SINGLE NUCLEOTIDE POLYMORPHISM DISCOVERY AND MAPPING IN BARLEY

Nils Rostoks, Sharon Mudie, Nicola Bonar, Linda Cardle, Joanne Russell, Luke Ramsay, David F. Marshall, Robbie Waugh

Scottish Crop Research Institute, Invergowrie, Dundee, DD2 5DA, Scotland, UK

Project description

Single Nucleotide Polymorphism (SNP) discovery is currently underway at the SCRI aiming at identification of SNP in up to 2000 genes associated with response to abiotic stresses with a focus on root-specific genes. Two barley EST unigene lists were compiled (>3000 non-redundant contigs) primarily based on microarray studies using Affymetrix Barley1 GeneChip (Close et al. 2004). We re-sequenced fragments of ca. 1400 genes from a core set of 8 cultivated and wild barley lines for SNP discovery (Oregon Wolfe Barleys, Steptoe, Morex, Lina, HS92, Golden Promise and Optic). A subset of 96 gene fragments was re-sequenced in additional 16 Western European elite spring and winter barley lines. The average polymorphism (SNP and indel) frequency was 1 per 113 bp for the core set and 1 per 90 bp for combined core and extended set of genotypes. 90% of polymorphisms and haplotypes were present already in the core set confirming it as a representative source for polymorphism discovery. On average, we observed 2.97 haplotypes per contig in the set of 24 genotypes, while the 21 cultivated lines displayed 2.1 haplotype per contig. To facilitate association of known barley traits with candidate genes, we are using SNPs as markers for linkage mapping. Here we present a barley linkage map with ca 300 sequence-based markers derived from the 3 doubled haploid mapping populations. The sequence and linkage mapping data were used to explore barley / rice synteny.

For up-to-date information on our SNP discovery and mapping project visit the SCRI Bioinformatics Web site:
http://bioinf.scri.sari.ac.uk/cgi-bin/barleysnp/barley_SNPs

SNP discovery

SNP discovery consisted of re-sequencing 3' regions of selected HarvEST assembly 21 contigs (http://harvest.ucr.edu). Primers have been designed for 1338 HarvEST contigs by a semi-automated procedure using Primer3.

PCR amplification was carried out on genomic DNA from core genotypes and the same primers were used in sequencing to generate >15000 sequencing reads (>5.5 MB) with phred quality >40. Sequences were assembled and mutations were called using Mutation Surveyor software (http://www.softgenetics.com). Analysis of the full data set is in progress.

Genetic Mapping

Oregon Wolfe Barley (Costa et al. 2001), Steptoe x Morex (Kleinhofs et al. 1993) and Lina x HS92 (Ramsay et al. 2000) doubled-haploid populations were used for linkage mapping. Sequence polymorphism detection in DH progeny was achieved by a variety of techniques, e.g., restriction enzyme or Cel I digestion, pyrosequencing, DHPLC, sequencing, etc.

SCRI polymorphism database

Polymorphism database contains pre-loaded sequences of all the HarvEST contigs, while other information, such as primer sequences, PCR product sequences, polymorphism reports and genetic mapping data, can be uploaded through an intuitive Web interface.

The database structure allows it to be used for generic project management of sequence-based polymorphic markers, such as SNPs, SSRs and other comparable data.

The access to the database will be available through the SCRI Bioinformatics web site in the near future.

International collaborations

The abiotic stress gene list was generated in collaboration with Dr. TJ Close (UC Riverside, CA, USA) and largely consists of unpublished microarray data from Dr. Close's lab. The genes for which SNP and linkage mapping data is generated at SCRI will also be prioritized for linking to BAC physical map within NSF funded project to Dr. Close.

Working lists of genes are also coordinated with other labs involved in SNP analysis and mapping to avoid duplicated efforts, in particular with Dr. A Graner and Dr. N Stein (IPK, Gatersleben, Germany) and with Dr. Kazuhiro Sato (Okayama University, Kurashiki, Japan).

References

Close et al. (2004) A new resource for cereal genomics: 22K barley GeneChip comes of age. Plant Physiol 134:960-968

Costa et al. (2001) Molecular mapping of the Oregon Wolfe Barleys: a phenotypically polymorphic doubled-haploid population. Theor Appl Genet 103 415-424

Kleinhofs et al. (1993) A molecular, isozyme and morphological map of the barley (Hordeum vulgare) genome. Theor Appl Genet 86:705-712

Kleinhofs and Graner (2001) An integrated map of the barley genome. In: Phillips RL, Vasil IK (eds) DNA-based markers in plants, 2nd edn. Kluwer Academic Publishers, Dordrecht, The Netherlands, pp 187-199

Ramsay et al. (2000) A simple sequence repeat-based linkage map of barley. Genetics 156:1997-2005

Rostoks et al. (2004) Single nucleotide polymorphism mapping of the barley genes involved in abiotic stresses. 9th International Barley Genetics Symposium, Brno, Czech Republic (http://www.ibgs.cz)

Acknowledgments

Funding is provided by BBSRC and SEERAD to RW.

Generous sharing of unpublished barley microarray data by Dr. TJ Close is gratefully acknowledged.

