



THE UNIVERSITY OF QUEENSLAND
A U S T R A L I A

Genome Diversity in *Triticum aestivum*

Kaitao Lai

*A thesis submitted for the degree of Doctor of Philosophy at
The University of Queensland in 2015
School of Agriculture and Food Sciences*

Abstract

The aims of this PhD research are to identify and characterise genome diversity in *Triticum aestivum* (bread wheat).

This research will establish a process for the identification of large numbers of single nucleotide polymorphisms (SNPs) and other genetic variations in *Triticum aestivum*.

Single nucleotide polymorphisms (SNPs) are the most abundant type of molecular genetic marker and can be used for producing high-resolution genetic maps, marker-trait association studies and marker assisted breeding. Large polyploid genomes, such as wheat, present a challenge for SNP discovery due to the potential presence of multiple homoeologues for each gene.

AutoSNPdb has been successfully applied to identify SNPs from Sanger sequence data for several species, including barley, rice and Brassica, but the volume of data required to accurately call SNPs in the complex genome of wheat has prevented its application to this important crop. DNA sequencing technology has been revolutionised by the introduction of next generation sequencing, and it is now possible to generate several million sequence reads in a timely and cost effective manner.

Wheat transcriptome sequence data has been generated using Roche 454 Life Sciences technology. This data has been applied for SNP discovery using a modified autoSNPdb method, which integrates SNP and gene annotation information with a graphical viewer. A total of 4,694,141 sequence reads from three bread wheat varieties were assembled to identify a total of 38,928 candidate SNPs. Each SNP is within an assembly complete with annotation, enabling the selection of polymorphism within genes of interest.

The discovery of large numbers of genomic SNPs across 16 Australian diverse bread wheat varieties has been completed using a novel algorithm SGSAutoSNP. More than 10x whole genome shotgun Illumina paired read sequence data was generated through a bioplatforms collaboration and the data mapped to the draft assemblies of chromosomes

7A, 7B and 7D. Over 4 million inter-varietal SNPs were identified. SNP density varied along the lengths chromosome syntenic builds as well as between genomes.

SNP density analysis and SNP transition/transversion ratio analysis provide insights into the evolution and breeding history of this important crop. Both the SNP density and SNP transition/transversion ratios across the D genome are significantly lower than across the A and B genomes. This difference reflects the evolutionary history of this crop. In addition, genes within low SNP density regions may have been selected during domestication and breeding. Furthermore, this SNP resource permits the application of high resolution skim based genotyping by sequencing (GBS), trait association and analysis of structural variation in populations.

An integrated database and portal for wheat genome resource, WheatGenome.info, has been established and published online. This portal provides a variety of web-based systems hosting wheat genome and genomic data, including genomic SNPs, to support wheat research and crop improvement.

This research provides approaches to understand the effect of sequence variation on the form and function of wheat growth and development and response to environment.

Declaration by author

This thesis is composed of my original work, and contains no material previously published or written by another person except where due reference has been made in the text. I have clearly stated the contribution by others to jointly-authored works that I have included in my thesis.

I have clearly stated the contribution of others to my thesis as a whole, including statistical assistance, survey design, data analysis, significant technical procedures, professional editorial advice, and any other original research work used or reported in my thesis. The content of my thesis is the result of work I have carried out since the commencement of my research higher degree candidature and does not include a substantial part of work that has been submitted to qualify for the award of any other degree or diploma in any university or other tertiary institution. I have clearly stated which parts of my thesis, if any, have been submitted to qualify for another award.

I acknowledge that an electronic copy of my thesis must be lodged with the University Library and, subject to the policy and procedures of The University of Queensland, the thesis be made available for research and study in accordance with the Copyright Act 1968 unless a period of embargo has been approved by the Dean of the Graduate School.

I acknowledge that copyright of all material contained in my thesis resides with the copyright holder(s) of that material. Where appropriate I have obtained copyright permission from the copyright holder to reproduce material in this thesis.

