

# Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Literature review</b>	<b>4</b>
2.1	Wheat Breeding . . . . .	4
2.2	Wheat Genetics . . . . .	4
2.3	Wheat Genomics . . . . .	4
2.4	Sequencing . . . . .	5
2.5	Sequence analysis . . . . .	5
2.6	Wheat online resources . . . . .	5
<b>3</b>	<b>Genetic map of <i>Yr15</i> with RNA-Seq</b>	<b>6</b>
3.1	Mapping population . . . . .	7
3.2	Sequencing and mapping . . . . .	8
3.3	SNP Calling . . . . .	9
3.4	Bulk Frequency Ratios . . . . .	9
3.5	<i>In silico</i> mapping . . . . .	9
3.6	Assay selection . . . . .	9
3.7	Genetic map . . . . .	9
3.8	Conclusions . . . . .	9
<b>4</b>	<b>PolyMarker: A fast polyploid primer design pipeline</b>	<b>11</b>
4.1	Pipeline . . . . .	11
4.2	Applications of PolyMarker . . . . .	15
4.2.1	KASP assays for public sets of SNPs . . . . .	16
4.2.2	SNPs in a mutant population . . . . .	16
4.2.3	Deletions on a mutant population . . . . .	16
4.2.4	PolyMarker public web service . . . . .	17
4.2.5	Genotyping of <i>Puccinia striiformis</i> f. sp. <i>tritici</i> isolates. . . . .	18

<i>CONTENTS</i>	2
4.3 Discussion . . . . .	18
<b>5 Gene expression (expVIP)</b>	<b>19</b>
5.1 Expression experiments (Introduction) . . . . .	19
5.2 Database design . . . . .	19
5.3 Analysis pipeline . . . . .	19
5.4 Graphical interface . . . . .	19
5.5 Conclusions . . . . .	19
<b>6 Conclusions and final remarks</b>	<b>20</b>
<b>A Supplemental tables</b>	<b>21</b>
A.1 Genetic map of <i>Yr15</i> with RNA-Seq supplemental tables.	22
A.2 PolyMarker supplemental tables. . . . .	24
<b>B Quality control</b>	<b>34</b>
B.1 Sequence read quality . . . . .	34
B.2 Sequence GC content . . . . .	37

# Chapter 1

## Introduction

It defines the objectives and the importance of the research. It focus on the the application of Next Generation Sequencing to molecular biology, wheat genetics and ultimately to breeding programs. It also mentions the current status of the wheat reference genome and other resources (genetic maps, markers) the need of tools to query them effectively.

# Chapter 2

## Literature review

It describes the current status of the wheat genome, genetics and other resources.

### 2.1 Wheat Breeding

An overview of how breeding is carried on currently, the different sources of genetic diversity and the relevance of fixing agriculturally important traits.

### 2.2 Wheat Genetics

The section describes alleles and the concept of gene, both as a locus in the genome (Quantitative Trait Locus, QTL) and an specific transcript (central dogma of molecular biology). Finally, it discusses traditional Mendelian inheritance and the effect of polyploidy.

### 2.3 Wheat Genomics

A description of the current status of the wheat genome (Mayer et al. (2014), Chapman et al. (2015)), the different available assemblies and approaches to sort the scaffolds (Genome Zipper, the various genetic maps).

## 2.4 Sequencing

The importance of the selection of the library preparation and the sequencing platforms available. A brief summary of RNA-Seq, Exome capture, Whole Genome Shotgun, etc. and on which cases are more suitable for different experiments. Mention the new technologies developed during the years of the PhD (Ren-Seq, PacBio?)

## 2.5 Sequence analysis

This section discusses the criteria to decide analysis done after sequencing, when to do re-alignments or *de novo* assemblies, how to do SNP calling in diploid and polyploid organisms and the bulk frequency ratios.

## 2.6 Wheat online resources

A compilation of the currently available resource for wheat genetics and genomics. MAS wheat, CerealsDB, Ensembl, etc.

# Chapter 3

## Genetic map of *Yr15* with RNA-Seq

Wheat breeding programs aim to improve the wheat lines available for production. One way to facilitate the selection of elite lines is to find molecular markers that can facilitate the selection of seeds without having to wait until the phenotype is observed. Bulk Segregant Analysis (BSA) consists on pooling the DNA of individuals with contrasting phenotypes (Michelmore et al., 1991) on a segregating population. The bulks show as heterozygous except for the region that is linked to the trait of interest. This approach can be used to identify SNPs using High Throughput Sequencing, such as: exome capture (Hodges et al., 2007), RNA-Seq (Pickrell et al., 2010), whole genome resquencing (Schneeberger et al., 2009), among others.

One of the traits desired in an elite line is the resistance to pathogens, such as *Puccinia striiformis* f. sp. *tritici*, the fungi responsible of yellow rust. A source of resistance genes is are introgressions from other species, suchas *Triticum diccoides*. In the University of Sydney a collection of Near Isogenic Lines (NILs) with introgressions to several Yellow Rust resistance genes on a susceptible background were developed (Wellings and McIntosh, 1998). On this chapter the NIL for the *Yr15* locus is used to produce a mapping population to improve diagnostic markers. The population was sequenced using RNA-Seq and a bioinformatic pipeline was developed to score Single Nucleotide Polymorphisms (SNPs) linked to the *Yr15* locus. Finally, the best candidate SNPs where selected to produce a genetic map which lead to a triplet of markers diagnostic to the target locus. The steps described in this chapter were first published

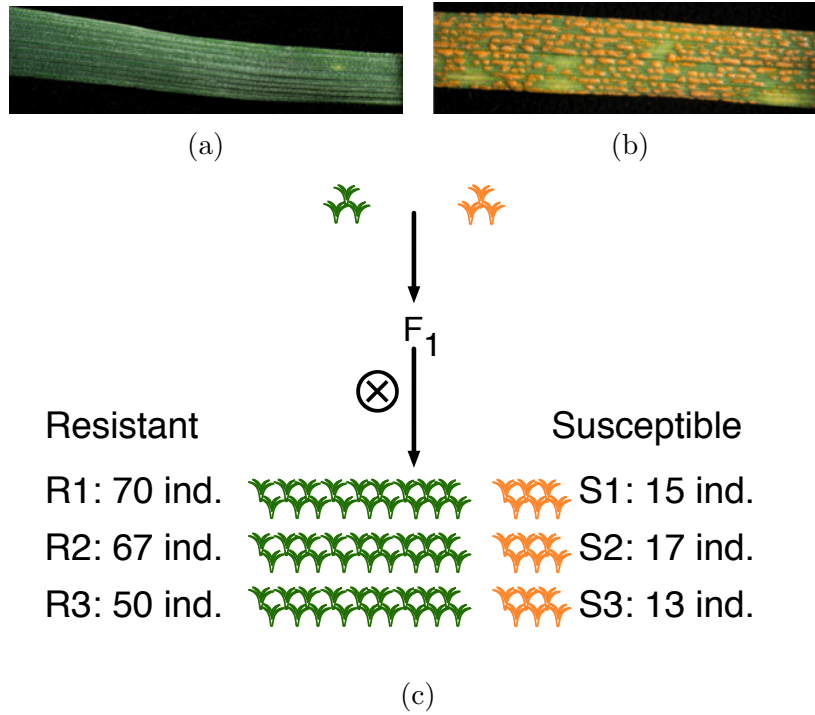


Figure 3.1: Response of (a) Avocet + *Yr15* and (b) Avocet when inoculated with *Puccinia striiformis* f. sp. *tritici* at the three leaf stage. (c) The phenotype of the  $F_2$  population was used to produce 6 bulks, 3 resistant and 2 susceptible. The RNA was pooled in bulks accordingly.

in Ramirez-Gonzalez et al. (2015b) and the results of this chapter are published in Ramirez-Gonzalez et al. (2015c).

### 3.1 Mapping population

The population was developed by crossing the resistant line ‘Avocet + *Yr15*’ (*Yr15*) (Wellings and McIntosh, 1998), Figure 3.1a, to the susceptible line Avocet (AVS), Figure 3.1b.  $F_2$  seeds from tree independent  $F_1$  plants were sown and tissue was collected, before the fungal inoculation to avoid the effect of the response on the gene expression. The plants were challenged at the three leaf stage as it is known that *Yr15* confers resistance in seedlings (Gerechter-Amitai et al., 1989). The expected segregation on an  $F_2$  population is 3:1 (resistant:susceptible), since *Yr15* is a dominant gene. From the 232 plants in the  $F_2$  population that germinated, 187 were resistant and 45 were susceptible, which deviates slightly from the expected ratio ( $\chi^2 = 0.049$ ). Segregation distortion has been shown for the same *Yr15* donor (Randhawa et al., 2009), however the

Table 3.1: Arrangement and number of sequenced base pairs per sample.

Library	name	Bar code	Lane	Reads ( $\times 10^8$ bp)
LIB1715	Bulk R1	ATCACG	1	0.77
LIB1716	Bulk R2	TAGCTT	1	1.20
LIB1717	Bulk R3	ACTTGA	2	0.96
LIB1718	Bulk S1	GGCTAC	2	1.64
LIB1719	Bulk S2	CGTACG	2	1.49
LIB1720	Bulk S3	GTGGCC	1	1.88
LIB1721	AvocetS	N/A	3	4.13
LIB1722	AvocetS + <i>Yr15</i>	N/A	4	3.99

decreased number of susceptible plants can be explained by escapes in the virulence essays (i.e. plants scored as resistant without the *Yr15* locus). For this study we extracted DNA from individual plants in the  $F_2$  population and we bulked RNA on 6 different bulks: 3 resistant and, 3 susceptible, Figure 3.1c.

## 3.2 Sequencing and mapping

RNA-Seq was used to avoid sequencing the non-coding regions and reduce the search space. The sequencing of the bulks and the parents were done on a single Illumina Hi-Seq2000 each. The bulks were multiplexed and sequenced on a third of a lane each, as shown on Table 3.1. To ensure that the quality of the sequencencing was good, **fastqc-0.10** (Babraham Bioinformatics, 2012) was run with its default parameters in each one of the fastq files. The GC content was around 52% in all the samples (Appendx B.2), which is expected as the sample should be of coding regions, and for wheat the reported GC content in genes is around 55%. The quality of the reads is fairly consistent, in general dropping after the base 80 across the samples (Appendix B.1).

When the analysis was started, the draft genome and the corresponding annotation where not not release yet, hence gene models where used. All the samples where aligned to the Unigenes v60 (56,954 genes) and the gene models from UCW (Krasileva et al., 2013) using **BWA 0.5.9** (Li and Durbin, 2009). The alignment provided showed that a few genes were overly expressed, however we still have have 22,107 and 36,808 genes, on the Unigenes and the UCW gene set respectivley, with a coverage greater than 20x in the progenitor with *Yr15*, Supplemental Table A.1.



Both gene sets performed similarly in terms of the percentage of genes with reads and percentage of aligned reads. Since each individual bulk has a lower coverage, the susceptible and resistant reads were merged *in silico* and mapped as the individual bulks.

### 3.3 SNP Calling

. Ruby implementation of the methodology described by Trick et al. (2012).

### 3.4 Bulk Frequency Ratios

Results of the simple SNP calls from the progenitors and how the score of the Bulk Frequency Ratios(BFR) improve the location of the SNPs.

### 3.5 *In silico* mapping

Mapping of the gene models to the IWGSC CSS Mayer et al. (2014) reference and the location of the SNPs using the genetic map from Wang et al. (2014).

### 3.6 Assay selection

. The selection criteria to decide which SNPs where selected to produce the genetic map: BFR>6, in the short arm of chromosome group 1 and from the *Yr15* progenitor.

### 3.7 Genetic map

The three versions of the genetic map: With a subset of the F<sub>2</sub> population

### 3.8 Conclusions

Remarks on how this technique can be used to do fine-mapping and that if I were to start the project now I would use exome capture or Ren-Seq.

The references have changed since we started we can use different techniques now (exome capture, ren-seq) The markers are now used by our collaborators.

# Chapter 4

## PolyMarker: A fast polyploid primer design pipeline

One of the main challenges of working with polyploid species is the design of genome specific molecular markers. This is particularly true when targeting conserved homoeologue regions, where a primer could bind to a pair, or triplet, of identical sequences. For that reason, designing primers for polyploids require to include bases that are specific to the target, in addition to the physicochemical properties of the primer. The traditional methodology to find primer candidates include a blast search and a local alignment, select the primer candidates manually, and finally, validate the primers with a tool, like **Primer3** (Rozen and Skaletsky, 2000). To reduce the time invested in designed primers I have developed PolyMarker (Ramirez-Gonzalez et al., 2015a), a pipeline to automate the primer design for polyploid organisms.

### 4.1 Pipeline

PolyMarker is an automated pipeline that takes as input a list of SNPs and a reference file and produces a list of primer triplets for SNP genotyping. The list of SNPs is first converted to a FASTA file with ambiguity codes (Cornish-Bowden, 1985). The template sequences are aligned with **exonerate** (Slater and Birney, 2005) to find the homoeologous regions to the target sequence. Then, the alignment between homoeologues is refined using **MAFFT** (Katoh and Standley, 2013). A list of candidate variations is produced and used as input for **Primer3** (Rozen and Skaletsky, 2000). Finally, the output of **Primer3** is parsed to find the best

