Designing genome specific primers in polyploid wheat

The polyploid nature of some wheat species makes the amplification of specific DNA sequences sometimes challenging. For instance, hexaploid (6x) bread wheat has a high level of coding sequence similarity (~95 %) between homoeologous genes on the A, B and D genomes, but this sequence conservation drops outsides the coding region. This makes sequence amplification of gene coding region from a specific genome challenging. However, we can take advantage of the sequence variation in the ~5 percent of coding sequences, as well as in the non-conserved intronic and intergenic region to make genome specific primers for genome-specific sequence amplification. This tutorial document introduces you to how to use homoeologous sequence variation to design genome-specific primers for PCR based application like gene amplification, qPCR assays, or molecular marker assays.

a) Principles of primer specificity.

To attain primer specificity, it is important to design primers with perfect complementarity to your target DNA sequence but with mismatches to non-target DNA sequence. More importantly, the mismatches between you primer sequence and the non-target DNA sequence should be located at (or very close to) the 3' end of the primer. This mismatch at the 3' end stalls DNA extension by DNA polymerase on the non-target sequence and this leads to preferential amplification of the target sequence over the non-target sequences.

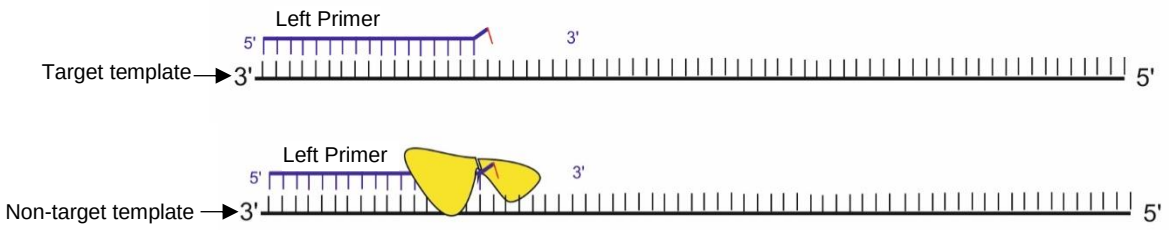

Figure 1: Effect of mispairing on template extension during PCR.
DNA base mispairing at the 3' end of a primer can stall DNA polymerase activity.

b) Designing genome specific primers for PCR amplification

1. Retrieve homoeologous (corresponding A, B and D genome) sequences of your target DNA.

   You can do this through the "homoeologue" link in the gene based display of the Ensembl Plants summary page for your target gene (see "Ensembl plants primer" for more info). Alternatively you can blast nucleotide sequence of your target template DNA in Ensembl Plants to access the link to homoeologous gene.

2. Align the homoeologous sequences using any suitable online multiple sequence alignment tool or sequence editor (e.g. Clustal Omega).

3. Identify homoeologous SNP (Single nucleotide Polymorphism) in your alignment and target SNP that is specific to your target genome.

   Homoeologous SNPs are SNPs between the A, B and D genome. In Figure 2*, SNP 1, 2 and 3 are homoeologous SNPs that are specific to the A, B and D genome sequences respectively.*

Figure 2: Selecting Genome specific SNP from sequence alignment.
Genome Specific SNP distinguishes a genome of interest from the other genome. This can be used to design genome specific primers for preferentially amplification of target genes of interest.

4. Using a primer design programme (e.g Primer3plus), design separately forward (left) and reverse (right) primers with your target genome-specific SNP the 3' end of the primers. Follow the steps below to do this:

   Forward primer: Input 30 – 40 bases inclusive and **upstream** of your target genome-specific SNP (e.g SNP 1 - cctgcgagcggcacg**A** for the A genome in Fig. 2) into a primer design programme and set design parameters to return as many primers as possible from the input sequence.

   Reverse primer: input 30 – 40 bases inclusive and **downstream** of your target genome-specific SNP (e.g SNP 1 - **A**cttcttccaggtcgtcaa for the A genome in Fig 2.) into a primer design programme and set design parameters to return as many primers as possible from the input sequence.

5. Set primer design parameters to optimise for primer length (18-25), Tm (55-62 °C), GC content (40-60 %), and to returns as many primers as possible from the input sequence.

6. Select primers with the genome specific SNP at the 3' end or close to (2-3 bases from) the 3' end.

   If using primer3plus or similar programme, you can change the case of the SNP nucleotide to lower case in the input sequence. This help in easily identifying the genome specific primers with SNP base (lower case) at the 3' end.

c) Designing genome specific KASP assays primers

Kompetitive allele specific PCR (KASP) is a gel-free fluorescent-based PCR assay for genotyping SNP variation between varieties. KASP is increasingly becoming the preferred SNP genotyping platform because of its affordability, ease-of-use, and ability to genotype to varying scales (high, medium or low genotyping scale). KASP uses fluorescent labelled primers targeted at SNP sites that differ between varieties (varietal SNPs) to differentially amplify short sequence fragments from these varieties.

