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 an 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 discuses 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 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 realignments or *de novo* assemblies, how to do SNP calling in diploid and polyploid organisims and the bulk frequency ratios.

2.6 Wheat online resources

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

Chapter 3

Genetic mapping of *Yr15*

This section describes in detail than the paper of Ramirez-Gonzalez et al. (2014)

3.1 (Introduction) *Yr15*

Breeding importance of *Yr15* and original source (an introgression of *T. diccocoides*).

3.2 Segregating population and resistance essays

A description of the starting material and how the population was generated.

3.3 Sequencing and mapping

RNA-Seq and the decision to call SNPs on gene models rather than the whole reference. Details of the mapping against the Wheat UniGenes Pontius et al. (2002) and the UCW. Krasileva et al. (2013) gene models.

3.4 SNP Calling

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

3.5 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.

3.6 *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).

3.7 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.

3.8 Genetic map

The three versions of the genetic map: With a subset of the F_2 population

3.9 Assembly of the transcriptome

A comparison between thef known unigenes and the transcript from the progenitors. Since *Yr15* comes from an introgression with *T. diccocoides*, some novel transcripts can be extracted. Analysis of the gels from Mitaly?

3.10 Conclusions

Remarks on how this techinque 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.

Chapter 4

PolyMarker: A fast polyploid primer design pipeline

4.1 Introduction

Explain how the SNP markers are designed without the tool and an overview.

4.2 Global alignment

Search of the contigs with the sequence in the CSS reference and the importance of being able to distinguish between homoeologous regions.

4.3 Local alignment

Once the region with the primer has been selected, make a local alignment. This section discusses why the local alignment is needed.

4.4 Primer design tools

In this section, the principles of *in silico* primer design are discussed, and why not simply selecting a genomic variation is enough (thermal stability, primers folding on themselves)

4.5 Primer selection algorithms

Different algorithms to select the best primer:

4.5.1 Regular markers

Algorithm to select the two primers with a geneome-specific variation. For amplicons/capillary sequencing.

4.5.2 KASP markers

For KASP markers, the product should be as short as possible with the mutation in the first three bases.