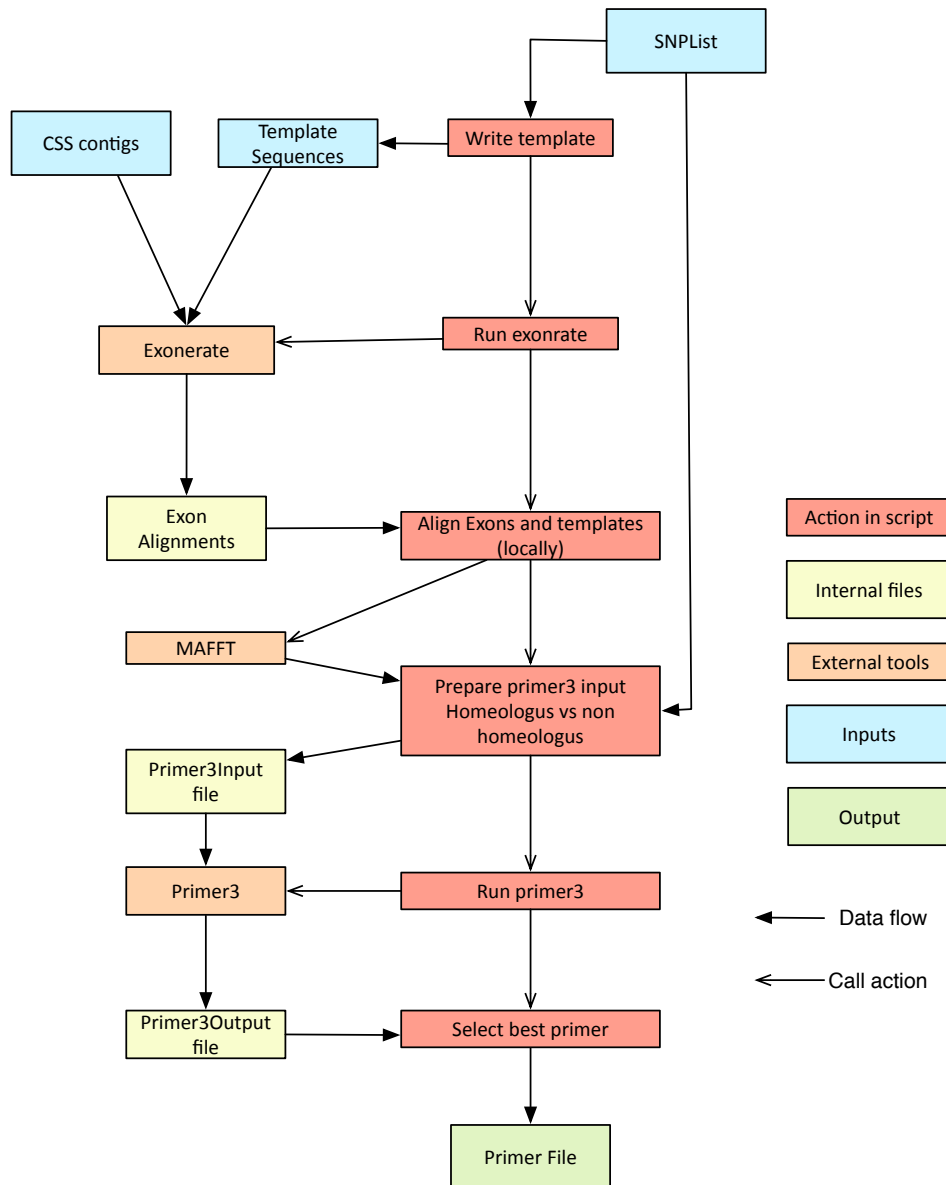


Figure 4.1: Steps and tools called by PolyMarker. The colour of the boxes represent: the step is an action inside the script (red); actions of the script (orange); temporary files (yellow); inputs (blue) and; output (green)

primer pair that contains a the targeted SNP and a base that is specific to the target genome (Figure 4.1). The pipeline is written as a Ruby script, using parsers and wrappers from BioRuby (Goto et al., 2010) and bio-samtools (Etherington et al., 2015; Ramirez-Gonzalez et al., 2012). The software is open source and released as a biogem (Bonnal et al., 2012), `bio-polyploid-tools`, the source code is available in github: <https://github.com/TGAC/bioruby-polyploid-tools>.

The PolyMarker input consist on SNP list with: unique name for the marker, the target chromosome and the sequence for the marker. The alternative alleles are surrounded by square brackets within the sequence. PolyMarker can take a list of several markers and design them in batch, Figure 4.2a. A FASTA file is produced with all the template sequences, with the alternative alleles substituted by the IUAPC ambiguity codes (Cornish-Bowden, 1985). The flanking sequence surrounding the SNP is limited by default to 100bp to reduce the search time and avoid missing regions that diverge near the SNP, as when the variation is near an intron-exon junction.

The template sequences are aligned to the reference using `exonerate` (Slater and Birney, 2005), Figure 4.2b. The alignment is refined with the `--model est2genome` option, to allow the search of sequences coming from transcripts, a common source of SNPs (Allen et al., 2011). The `exonerate` output is formatted with the `--ryo` (roll your own format) to get an output easy to parse. All the hits that contain the SNP are extracted from the reference with a flanking sequence that extend out of the hit, by default, to 100bp on each side of the SNP, Figure 4.2c. The size of the flanking sequence can be set to different sizes to allow the design of different types of primers. Different homoeologues may contain small indels, Figure 4.2d. To enable a comparasion base-per-base, a local alignment with `MAFFT` (Kato and Standley, 2013) is produced, Figure 4.2e.

PolyMarker searches across each base in the local alignment to identify the variations across homoeologues and the target marker. A mask is produced to highlight the bases with a variations, Figure 4.2f, on the following categories:

<b>Specific</b>	Homoeologous polymorphism which is only present in the target genome (upper case).
-----------------	--

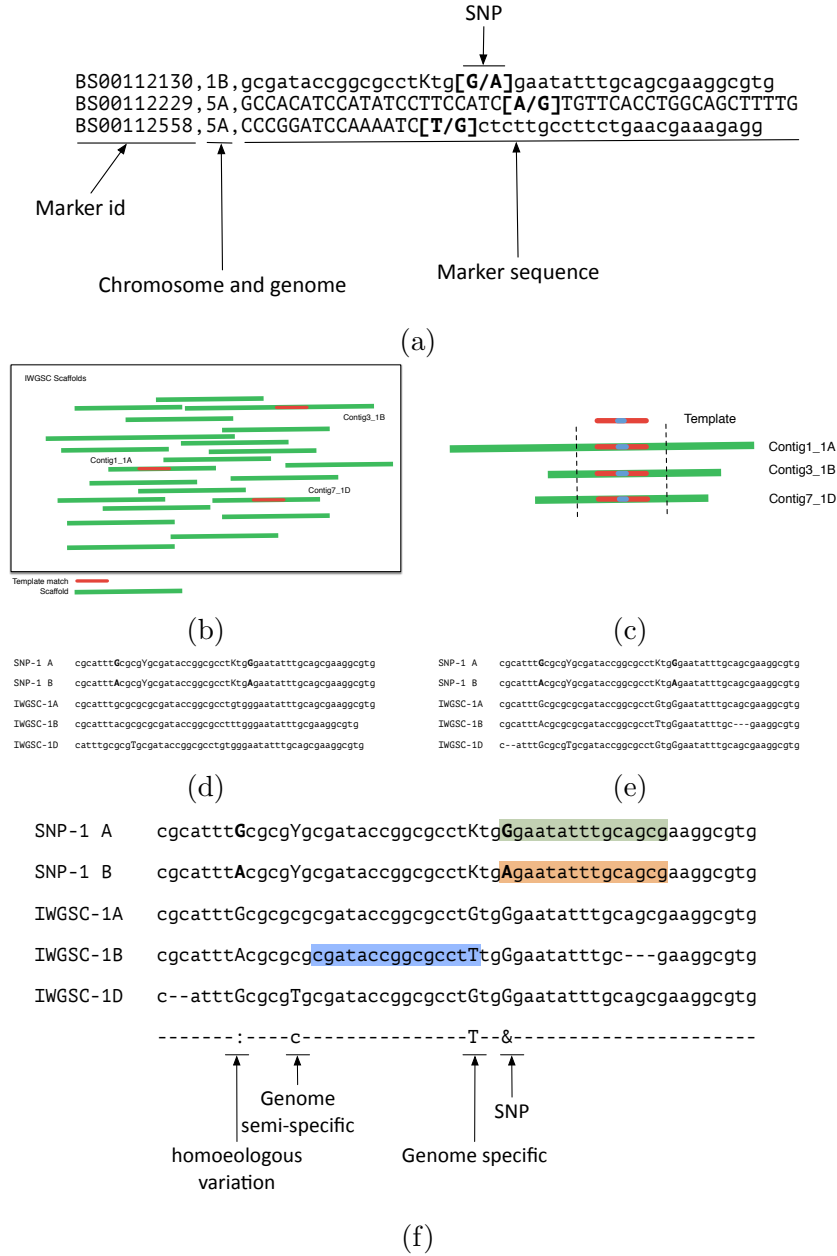


Figure 4.2: Alignments done by PolyMarker. (a) input. The alternative alleles are surrounded by brackets. (b) Global search of templates in the reference contigs. (c) Selected regions around the SNP on every chromosome. (d) Sequence of found regions around the SNP. (e) Local alignment on regions around the SNP detects indels. (f) Alignment with mask and primer candidates.

<b>Semi-specific</b>	Homoeologous polymorphism which is found in 2 of the 3 genomes, hence it discriminates against one of the off-target genomes or when not all the homoeologous sequences were found (lower case).
<b>Non-specific</b>	No variation is found across homoeologues (-).
<b>Homoeologous</b>	The target SNP is present across different chromosomes, so candidate SNP markers on this category are not expected to be reliably identify the allele (:).
<b>Non-homoeologous</b>	The target SNP is not present across chromosomes, so it can be used to identify an allele (&).

PolyMarker was designed to produce SNP assays for KASP genotyping (LGC Genomics, 2013), which requires a common primer and two allele-specific primers. The common primer is selected to start on a position from a: Specific; Semi-specific or; Non-specific, on that priority. This means that the common primer will be as specific as possible in the region. For the allele-specific primers, the starting position of the primer is on the base with the SNP. To ensure that the stability of the candidate primers will be met, the putative starting positions are tested with *Primer3* (Rozen and Skaletsky, 2000).

PolyMarker was designed and validated with the markers described in section 3.7. For wheat, PolyMarker uses the contigs from Mayer et al. (2014), as deposited in Ensembl. As new releases of the wheat genome are made available, different parsers to assign the chromosome to each sequence can be added with little effort to PolyMarker.

## 4.2 Applications of PolyMarker

PolyMarker is not restricted to wheat or to KASP assays, the source code is flexible and can be extended for other types of analysis. On each of the following projects, PolyMarker has been adapted to design primers in species where KASP hasn't been used before, the primers are used for regular PCR amplification, or the use of KASP is not the conventional SNP calling.

### 4.2.1 KASP assays for public sets of SNPs

PolyMarker was used to design KASP assays for the 81,587 markers from (Wang et al., 2014), available on the PolyMarker website and in CerealsDB (Wilkinson et al., 2012). Of those markers, 40,267 were designed using the target chromosome using the genetic map published by the genetic map. Genes without a genetic position were aligned to scaffolds sorted by chromosome from the International Wheat Genome Sequencing Consortium (Mayer et al., 2014) with BLAT (Kent, 2002) and the best hit was selected as putative location. 97.5% of the assays were designed and 76% of them are semi-specific or specific, thereby improving their expected performance with respect to randomly designed primers (Table A.2). A subset of the designed assay was used to genotype a mapping population to find resistance to Fusarium head blight (Burt et al., 2015).

### 4.2.2 SNPs in a mutant population

PolyMarker was used to design primers to validate SNPs in a Targeted Induced Local Lesions in Genomes (TILLING) population, an approach to identify the function of genes by mutating them. To calibrate the SNP calling, KASP assays were designed to get the mutations from  $M_2$ ,  $M_3$  and,  $M_5$  mutants (King et al., 2015). Then primers were designed for the whole mutant population, consisting of 1,200 Cadenza (Hexaploid) and 1,535 Kronos (Tetraploid) wheat lines (Krasileva et al., submitted 2016). Genome-specific primers 172 and 80 SNP assays on 19 and 8  $M_4$  Cadenza and Kronos lines respectively. Of those, 71(85.5%) Kronos and 147(88.8%) of the Cadenza primers were valid assays (Tables A.5 and A.6).

### 4.2.3 Deletions on a mutant population

On some of the TILLING mutant lines long deletions were detected (Krasileva et al., submitted 2016). To validate the deletions is possible to use KASP assays to produce primers that amplify homoeologues. PolyMarker was modified to search for variations across homoeologues to select a common primer that will amplify two genomes, Figures 4.3a and 4.3b. On lines without the targeted deletion, the amplification correspond to an homozygous assay, Figure 4.3c. When a deletion is present



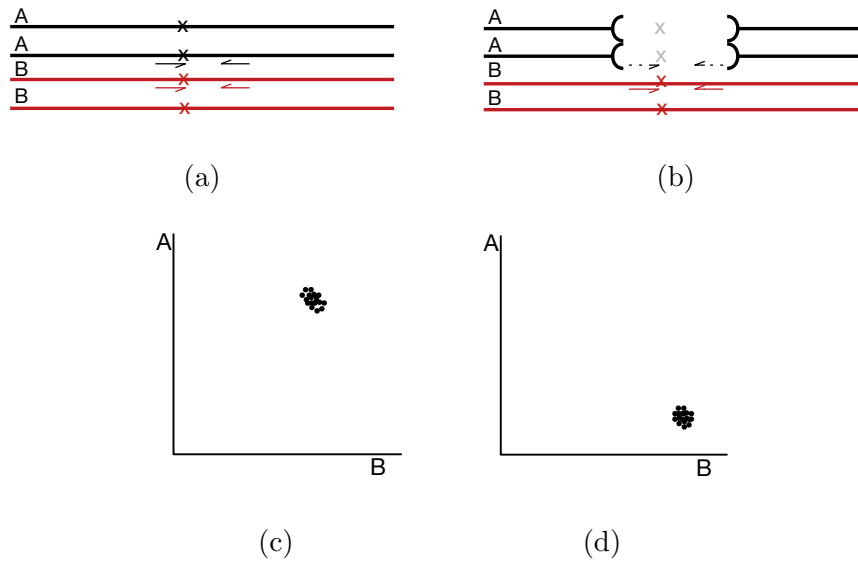


Figure 4.3: KASP assays to validate homozygous deletions. (a) Primer positions for wildtype. (b) Primer positions on homozygous deletion on  $M_4$  (c) Heterozygous amplification on wildtype, including both homoeologues. (d) Homozygous amplification on deletion line, only the non-deleted homoeologue is amplified.

the results of the assay look like an homozygous sample, with the intensity of the assay towards the the conserved homoeologue, Figure 4.3d. A set of KASP assays for the the deletions and mutations located on the same chromosome where designed to validate 11 homozygous deletions on  $m_4$  plants. In all cases the segregation of the mutations was as expected, except for a predicted heterozygous mutation that was called as homozygous. Also, all the KASP assays that contained a deletion were called homozygous, as expected. To ensure that the calls didn't come from a single cluster, 4 wildtype plants were genotyped and the markers for deletions where called as heterozygous. An example of a validated deletion, with the calls for each individual is shown on Table A.4.

#### 4.2.4 PolyMarker public web service

To make PolyMarker accessible to the community, a web server that allow the submission of SNPs was developed. The web interface consists on two virtual machines, one with a web facing interface that stores the queries, and a dedicated node to submit jobs to an HPC cluster. The on-line interface further simplifies the design of KASP assays, a process that used to take a couple of weeks now is done in a couple of hours.

Since the release of the public service in July 2014 until August 2016, 1,739 requests to PolyMarker have been done.

#### 4.2.5 Genotyping of *Puccinia striiformis* f. sp. *tritici* isolates.

In Hubbard et al. (2015), *Puccinia striiformis* f. sp. *tritici* (PST) isolates were sequenced and assigned to clusters, according to their genotype. The clusters are useful to monitor the changes in the pathogen population, which can be used to predict if certain wheat lines will be resistant to the isolates in the field. PolyMarker was used to design primers for PST, using the assembly PST-130 (Cantu et al., 2011). Out of 15 assays 11 can be used to identify to which cluster of isolates a sample is likely to belong, Supplemental Table A.3.