Publications during candidature

Peer-reviewed journal papers

1. **Kaitao Lai**, Michał T. Lorenc, Hong Lee, Paul J. Berkman, Philipp Emanuel Bayer, Paul Visendi, Pradeep Ruperao, Timothy L. Fitzgerald, Manuel Zander, Chon-Kit Kenneth Chan, Sahana Manoli, Jiri Stiller, Jacqueline Batley, David Edwards (2014) [Identification and characterisation of more than 4 million inter-varietal SNPs across the group 7 chromosomes of bread wheat](#). Plant Biotechnology Journal 13(1): 97-104.
2. **Kaitao Lai**, Chris Duran, Paul J. Berkman, Michał T. Lorenc, Jiri Stiller, Sahana Manoli, Matthew Hayden, Kerrie Forrest, Delphine Fleury, Ute Baumann, Manuel Zander, Annaliese Mason, Jacqueline Batley, David Edwards (2012) [Single Nucleotide Polymorphism Discovery from Wheat Next Generation Sequence Data](#). Plant Biotechnology Journal 10(6): 743–749.
3. **Kaitao Lai**, Michał Tadeusz Lorenc, David Edwards (2012) [Genomic Databases for Crop Improvement](#). Agronomy 2(1):62-73. Republished in book: [The Role of Bioinformatics in Agriculture](#) (2014). Ed. Kumar, S. Apple Academic Press (USA), pp. 189-203.
4. **Kaitao Lai**, Paul J Berkman, Michał Tadeusz Lorenc, Christopher Duran, Lars Smits, Sahana Manoli, Jiri Stiller, David Edwards (2012) [WheatGenome.info: An integrated database and portal for wheat genome information](#). Plant and Cell Physiology 53(2): e2.
5. Agnieszka A. Golicz, Philipp E. Bayer, Paula A. Martinez, **Kaitao Lai**, Michał T. Lorenc, Salman Alamery, Alice Hayward, Reece Tollenaere, Jacqueline Batley, David Edwards, Yan Long and Jinling Meng (2013) [Characterising diversity in the Brassica genomes](#). Acta Horticulturae, 1005 : 33-48.
6. Manuel Zander, Dhwani A. Patel, Angela Van de Wouw, **Kaitao Lai**, Michał T. Lorenc, Emma Campbell, Alice Hayward, David Edwards, Harsh Raman, Jacqueline Batley (2013) [Identifying genetic diversity of avirulence genes in *Leptosphaeria maculans* using whole genome sequencing](#). Functional & Integrative Genomics, 13(3): 295-308.
7. Paul J. Berkman, Paul Visendi, Hong C. Lee, Jiri Stiller, Sahana Manoli, Michał T. Lorenc, **Kaitao Lai**, Jacqueline Batley, Delphine Fleury, Hana Šimková, Marie Kubaláková, Song Weining, Jaroslav Doležel, David Edwards (2013) [Dispersion and domestication shaped the genome of bread wheat](#). Plant Biotechnology Journal, 11(5): 564-571.
8. Michał T. Lorenc, Satomi Hayashi, Jiri Stiller, Hong Lee, Sahana Manoli, Pradeep Ruperao, Paul Visendi, Paul J. Berkman, **Kaitao Lai**, Jacqueline Batley, David Edwards (2012) [Discovery of Single Nucleotide Polymorphisms in complex genomes using SGSAutoSNP](#). Biology 1(2): 370-382.
9. Hong Lee, **Kaitao Lai**, Michał T. Lorenc, Mike Imelfort, Chris Duran, David Edwards (2012) [Bioinformatics tools and databases for analysis of next-generation sequence data](#). Briefings in Functional Genomics 11(1): 12-24.
10. Paul J. Berkman, **Kaitao Lai**, Michał T. Lorenc, David Edwards (2012) [Next generation sequencing applications for wheat crop improvement](#). American Journal of Botany 99(2): 365-371.

