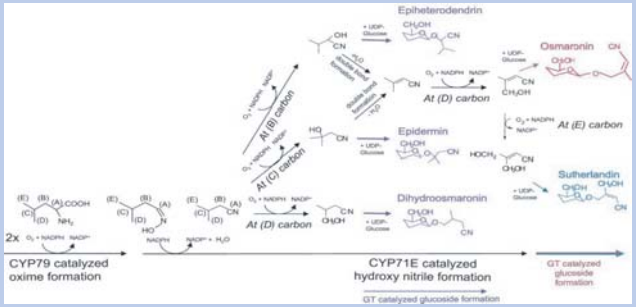
Haplotype analysis of cyp79 reveals a perfect correlation with producer and non-producer phenotype in the cultivated genepool of barley. This unbreakable association was also observed in wild barley accessions, indicating a direct involvement of cyp79 in determining the EPH phenotype. Future work will involve sequencing full-length alleles in both producers and non-producers and utilizing gene silencing and transgenic approaches to unambiguously link these genes with EPH production.

Cyanogenesis in barley: a problem for Scotch Whisky

Many plant species produce cyanogenic glucosides which, through release of the breakdown product hydrogen cyanide (HCN), have been implicated in the natural defence response to herbivores and pathogens. These compounds are also believed to have a role in nitrogen storage and osmoregulation. In certain cultivars of barley a leucine-derived cyanogenic glucoside, epiheterodendrin (EPH), is present at high levels in young seedling (malt) leaf tissue. Fermentation of malted barley leads to hydrolysis of EPH, via action of yeast-derived β-glucosidase and subsequent heating during distillation, to form HCN. A reaction within the distillate between HCN and ethanol, in the presence of copper and oxygen, leads to trace but significant levels of the potentially carcinogenic compound ethyl carbamate. Low ethyl carbamate varieties of barley are therefore high priority for the Scotch Whisky industry, indicating a necessary requirement for both detailed biochemical and molecular characterisation of EPH pathways, and development of an unambiguous marker for varietal selection by malting barley breeders.

Epiheterodendrin biosynthetic pathway

Cytochrome P450 enzymes form the largest family of plant proteins, with over 1,000 members identified, catalysing a wide array of both simple and complex reactions, generating a diverse range of natural plant products. EPH has been proposed to be derived from the amino acid leucine through action of two cytochrome P450 enzymes, cyp79 and cyp71E (Nielsen et al., 2002). By screening the HarvEST barley database, which contains around 400,000 expressed sequence tags (ESTs), we identified a small number of barley genes whose derived amino acid sequences showed significant homologies to previously characterised cyp79 and cyp71E proteins. Using these barley homologues we were able isolate and characterise full-length P450 gene sequences and provide molecular evidence for their implication in EPH production.



Characterisation of EPH genes from barley

Barley ESTs with homologies to cyp79 and cyp71 were identified, representing partial 3' cDNA sequences, and the corresponding gene fragments were subsequently isolated using PCR. These gene probes were used to screen a barley (cv. Morex) bacterial artificial chromosome (BAC) library to isolate full-length gene sequences. Fingerprinting of the BAC clones, utilising SNaPshot-based fluorescent labeling of restriction-digested fragments, clustered the clones into two distinct groups. Single BAC clones for each gene were subcloned and sequenced, revealing full-length genes encoding cyp79 and cyp71, which contain two and three exons respectively and also indicate the presence of putative retroelement sequences, which are common features of the barley genome.

Using polymorphic barley populations, both genes genetically mapped to the same position, which was also confirmed as the Eph locus on chromosome 1H, first identified by Swanston et al. (1999).

Expression of cyp79 and cyp71 genes were determined by real-time RT-PCR. Both genes were clearly up-regulated in EPH producers compared to non-producers, and showed similar patterns of temporal decrease in expression levels between 3 day and 4 day old seedling leaf material.

Development of EPH molecular marker

Central to selection of barley varieties with low levels of potential ethyl carbamate generation is the development of a robust unambiguous molecular marker. Previously at SCRI, a simple-sequence repeat (SSR) marker (BMAC213) was generated which showed good, but not complete, association with the EPH phenotype, as determined by the biochemical assay (Swanston et al., 1999). The SNP haplotype identified for the cyp79 gene shows complete (100%) association with the EPH phenotype, clearly distinguishing producers from non-producers and was therefore used as the basis for development of a multiplex PCR-based screen. This assay is currently being used for efficient, reliable and cost-effective marker-assisted varietal selection in malting barley breeding programmes.

Acknowledgements

This work was funded through SEERAD. Special thanks to Stuart Swanston (SCRI), Bill Thomas (SCRI) and James Brosnan (SWRI) for useful discussions.

References

Caldwell et al., 2005. Genetics, 10: epub;
Nielsen et al., 2002. Plant Physiology, 129: 1066-1075;
Swanston et al., 1999. Molecular Breeding, 5: 103-109