### 4.3 Discussion

PolyMarker is a tool that was born as part of the validation of the SNPs found in Chapter 3. Originally, the primer design was ought to be done manually, a slow, error-prone and, repetitive process. The steps require the use of several bioinformatics tools, but once I figured out the steps I decided to automate the process. Since designing genome-specific primers is a common task in wheat research and breeding, the community showed interest on the tool and I decided to refine it and make it open source. PolyMarker has been used successfully in several projects and it even allowed the novel use of KASP assays to validate long deletions in polyploids.

The current web interface of PolyMarker is limited to KASP assays, however the command line version is more flexible and has been used to design primers for PCR amplicons, capillary sequencing and on other organisms. The ideas behind PolyMarker had been taken by other projects like the scripts described in Ma et al. (2015) and the corresponding web interface, GSP (Wang et al., 2016). As new references of wheat come available, PolyMarker should be updated to work with pseudomolecules and the web interface updated accordingly.

# Chapter 5

## Gene expression (expVIP)

### 5.1 Expression experiments (Introduction)

Describe the list of previously published expression experiments and how they can potentially be used as a framework for new experiments.

### 5.2 Database design

Description of how the database was designed and the flexibility given by having the factors and units as variables

### 5.3 Analysis pipeline

Implementation of the pipeline, from running kallisto to load the data in the database

### 5.4 Graphical interface

How the expression can be displayed filtered, and sorted

### 5.5 Conclusions

The use of previously published studies is a valuable resource. Also, mention that despite the fact that there are several expression/gene browsers, none of them allow comparisons between species and don't consider polyploids.

# Chapter 6

## Conclusions and final remarks

This section wraps up by showing the relationship and importance of a comprehensive approach to data analysis, from the field, genetics, molecular biology and genomics. I will also remark how the technology and the resources have changed in the last 4 years. As at the references used at beginning where superseded during the PhD.

# Appendix A

## Supplemental tables

## A.1 Genetic map of *Yr15* with RNA-Seq supplemental tables.

Table A.1: Number of genes with a coverage over 20x, 10x and at least one read ( $>0x$ ).

Coverage	Reference	R1	R2	R3	Bulks			Bulk mixes			Progenitors		
					S1	S2	S3	R1+R2	S1+S2	R1+R2+R3	S1+S2+S3	Yr15	AVS
20x	UCW	16,434	27,871	27,223	32,287	28,669	34,898	33,968	41,019	40,985	47,507	36,808	42,248
	UniGene v60	17%	30%	29%	34%	30%	37%	36%	44%	44%	50%	39%	45%
10x	UCW	9,643	16,182	15,222	19,549	17,397	20,567	20,219	25,270	24,598	29,052	22,107	25,842
	UniGene v60	17%	28%	27%	34%	31%	36%	36%	44%	43%	51%	39%	45%
>0x	UCW	27,371	38,282	37,777	42,658	38,999	44,610	43,266	49,473	49,182	54,781	46,356	50,760
	UniGene v60	29%	41%	40%	45%	41%	47%	46%	53%	52%	58%	49%	54%
>0x	UCW	16,201	22,948	22,130	26,200	24,130	26,914	26,318	30,579	29,857	33,557	28,044	31,095
	UniGene v60	28%	40%	39%	46%	42%	47%	46%	54%	52%	59%	49%	55%
>0x	UCW	68,302	72,484	72,957	74,694	73,290	75,201	74,397	77,093	76,715	78,796	76,275	77,080
	UniGene v60	73%	77%	77%	79%	78%	80%	79%	82%	81%	84%	81%	82%
>0x	UCW	40,717	42,489	42,595	43,625	43,059	43,748	43,393	44,655	44,364	45,392	43,732	44,596
	UniGene v60	71%	75%	75%	77%	76%	77%	76%	78%	78%	80%	77%	78%

Table A.2: Count of KASP assays designed for the 40,267 SNP markers located in the genetic map from Wang et al. (2014). 4,228 assays did not align to the target chromosome. Not designed: Primer3 could not find viable primers flanking the SNP.

	Homoeologous variant	Varietal SNP	Percentage
Non-specific	1,765	5,857	21.15%
Semi-specific	7,942	6,907	41.20%
Specific	6,813	5,957	35.43%
Not designed	242	556	2.21%
Total	16,762	19,277	36,039

## A.2 PolyMarker supplemental tables.



Table A.3: PolyMarker used to genotype PST

Assay	Contig	Position	X	Y	Cluster I isolates		Cluster II isolates		Cluster III isolates		Cluster IV isolates	
					13/26	13/123	CL1	T-13/3	13/09	13/23	13/182	13/36
1	PST130.14470	268	C	T	X:Y	X:Y	X:X	X:X	X:X	X:X	X:X	X:X
2	PST130.8160	11876	C	T	Y:Y	Y:Y	X:Y	X:Y	X:Y	X:Y	X:Y	X:Y
3	PST130.14628	1712	A	C	X:Y	-	X:X	X:X	X:X	X:X	X:X	X:X
4	PST130.14898	503	G	A	X:X	X:X	X:Y	X:Y	X:Y	X:Y	X:Y	X:Y
5	PST130.28344	2372	A	G	Y:Y	Y:Y	X:Y	X:Y	Y:Y	Y:Y	Y:Y	Y:Y
6	PST130.7634	3463	A	C	Y:Y	Y:Y	X:Y	X:Y	Y:Y	Y:Y	Y:Y	Y:Y
7	PST130.7629	11699	G	A	Y:Y	Y:Y	X:Y	X:Y	Y:Y	Y:Y	Y:Y	Y:Y
8	PST130.10943	2979	C	T	X:Y	Y:Y	X:Y	X:X	X:X	X:X	X:Y	X:Y
9	PST130.10126	6216	G	T	Y:Y	Y:Y	X:X	X:X	X:X	X:X	Y:Y	Y:Y
10	PST130.22010	172	C	T	Y:Y	Y:Y	Y:Y	X:Y	X:Y	X:Y	X:Y	X:Y
11	PST130.16961	1098	C	T	X:X	X:X	X:Y	X:Y	Y:Y	Y:Y	X:Y	X:Y
12	PST130.6915	2710	A	T	Y:Y	Y:Y	Y:Y	Y:Y	X:Y	X:Y	Y:Y	Y:Y
13	PST130.12479	1428	C	T	X:X	X:X	Y:Y	X:X	X:X	X:X	Y:Y	X:X
14	PST130.7634	3883	C	G	X:X	X:X	X:Y	X:Y	X:X	X:X	X:Y	X:X
15	PST130.14470	456	T	C	Y:Y	Y:Y	X:Y	X:Y	Y:Y	Y:Y	Y:Y	Y:Y

Table A.4: Validation of homozygous deletions on line Cadenza0423.

Marker	Deletion	chr	cM	1	2	3	4	5	6	7	8	9	10	11	12	C	C	C	C	Result
5BS_2297308_Cadenza0423.12664_C12664T	-	5B	4.551	X	X	-	X	X	X	X	X	X	X	-	X	Y	Y	Y	Y	HOM Mutation
5BL_10812849_Cadenza0423.5664_G5664T	-	5B	38.769	X	X	-	X	X	X	X	X	X	X	-	X	Y	Y	Y	Y	HOM Mutation
5BL_10825062_Cadenza0423.7917_G7917A	-	5B	38.769	X	X	-	X	X	X	X	X	X	X	-	X	Y	Y	Y	Y	HOM Mutation
IWGSC_CSS_5BL_scaff.10847976:27068-27231	+	5B	38.769	X	X	-	X	X	X	X	X	X	X	-	X	H	H	H	H	Hom Deletion
IWGSC_CSS_5BL_scaff.10847976:28118-28674	+	5B	38.769	X	X	-	X	X	X	X	X	X	X	-	X	H	H	H	H	Hom Deletion
IWGSC_CSS_5BL_scaff.10865441:15863-15946	+	5B	38.769	X	X	-	X	X	X	X	X	X	X	-	X	H	H	H	H	Hom Deletion
5BL_10837222_Cadenza0423.4616_G4616A	-	5B	39.905	X	X	-	X	X	X	X	X	X	X	-	X	Y	Y	Y	Y	HOM Mutation
5BL_10891320_Cadenza0423.18847_C18847T	-	5B	45.594	Y	Y	-	Y	H	X	X	Y	H	Y	-	H	Y	Y	Y	Y	HET Mutation

Table A.5: Validation of mutations on  $M_4$  on Cadenza

IWGSC contig	Line	Pos	WT	Mut	Predicted	$M_4$	Primer 1 (Cadenza)	Primer 2 (mutant)	Common Primer
IWGSC_CSS_3B_scaff_10445294	Cadenza1772	6019	C	T	het	het	caggatAgtGggactgtcaaaG	caggatAgtGggactgtcaaaA	ggagacGGctGtggacatT
IWGSC_CSS_3DL_scaff_6955403	Cadenza1772	2418	C	T	het*	hom	tcagCggattgtcgggatG	tcagCggattgtcgggatA	tgteCatgaaTcttgtccacG
IWGSC_CSS_4AL_scaff_7106846	Cadenza1772	11277	G	A	hom	hom	tgggatccatgcctacactG	tgggatccatgcctacactA	gatggTgatttgcgcctA
IWGSC_CSS_4AS_scaff_5991335	Cadenza1772	15710	G	A	hom	hom	ctggccctgcgctgctaC	ctggccctgcgctgctaT	gtggaaGttcagaaggaccaG
IWGSC_CSS_4BS_scaff_4956646	Cadenza1772	252	G	A	het*	hom	gcaggttgacttcccgaG	gcaggttgacttcccgaA	tGaggtacgaGcTaaagAaagC
IWGSC_CSS_4DS_scaff_1715962	Cadenza1772	1225	G	A	hom	hom	cagctgtggTatctcaactgG	cagctgtggTatctcaactgA	CcCtGaaACACcGtttggAT
IWGSC_CSS_5AL_scaff_2763407	Cadenza1772	2119	G	A	hom	hom	gcgacGaacctcgagatctG	gcgacGaacctcgagatctA	gaTggcaAtcgtCgtgcA
IWGSC_CSS_5AS_scaff_1548786	Cadenza1772	12625	C	T	het	het	AtaggcacattgctagactgaG	AtaggcacattgctagactgaA	ggattgggtgttcacgC
IWGSC_CSS_5BL_scaff_10849226	Cadenza1772	2289	C	T	het*	hom	cctgacatcattgttcacgatC	cctgacatcattgttcacgatT	cactccgaggtgtccatgaT
IWGSC_CSS_5BS_scaff_2270737	Cadenza1772	2262	G	A	hom	—	attcCTgtgttggtggCaaatgaG	attcCTgtgttggtggCaaatgaA	taaGcaciaAccctccagctgG
IWGSC_CSS_1AL_scaff_3022915	Cadenza1661	891	C	T	hom	hom	ccacagtgcgactcctattgaCG	ccacagtgcgactcctattgaCA	atgtctgattcGtcGtagtcC
IWGSC_CSS_1AS_scaff_3297240	Cadenza1661	1970	C	T	het	het	catcccgccGtttctcC	catcccgccGtttctcT	gtctcgccgatgaagagcT
IWGSC_CSS_1BL_scaff_3828996	Cadenza1661	1340	G	A	hom	hom	agccggatgttagtgttaacC	agccggatgttagtgttaacT	agcagcttgTcgcgttaaC
IWGSC_CSS_1DS_scaff_1884529	Cadenza1661	10575	G	A	hom	hom	aCagatacaAttgtcatgcaggC	aCagatacaAttgtcatgcaggT	acctgggTTgtccaatacttC
IWGSC_CSS_2AL_scaff_6318370	Cadenza1661	19142	C	T	het	—	cgtggcCgaatCtcGacG	cgtggcCgaatCtcGacA	ttcttggggagccgggC
IWGSC_CSS_2AS_scaff_5213460	Cadenza1661	1358	G	A	hom	hom	gtcacgaaCccgctcagA	gtcacgaaCccgctcagA	aggaaagagagaaaagaGcG
IWGSC_CSS_2BS_scaff_5179331	Cadenza1661	5604	G	A	het	het	actctcgtcaagaactgatacaG	actctcgtcaagaactgatacaA	gcaGagaatgttcttgaacT
IWGSC_CSS_2DS_scaff_5341235	Cadenza1661	4673	G	A	het	het	ggtaggatctcggagctG	ggtaggatctcggagctA	gcgcggtcgtacaggttG
IWGSC_CSS_3AL_scaff_4250995	Cadenza1661	7046	G	A	hom	hom	cCaagaaacgggtgggtccaG	cCaagaaacgggtgggtccaA	ctgcagctgtccatcatcgT
IWGSC_CSS_3BS_scaff_10404421	Cadenza1661	4303	G	A	het	het	ccttcgtcgaCaggacctG	ccttcgtcgaCaggacctA	GCcagtaactCacAtgtctC
IWGSC_CSS_5DL_scaff_2390496	Cadenza1538	2125	C	T	hom	het	gcagttttatcctcagtagtcttgG	gcagttttatcctcagtagtcttgA	ttctgagaaTgtaagtgcGatG
IWGSC_CSS_6AL_scaff_5753680	Cadenza1538	3920	C	T	hom	hom	tgctccaaatttgagcaciaTaaC	tgctccaaatttgagcaciaTaaT	aaatgcaaggggttaagtttttG
IWGSC_CSS_6AS_scaff_4425792	Cadenza1538	4307	G	A	hom	het	agatgcttgtCggGccaG	agatgcttgtCggGccaA	gctgaagcaacgcgatcaaT
IWGSC_CSS_6BS_scaff_3003630	Cadenza1538	6933	C	T	het	het	ggcagtaagtgtgtgctgagC	ggcagtaagtgtgtgctgagT	tTgaCttctggtttgtggcA
IWGSC_CSS_6DL_scaff_3246988	Cadenza1538	9186	G	A	het	het	gctaaagaagagcttgagagaattC	gctaaagaagagcttgagagaattT	aatttctgaagagaggtgtgtatG
IWGSC_CSS_7AL_scaff_4480114	Cadenza1538	3446	C	T	het	—	gatatctcccacacggcgG	gatatctcccacacggcgA	tgagccactcttcagtttT
IWGSC_CSS_7AS_scaff_4193541	Cadenza1538	8359	C	T	hom	het	agcaattctttggctatcaattagC	agcaattctttggctatcaattagT	tcactGtcttaactctactgctG
IWGSC_CSS_7BL_scaff_6721572	Cadenza1538	9223	C	T	het	het	gctCaggaggagagacaagaaG	gctCaggaggagagacaagaaA	tgctatgaagaattccgacctC
IWGSC_CSS_7BS_scaff_3152545	Cadenza1538	3960	G	A	hom	—	tcagcaaaatcacctgcCgC	tcagcaaaatcacctgcCgT	gCtgcccatcatcgtttaT
IWGSC_CSS_7DS_scaff_3963838	Cadenza1538	2913	G	A	het	het	tCgttgcaagcCttTtgtgT	tCgttgcaagcCttTtgtgT	agaGttaTcaageTactgtcacA
IWGSC_CSS_1AL_scaff_3903380	Cadenza1469	6193	G	A	hom	hom	ctcttcAgagatgaacgcgA	ctcttcAgagatgaacgcgA	tcGtGagatgGtggtttGTTA
IWGSC_CSS_1AS_scaff_3287728	Cadenza1469	3817	C	T	het*	hom	ccgaccaAttcactaacggG	ccgaccaAttcactaacggA	accctctttcccAgacatgaT
IWGSC_CSS_1BL_scaff_3815304	Cadenza1469	513	G	A	hom	hom	aacatttgctTaCcaaaacGC	aacatttgctTaCcaaaacGT	acacagcaagttataatgCAAAGC
IWGSC_CSS_1DL_scaff_2266648	Cadenza1469	5926	C	T	het	het	caacatgagacacacaccttC	caacatgagacacacaccttT	gtcaacgcgtgaggattgtC
IWGSC_CSS_1DS_scaff_1906671	Cadenza1469	3697	C	T	hom	hom	tggTGTgagacacttggcgA	tggTGTgagacacttggcgA	catggcgacaccAcctG
IWGSC_CSS_2AL_scaff_6337088	Cadenza1469	7334	G	A	het*	hom	acaatgccAagttgacaggttG	acaatgccAagttgacaggttA	gggagtggtgttCagaacaT

