



Discovery and development of exome-based, co-dominant single nucleotide polymorphism markers in hexaploid wheat (*Triticum aestivum* L.)

Keith J. Edwards, University of Bristol, School of Biological Sciences, University of Bristol, Bristol, UK
LGC, Unit 1 - 2 Trident Industrial Estate, Pindar Road, Hoddesdon, Herts EN11 0WZ, UK

Summary

Wheat is a key food source both for humans and animals, but the modern varieties are reducing in genetic diversification. Introducing genes from related species could increase its resistance to pests and diseases, and increase crop yields. Improvements in next-generation sequencing (NGS) and our understanding of wheat's complex genome will help the process to identify molecular markers for useful wheat characteristics, to improve this development of novel wheat cultivars.

Kompetitive Allele Specific PCR (KASP™) genotyping technology is a one-step system that can validate SNPs (single nucleotide polymorphisms) that have been identified by NGS, and is particularly suited to use with polyploid species. Researchers led by the University of Bristol used KASP to validate varietal SNPs mined from UK and European cultivars of wheat. The SNPs were derived from a publicly-available wheat expressed sequence tag (EST) database and unique sequences generated by the team using NGS.

The team created the first large scale assembly of genotyping and genetic map information for elite UK wheat varieties based on individual SNP markers, and were the first to demonstrate the use of KASP-based technology to genotype wheat cultivars and generate an extensive linkage map. The research goes on, to find more SNPs in areas where the information is sparse.

Introduction

The global population is growing, and it is estimated that it will reach nine billion by 2050. There is an increasing demand for cereals, both as a food source and as feed for animals⁽¹⁾. As wheat is among the three most important food sources and the main cereal crop for temperate climates⁽²⁾, increasing the yield of wheat crops will be very important⁽³⁾.

Modern bread wheat has a complex genome – it is allohexaploid (AABBDD), formed by introgression between the diploid (DD) genome of *Aegilops tauschii* and the tetraploid (AABB) genome of *Triticum turgidum*⁽⁴⁾. Modern wheat has lost a large proportion of its genetic diversity⁽⁵⁾, and introducing genes from its wild relatives could increase stress, pest and disease resistance⁽⁶⁾. Finding and using molecular markers such as SNPs will enable precision breeding of important desirable traits.

This application note describes the use of KASP genotyping technology from LGC to validate varietal SNPs mined from UK wheat germplasm⁽⁷⁾.

The map positions of the different SNP marker types were investigated to determine whether any bias in genetic location was introduced by using co-dominant SNP assays in two double haploid (DH) mapping populations developed from UK cultivars Avalon x Cadenza (A x C) and Savannah x Rialto (S x R). Of the 3214 SNP markers developed to date, 2109 were identified as polymorphic between Avalon and Cadenza, of which 1807 were placed on the Avalon x Cadenza map.

‘KASP genotyping technology is a one-step system that can validate SNPs that have been identified by NGS, and is particularly suited to use with polyploid species.’

Materials and methods

Plant material

For SNP genotyping, the researchers selected 23 different wheat varieties to represent the UK wheat germplasm, harvesting and freezing the root and leaf tissues at six weeks.

Preparation of normalised cDNA libraries

For NGS, the researchers extracted and purified total RNA from root and leaf tissue of five wheat varieties (Avalon, Cadenza, Rialto, Savannah, Recital) and then synthesised, purified and normalised complementary DNA (cDNA).

Next-generation sequencing

The University of Bristol Transcriptomics Facility processed and sequenced 5 µg of normalised cDNA for each variety.

SNP discovery

Putative varietal SNPs came from two sources:

- The publicly-available wheat expressed sequence tags (EST) data from NCBI (National Center for Biotechnology Information) – the researchers selected 213 SNPs with a high probability of being varietal for validation.
- Combining the NCBI wheat sequence data set with the next-generation sequencing data from the five wheat varieties sequenced as part of the study. The resulting reference data set of 91,368 sequences was mined for SNPs with at least two alternative bases at a reference position.