11. Paul J. Berkman, Adam Skarszewski, Sahana Manoli, Michał T. Lorenc, Jiri Stiller, Lars Smits, **Kaitao Lai**, Emma Campbell, Marie Kubaláková, Hana Šimková, Jacqueline Batley, Jaroslav Doležel, Pilar Hernandez, and David Edwards (2012) [Sequencing wheat chromosome arm 7BS delimits the 7BS/4AL translocation and reveals homoeologous gene conservation](#). Theoretical and Applied Genetics 124(3): 423-432.
12. Paul J. Berkman, Adam Skarszewski, Michał T. Lorenc, **Kaitao Lai**, Chris Duran, Edmund Y.S. Ling, Jiri Stiller, Lars Smits, Michael Imelfort, Sahana Manoli, Megan McKenzie, Marie Kubaláková, Hana Šimková, Jacqueline Batley, Delphine Fleury, Jaroslav Doležel and David Edwards (2011) [Sequencing and assembly of low copy and genic regions of isolated *Triticum aestivum* chromosome arm 7DS](#). Plant Biotechnol J 9(7), 768-775.
13. Daniel J Marshall, Alice Hayward, Dominic Eales, Michael Imelfort, Jiri Stiller, Paul J Berkman, Terry Clark, Megan McKenzie, **Kaitao Lai**, Chris Duran, Jacqueline Batley and Dave Edwards (2010) [Targeted identification of genomic regions using TAGdb](#). Plant Methods 6: 19.

Book chapters

1. **Kaitao Lai**, Michał Tadeusz Lorenc, David Edwards (2015) Molecular Marker Databases. In: [Plant Genotyping: Methods and Protocols](#), Methods in Molecular Biology, Vol. 1245. Ed. Batley, J. Humana Press (Springer) (USA) Publishing in progress.
2. **Kaitao Lai**, David Edwards (2015) WheatGenome.info: A resource for wheat genomics resource. In: Plant Bioinformatics: Methods and Protocols, Methods in Molecular Biology. Ed. Edwards, D. Humana Press (Springer) (USA) Waiting for submission.

Conference Posters

3. **Kaitao Lai**, Michal Lorenc, Hong Lee, Paul Berkman, Paul Visendi Muhindira, Philipp Bayer, Pradeep Ruperao, Kenneth Chan, Sahana Manoli, Jiri Stiller, Jacqueline Batley, Dave Edwards. Genetic variation across Australian bread wheat varieties, poster for International Plant and Animal (PAG) XXI, 14th January, San Diego, California, USA. Poster is also for Plant and Animal (PAG) Asia 2013, 17th and 19th March, 2013, Singapore.
4. Manuel Matias Zander, Dhwani A. Patel, Angela Van de Wouw, **Kaitao Lai**, Michal T. Lorenc, Alice C Hayward, Agnieszka Golicz, Philipp Bayer, David Edwards, Jacqueline Batley. Understanding Genome Structure and Diversity of the Plant Pathogen, *Leptosphaeria maculans*, poster for International Plant and Animal (PAG) XXI, 13rd January, 2014, San Diego, California, USA.
5. Philipp Emanuel Bayer, Paul Visendi, Kenneth Chan, Michal Lorenc, **Kaitao Lai**, Jessica Dalton-Morgan, Yan Long, Jinling Meng, Jacqueline Batley, David Edwards, Skim Based Genotyping by Sequencing in Complex Crop Genomes, poster for

International Plant and Animal (PAG) XXI, 14th January, 2013, San Diego, California, USA.

6. Paul Berkman, Paul Visendi Muhindira, Michal Lorenc, **Kaitao Lai**, Hong Lee, Mike Imelfort, Sahana Manoli, Pradeep Ruperao, Chris Duran, Hana Šimková, Marie Kubaláková, Emma Campbell, Pilar Hernandez, Jiri Stiller, Jaroslav Doležel, Jacqueline Batley and David Edwards, Gene content, loss, conservation and genetic variation among *Triticum aestivum* group 7 chromosomes, the Plant and Animal Genome XX Conference, 14th-18th January, 2012, San Diego, California, USA.
7. Hong Lee, Paul J. Berkman, **Kaitao Lai**, Michal Lorenc, Mike Imelfort, Sahana Manoli, Pradeep Ruperao, Chris Duran, Emma Campbell, Alice C Hayward, Jessica Dalton-Morgan, Jiri Stiller, Jacqueline Batley and David Edwards, Finding Function in Complex Crop Genomes, the Plant and Animal Genome XX Conference, 14th and 18th January, 2012, San Diego, California, USA.
8. Paul J. Berkman, Adam Skarszewski, Michał T. Lorenc, **Kaitao Lai**, Chris Duran, Edmund Y.S. Ling, Jiri Stiller, Lars Smits, Michael Imelfort, Sahana Manoli, Megan McKenzie, Marie Kubaláková, Hana Šimková, Jacqueline Batley, Delphine Fleury, Jaroslav Doležel and David Edwards, Applying second-generation sequencing technology in the assembly and analysis of the wheat group 7 chromosomes, ITMI 2010, September 2010.