IWGSC contig	Line	Pos	WT	Mut	Predicted	$M_4$	Primer 1 (Cadenza)	Primer 2 (mutant)	Common Primer
IWGSC_CSS_2BL_scaff.7972799	Cadenza1469	8995	C	T	het	hom	gTgCtcctcGgcaccccttC	gTgCtcctcGgcaccccttT	gatccgGgcaaacacgTG
IWGSC_CSS_2DL_scaff.9832343	Cadenza1469	3262	G	A	het	het	TtgtctaAcagcacCGcagG	TtgtctaAcagcacCGcagA	agatctcggtcagcctttcT
IWGSC_CSS_2DS_scaff.5327939	Cadenza1469	3889	G	A	het	het	ttttTgccttatgtgactctagtaC	ttttTgccttatgtgactctagtaT	gaggccatcacagatagcG
IWGSC_CSS_3B_scaff.10395219	Cadenza1469	1292	G	A	hom	—	agggtccttgctgcttgctgG	agggtccttgctgcttgctgA	cctctctgggggctttataC
IWGSC_CSS_3B_scaff.10592217	Cadenza0580	2994	C	T	het	—	acagcagtatcaagccctcC	acagcagtatcaagccctcT	tgatactgttgTggCggagG
IWGSC_CSS_3DS_scaff.2596771	Cadenza0580	1037	G	A	het	het	tggttatgCAcaggataatCagG	tggttatgCAcaggataatCagA	tggcaaatgtgatgtcattaggT
IWGSC_CSS_4AL_scaff.7093953	Cadenza0580	9881	C	T	hom	hom	GacaggaagccggtaacaC	GacaggaagccggtaacaT	ctccAgcaggcatgggaT
IWGSC_CSS_4BL_scaff.7037448	Cadenza0580	1837	C	T	hom	hom	CgttgaaaaGctgcaagaacttaaC	CgttgaaaaGctgcaagaacttaaT	cagttcttccTtCaGagcagataT
IWGSC_CSS_4BS_scaff.4929479	Cadenza0580	10668	G	A	hom	—	tggattttcccgcactgttC	tggattttcccgcactgttT	gtaaacaggcatttcaagagtcA
IWGSC_CSS_4DL_scaff.14359838	Cadenza0580	1408	G	A	hom	—	gCtcAttcagggatTGTcCtaTatG	gCtcAttcagggatTGTcCtaTatA	tgaCagaacagttggctacacT
IWGSC_CSS_4DS_scaff.2276484	Cadenza0580	8034	G	A	hom	hom	gccgtggttgatggAgaG	gccgtggttgatggAgaA	cgtccagattactgatacttgcA
IWGSC_CSS_5AL_scaff.2756579	Cadenza0580	5278	G	A	het	het	tgaatggatttttctgccgttC	tgaatggatttttctgccgttT	ggAAatCCTATgCagaAgAaaCTG
IWGSC_CSS_5BL_scaff.10787208	Cadenza0580	10627	G	A	het	—	gcctctcacatcgcgagaC	gcctctcacatcgcgagaT	acgatgtcAggtggGcgT
IWGSC_CSS_5BS_scaff.2282179	Cadenza0580	5267	G	A	het	—	tgatgggctacgacgtgC	tgatgggctacgacgtgT	tcggcgcccttgaaAtcC
IWGSC_CSS_5DL_scaff.4498073	Cadenza0423	4937	C	T	hom	hom	gcaccctctggttggtcatC	gcaccctctggttggtcatT	tgacagcaAagcagccG
IWGSC_CSS_5DS_scaff.2738970	Cadenza0423	2319	C	T	het	—	cgtgaggtgggtgatttG	cgtgaggtgggtgatttT	tggaaactagtacactgcagtTC
IWGSC_CSS_6AL_scaff.5757109	Cadenza0423	2788	G	A	hom	hom	caggaGcctggcaataaaGG	caggaGcctggcaataaaGA	ctttcGagtcctcttagtttcG
IWGSC_CSS_6AS_scaff.4387871	Cadenza0423	2543	G	A	hom	hom	gcatgctaacaggcgaaaaG	gcatgctaacaggcgaaaaA	ctcatgctcctgatcttaaggtC
IWGSC_CSS_6BL_scaff.4271391	Cadenza0423	4660	C	T	hom	hom	tacgtgcatgatgtggtagtctgaC	tacgtgcatgatgtggtagtctgaT	gtttgaaagtgcacagatgTaccA
IWGSC_CSS_6DS_scaff.1880206	Cadenza0423	9159	G	A	het	het	ctgCgaaggctccacaaG	ctgCgaaggctccacaaA	ggatgagaagtttgcattgctC
IWGSC_CSS_7AS_scaff.4227506	Cadenza0423	952	G	A	het	—	ccatgtgtttccaatgttagagC	ccatgtgtttccaatgttagagT	tgccctagctggtatgcT
IWGSC_CSS_7BL_scaff.6681782	Cadenza0423	1486	C	T	hom	hom	agtaagCGtgacagcaatggG	agtaagCGtgacagcaatggA	AtgtctTtgGtggaagtacatcA
IWGSC_CSS_7BS_scaff.3160328	Cadenza0423	7801	C	T	het	het	tgttaaatGatacagCctgcagC	tgttaaatGatacagCctgcagT	tggaaatgggtCgttgtttT
IWGSC_CSS_7DS_scaff.407428	Cadenza0423	2051	G	A	het	het	gtcGCgccatcctgacaG	gtcGCgccatcctgacaA	actcatcAggtcagcccaA
IWGSC_CSS_3AL_scaff.442479	Cadenza0364	3198	C	T	het	het	gagtcATaagttggtaagattggC	gagtcATaagttggtaagattggT	GCaGaTaaCaacaggatcacG
IWGSC_CSS_3AL_scaff.4447942	Cadenza0364	11917	G	A	het	het	gtcataaaagattgctcctgtgaaG	gtcataaaagattgctcctgtgaaA	ctcGgatgtgggaggaagA
IWGSC_CSS_3AS_scaff.1557483	Cadenza0364	2547	C	T	het	het	aaagtcacatcatgcttaccataaG	aaagtcacatcatgcttaccataaA	cgaataccaacgcctcatcA
IWGSC_CSS_3AS_scaff.2648747	Cadenza0364	2688	G	A	het	het	tggAagcAcaaggggccC	tggAagcAcaaggggccT	GccgccgatggagactcG
IWGSC_CSS_3AS_scaff.3304956	Cadenza0364	1017	G	A	het	het	gtcccttgacacagctttG	gtcccttgacacagctttA	cctgctggactacaactcaaT
IWGSC_CSS_3AS_scaff.3321091	Cadenza0364	4585	C	T	het	het	caagaatgATgctgatgttggaG	caagaatgATgctgatgttggaA	acatgctgaatgccgaatC
IWGSC_CSS_3AS_scaff.3371333	Cadenza0364	538	G	A	het	het	gggaaaCgAgAcgagcgG	gggaaaCgAgAcgagcgA	ccgtgcttctctaccctT
IWGSC_CSS_3AS_scaff.3371815	Cadenza0364	1061	C	T	het	het	atccccacggcacagagG	atccccacggcacagagA	aAttggcccttggtgattcC
IWGSC_CSS_3AS_scaff.3440912	Cadenza0364	4498	G	A	het	het	ccgtaaaaactttctgtgcttgC	ccgtaaaaactttctgtgcttgT	atActgacaaactacatgatgtgC
IWGSC_CSS_3B_scaff.10343586	Cadenza0364	2242	G	A	het	—	gggttcTgTcctctcttccactG	gggttcTgTcctctcttccactA	tgtgttgaaccgcgaagcA
IWGSC_CSS_3AL_scaff.442479	Cadenza0364	3198	C	T	het	het	gagtcATaagttggtaagattggC	gagtcATaagttggtaagattggT	GCaGaTaaCaacaggatcacG
IWGSC_CSS_3AL_scaff.4447942	Cadenza0364	11917	G	A	het	het	gtcataaaagattgctcctgtgaaG	gtcataaaagattgctcctgtgaaA	ctcGgatgtgggaggaagA
IWGSC_CSS_3AS_scaff.1557483	Cadenza0364	2547	C	T	het	het	aaagtcacatcatgcttaccataaG	aaagtcacatcatgcttaccataaA	cgaataccaacgcctcatcA
IWGSC_CSS_3AS_scaff.2648747	Cadenza0364	2688	G	A	het	het	tggAagcAcaaggggccC	tggAagcAcaaggggccT	GccgccgatggagactcG
IWGSC_CSS_3AS_scaff.3304956	Cadenza0364	1017	G	A	het	het	gtcccttgacacagctttG	gtcccttgacacagctttA	cctgctggactacaactcaaT
IWGSC_CSS_3AS_scaff.3321091	Cadenza0364	4585	C	T	het	het	caagaatgATgctgatgttggaG	caagaatgATgctgatgttggaA	acatgctgaatgccgaatC

IWGSC contig	Line	Pos	WT	Mut	Predicted	$M_4$	Primer 1 (Cadenza)	Primer 2 (mutant)	Common Primer
IWGSC_CSS_3AS_scaff_3371333	Cadenza0364	538	G	A	het	het	gggaaaCgAgAcgagcgG	gggaaaCgAgAcgagcgA	cctgtccttctcaccctT
IWGSC_CSS_3AS_scaff_3371815	Cadenza0364	1061	C	T	het	het	atccccacggcacagagG	atccccacggcacagagA	aAttgcccttggtgattcC
IWGSC_CSS_3AS_scaff_3440912	Cadenza0364	4498	G	A	het	het	ccgtaaaaactttctgtgcttgC	ccgtaaaaactttctgtgcttgT	atActgacaaactacatgatgtgC
IWGSC_CSS_3B_scaff_10343586	Cadenza0364	2242	G	A	het	—	ggttcTgTcctctcttccactG	ggttcTgTcctctcttccactA	tgtgttgaaccgcaagcA
IWGSC_CSS_5DL_scaff_242342	Cadenza0281	2433	C	T	hom	hom	catggCgacggtGtcctG	catggCgacggtGtcctA	aAccctcatTTtggCTACTtCT
IWGSC_CSS_5DL_scaff_4538822	Cadenza0281	1208	G	A	hom	—	acgtcagaacaaccgtttgaC	acgtcagaacaaccgtttgaT	ttaaattggttggcgccacC
IWGSC_CSS_6AL_scaff_5813297	Cadenza0281	4532	C	T	hom	—	gggagaggggacgtctcgG	gggagaggggacgtctcgA	ttctctgcacaacgattccG
IWGSC_CSS_6AS_scaff_4378990	Cadenza0281	6748	C	T	hom	hom	cccaggttctgctcttttcC	cccaggttctgctcttttcT	caagtatcaagaaatgaaggGgT
IWGSC_CSS_6BL_scaff_4360781	Cadenza0281	5426	C	T	het	het	aCtactcaaatggcttGgtgtaG	aCtactcaaatggcttGgtgtaA	tcagtccaacatgTcaagagatT
IWGSC_CSS_7AL_scaff_4488310	Cadenza0281	3808	G	A	hom	hom	gttctctttagtagcagccG	gttctctttagtagcagccA	ggcgctttcttcggcctA
IWGSC_CSS_7BL_scaff_6696509	Cadenza0281	9232	G	A	het	het	gctctaggGgtggcaaAagG	gctctaggGgtggcaaAagA	ggcttGaGgtcGcagtgT
IWGSC_CSS_7BS_scaff_3143575	Cadenza0281	1866	C	T	het	het	agatgttgagagggcgcttC	agatgttgagagggcgcttT	gcttggAtgtgggcaagtT
IWGSC_CSS_7DL_scaff_3346250	Cadenza0281	1663	G	A	het	het	acgtgcagcaacatcctaaC	acgtgcagcaacatcctaaT	TttcccaccaggccaagA
IWGSC_CSS_7DS_scaff_3933917	Cadenza0281	1243	C	T	het	het	tgCtgagcCttTcaccttgC	tgCtgagcCttTcaccttgT	agaggtttggttccatcGG
IWGSC_CSS_3B_scaff_10626860	Cadenza0148	7847	G	A	het	het	gcagctctgggaaggagA	gcagctctgggaaggagA	gttaatgtacCTcctagctcG
IWGSC_CSS_3DL_scaff_6915683	Cadenza0148	6904	C	T	het	het	cgtcaaCctgtgggcaattG	cgtcaaCctgtgggcaattA	tcatgctcataatgTcatagggT
IWGSC_CSS_4AS_scaff_5929057	Cadenza0148	4238	G	A	hom	hom	gcgcaacgtagCacctacC	gcgcaacgtagCacctacT	ttatctggtgaagtgaacggttCA
IWGSC_CSS_4AS_scaff_5950625	Cadenza0148	10590	C	T	het	het	agaTattCaaaTcggtggAttggC	agaTattCaaaTcggtggAttggT	cctgCtccctcacgtcC
IWGSC_CSS_4AS_scaff_5967119	Cadenza0148	11626	C	T	hom	hom	cgtGgacaccccgagctG	cgtGgacaccccgagctA	gacagcgacatgcacgaC
IWGSC_CSS_4DL_scaff_14455742	Cadenza0148	1946	C	T	hom	hom	gCctgagggagatcgcgC	gCctgagggagatcgcgT	aaccgGtAaCTGtGgGcA
IWGSC_CSS_4DS_scaff_2318993	Cadenza0148	4000	C	T	hom	hom	tccagtttgacacagattgaatggG	tccagtttgacacagattgaatggA	tgagaTtctgtttctttcacAttG
IWGSC_CSS_5AL_scaff_2750707	Cadenza0148	4603	G	A	het	het	ccttgggtgtagccatttcaagTaG	ccttgggtgtagccatttcaagTaA	ccaggaTgcAgtgcaatatttcaagG
IWGSC_CSS_5BL_scaff_10794137	Cadenza0148	9235	C	T	hom	hom	gaagctgcttctcgcttG	gaagctgcttctcgcttA	agtatcccttccatataagcagtG
IWGSC_CSS_5BS_scaff_1646558	Cadenza0148	2916	C	T	het	het	gccGtacactcacctAtcctttG	gccGtacactcacctAtcctttA	gcaaTgtccacttAtcatcccT
IWGSC_CSS_1AL_scaff_3883106	Cadenza0110	27536	C	T	het	het	accttccatcactggctgG	accttccatcactggctgA	gtgaagaacaacaggttgaagC
IWGSC_CSS_1BL_scaff_3812829	Cadenza0110	10770	G	A	het*	hom	ccccactccattccagA	ccccactccattccagA	gGatgtgttctgtgctggaA
IWGSC_CSS_1DL_scaff_2266648	Cadenza0110	6156	G	A	het	het	actgcgtggttatgggacC	actgcgtggttatgggacT	ccccatcactgaacacaacA
IWGSC_CSS_1DS_scaff_1889435	Cadenza0110	8826	C	T	hom	hom	aaccatgaattactcggacagG	aaccatgaattactcggacagA	gcctgaagaattgtatcaaaacaG
IWGSC_CSS_2AS_scaff_5268634	Cadenza0110	4636	G	A	het	het	gatccatgtgattggcatgtttG	gatccatgtgattggcatgtttA	TgctgtTggatagcagttacT
IWGSC_CSS_2BL_scaff_7965110	Cadenza0110	15801	C	T	hom	hom	cattgaagcAtacacAattgcAtaC	cattgaagcAtacacAattgcAtaT	gccagagatccagataaggTttA
IWGSC_CSS_2DL_scaff_9852812	Cadenza0110	13788	G	A	hom	hom	atttttgtatggtctcaatcttcgC	atttttgtatggtctcaatcttcgT	gaacgtTcattctgtactgtcT
IWGSC_CSS_2DS_scaff_5371379	Cadenza0110	2166	C	T	hom	hom	agacacaaaactagtGatgcgC	agacacaaaactagtGatgcgT	gctgctgagaatgttTtgtatttG
IWGSC_CSS_3AL_scaff_4384278	Cadenza0110	1276	C	T	het	het	agcTgaactgccccTgtaG	agcTgaactgccccTgtaA	agggaacctCgGtgatgaA
IWGSC_CSS_3AS_scaff_3340122	Cadenza0110	1467	C	T	hom	hom	attcctAgtgttgcggaacatG	attcctAgtgttgcggaacatA	gagaagactagaaagttttcAgcaT
IWGSC_CSS_5DL_scaff_4554222	Cadenza2103	6528	C	T	het*	hom	gctgccctacaaagaaacaaattG	gctgccctacaaagaaacaaattA	aTcccaactatCGaTtttgctataC
IWGSC_CSS_6AL_scaff_5833640	Cadenza2103	7346	C	T	hom	hom	aagaaaagccacaatggtttctC	aagaaaagccacaatggtttctT	aCTctgTcagtgtttcccagC
IWGSC_CSS_6AS_scaff_4429974	Cadenza2103	3867	G	A	hom	hom	GagatgaAttatttgagcatgtggC	GagatgaAttatttgagcatgtggT	ggttccgggtgcataagT
IWGSC_CSS_6DL_scaff_3307626	Cadenza2103	4970	C	T	hom	hom	tgcagatgttgcctgtgtaG	tgcagatgttgcctgtgtaA	tgtagaaggtgattttgtactGtC
IWGSC_CSS_6DS_scaff_2059604	Cadenza2103	5224	G	A	het	—	gctcaatgcatgcTgagtgG	gctcaatgcatgcTgagtgA	tgtcaagtattattttcctgctcG
IWGSC_CSS_7AL_scaff_4552322	Cadenza2103	1412	C	T	het	het	gcaaaggcTgatactccaacaG	gcaaaggcTgatactccaacaA	ggcAAGccAgtataaaagtaaGC

