

Designing genome specific primers in polyploid wheat

The polyploid nature of some wheat 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. This makes sequence amplification of 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. This tutorial document introduces you to how to use homoeologous sequence variation to design genome-specific primers for PCR based applications 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. Importantly, the mismatches between you primer sequence and the non-target DNA sequence should be located at the 3' end of the primer, ideally within the three most 3' bases. 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.

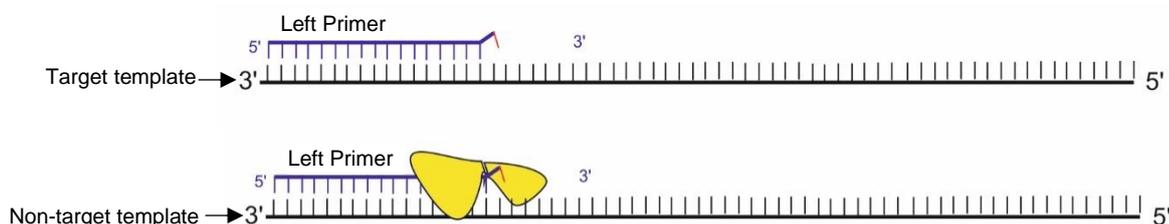


Figure 1: Effect of mispairing on template extension during PCR. DNA base mispairing at the 3' end of a primer can stall DNA polymerase activity.

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 “homoeolog” link in the gene based display of the EnsemblPlants summary page for your target gene (see “[EnsemblPlants primer](#)” for more info). Alternatively you can BLAST the nucleotide sequence of your target template DNA in EnsemblPlants to access the link to the homoeologous gene(s).

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

- 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.

```

A   CCTGCGAGCGGCACGACTTCTTCCAGGTCGTCAACCACGGCATCGACGCCGAGCTGCTGG
B   CCTGCGAGCGGCACGGCTTCTTCCAGGTCGTGAACCACGGCATCGACGCCGAGCTGCTGG
D   CCTGCGAGCGGCACGGCTTCTTCCAGGTCGTCAACCACGGTATCGACGCCGAGCTGCTGG
*****
                SNP 1                SNP 2                SNP 3

```

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.

- 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**cttcttccaggctcg**t**caa 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.

- 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.
- 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 fluorescence-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. For more information about KASP genotyping, see [here](#).

Distinguishing between varietal SNPs and Homoeologous SNPs

Homoeologous SNPs have been introduced earlier in this document as single nucleotide variations that exist between genomes (A, B and D). 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 because 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.

```
Variety 1
A-Genome CCTGCGAGCGGCACGACTTCTTCCAGGTCGTCAACCACGGTATCGACGCCGAGCTGCTGG
B-Genome CCTGCGAGCGGCACGGCTTCTTCCAGGTCGTCAACCACGGCATCGACGCCGAGCTGCTGG
D-Genome CCTGCGAGCGGCACGGCTTCTTCCAGGTCGTCAACCACGGCATCGACGCCGAGCTGCTGG
*****

Variety 2
A-Genome CCTGCGAGCGGCACGGCTTCTTCCAGGTCGTCAACCACGGTATCGACGCCGAGCTGCTGG
B-Genome CCTGCGAGCGGCACGGCTTCTTCCAGGTCGTCAACCACGGCATCGACGCCGAGCTGCTGG
D-Genome CCTGCGAGCGGCACGGCTTCTTCCAGGTCGTCAACCACGGCATCGACGCCGAGCTGCTGG
*****

SNP 1 SNP 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 reference genome of your choice (including IWGSC RefSeqv1.0; see “[Genome assemblies](#)” for more information) 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. Visit www.polymarker.tgac.ac.uk to access Polymarker user interface (**Error! Reference**

Figure 4: Polymarker User Interface

source not found.)

2. Prepare input SNP data. Input data can either be saved as a comma separated values (csv) file and uploaded, or manually entered into the user interface. The input data must contain three fields:
 - a. **Gene name:** A user defined gene or SNP name or I.D.
 - b. **Chromosome location.** The target chromosome i.e. 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...

3. Optionally, enter your email address in the email box on the bottom left-hand side of the page. The status of your jobs and link to your results will be sent to this email.
4. Select the reference assembly to use for the alignment of corresponding homoeologous sequence of the input SNP sequence(s).
5. 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
Chr	The target chromosome for which the primers were designed.
AIn	Total number of matching sequences retrieved for the SNP sequence. Typically this should be 3 (i.e. one hit for each genome) although this may vary in cases where there have been gene losses or duplications, for example
Type	This describes whether the input SNP is homoeologous (polymorphic between the genomes) or non-homoeologous (monomorphic in all the genomes). Non-homoeologous SNPs are preferred.
A	Allelic primer sequence having one of the SNP alleles 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. This could be <i>Specific:</i> The common primer is genome specific to the target chromosome <i>Semi-specific:</i> The common primer discriminates between the target genome and one, but not both, of the other two genomes. <i>Non-specific:</i> No variation could be found between the target genome and non-target genome to design gene specific common primer.
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: gaaggtgaccaagttcatgctcctgcgagcggcagA.
VIC labelled primer: gaaggtcggagtcaacggattcctgcgagcggcagG

Pre-designed KASP primers for known markers

In addition to the ability to design new genome-specific markers, the PolyMarker website also contains pre-designed KASP assay primers for known. This includes markers from the high density SNP arrays like iSelect and Axiom chips (See "[Variation Data](#)" for more information on these). These files can be downloaded through links found under the "Primers for wheat SNP-chips" option at the top of the PolyMarker homepage. It is important to note that the primers may not have been experimentally validated before.

Pre-designed KASP primers 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 many of these mutations and can be found on *EnsemblPlants*.

d) Design of homoeolog-specific primers for full gene cloning

Coulton and Edwards recently published AutoCloner, an automated primer design pipeline for full-gene cloning in wheat and other polyploids. Although there already exists a tool for designing primers for use in Kompetitive allele specific PCR (KASP) assays in polyploids, PolyMarker, this only considers flanking sequences of 100 bases either side of a varietal SNP, meaning that it cannot be used to clone entire genes. AutoCloner was created to fill this gap, allowing researchers to easily clone genes in polyploids regardless of their bioinformatic skill set.

Website interface: www.autocloner.com

Paper: <https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-020-03601-7>