SNP validation

For each putative varietal SNP, KASP assays (two allele-specific forward primers and one common reverse primer) were designed by LGC (see call out box).

Genotyping reactions were performed in a thermal cycler in a final volume of 5 µL, containing 1x KASP Master mix, the SNP-specific KASP assay mix and 10-20 ng of genomic DNA.

Importantly, the cycling conditions were the same for every SNP assay:

- 15 min at 94°C
- 10 touchdown cycles of 20 sec at 94°C
- 60 sec at 65–57°C (dropping 0.8°C per cycle)
- 26 – 35 cycles of 20 sec at 94°C
- 60 sec at 57°C

After detecting the fluorescence and analysing the data, the team calculated polymorphic information content for each marker, and carried out hierarchical cluster analysis.

Genetic map construction

The software program MapDistro v. 1.7 was used to place the SNP markers into the previously established genetic map derived from 190 Avalon x Cadenza doubled haploid lines. The Kosambi mapping function was used to calculate map distances.

Results and discussion

This was the first large scale assembly of genotyping and genetic map information for elite UK wheat varieties based on individual SNP markers.

The mining for SNPs specifically targeted varietal SNPs in exon sequences in order to generate molecular markers with a potential link to QTL's, and to include both synonymous and non-synonymous point mutations.

SNP discovery

By screening the publicly-available wheat EST database the researchers identified around 3500 putative varietal SNPs in 8668 sequences. Sequencing the normalised whole-seedling cDNA from the five wheat varieties grown for the purpose generated 24 - 45 million, 75-base paired-end reads for each line, and SNP discovery led to 14,078 putative SNPs in 6255 distinct reference sequences (2.7 megabases). This is an average of five varietal SNPs per kilobase in the reference sequences that have one or more SNP.

Both techniques identified a similar number of varietal SNPs per kilobase, and matched previous work on varietal SNPs in wheat^(9; 10).

SNP validation and characterisation

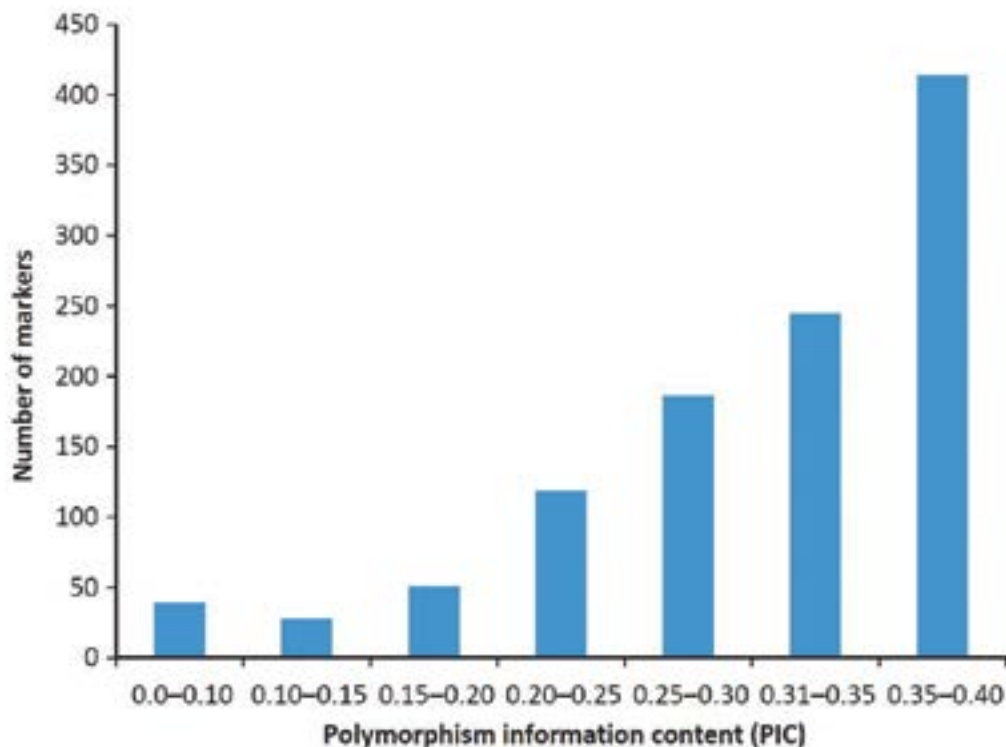
The researchers selected 1659 putative SNPs for validation with genomic DNA (213 from sequences from the NCBI database; 1446 from the NGS process). The NGS SNPs were selected based on their predicted polymorphism level and validated using the KASP genotyping platform on 21 hexaploid wheat varieties, with a diploid and a tetraploid variety.