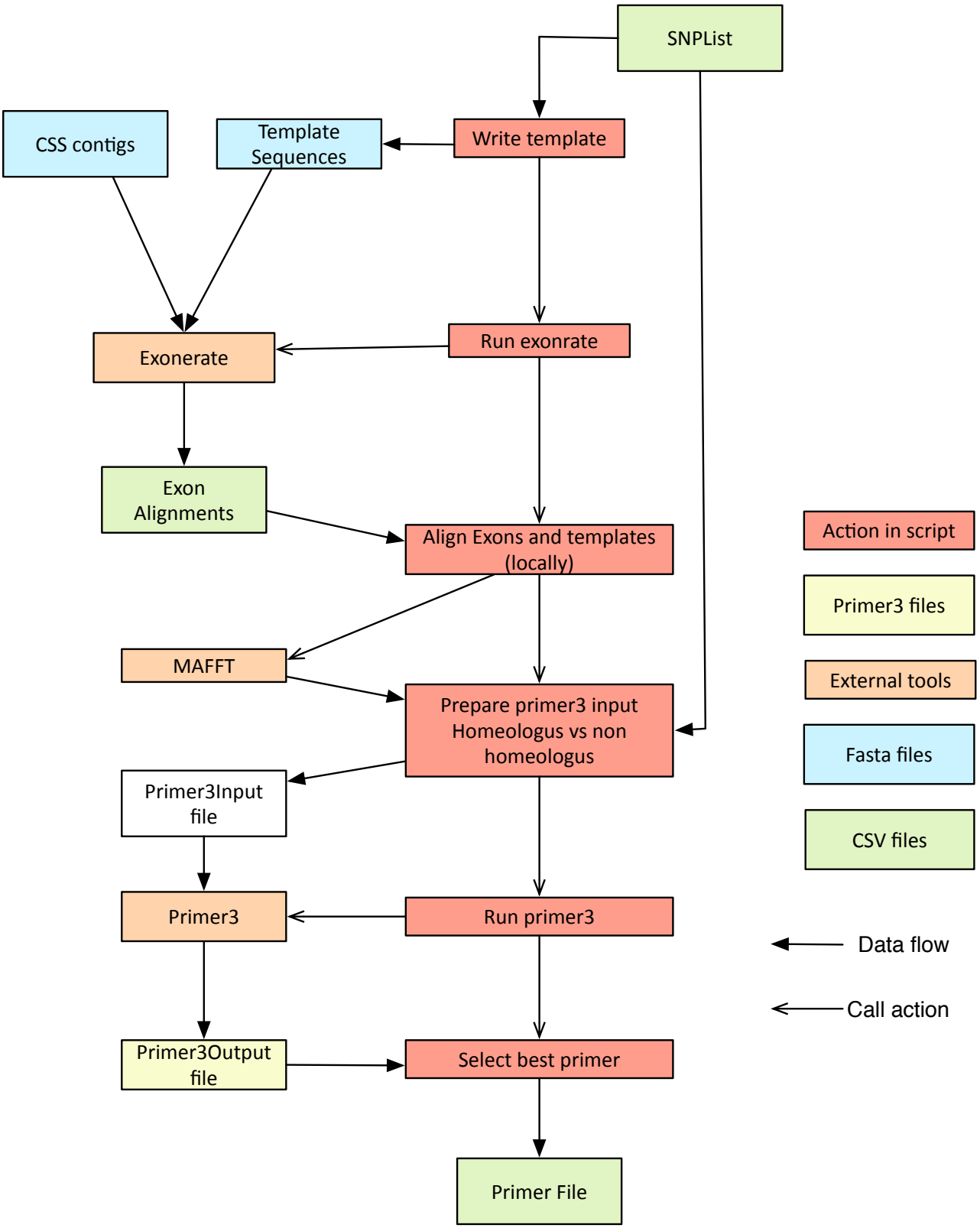
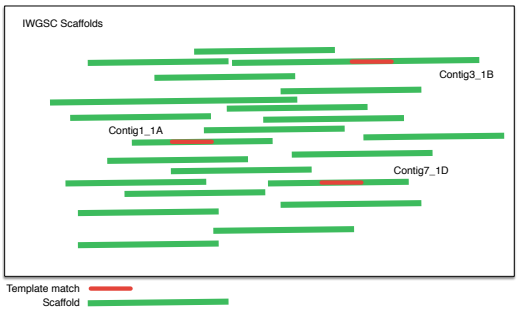
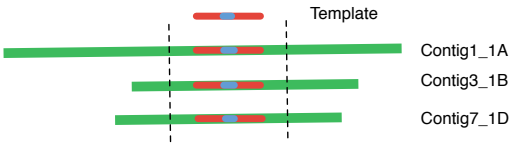


Figure 4.1: PolyMarker Pipeline



(a) Global search.