IWGSC contig	Line	Pos	WT	Mut	Predicted	$M_4$	Primer 1 (Cadenza)	Primer 2 (mutant)	Common Primer
IWGSC_CSS.7BS_scaff.3147455	Cadenza2103	4607	G	A	het	—	gcaccttaggatgtgagTtatgC	gcaccttaggatgtgagTtatgT	gcatgtagggtttatttgactgttA
IWGSC_CSS.7DL_scaff.3382467	Cadenza2103	3473	C	T	hom	—	GGTtctgCaGTTTCATAActcatC	GGTtctgCaGTTTCATAActcatT	attgaatcaactgatacGaaGactC
IWGSC_CSS.3B_scaff.10457010	Cadenza0277	10599	G	A	het	het	aaccttggccgcagaacaC	aaccttggccgcagaacaT	actggtcgcacgagaggG
IWGSC_CSS.3B_scaff.10593852	Cadenza0277	10124	C	T	het	het	tgacaggggacgctatacaG	tgacaggggacgctatacaA	gtctaaCTtACattAcccatcagC
IWGSC_CSS.3DS_scaff.2583390	Cadenza0277	663	G	A	hom	hom	actgcactcatacaatActtCtgC	actgcactcatacaatActtCtgT	tcCacctggacagcaagtG
IWGSC_CSS.4AL_scaff.7093953	Cadenza0277	10004	C	T	hom	hom	ccttgattcaatggaTtgTtttgG	ccttgattcaatggaTtgTtttgA	ttcccaaaaTaaaaggaagagC
IWGSC_CSS.4AL_scaff.7176064	Cadenza0277	6220	C	T	het	het	gtgccgtaTtcCgcctgG	gtgccgtaTtcCgcctgA	atgttcgaggggatgggG
IWGSC_CSS.4DL_scaff.14122349	Cadenza0277	1010	C	T	hom	hom	gtcgctgctgCttgtgaG	gtcgctgctgCttgtgaA	ggaacaggcccaaggagG
IWGSC_CSS.5AL_scaff.2736916	Cadenza0277	4296	G	A	het	het	aagaactATgAaaGtaacacacgaC	aagaactATgAaaGtaacacacgaT	ttcGcTttTaagGcAttCtcG
IWGSC_CSS.5BL_scaff.10883744	Cadenza0277	2080	C	T	hom	hom	gcctctttCtgttTagcctcaG	gcctctttCtgttTagcctcaA	cgacaaggttcgtgatTgcA
IWGSC_CSS.1AL_scaff.3932013	Cadenza0548	11765	C	T	hom	hom	accgccaaCccaagacaG	accgccaaCccaagacaA	cccatTAcccGgcGtgcAacG
IWGSC_CSS.1BS_scaff.3417505	Cadenza0548	373	C	T	het	het	gtggtgaggaGGgtgGaG	gtggtgaggaGGgtgGaA	tggtcgcGccagttgttgA
IWGSC_CSS.2AS_scaff.5305619	Cadenza0548	2786	C	T	hom	hom	atacagatgccttAAGtggTtC	atacagatgccttAAGtggTtT	ggaagacaATGctccaggaC
IWGSC_CSS.2AS_scaff.5306489	Cadenza0548	46953	T	G	het	wt	aggttcattgctcatagaagGT	aggttcattgctcatagaagGG	aggctaTAGactcctgtACAgT
IWGSC_CSS.2BL_scaff.7984123	Cadenza0548	11660	G	A	het	het	cattgtggcatagtaatcagtacaG	cattgtggcatagtaatcagtacaA	aatacattgaggaatacaagccC
IWGSC_CSS.2DL_scaff.9907477	Cadenza0548	1363	C	T	hom	hom	tgctccctttgccagaaC	tgctccctttgccagaaT	ggcaaacctgatgtggcatC
IWGSC_CSS.2DS_scaff.5330886	Cadenza0548	5449	G	A	hom	hom	gcattgctcattataactgaacGtG	gcattgctcattataactgaacGtA	catgctgctctctctggacC
IWGSC_CSS.3AL_scaff.4449951	Cadenza0548	633	C	T	het	het	tccaaacctaacagcttaacactaG	tccaaacctaacagcttaacactaA	gtctgcagTGCaatgtgC
IWGSC_CSS.3B_scaff.10479889	Cadenza0097	3339	C	T	hom	—	ttgTttctGgagaagatgcCG	ttgTttctGgagaagatgcCA	ggtgctcattcaAcGgcA
IWGSC_CSS.3B_scaff.10562262	Cadenza0097	7819	C	T	het	het	agaggggtgctatccatAttgG	agaggggtgctatccatAttgA	agcgatgccaaaggcttcC
IWGSC_CSS.4AL_scaff.7040796	Cadenza0097	10772	G	A	hom	hom	acacaacattgccaccagaG	acacaacattgccaccagaA	CAatCgattgcttgcTtctC
IWGSC_CSS.4AL_scaff.7063488	Cadenza0097	6360	C	T	het	het	gcctctcacCttAattgaaagctgC	gcctctcacCttAattgaaagctgT	aggcagtgaggatgtggaagttT
IWGSC_CSS.4AL_scaff.7091701	Cadenza0097	5050	G	A	het	het	catgagcatctgggaggaaaatG	catgagcatctgggaggaaaatA	agcaagggaAtaatgaacggaaA
IWGSC_CSS.4DS_scaff.1845841	Cadenza0097	7110	G	A	hom	hom	aatgTAGctccccatacCgG	aatgTAGctccccatacCgA	actgaacTgcaatcgtTtatggA
IWGSC_CSS.5AL_scaff.2767581	Cadenza0097	3737	G	A	het	het	gagaggtcctcactAtcggC	gagaggtcctcactAtcggT	cgTcatcacaatattgtcggG
IWGSC_CSS.5BL_scaff.10784643	Cadenza0097	1568	C	T	hom	hom	agaaaTAcatggatggatggaCG	agaaaTAcatggatggatggaCA	catctcCCttcaCgGaaaG
IWGSC_CSS.1AL_scaff.3952258	Cadenza2092	8107	C	T	het	—	tgagtagagaaattgacagtgtgG	tgagtagagaaattgacagtgtgA	tgccaccattgacatgagaG
IWGSC_CSS.1BL_scaff.3858008	Cadenza2092	10278	G	A	hom	hom	tttgagcaggcaggatcgC	tttgagcaggcaggatcgT	actcaggcctatacActattC
IWGSC_CSS.1DL_scaff.2265172	Cadenza2092	9094	C	T	hom	hom	tgcaTGTcatttgttcttatcagC	tgcaTGTcatttgttcttatcagT	agtgccaacttccGttcatC
IWGSC_CSS.2AL_scaff.6435867	Cadenza2092	16201	G	A	hom	hom	tttctgTaccttaacgtcaattgaC	tttctgTaccttaacgtcaattgaT	gtgaggatgatgaggaagacC
IWGSC_CSS.2AL_scaff.6439430	Cadenza2092	25101	C	T	het	—	caagaaagggCagCtCagC	caagaaagggCagCtCagT	tcGttAcTctttcActggtgaA
IWGSC_CSS.2DL_scaff.9760848	Cadenza2092	4733	C	T	het	het	gcaccatgggtctcaggtaC	gcaccatgggtctcaggtaT	tcagtcagtttGCTCtGCTG
IWGSC_CSS.3AL_scaff.4407012	Cadenza2092	2785	C	T	hom	hom	acatatAgtgttctcatccaccatC	acatatAgtgttctcatccaccatT	acctctcatgttaaataggtttgT
IWGSC_CSS.3AS_scaff.3441108	Cadenza2092	541	G	A	het	het	GtgatgaccttgagacGgaC	GtgatgaccttgagacGgaA	aggcaTgacaaCgcgcaA
IWGSC_CSS.3B_scaff.10449827	Cadenza1551	4779	G	A	hom	hom	ggcaaggtcaagaacGgtC	ggcaaggtcaagaacGgtT	aCagaGtgggttagaggcaG
IWGSC_CSS.3B_scaff.10550638	Cadenza1551	3250	C	T	het	het	ctccttcacttgttgcggC	ctccttcacttgttgcggT	gcaacATtTgatactgcaagG
IWGSC_CSS.3DL_scaff.6945816	Cadenza1551	589	C	T	hom	hom	agcatctcacctgcaacCaataC	agcatctcacctgcaacCaataT	TgtgccCTctgaAtattttcaTG
IWGSC_CSS.3DL_scaff.6954177	Cadenza1551	3508	C	T	het	het	tgtagcatcacattaactttctcG	tgtagcatcacattaactttctcA	gcttggtataaacCttacgacA
IWGSC_CSS.4AS_scaff.5938272	Cadenza1551	19080	G	A	hom	hom	agAcCccgAtcgccatgG	agAcCccgAtcgccatgA	GggAgatAcaggtaaaActcTtcG
IWGSC_CSS.4AS_scaff.5977594	Cadenza1551	11092	C	T	het	het	gccttgattcggaacaacaaaC	gccttgattcggaacaacaaaT	gcgtctctcagtcctgcA