Of these 1659 SNPs, 1114 (67%) were polymorphic between the different varieties. This conversion rate of around 67% from the KASP system is relatively high for a complex polyploid species⁽¹¹⁾, but still has potential to be increased using primer design and PCR optimisation if necessary. 70 (4%) were monomorphic in the hexaploid varieties, but polymorphic between the hexaploid varieties and the diploid and/or tetraploid, which suggests that these markers represent intravarietal homoeologous SNPs.

The primers can be redesigned as homoeologous-specific KASP primers, to discriminate when homoeologous copies do or do not contain varietal SNPs. Screening markers from this category against the Chinese Spring nullisomic lines confirmed that this was the case in 26 of 28. As intravarietal homoeologous SNPs normally account for around 74% of all SNPs in wheat⁽¹⁰⁾, this suggests that these data sets have been enriched for varietal SNPs. However, as the process is time- and work-intensive, this is probably only useful for specific projects, such as investigating homoeologs tightly linked to specific loci of interest.

475 (29%) failed to generate a useful amplification signal, and this could be improved by using optimised PCR conditions and alternative primers.

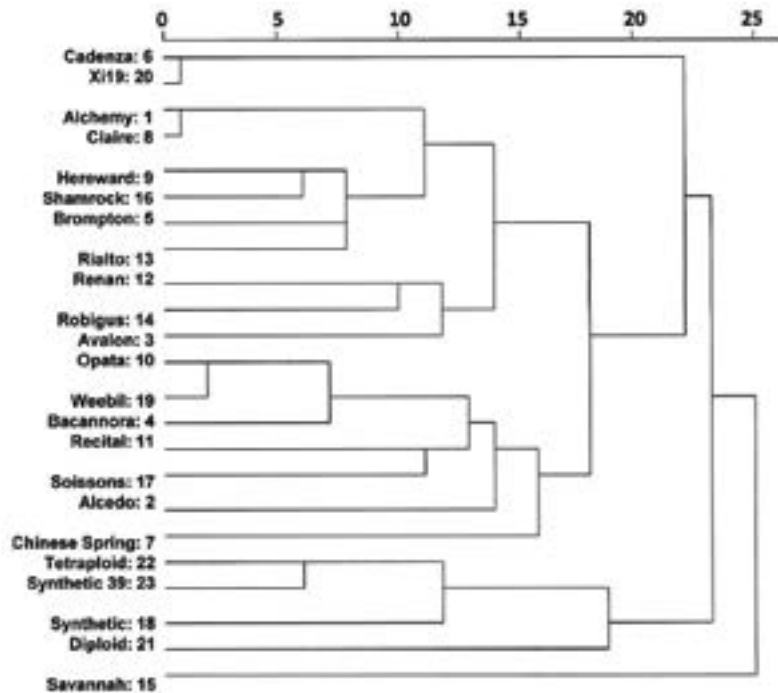
Figure 1: Distribution of PIC of SNPs among the 23 wheat varieties.