Presentations

1. **Kaitao Lai**, Michal Lorenc, Chris Duran, Satomi Hayashi, Jiri Stiller, Hong Lee, Paul Visendi, Pradeep Ruperao, Sahana Manoli, Jacqueline Batley, Dave Edwards. SNP discovery in complex genomes using SGSautoSNP and autoSNPdb, computer demo for International Plant and Animal (PAG) XXI, 13th January, 2013, San Diego, California, USA.
2. **Kaitao Lai**, Michal Lorenc, Hong Lee, Paul Berkman, Paul Visendi Muhindira, Philipp Bayer, Kenneth Chan, Pradeep Ruperao, Sahana Manoli, Jiri Stiller, Jacqueline Batley, Dave Edwards. Characterise genetic variations across Australian bread wheat varieties, 2012 UQ seminar series: computational biology, 9th August 2012, IMB, UQ, Brisbane, Australia.
3. **Kaitao Lai**, Nikki Appleby, Paul Berkman, Chris Duran, Michal Lorenc, Sahana Manoli, Jiri Stiller, Matt Hayden, Annalise Mason, Jacqueline Batley, David Edwards. Single nucleotide polymorphism discovery in wheat transcriptome, 25th October 2011, ACPFG Joint Research Meeting, Adelaide, Australia.

Publications included in this thesis

1. Paul J. Berkman, **Kaitao Lai**, Michał T. Lorenc, David Edwards (2012) [Next generation sequencing applications for wheat crop improvement](#). American Journal of Botany 99(2): 365-371. – Incorporated partially in Chapter 1.

Contributor	Statement of contribution
Paul J. Berkman	Designed conception (35%) Wrote and edited paper (30%)
Kaitao Lai (Candidate)	Designed conception (20%) Wrote and edited paper (20%)
Michał T. Lorenc	Designed conception (10%) Wrote and edited paper (20%)
David Edwards	Designed conception (35%) Wrote and edited paper (30%)

2. Hong Lee, **Kaitao Lai**, Michał T. Lorenc, Mike Imelfort, Chris Duran, David Edwards (2012) [Bioinformatics tools and databases for analysis of next-generation sequence data](#). *Briefings in Functional Genomics* 11(1): 12-24. – Incorporated partially in Chapter 1.

Contributor	Statement of contribution
Hong Lee	Designed conception (25%) Wrote and edited paper (25%)
Kaitao Lai (Candidate)	Designed conception (20%) Wrote and edited paper (25%)
Michał T. Lorenc	Designed conception (15%) Wrote and edited paper (15%)
Mike Imelfort	Designed conception (10%) Wrote and edited paper (5%)
Chris Duran	Designed conception (5%) Wrote and edited paper (5%)
David Edwards	Designed conception (25%) Wrote and edited paper (25%)

3. **Kaitao Lai**, Michał Tadeusz Lorenc, David Edwards (2015) Molecular Marker Databases. In: [Plant Genotyping: Methods and Protocols](#), Methods in Molecular Biology, Vol. 1245. Ed. Batley, J. Humana Press (Springer) (USA) Publishing in progress. – Incorporated partially in Chapter 1.

Contributor	Statement of contribution
Kaitao Lai (Candidate)	Designed conception (35%) Wrote and edited paper (40%)
Michał T. Lorenc	Designed conception (30%) Wrote and edited paper (20%)
David Edwards	Designed conception (35%) Wrote and edited paper (40%)

4. **Kaitao Lai**, Chris Duran, Paul J. Berkman, Michał T. Lorenc, Jiri Stiller, Sahana Manoli, Matthew J. Hayden, Kerrie L. Forrest, Delphine Fleury, Ute Baumann, Manuel Zander, Annaliese S. Mason, Jacqueline Batley, David Edwards (2012) [Single Nucleotide Polymorphism Discovery from Wheat Next Generation Sequence Data](#). Plant Biotechnology Journal 10(6): 743–749. – Incorporated completely as Chapter 2.