IWGSC contig	Line	Pos	WT	Mut	Predicted	$M_4$	Primer 1 (Cadenza)	Primer 2 (mutant)	Common Primer
IWGSC_CSS_5AL_scaff_2671035	Cadenza1551	5859	C	T	het	het	cggatgatattTttagacttcgacG	cggatgatattTttagacttcgacG	ggcagttcagcGacccatT
IWGSC_CSS_5BL_scaff_10889480	Cadenza1551	2530	G	A	hom	hom	gagcttaactcgagatggaG	gagcttaactcgagatggaA	tccatgCAacGccttgG
IWGSC_CSS_3B_scaff_10528396	Cadenza2088	8059	G	A	hom	—	ctttccgctccgtaagcaataG	ctttccgctccgtaagcaataA	gtgcaactgttcaggcctgA
IWGSC_CSS_3B_scaff_10637573	Cadenza2088	16815	G	A	het	het	agcaagcttaccGgtctgC	agcaagcttaccGgtctgT	cgagcAactacgagcagctT
IWGSC_CSS_4AL_scaff_7086469	Cadenza2088	6697	G	A	het	het	gccgtctacttcaacgcG	gccgtctacttcaacgcA	ccaGaggcttgTGCattttT
IWGSC_CSS_4AL_scaff_7126302	Cadenza2088	3627	G	A	hom	hom	gttcaaaaacaagtggtAatttgC	gttcaaaaacaagtggtAatttgT	cacaaggatatgaagcTctctagA
IWGSC_CSS_4BL_scaff_7041808	Cadenza2088	10234	G	A	hom	hom	tcaatggatgagggtgcttC	tcaatggatgagggtgcttT	ccatagcagcatcagccacA
IWGSC_CSS_5AL_scaff_2794167	Cadenza2088	13162	G	A	het	—	agtattcaggacaagcatCttCaG	agtattcaggacaagcatCttCaA	caatgaacctctcgaagaaGaG
IWGSC_CSS_5BL_scaff_10889232	Cadenza2088	3885	G	A	het	het	cTcaaccacaatgggcaAatC	cTcaaccacaatgggcaAatT	tccttcatcaatcatcaattgtgG
IWGSC_CSS_5BS_scaff_2267405	Cadenza2088	11113	C	T	hom	hom	ctttgatgatcctaggcctctTG	ctttgatgatcctaggcctctTA	tgatttggTctggtAgagtttGA
IWGSC_CSS_3B_scaff_10475354	Cadenza1409	2203	G	A	hom	hom	agCGaacaagagGtcaaacG	agCGaacaagagGtcaaacA	ctgaaacacaCtagaCAattAccG
IWGSC_CSS_3B_scaff_10674115	Cadenza1409	4555	C	T	het	het	gcttcagtgcatgccttcaG	gcttcagtgcatgccttcaA	cttcacaccGagataatGtattG
IWGSC_CSS_4AL_scaff_7153568	Cadenza1409	13073	C	T	hom	hom	tccgaccgAtcaaccttgG	tccgaccgAtcaaccttgA	gaccggaactcctcggcC
IWGSC_CSS_4DL_scaff_14314966	Cadenza1409	2010	G	A	het	hom	gtaggccccctctCAGgG	gtaggccccctctCAGgA	cggcgTcaCaAgttgCcT
IWGSC_CSS_4DS_scaff_2324074	Cadenza1409	7606	G	A	het	het	tGcatgaaaatgtgtGcaGaG	tGcatgaaaatgtgtGcaGaA	gggtaAgttcAaaactGaagtgaG
IWGSC_CSS_5AS_scaff_1517889	Cadenza1409	3561	G	A	het	het	tctcgacatcttcccggtgaC	tctcgacatcttcccggtgaT	gtgccttgaacattgcttattA
IWGSC_CSS_5AS_scaff_1523866	Cadenza1409	8054	G	A	hom	—	ggatgatctaccgcaGgaC	ggatgatctaccgcaGgaT	tctcgagCcTctctcA
IWGSC_CSS_5BL_scaff_10917655	Cadenza1409	19073	G	A	hom	hom	caaatgacatgcaaaagaagttgC	caaatgacatgcaaaagaagttgT	cgcttcatcactacaAaatatgtcT
IWGSC_CSS_1AL_scaff_3886649	Cadenza1599	5204	C	T	het	het	tgaatgccaaccacaatGcC	tgaatgccaaccacaatGcT	ggactgactgtgacatatttaG
IWGSC_CSS_1BL_scaff_3810267	Cadenza1599	6634	C	T	hom	hom	ccCaggaaatgagcacctC	ccCaggaaatgagcacctT	cgcaggcgaagatgtgaTtG
IWGSC_CSS_1DL_scaff_2291677	Cadenza1599	12856	C	T	hom	hom	GgtagacaagtcgccgaG	GgtagacaagtcgccgaA	cctcctccttcaacGCcG
IWGSC_CSS_2AL_scaff_6354492	Cadenza1599	7566	G	A	het	het	gGagaatgcaCAgtAacTtctgG	gGagaatgcaCAgtAacTtctgA	ttccgaagaaccacaTccTG
IWGSC_CSS_2AS_scaff_5282937	Cadenza1599	9736	G	A	het	het	gctgtagattttatagctgctatgC	gctgtagattttatagctgctatgT	cacCagaattgttCactgatttTC
IWGSC_CSS_2BL_scaff_7952427	Cadenza1599	19249	G	A	hom	hom	cgTccctCcttagcacgaC	cgTccctCcttagcacgaT	aTcactcattagcgcgAG
IWGSC_CSS_2DL_scaff_9897981	Cadenza1599	5627	C	T	het	het	cttgggtgctTgattgcttactC	cttgggtgctTgattgcttactT	gTttgctCtctgactctTtgtG
IWGSC_CSS_3AL_scaff_4446105	Cadenza1599	1765	G	A	hom	—	aaatgctttcctaCcgctagtG	aaatgctttcctaCcgctagtA	ttctAgaggcaatagctTatatgcT

Table A.6: Validation of mutations on  $M_4$  on Kronos

IWGSC contig	Line	Pos	WT	Mut	Predicted	$M_4$	Primer 1 (Kronos)	Primer 2 (mutant)	Common Primer
IWGSC_CSS_1AS_scaff_3284790	Kronos3085	7449	G	A	Het	Het	ccacaccttgagcctcgC	ccacaccttgagcctcgT	gtgattttgccaggggagA
IWGSC_CSS_1BL_scaff_3897513	Kronos3085	1515	C	T	Het	Het	gcttcactGggtcctgC	gcttcactGggtcctgT	acAaggactgcttcagaGaC
IWGSC_CSS_2AL_scaff_6434745	Kronos3085	3424	C	T	Het	Het	cctcGgttttgcaaatttctatgC	cctcGgttttgcaaatttctatgT	gGCaaTggcataacaacagatA
IWGSC_CSS_3AS_scaff_3408995	Kronos3085	732	C	T	Het	Het	aggccatttcgaattccgC	aggccatttcgaattccgT	ggTgttaTccagAacctgagTG
IWGSC_CSS_3B_scaff_10708748	Kronos3085	2675	G	A	Het	Het	gttgcatgcttcacccagG	gttgcatgcttcacccagA	gtaacaatctgagttcgtagcaC
IWGSC_CSS_4AL_scaff_7132733	Kronos3085	1799	C	T	Hom	Hom	caccctgtagtgaccctC	caccctgtagtgaccctT	aCcGcctaGaaagaaagcttC

IWGSC contig	Line	Pos	WT	Mut	Predicted	$M_4$	Primer 1 (Kronos)	Primer 2 (mutant)	Common Primer
IWGSC.CSS_5AS_scaff.1534693	Kronos3085	4605	C	T	Het	Het	cagcttctctggccctcAtC	cagcttctctggccctcAtT	gtaCctcagcAgtcaTgagAG
IWGSC.CSS_6AS_scaff.4361911	Kronos3085	8857	G	A	Het	Het	tcacgaaagacgacttcaacctcC	tcacgaaagacgacttcaacctcT	catgagggtctgcatctccatcA
IWGSC.CSS_6BS_scaff.3008326	Kronos3085	1528	G	A	Het	Het	ccatgttgacttggtggtgC	ccatgttgacttggtggtgT	ggaagcatggCaagtgcA
IWGSC.CSS_7AS_scaff.4214385	Kronos3085	27835	C	T	Hom	Hom	cgtaccttcggtgggaaagG	cgtaccttcggtgggaaagA	ctcttggtcagctgataaagacT
IWGSC.CSS_1AL_scaff.3929964	Kronos3191	1336	C	T	Het	Het	tttcggccataacctgacatC	tttcggccataacctgacatT	attgcctcagcttcttgcaG
IWGSC.CSS_1BL_scaff.3899789	Kronos3191	7925	C	T	Het	Het	actctcacTggcagcagC	actctcacTggcagcagT	caacgtggtgcccatcGtA
IWGSC.CSS_2AL_scaff.6426728	Kronos3191	1481	G	A	Hom	Hom	gaaActgccgcagctCgC	gaaActgccgcagctCgT	ccaGcaGctcgtgagaaA
IWGSC.CSS_2BL_scaff.7960273	Kronos3191	690	C	T	Hom	Hom	gccattcatccttaggcgC	gccattcatccttaggcgT	acatgcaattgctgatgactG
IWGSC.CSS_3AS_scaff.3286603	Kronos3191	2975	G	A	Het*	Hom	ccgtgtggtttgttggtgG	ccgtgtggtttgttggtgA	gaaaggaacgtgTcaTgcaG
IWGSC.CSS_5AL_scaff.2694249	Kronos3191	2399	C	T	Het	Het	gccttcagatagagccGC	gccttcagatagagccGT	cgccacatcgacattctcG
IWGSC.CSS_5BL_scaff.10923577	Kronos3191	3713	C	T	Het	Het	gtggattgcctgagcttgC	gtggattgcctgagcttgT	tgggtggcctcttgggaC
IWGSC.CSS_6AL_scaff.5823017	Kronos3191	13225	C	T	Hom	Hom	ccctttcagcctctggaG	ccctttcagcctctggaA	ttcgagaaggcccatcgA
IWGSC.CSS_6BS_scaff.2955394	Kronos3191	1622	C	T	Het*	Hom	gtggagatgaaggtctagcaaG	gtggagatgaaggtctagcaaA	gatactcTgcaatgggtgT
IWGSC.CSS_7BL_scaff.6739382	Kronos3191	12261	G	A	Hom	Hom	gagacaagctttgaattgctcC	gagacaagctttgaattgctcT	CgagtgaacctTcatttcccG
IWGSC.CSS_1AS_scaff.3276389	Kronos3288	9720	C	T	Hom	Hom	aCcaGcaggaccAatgtctC	aCcaGcaggaccAatgtctT	atgatgcaacctcagccaT
IWGSC.CSS_2AL_scaff.6367515	Kronos3288	6976	G	A	Het	Het	caggtcgagTgtctccgG	caggtcgagTgtctccgA	ggggtgatCtggaaaggG
IWGSC.CSS_2AL_scaff.6422019	Kronos3288	4523	G	A	Het	Het	cgctaggtccctgcatagG	cgctaggtccctgcatagA	acgcAcgtgaagccgtcA
IWGSC.CSS_3AL_scaff.4284850	Kronos3288	7901	C	T	Hom	Hom	tgcccttggacaacatcgG	tgcccttggacaacatcgA	tgtcAgtcagcagaccaG
IWGSC.CSS_4AS_scaff.5962359	Kronos3288	13049	G	A	Het	Hom	ccatcaagaagtacgagttcgaC	ccatcaagaagtacgagttcgaT	accatgccagcttgcA
IWGSC.CSS_6AL_scaff.5778773	Kronos3288	6853	G	A	Het	Het	gagtgaccttcccgtctttC	gagtgaccttcccgtctttT	ggagaacagctactcggcT
IWGSC.CSS_6AS_scaff.4392100	Kronos3288	3434	C	T	Het	Het	atggaagcacaggtgaccG	atggaagcacaggtgaccA	ggAagcgaaagtgaacaaacA
IWGSC.CSS_7BL_scaff.6744240	Kronos3288	9772	G	A	Het	Het	agctgttcttctcctacttcaaG	agctgttcttctcctacttcaaA	caggtcgttcttgagctcC
IWGSC.CSS_1AL_scaff.3887185	Kronos3413	9708	C	T	Hom	Hom	gcacgcctttatcgaggtaaaG	gcacgcctttatcgaggtaaaA	AgaacagcagagcgcaA
IWGSC.CSS_2BS_scaff.3381362	Kronos3413	5160	C	T	Het*	Hom	caacttctgggctgtagtG	caacttctgggctgtagtA	tgAgaattctgacGcaaaagaC
IWGSC.CSS_3AS_scaff.3296605	Kronos3413	6154	G	A	Het	Het	ctggtcacgggctctagC	ctggtcacgggctctagT	cagcactgagagacatggaC
IWGSC.CSS_3BL_scaff.10693516	Kronos3413	12632	C	T	Het	Het	ctaggcttgacaaaacaggC	ctaggcttgacaaaacaggT	agcttgcatctatgggcatT
IWGSC.CSS_5AS_scaff.1547699	Kronos3413	2686	G	A	Het	Het	gCtacaaccttcaccaatcgC	gCtacaaccttcaccaatcgT	gacggctttgaagtgtcatC
IWGSC.CSS_5BL_scaff.10856077	Kronos3413	5853	G	A	Het	Het	agagcttcaccccatgctC	agagcttcaccccatgctT	acgCacatttAatagctgaagC
IWGSC.CSS_6AL_scaff.5750718	Kronos3413	11046	G	A	Hom	Hom	cacgcTtcccgaacttcttataG	cacgcTtcccgaacttcttataA	AgacgatgtgatcaggattcaG
IWGSC.CSS_7AL_scaff.4433177	Kronos3413	3511	C	T	Het	Het	GaTgctccGtcaggctgG	GaTgctccGtcaggctgA	cactactggacaagctcttgG
IWGSC.CSS_7BL_scaff.6742567	Kronos3413	667	C	T	Het	Het	gttgcttgctggcgagaC	gttgcttgctggcgagaT	cattttgcacctgtgtcTG
IWGSC.CSS_1AL_scaff.3976389	Kronos3935	10941	C	T	Hom	Hom	gggtgaggagatcggCgatG	gggtgaggagatcggCgatA	cagtcactcatagagaggtcaG
IWGSC.CSS_1BL_scaff.3873362	Kronos3935	1392	G	A	Het	Het	cagatctgaagcctaGcacatG	cagatctgaagcctaGcacatA	actaccagaatcagcacaaaaAC
IWGSC.CSS_2BL_scaff.7882382	Kronos3935	2721	C	T	Het	Het	gcaagctaagatgtaccgtagC	gcaagctaagatgtaccgtagT	gccacagttaggagaaagactT
IWGSC.CSS_3AL_scaff.4242376	Kronos3935	2410	C	T	Het	Het	agaacccaaaacccgTacttaG	agaacccaaaacccgTacttaA	gtagGgtCcatCTaaagcttG
IWGSC.CSS_3B_scaff.10485067	Kronos3935	3349	C	T	Hom	Hom	gcttgagcaactactccaactG	gcttgagcaactactccaactA	gcaatttcccttaTccgcagT
IWGSC.CSS_4AS_scaff.5984153	Kronos3935	6006	G	A	Het	Het	agCaggctctggccaagttG	agCaggctctggccaagttA	cgaatGtatgaGtaggcgcT
IWGSC.CSS_4BL_scaff.7019402	Kronos3935	9081	C	T	Het	Het	tgcaatcatgtagtgcctgG	tgcaatcatgtagtgcctgA	agcatgatccctagaaCCataC
IWGSC.CSS_5BL_scaff.10842786	Kronos3935	3304	G	A	Het	Het	tggttcccGaaacctgaaC	tggttcccGaaacctgaaT	cgcatacttgaacaTGagcAC
IWGSC.CSS_6BS_scaff.3045205	Kronos3935	2293	G	A	Het	Het	aaggaccaagcccaactctcG	aaggaccaagcccaactctcA	agtgatcaagcccaatgtgcA