The polymorphism information content (PIC) values for the validated markers were 0.08-0.975, with an average of 0.300, comparable to other results from wheat^(11; 12),

and no significant differences between markers from different genomes or homoeologous groups (Figure 1).

Figure 2: Genetic relationships between the 23 wheat lines screened with 1114 KASP probes.



Hierarchical cluster analysis suggests possible genetic relationships between the 23 lines (Figure 2).

Genetic map construction

This study produced the first SNP-based linkage map, based on 500 transcript-linked varietal SNPs and 574 existing markers. The total map length is similar to maps previously created^(13; 14; 15) using a variety of markers, and showed similar map lengths for the A, B and D genome. There was a significant difference in the distribution of markers between the three genomes, and the markers were not evenly distributed between the homoeologous chromosome groups, with markers clustered around the centromeres, particularly on the A and B chromosomes.

There were fewer SNP markers on the D chromosomes, both compared with the A and B chromosomes and with other marker studies. This is probably because SNP markers target genic regions, and this reflects the lack of genetic diversity in modern wheat varieties. Future SNP discovery studies should target the D chromosomes to counterbalance this lack of diversity^(12; 16).

The researchers compared the genetic map locations of the 500 SNP loci with their predicted map location based on physical mapping experiments. Overall, the correlation was relatively high, and is useful for identifying and mapping deletions. Discrepancies, for example SNPs mapping further apart than expected

based on linkage group information, suggests that the relative positions of ancestral markers have been altered through chromosomal rearrangements.

Some SNPs were similar to sequences in the NCBI non-redundant protein sequence database but mapped to multiple regions of the genome. This suggests that KASP-based genotyping has high enough levels of sensitivity to map individual paralogous genes in a complex polyploid genome such as the wheat genome.

Conclusion

According to the researchers, this is the first report of a public linkage map for hexaploid wheat containing several hundred individual SNP markers, and the first demonstration of KASP-based technology to genotype wheat varieties and generate a linkage map.

Creation of the linkage map took 102,220 individual KASP reactions, using 538 probes on 190 plants. The reactions were carried out within 24 hours, using simple microplate technology, and read with a standard fluorescence resonance energy transfer (FRET)-capable plate reader. The team behind the study is confident that this could lead to fast and cost-effective genotyping of thousands of plants with a large and flexible number of markers.

The next step will be to find more SNPs, particularly focusing on the D genome and the homoeologous group 4 chromosomes, which would make genome-wide association studies (GWAS) possible in wheat. This project has also shown that it's not just the SNP markers that are important, but also the information associated with these, such as the surrounding sequences. This additional information means that the SNP data can be used in a variety of genotyping platforms, now and in the future. Making this information freely available will provide the power needed to develop new varieties of wheat and meet the global food needs in the future.