Distinguishing between varietal SNPs and Homoeologous SNPs

Homoeologous SNPs have been introduced earlier in this document as single nucleotide variations that exist between the genomes (A, B and D) sequences. Varietal SNPs, on the other hand, are single nucleotide variations between the sequences of different wheat lines (varieties/cultivars, landraces, or even species) on the same genome. In Figure 3, SNP 1 is a varietal SNP between the A genome sequence of variety 1 and variety 2, as the A-G SNP found in the variety 1 is not found in variety 2. However, SNP 2 is a homoeologous SNP that distinguishes the A genome from the B and D genomes in both variety 1 and variety 2.

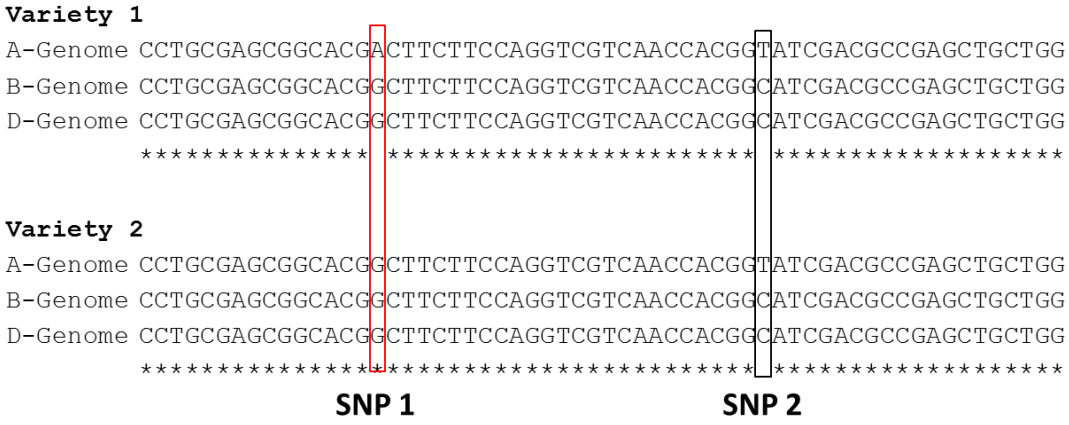

Figure 3: Varietal and Homoeologous SNP in polyploid wheat
A varietal SNP distinguishes between different wheat lines (SNP 1) while a homoeologous SNP distinguishes between genomes (SNP 2)

Using PolyMarker for high-throughput genome-specific KASP marker design

For each KASP assay, two allele-specific primers (one primer for each allele of the varietal SNP) and one common genome specific primer are required. The principles previously described for designing genome specific primer can also be used to design these primers. However this can be tedious when a large number of KASP assays are to be designed. PolyMarker - a fast and accurate primer design pipeline for polyploid wheat - can be used to automate this process.

PolyMarker generates a multiple alignment between the target SNP sequence and the International Wheat Genome Sequence Consortium (IWGSC) chromosome survey sequences for each of the three wheat genomes. It then generates a mask that helps identify varietal and homoeologous SNPs and uses those to design genome specific diagnostic markers. To use PolyMarker, follow the steps below.

1. Prepare input SNP data and save in a comma separated file format. This should contain three fields:

   a. ***Gene name***: A user defined gene or SNP name or I.D.
   b .***Chromosome location***. The target chromosome and genome to amplify from.
   c .***SNP sequence***: the target varietal SNP allele in square bracket([A/C]) along with 100 bp of flanking sequence on either side.
   E.g TaGa20Ox1, 4A, …CCTGCGAGCGGCACG[a/g]CTTCTTCCAGGTCGT…