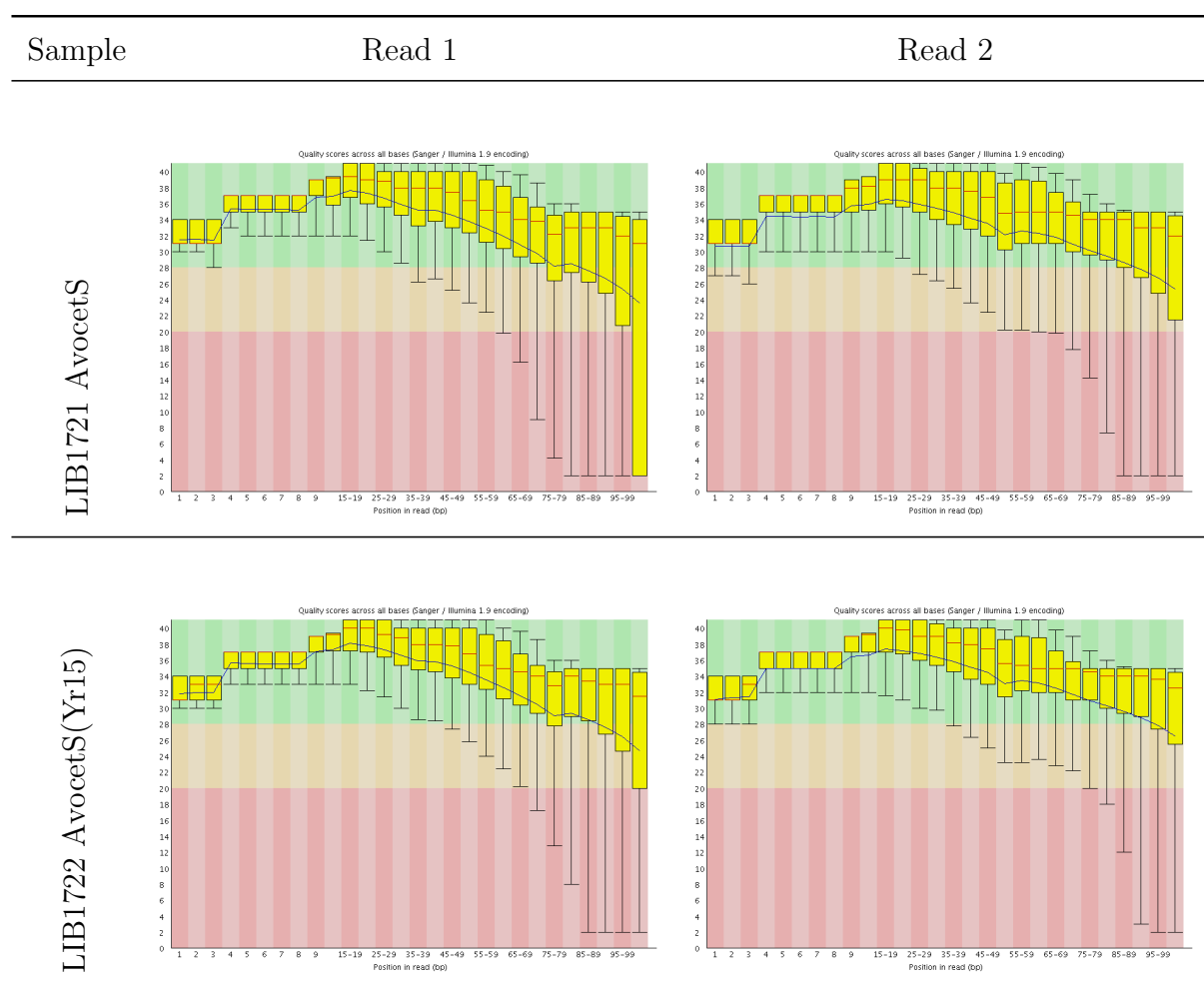


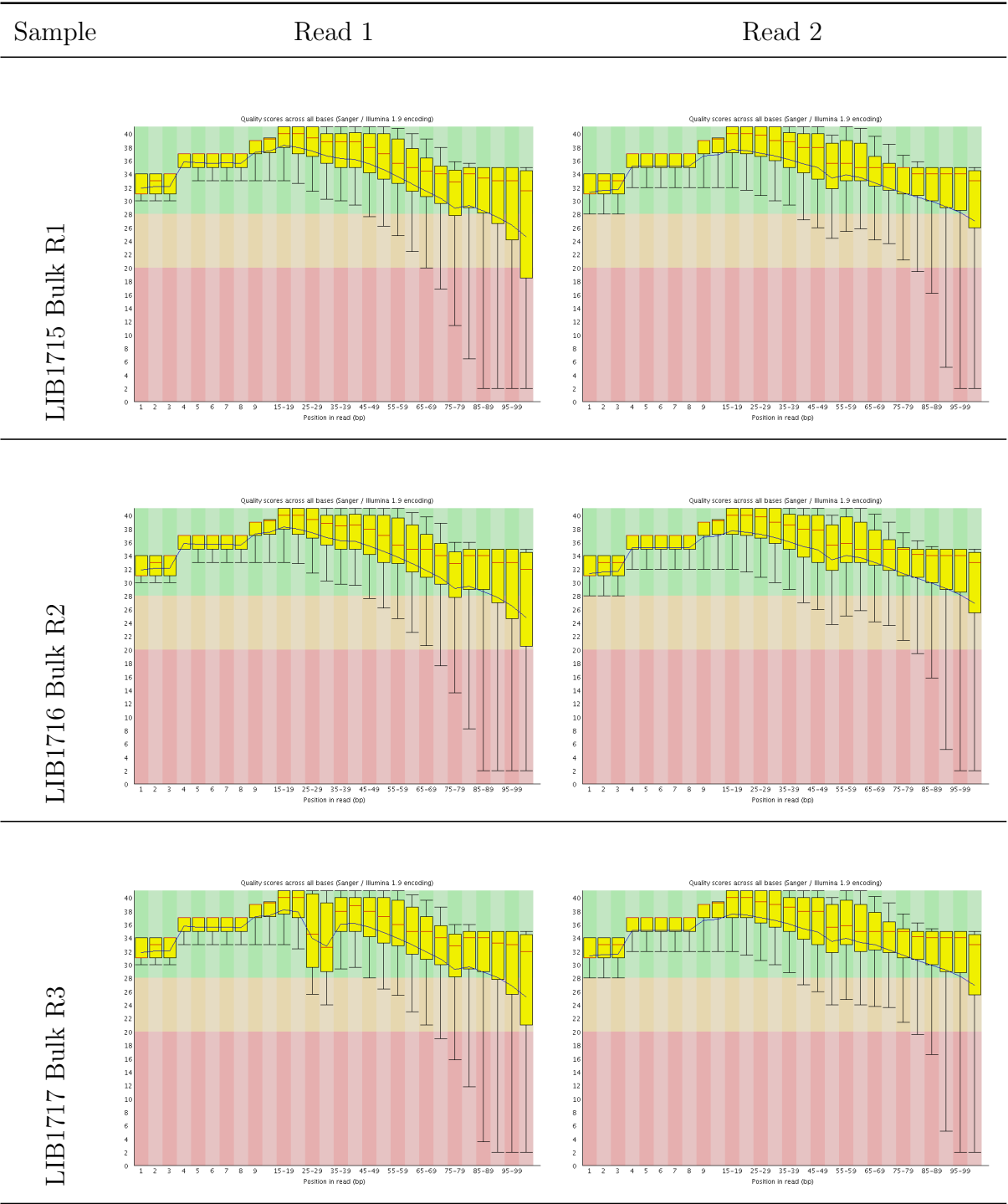
IWGSC contig	Line	Pos	WT	Mut	Predicted	$M_4$	Primer 1 (Kronos)	Primer 2 (mutant)	Common Primer
IWGSC.CSS_7AL_scaff.4555249	Kronos3935	4487	C	T	Het	Het	cAgtgctcgagatggcgC	cAgtgctcgagatggcgT	cCttgcaacctctctgatT
IWGSC.CSS_1BL_scaff.3918498	Kronos4240	6096	G	A	Het	Het	ttgcatgcccccaagaagaG	ttgcatgcccccaagaagaA	tgggcgaactgggtaatgtgG
IWGSC.CSS_2BS_scaff.5131713	Kronos4240	5900	G	A	Het	Het	cctttatcgaggaaaagagacacC	cctttatcgaggaaaagagacacT	caccattgtagggttcctTttC
IWGSC.CSS_5AL_scaff.2769540	Kronos4240	9626	C	T	Het	Het	tgCagtgtgggaaacggaG	tgCagtgtgggaaacggaA	catgagtGagatcttctctgcT
IWGSC.CSS_5BL_scaff.10871091	Kronos4240	7062	G	A	Het	Het	gccaaaggAaccataacctgC	gccaaaggAaccataacctgT	GgactcttggcAaccggA
IWGSC.CSS_6AL_scaff.5800333	Kronos4240	2360	G	A	Het	Het	cgacaggattgtgagCgC	cgacaggattgtgagCgT	tcagatgctgcaagattcatcT
IWGSC.CSS_7BL_scaff.6716931	Kronos4240	2613	G	A	Het	Het	gGtgGgtattTgcttgggtgaG	gGtgGgtattTgcttgggtgaA	tgGtggactcgacaGtGtA
IWGSC.CSS_2BL_scaff.8029221	Kronos4346	2860	G	A	Het	Het	tgcttccgctcttgctcC	tgcttccgctcttgctcT	atTtgcATCgAtcgggcC
IWGSC.CSS_3B_scaff.10460714	Kronos4346	14359	C	T	Hom	Hom	ctaccttgccatgcgacatG	ctaccttgccatgcgacatA	agcaccccgactctttgacG
IWGSC.CSS_4AS_scaff.5989735	Kronos4346	6404	G	A	Hom	Hom	acgcatgctaacatcagcG	acgcatgctaacatcagcT	actcaagataccaCcgcacG
IWGSC.CSS_5BL_scaff.7648030	Kronos4346	6893	C	T	Het	Het	taccttttctactggcagG	taccttttctactggcagA	ttttcagaggaaacacaggtatcA
IWGSC.CSS_6AL_scaff.5755840	Kronos4346	778	C	T	Het	Het	atcgagtaagctgtcacCgC	atcgagtaagctgtcacCgT	acctgcatgtcaCatccaC
IWGSC.CSS_6BS_scaff.2972151	Kronos4346	7876	G	A	Hom	Hom	gcagcaatgtcActgtttgG	gcagcaatgtcActgtttgA	gcttggactgggcattttatG
IWGSC.CSS_7AL_scaff.4542983	Kronos4346	18700	G	A	Het	Het	gcagggctAccggatacC	gcagggctAccggatacT	catctgccGgttaaacatgC
IWGSC.CSS_7BS_scaff.3098098	Kronos4346	5183	C	T	Het	Het	gCgatatgggtacttgcaatgaG	gCgatatgggtacttgcaatgaA	ttacattgcttataGTtgCcgG
IWGSC.CSS_1AS_scaff.3259804	Kronos4485	219	C	T	Het	Het	gtcggcacacccttgC	gtcggcacacccttgT	gcttctttaaggaggcgA
IWGSC.CSS_2AL_scaff.6315418	Kronos4485	10490	G	A	Hom	Hom	gccccctctcaaCcttctcagC	gccccctctcaaCcttctcagT	ttcagacgtCGaggaaatttcC
IWGSC.CSS_2BS_scaff.5181092	Kronos4485	3742	G	A	Het	Het	TggccagcacacctgcaG	TggccagcacacctgcaA	tggacgatgagTgatggAaaT
IWGSC.CSS_3B_scaff.10425015	Kronos4485	2372	C	T	Het	Het	gctactgaagttggCtcGG	gctactgaagttggCtcGA	cttcacatccttgggggTtC
IWGSC.CSS_3B_scaff.10775915	Kronos4485	4701	C	T	Het	Het	ccaagggctgcagagagG	ccaagggctgcagagagA	agacctcacgatGtcctcC
IWGSC.CSS_5AL_scaff.2754304	Kronos4485	2301	G	A	Het	Het	taaccTgccatcgcccG	taaccTgccatcgcccA	cattgGccagccaTgacT
IWGSC.CSS_5BL_scaff.10919959	Kronos4485	1867	C	T	Hom	Hom	gatgccctttgtggagaagG	gatgccctttgtggagaagA	tcttgttcccgaacatgtcA
IWGSC.CSS_7AS_scaff.4245431	Kronos4485	3402	G	A	Hom	Hom	aaggcgctgtgtttcC	aaggcgctgtgtttcT	agtaagtggAacagctaagatcaT
IWGSC.CSS_7BL_scaff.6667357	Kronos4485	641	C	T	Het	Het	gatcAgctgctcattcgagG	gatcAgctgctcattcgagA	ttccctgtcaattgatgccC