References

1. Foresight. The Future of Food and Farming: Final Project Report. s.l. : The Government Office for Science, 2011.
2. Wheat. Shewry, P R. 2009, J Exp Bot, Vol. 60, pp. 1537-1553.
3. Reaping the benefits: Science and the sustainable intensification of global agriculture. s.l. : The Royal Society, 2009.
4. Genome plasticity a key factor in the success of polyploid wheat under domestication. Dubcovsky, J and Dvorak, J. Jun 2007, Science, Vol. 316, pp. 1862-1866.
5. Grinding up wheat: a massive loss of nucleotide diversity since domestication. Haudry, A, *et al.* Jul 2007, Mol Biol Evol, Vol. 24, pp. 1506-1517.
6. Raising yield potential of wheat. I. Overview of a consortium approach and breeding strategies. Reynolds, M, *et al.* Jan 2011, J Exp Bot, Vol. 62, pp. 439-452.
7. Transcript-specific, single-nucleotide polymorphism discovery and linkage analysis in hexaploid bread wheat (*Triticum aestivum* L.). Allen, A M, *et al.* Dec 2011, Plant Biotechnol J, Vol. 9, pp. 1086-1099.
8. Sambrook, J, Fritsch, E F and Maniatis, T. Molecular Cloning: A Laboratory Manual. 2nd edition. Cold Spring Harbor : Cold Spring Harbor Laboratory Press, 1989.
9. DNA sequence polymorphisms and their application to bread wheat quality. Ravel, Catherine, *et al.* 3, December 2007, Euphytica, Vol. 158, pp. 331-336.
10. A genome-wide analysis of single nucleotide polymorphism diversity in the worlds major cereal crops. Barker, G L and Edwards, K J. May 2009, Plant Biotechnol J, Vol. 7, pp. 318-325.
11. Multiplex single nucleotide polymorphism (SNP)-based genotyping in allohexaploid wheat using padlock probes. Edwards, K J, *et al.* May 2009, Plant Biotechnol J, Vol. 7, pp. 375-390.
12. Population- and genome-specific patterns of linkage disequilibrium and SNP variation in spring and winter wheat (*Triticum aestivum* L.). Chao, S, *et al.* 2010, BMC Genomics, Vol. 11, pp. 727-727.
13. A high-density microsatellite consensus map for bread wheat (*Triticum aestivum* L.). Somers, D J, Isaac, P and Edwards, K. Oct 2004, Theor Appl Genet, Vol. 109, pp. 1105-1114.
14. Diversity arrays technology (DArT) for high-throughput profiling of the hexaploid wheat genome. Akbari, M, *et al.* Nov 2006, Theor Appl Genet, Vol. 113, pp. 1409-1420.
15. A high-density genetic map of hexaploid wheat (*Triticum aestivum* L.) from the cross Chinese Spring x SQ1 and its use to compare QTLs for grain yield across a range of environments. Quarrie, S A, *et al.* Mar 2005, Theor Appl Genet, Vol. 110, pp. 865-880.
16. Sequence polymorphism in polyploid wheat and their d-genome diploid ancestor. Caldwell, K S, *et al.* Jun 2004, Genetics, Vol. 167, pp. 941-947.

About LGC

LGC is an international science-based company and market leader in the laboratory services, measurement standards, genomics, reference materials and proficiency testing marketplaces. LGC operates in a variety of markets – including, but not confined to, Food & Agriculture, Government, Pharmaceuticals and Biopharmaceuticals and Sports - which underpin the safety, health and security of the public and the regulation of industry, for both private and public sector clients.

With headquarters in Teddington, South West London, LGC employs over 2,000 staff, operating out of 22 countries worldwide. Its operations are extensively accredited to international quality standards such as ISO/IEC 17025.

Set up in 1842 as the Laboratory of the Government Chemist, for more than 100 years LGC has held the unique function of the Government Chemist in the UK. LGC was privatised in 1996 and is now majority-owned by funds managed by Bridgepoint.

For further information:

Our genomics solutions (www.lgcgenomics.com) provide high quality services and products for DNA and genetic analysis, and sample preparation including:

- Genotyping services, assays and reagents
- Sanger and next-generation sequencing services
- DNA and RNA extraction products
- Instruments, reagents and consumables for molecular biology

Twitter: twitter.com/LGCGenomics

Facebook: www.facebook.com/LGC.Genomics

LinkedIn: www.linkedin.com/company/lgc-genomics

Follow LGC on:

twitter: twitter.com/LGCGroup

LGC's Science blog: www.blog.lgcgroup.com

Web: www.lgcgroup.com

KASP chemistry: How it works

The Kompetitive Allele Specific PCR (KASP) genotyping system, is a PCR-based genotyping technology. The KASP assay uses just two components, along with the test DNA carrying the SNP of interest:

- KASP Assay mix – two different, allele-specific competing forward primers with unique tail sequences and one reverse primer
- KASP Master mix – the universal FRET cassette reporter system, Taq polymerase, dNTPs, ROX passive reference dye, and MgCl₂ in an optimised buffer solution.