Thanks are due to Dr. P Hedley for help with barley microarray experiment, I Druka for help with microarray analysis, Dr. H Liu for providing the root tissue samples and Dr. WTB Thomas for the list of elite spring and winter barleys

Barley genotypes

A core set consisting of 8 genotypes represented diverse wild and cultivated barley lines from Western Europe and North America which have been used to create genetics and genomics resources for barley, such as 3 mapping populations, BAC library and EST sequences (see table below; parents of mapping populations used in this study are highlighted).

An extended set added 16 elite Western European barley cultivars representing both spring and winter varieties that have dominated the UK barley harvest during past 10 years.

Barley line	Genotype set	Species	Growth habit	Ear
OWB D	Core	H. vulgare ssp. spontaneum (?)	Spring	2 row
OWB R	Core	H. vulgare ssp. vulgare	Spring	6 row
Steptoe	Core	H. vulgare ssp. vulgare	Spring	6 row
Morex	Core	H. vulgare ssp. vulgare	Spring	6 row
Lina	Core	H. vulgare ssp. vulgare	Spring	2 row
HS92	Core	H. vulgare ssp. spontaneum	Winter	2 row
Golden Promise	Core	H. vulgare ssp. vulgare	Spring	2 row
Optic	Core	H. vulgare ssp. vulgare	Spring	2 row
Barke	Extended	H. vulgare ssp. vulgare	Spring	2 row
Triumph	Extended	H. vulgare ssp. vulgare	Spring	2 row
Chariot	Extended	H. vulgare ssp. vulgare	Spring	2 row
Atem	Extended	H. vulgare ssp. vulgare	Spring	2 row
Blenheim	Extended	H. vulgare ssp. vulgare	Spring	2 row
Prisma	Extended	H. vulgare ssp. vulgare	Spring	2 row
Camargue	Extended	H. vulgare ssp. vulgare	Spring	2 row
Igri	Extended	H. vulgare ssp. vulgare	Winter	2 row
Pastoral	Extended	H. vulgare ssp. vulgare	Winter	2 row
Marinka	Extended	H. vulgare ssp. vulgare	Winter	2 row
Halcyon	Extended	H. vulgare ssp. vulgare	Winter	2 row
Regina	Extended	H. vulgare ssp. vulgare	Winter	2 row
Panda	Extended	H. vulgare ssp. vulgare	Winter	2 row
Fighter	Extended	H. vulgare ssp. vulgare	Winter	2 row
Puffin	Extended	H. vulgare ssp. vulgare	Winter	2 row
Franka	Extended	H. vulgare ssp. vulgare	Winter	6 row

Barley genes

A 2508 member gene list was assembled in collaboration with Dr. TJ Close (UC Riverside, CA, USA) and consists primarily from genes differentially expressed during drought, salt and low temperature stresses in barley seedlings. It also includes barley homologues of stress-responsive genes in other plant species (Rostoks et al. 2004). The list is available at
http://bioinf.scri.sari.ac.uk/cgi-bin/barleysnp/2508_contig_list

Additionally, 862 differentially expressed genes were identified in barley roots during drought, salt, low nitrogen and waterlog stresses at SCRI using Affymetrix barley microarray (Rostoks et al. 2004). The list is available at
http://bioinf.scri.sari.ac.uk/cgi-bin/barleysnp/862_contig_list

Preliminary SNP analysis in the core set

Preliminary sequence analysis was carried out on a 189 contig subset in the core set of genotypes (Rostoks et al. 2004).

Average number of polymorphisms	1 per 113 bp
Number of haplotypes per contig	2.61 per contig

SNP analysis in the core and extended set

A set of 92 contigs was analyzed in additional 16 genotypes representing Western European cultivated germplasm (extended set). 90% of polymorphisms and haplotypes were present in the core set confirming it as a rich and representative source for sequence polymorphism discovery in barley.

Average number of polymorphisms	1 per 90 bp
Average number of haplotypes	2.97 per contig
Number of haplotypes in cultivated barley (excluding HS92 and OWB)	2.1 per contig
Number of haplotypes in HS92 and OWB	1.99 per contig
Number of haplotypes in spring barley (excluding HS92 and OWB)	1.87 per contig
Number of haplotypes in winter barley (excluding HS92 and OWB)	1.64 per contig

Integrated linkage map of 301 barley sequence-based loci (SNPs, SSRs, indels). Mapping was done in Oregon Wolfe Barley Dominant x Recessive, Steptoe x Morex minimapper and Lina x HS92 DH populations. Linkage maps for individual populations and the integrated map shown below was calculated using JoinMap 3.0. Calculations included 1026 loci, however, only 301 SNP loci and 105 bin markers (Kleinhofs and Graner, 2001; shown in bold) are shown. Names of the loci consist of the HarvEST assembly 21 contig number preceded by letters indicating type of the contig (abc - Affymetrix Barley Contig) or type of the polymorphism (scsnp, scssr or scind).

Syntenous relationships with rice were explored by comparing HarvEST contig sequences with TIGR rice pseudomolecule 2 gene models using blastx and blastn. Rice blastx matches are aligned along the barley chromosomes. All rice matches have E-value better than e^{-20}. The top blastx hit is shown along with the second hit, if they were highly similar.