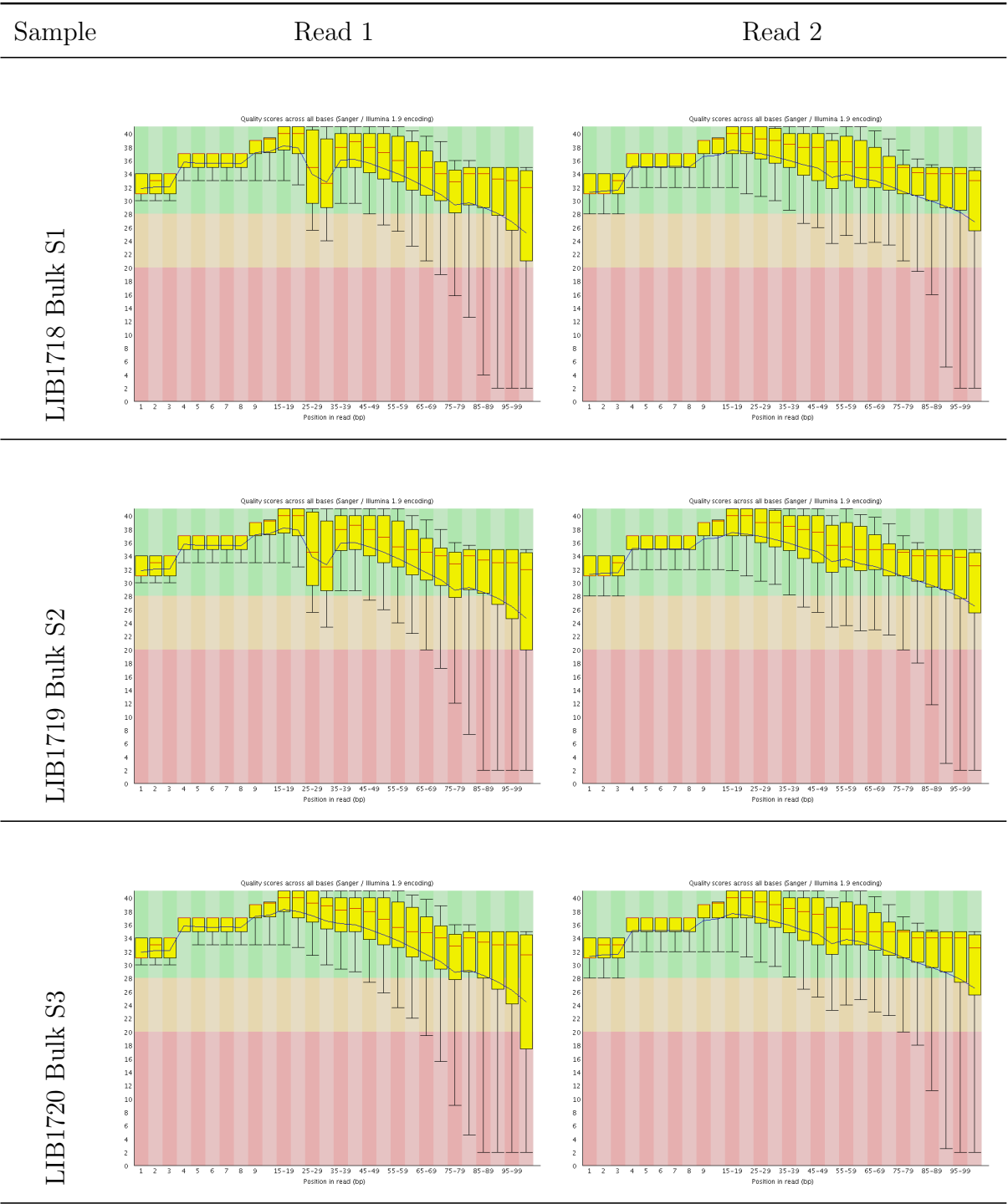
# Appendix B

## Quality Control

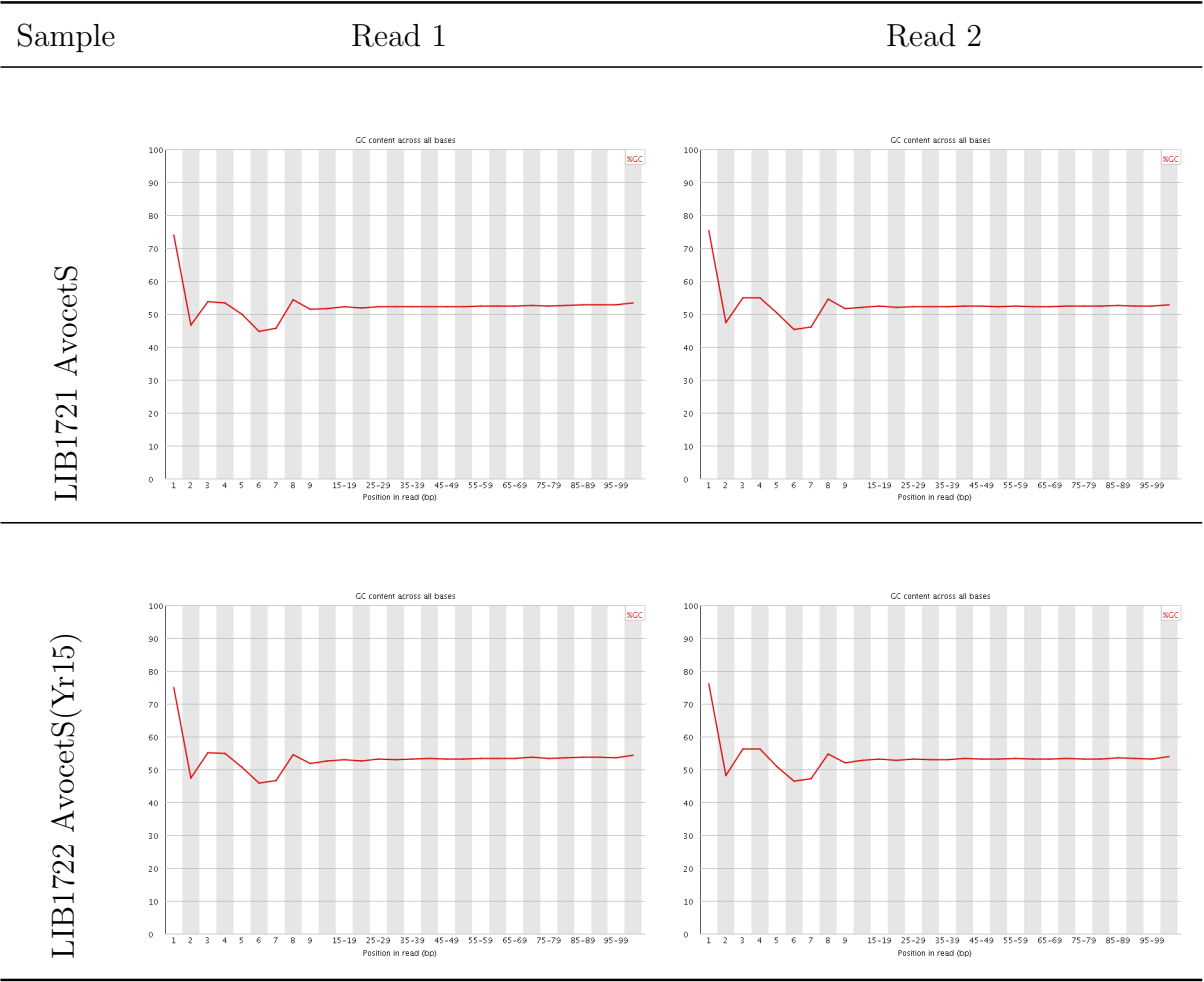
### B.1 Sequence read quality







B.2 Sequence GC content



Sample	Read 1	Read 2
LIB1715 Bulk R1		
LIB1717 Bulk R3		

Sample	Read 1	Read 2
LIB1718 Bulk S1		
LIB1720 Bulk S3		

# Bibliography

- Alexandra M Allen, Gary L a Barker, Simon T Berry, et al. Transcript-specific, single-nucleotide polymorphism discovery and linkage analysis in hexaploid bread wheat (*Triticum aestivum* L.). *Plant biotechnology journal*, 9(9):1086–99, December 2011. ISSN 1467-7652. doi: 10.1111/j.1467-7652.2011.00628.x.
- Babraham Bioinformatics. FastQC A Quality Control tool for High Throughput Sequence Data, 2012. URL <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
- Raoul J P Bonnal, Jan Aerts, George Githinji, et al. Biogem: An effective tool-based approach for scaling up open source software development in bioinformatics. *Bioinformatics*, 28(7):1035–1037, April 2012. ISSN 13674803. doi: 10.1093/bioinformatics/bts080.
- C. Burt, A. Steed, N. Gosman, et al. Mapping a type 1 fhb resistance on chromosome 4as of triticum macha and deployment in combination with two type 2 resistances. *Theoretical and Applied Genetics*, 128(9): 1725–1738, 2015. ISSN 1432-2242. doi: 10.1007/s00122-015-2542-9.
- Dario Cantu, Manjula Govindarajulu, Alex Kozik, et al. Next generation sequencing provides rapid access to the genome of *Puccinia striiformis* f. sp. *tritici*, the causal agent of wheat stripe rust. *PLoS ONE*, 6(8): 1–8, 08 2011. doi: 10.1371/journal.pone.0024230.
- Jarrold a Chapman, Martin Mascher, Aydn Buluç, et al. A whole-genome shotgun approach for assembling and anchoring the hexaploid bread wheat genome. *Genome Biology*, 16(1):1–17, 2015. ISSN 1465-6906. doi: 10.1186/s13059-015-0582-8.
- A Cornish-Bowden. Nomenclature for incompletely specified bases in



- nucleic acid sequences: recommendations 1984. *Nucleic acids research*, 13(9):3021–30, May 1985. ISSN 0305-1048.
- Graham J Etherington, R. H. Ramirez-Gonzalez, and D. MacLean. bio-samtools 2: a package for analysis and visualization of sequence and alignment data with SAMtools in Ruby. *Bioinformatics*, pages 1–2, 2015. ISSN 1367-4803. doi: 10.1093/bioinformatics/btv178.
- Z. K. Gerechter-Amitai, C. H. van Silfhout, Adriana Grama, and Frida Kleitman. Yr15 — a new gene for resistance to *puccinia striiformis* in *triticum dicoccoides* sel. g-25. *Euphytica*, 43(1):187–190, 1989. ISSN 1573-5060. doi: 10.1007/BF00037912.
- Naohisa Goto, Pjotr Prins, Mitsuteru Nakao, et al. BioRuby: bioinformatics software for the Ruby programming language. *Bioinformatics (Oxford, England)*, 26(20):2617–9, October 2010. ISSN 1367-4811. doi: 10.1093/bioinformatics/btq475.
- Emily Hodges, Zhenyu Xuan, Vivekanand Baliya, et al. Genome-wide in situ exon capture for selective resequencing. *Nat Genet*, 39(12):1522–1527, Dec 2007. ISSN 1061-4036. doi: 10.1038/ng.2007.42.
- Amelia Hubbard, Clare M Lewis, Kentaro Yoshida, et al. Field pathogenomics reveals the emergence of a diverse wheat yellow rust population. *Genome Biology*, 16(1):1–15, 2015. ISSN 1465-6906. doi: 10.1186/s13059-015-0590-8.
- Kazutaka Katoh and Daron M Standley. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution*, 30(4):772–80, April 2013. ISSN 1537-1719. doi: 10.1093/molbev/mst010.
- W. J. Kent. BLAT—The BLAST-Like Alignment Tool. *Genome Research*, 12(4):656–664, March 2002. ISSN 1088-9051. doi: 10.1101/gr.229202.
- Robert King, Nicholas Bird, Ricardo Ramirez-Gonzalez, et al. Mutation scanning in wheat by exon capture and next-generation sequencing. *PLoS ONE*, 10(9):1–18, 09 2015. doi: 10.1371/journal.pone.0137549.
- Ksenia Krasileva, Hans Vasquez-Gross, Tyson Howell1, et al. Uncovering hidden variation in young polyploid wheat genomes. submitted 2016.

- Ksenia V Krasileva, Vince Buffalo, Paul Bailey, et al. Separating homeologs by phasing in the tetraploid wheat transcriptome. *Genome biology*, 14(6):R66, June 2013. ISSN 1465-6914. doi: 10.1186/gb-2013-14-6-r66.
- LGC Genomics. <http://www.lgcgroup.com/services/genotyping/>, 2013. URL <http://www.lgcgroup.com/services/genotyping/>.
- Heng Li and Richard Durbin. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)*, 25(14):1754–60, July 2009. ISSN 1367-4811. doi: 10.1093/bioinformatics/btp324.
- Jian Ma, Jiri Stiller, Zhi Zheng, et al. A high-throughput pipeline for detecting locus-specific polymorphism in hexaploid wheat (*triticum aestivum* l.). *Plant Methods*, 11(1), aug 2015. doi: 10.1186/s13007-015-0082-6.
- K. F. X. Mayer, J. Rogers, J. Dole el, et al. A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. *Science*, 345(6194):1251788–1251788, July 2014. ISSN 0036-8075. doi: 10.1126/science.1251788.
- R. W. Michelmore, I. Paran, and R. V. Kesseli. Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions by using segregating populations. *Proceedings of the National Academy of Sciences*, 88(21):9828–9832, November 1991. ISSN 0027-8424. doi: 10.1073/pnas.88.21.9828.
- Joseph K. Pickrell, John C. Marioni, Athma A. Pai, et al. Understanding mechanisms underlying human gene expression variation with rna sequencing. *Nature*, 464(7289):768–772, Apr 2010. ISSN 0028-0836. doi: 10.1038/nature08872.
- R. H. Ramirez-Gonzalez, Cristobal Uauy, and Mario Caccamo. PolyMarker: A fast polyploid primer design pipeline. *Bioinformatics*, pages 2–3, 2015a. ISSN 1367-4803. doi: 10.1093/bioinformatics/btv069.
- Ricardo H Ramirez-Gonzalez, Raoul Bonnal, Mario Caccamo, and Daniel Maclean. Bio-samtools: Ruby bindings for SAMtools, a library for

- accessing BAM files containing high-throughput sequence alignments. *Source code for biology and medicine*, 7(1):6, January 2012. ISSN 1751-0473. doi: 10.1186/1751-0473-7-6.
- Ricardo H. Ramirez-Gonzalez, Vanesa Segovia, Nicholas Bird, Mario Caccamo, and Cristobal Uauy. *Next Generation Sequencing Enabled Genetics in Hexaploid Wheat*, pages 201–209. Springer Japan, Tokyo, 2015b. ISBN 978-4-431-55675-6. doi: 10.1007/978-4-431-55675-6\_22. URL [http://dx.doi.org/10.1007/978-4-431-55675-6\\_22](http://dx.doi.org/10.1007/978-4-431-55675-6_22).
- Ricardo H. Ramirez-Gonzalez, Vanesa Segovia, Nicholas Bird, et al. Rna-seq bulked segregant analysis enables the identification of high-resolution genetic markers for breeding in hexaploid wheat. *Plant Biotechnology Journal*, 13(5):613–624, 2015c. ISSN 1467-7652. doi: 10.1111/pbi.12281.
- Harpinder S Randhawa, Jasdeep S Mutti, Kim Kidwell, et al. Rapid and targeted introgression of genes into popular wheat cultivars using marker-assisted background selection. *PloS one*, 4(6):e5752, January 2009. ISSN 1932-6203. doi: 10.1371/journal.pone.0005752.
- Steve Rozen and Helen Skaletsky. Primer3 on the WWW for general users and for biologist programmers. *Methods in molecular biology (Clifton, N.J.)*, 132:365–86, January 2000. ISSN 1064-3745.
- Korbinian Schneeberger, Stephan Ossowski, Christa Lanz, et al. Shoremap: simultaneous mapping and mutation identification by deep sequencing. *Nat Meth*, 6(8):550–551, Aug 2009. ISSN 1548-7091. doi: 10.1038/nmeth0809-550.
- Guy St C Slater and Ewan Birney. Automated generation of heuristics for biological sequence comparison. *BMC bioinformatics*, 6:31, January 2005. ISSN 1471-2105. doi: 10.1186/1471-2105-6-31.
- Martin Trick, Nikolai Adamski, Sarah G Mugford, et al. Combining SNP discovery from next-generation sequencing data with bulked segregant analysis (BSA) to fine-map genes in polyploid wheat. *BMC plant biology*, 12(1):14, January 2012. ISSN 1471-2229. doi: 10.1186/1471-2229-12-14.

- Shichen Wang, Debbie Wong, Kerrie Forrest, et al. Characterization of polyploid wheat genomic diversity using a high-density 90 000 single nucleotide polymorphism array. *Plant biotechnology journal*, 12(6): 787–796, March 2014. ISSN 1467-7652. doi: 10.1111/pbi.12183.
- Yi Wang, Vijay K. Tiwari, Nidhi Rawat, et al. GSP: a web-based platform for designing genome-specific primers in polyploids. *Bioinformatics*, page btw134, mar 2016. doi: 10.1093/bioinformatics/btw134.
- C.R. Wellings and Robert A. McIntosh. Host-pathogen studies of wheat stripe rust in australia. In A.E Slinkard, editor, *Proceedings 9th International Wheat Genetics Symposium*, pages 336–338. University of Saskatchewan, Saskatoon, SK, Canada, 1998.
- Paul a Wilkinson, Mark O Winfield, Gary L a Barker, et al. CerealsDB 2.0: an integrated resource for plant breeders and scientists. *BMC bioinformatics*, 13(1):219, January 2012. ISSN 1471-2105. doi: 10.1186/1471-2105-13-219.