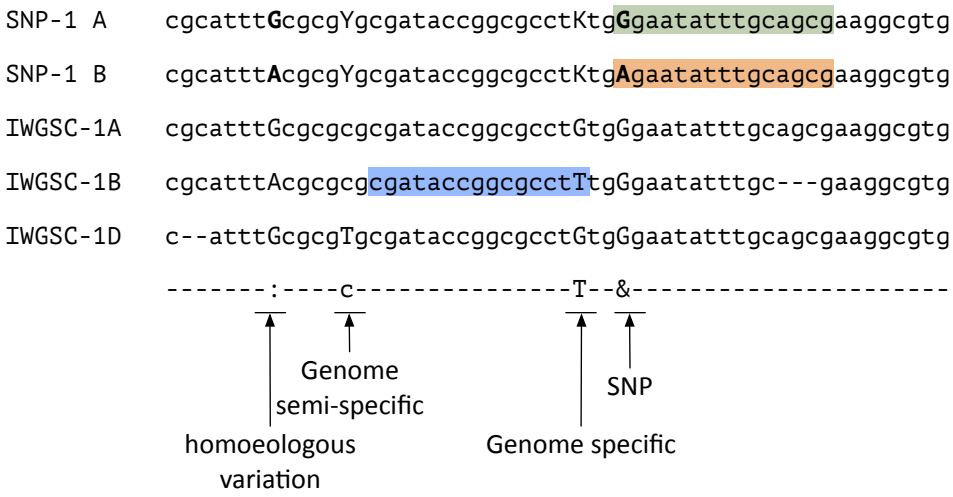
(b) Selected regions around the SNP on every chromosome.

```
SNP-1 A     cgcatttGcgcgYgcgataccggcgcctKtgGgaatatttgcagcgaaggcgtg
SNP-1 B     cgcatttAcgcgYgcgataccggcgcctKtgAgaatatttgcagcgaaggcgtg
IWGSC-1A    cgcatttgcgcgcgcgataccggcgcctgtgggaatatttgcagcgaaggcgtg
IWGSC-1B    cgcatttacgcgcgcgataccggcgcctttgggaatatttgcgaaggcgtg
IWGSC-1D    catttgcgcgTgcgataccggcgcctgtgggaatatttgcagcgaaggcgtg
```

(c) Sequence of found regions around the SNP.

```
SNP-1 A     cgcatttGcgcgYgcgataccggcgcctKtgGgaatatttgcagcgaaggcgtg
SNP-1 B     cgcatttAcgcgYgcgataccggcgcctKtgAgaatatttgcagcgaaggcgtg
IWGSC-1A    cgcatttGcgcgcgcgataccggcgcctGtgGgaatatttgcagcgaaggcgtg
IWGSC-1B    cgcatttAcgcgcgcgataccggcgcctTtgGgaatatttgc---gaaggcgtg
IWGSC-1D    c--atttGcgcgTgcgataccggcgcctGtgGgaatatttgcagcgaaggcgtg
```

(d) Local alignment on regions around the SNP to detect indels.



(e) Alignment with mask and primer candidates.

Figure 4.2: Alignments done by PolyMarker.

4.5.3 Deletion algorithms

Algorithm to produce KASP for deletions in polyploids.

4.6 Designed markers

Details of the generated primers for the 80k iSelect chip and the 820k axiom chip. This section also include counts on how many are genome specific, semi-specific and non specific. Also an analysis of how many are repeated or map to more than one chromosome perfectly.

4.7 Conclusions

Remarks on the importance of getting the primers right, and the time saved by automating the primer selection. Also mention other primer design tools that have been inspired by polymarker: Ma et al. (2015), Wang et al. (2016)

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.

Bibliography

Jarrod 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. URL http://genomebiology.com/2015/16/1/26.

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. URL http://genomebiology.com/2013/14/6/R66 http://www.ncbi.nlm.nih.gov/pubmed/23800085.

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. URL http://dx.doi.org/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. URL http://www.sciencemag.org/cgi/doi/10.1126/science.1251788.

JU Pontius, L Wagner, and GD. Schuler. UniGene: A Unified View of the Transcriptome. In *The NCBI Handbook [Internet]*, chapter 21. National Center for Biotechnology Information (US), October 2002. URL http://www.ncbi.nlm.nih.gov/books/NBK21083/.

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*, 12(9):1–12, November 2014. ISSN 14677644. doi: 10.1111/pbi.12281. URL http://doi.wiley.com/10.1111/pbi.12281.

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. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3296661.

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. URL http://www.ncbi.nlm.nih.gov/pubmed/24646323.

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. URL http://dx.doi.org/10.1093/bioinformatics/btw134.