2. Visit www.polymarker.tgac.ac.uk to access Polymarker user interface (Figure 4)

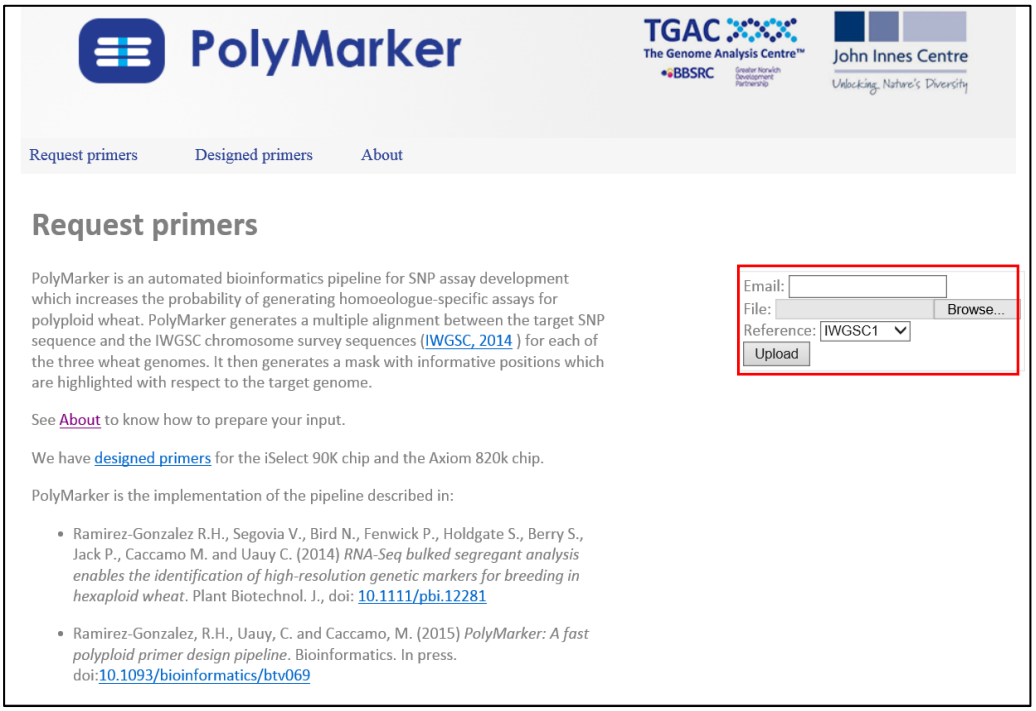

Figure 4: Polymarker user interface

3. Enter your email address in the email box on the right hand side of the page (red box in Figure 4). Progress status and link to the result of your jobs will be sent to this email.
4. Upload the input SNP data field prepared in 1 above.
5. Select the reference assembly to use for the alignment of corresponding homoeologous sequence of the input SNP sequence(s).
6. Click upload.

Understanding PolyMarker output

The result from PolyMarker has the output fields described in the table below.

Table 1: Description of Polymarker output fields

Fields	Description
ID	The gene or SNP name supplied in the input data
SNP	The SNP alleles and position in the input SNP sequence.
Chr	The target chromosome for which the primers were designed.
CTotal	Total number of matching IWGSC contig retrieved for the SNP sequence. Typically this should be 3 (i.e. one contig for each genome). However, in certain cases, the IWGSC reference might not have homoeologous sequence for all the genomes.
Contig regions	The retrieved IWGSC contig names and the position in the contig that matches the SNP sequences
SNP type	This describes whether the input SNP is homoeologous (polymorphic between the genomes) or non-homoeologous (monomorphic in all the genome). Non-homoeologous SNPs are preferred.
A*	Allelic primer sequence having one of the SNP allele at the 3' end
B*	Allelic primer sequence having the other SNP allele at the 3' end
Common	Common primer sequence
Primer type	Describes the type of genome specificity achieved with the designed primer. That is how specific is the designed primer to the target genome. This could be *Specific*: The common primer is genome specific to the target chromosome *Semi-specific*: The common primer discriminates between the target genome and one, but not both, of the other two genomes. *Non-specific*: No variation could be found between the target genome and non-target genome to design gene specific common primer.
Product size	The size of the amplicon expected from using the primer pair
Error	Report on the errors that might be encountered during the design process

**The diagnostic allelic primers designed by PolyMarker do not contain probe sequences for fluorescent dyes (FAM and HEX/VIC) used in KASP. These must be added to the 5' end of the primers. Add the FAM probe sequence (gaaggtgaccaagttcatgct) to the 5' end of A primers and VIC/HEX probe (gaaggtcggagtcaacggatt) to the 5' end of the B primer or vice versa. E.g*

FAM labelled primer: **gaaggtgaccaagttcatgct***cctgcgagcggcacgA.*

VIC labelled primer: **gaaggtcggagtcaacggatt***cctgcgagcggcacgG*

Ready-made KASP primer for Known markers

In addition to the ability to design new genome-specific markers, the PolyMarker website also contains already designed KASP assay primers for known markers which you can take advantage of. This includes markers from the high density SNP arrays like iSelect and Axiom chip (See "Variation Data" for more information on these). To downloaded files containing these primer sequences, follow steps below.

1. Click on the "Designed Primer" option on PolyMarker main menu bar
2. Select the link to the files you are interest in to download a zipped folder.
   Designed primers are currently arrange under three files described below
   Wheat iSelect 90K Located against CSS: Primers for all the probe sequence on the iSelect SNP array. Chromosome and scaffold information for the probe sequences was first obtained by BLAST against IWGSC Chinese Spring flow-sorted scaffold assembly (See "Wheat Genome" for more information) and the primers were designed to be specific (if possible) to his scaffold sequence.
   Wheat iSelect 90 K genetically mapped: This contains a subset of the markers above that have been genetically mapped in addition to the chromosome and scaffold information.
   Axiom Located against CSS: Primers for all the probe sequences on the Axiom SNP array. These primers were designed to be specific to the CSS scaffold that best hit the Axiom SNP probe sequence.
3. Unzip the folder to access the file. Note that these file can have very big size which might be difficult to open and explore in Excel.

Ready-made KASP primer for Wheat TILLING Mutants

The section on "TILLING mutant resources" introduces you to the EMS-mutant populations (Tetraploid and Hexaploid) available for forward and reverse genetics in wheat. Mutations within and around the gene-coding region in these mutant lines have been identified and annotated. To track these mutations in segregating populations, genome-specific KASP assay primers have already been designed for most of these mutations and can be found on the Wheat TILLING website at www.wheat-tilling.com. See "Selecting TILLING Mutants" for a tutorial on using the Wheat TILLING website.