The primers are unique to each assay and are not labelled. The labelled FRET cassette reporter system, part of the Master mix, includes the label.

In the first round of PCR, one of the allele-specific forward primers matches with the target SNP and, with the common reverse primer, amplifies the target region.

In the second round of PCR, the complement of the allele-specific tail is created.

In subsequent rounds of PCR, the levels of the allele-specific tail climb. The FRET cassette reporter, with the quenched fluorescent dye, is complementary to the tail sequence, and therefore binds to the tail. This means that the dye is no longer quenched and a fluorescent signal is generated, which can then be measured.

Key benefits

Simplicity

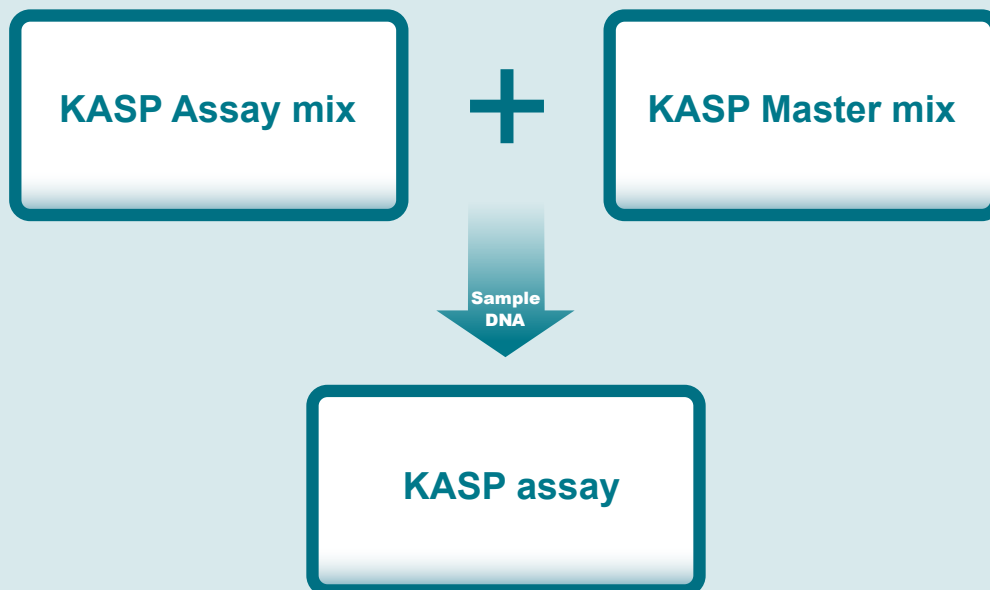
- Accuracy >99.8% based on independent assessment
- Industry leading SNP and InDel assay conversion rate

Flexibility

- Flexible primer design, increasing the chance of creating a successful assay.
- Supports low-, medium- and high-throughput studies
- Platform independent and compatible with most qPCR machines and FRET capable readers.

Cost

KASP uses a universal reporting system where labelled components are present in the master mix; this eliminates the need for expensive labelled assay-specific primers or probes.



Key words: PCR, genotyping, fluorescence, allele-specific, polyploid.

www.lgcgenomics.com

LGC

Germany

Ostendstr. 25 • TGS Haus 8
12459 Berlin • Germany

Tel: +49 (0)30 5304 2200
Fax: +49 (0)30 5304 2201
Email: info.de@lgcgroup.com

United Kingdom

Unit 1-2 Trident Industrial Estate • Pindar Road
Hoddesdon • Herts • EN11 0WZ • UK

Tel: +44 (0)1992 470 757
Fax: +44 (0)1438 900 670
Email: info.uk@lgcgroup.com

United States

100 Cummings Center • Suite 420H
Beverly • MA 01915 • USA

Tel: +1 (978) 232 9430
Fax: +1 (978) 232 9435
Email: info.us@lgcgroup.com