Contributor	Statement of contribution
Kaitao Lai (Candidate)	Designed conceptions (25%) Coding (40%) Database implementation (40%) Analysis and interpretation of data (30%) Wrote and edited paper (30%)
Chris Duran	Designed conceptions (15%) Coding (40%) Database implementation (20%) Analysis and interpretation of data (10%) Wrote and edited paper (5%)
Paul J. Berkman	Designed conceptions (10%) Coding (10%) Database implementation (10%)

	Analysis and interpretation of data (10%) Wrote and edited paper (5%)
Michał T. Lorenc	Designed conceptions (5%) Coding (10%) Database implementation (10%) Analysis and interpretation of data (5%) Wrote and edited paper (5%)
Jiri Stiller	Designed conceptions (5%) Database implementation (10%) Analysis and interpretation of data (5%) Wrote and edited paper (5%)
Sahana Manoli	Database implementation (10%) Analysis and interpretation of data (10%)
Matthew J. Hayden	Data Generation (100%) Analysis and interpretation of data (5%)
Kerrie L. Forest	Analysis and interpretation of data (5%) Wrote and edited paper (5%)
Delphine Fleury	Analysis and interpretation of data (5%) Wrote and edited paper (5%)
Ute Baumann	Analysis and interpretation of data (5%) Wrote and edited paper (5%)
Manuel Zander	Performed validation and analysis (60%) Analysis and interpretation of data (5%)
Annaliese S. Mason	Performed validation and analysis (40%) Analysis and interpretation of data (5%)
Jacqueline Batley	Designed conceptions (15%) Wrote and edited paper (5%)
David Edwards	Designed conceptions (25%) Wrote and edited paper (30%)

5. **Kaitao Lai**, Michał T. Lorenc, Hong Lee, Paul J. Berkman, Philipp Emanuel Bayer, Paul Visendi Muhindira, Pradeep Ruperao, Timothy L. Fitzgerald, Manuel Zander, Chon-Kit Kenneth Chan, Sahana Manoli, Jiri Stiller, Jacqueline Batley, David Edwards (2014) Identification and characterisation of more than 4 million inter-varietal SNPs across the group 7 chromosomes of bread wheat, Plant Biotechnology Journal 13(1): 97-104. – Incorporated completely as Chapter 3.

Contributor	Statement of contribution
Kaitao Lai (Candidate)	Designed conception (25%) Coding (40%) Database implementation (10%) Analysis and interpretation of data (20%) Wrote and edited paper (30%)
Michał T. Lorenc	Designed conception (10%) Coding (30%) Database implementation (10%) Analysis and interpretation of data (10%)
Hong Lee	Designed conception (10%) Coding (20%) Analysis and interpretation of data (5%) Wrote and edited paper (10%)
Paul J. Berkman	Designed conception (10%) Coding (5%) Analysis and interpretation of data (5%) Wrote and edited paper (5%)
Philipp Emanuel Bayer	Designed conception (5%) Coding (5%) Analysis and interpretation of data (5%) Wrote and edited paper (5%)
Paul Visendi Muhindira	Analysis and interpretation of data (5%) Wrote and edited paper (5%)
Pradeep Ruperao	Analysis and interpretation of data (5%)

	Wrote and edited paper (5%)
Timothy L. Fitzgerald	Analysis and interpretation of data (5%) Wrote and edited paper (5%)
Manuel Zander	Performed validation and analysis (100%) Analysis and interpretation of data (5%)
Chon-Kit Kenneth Chan	Database implementation (80%) Analysis and interpretation of data (5%)
Sahana Manoli	Analysis and interpretation of data (5%)
Jiri Stiller	Analysis and interpretation of data (5%)
Jacqueline Batley	Designed conceptions (15%) Wrote and edited paper (5%)
David Edwards	Designed conception (25%) Analysis and interpretation of data (20%) Wrote and edited paper (30%)

6. **Kaitao Lai**, Paul J Berkman, Michał Tadeusz Lorenc, Christopher Duran, Lars Smits, Sahana Manoli, Jiri Stiller, David Edwards (2012) [WheatGenome.info: An integrated database and portal for wheat genome information](#). Plant and Cell Physiology 53(2): e2. – Incorporated completely as Chapter 4.

Contributor	Statement of contribution
Kaitao Lai (Candidate)	Designed conception (25%) Database implementation (20%) Wrote and edited paper (30%)
Paul J. Berkman	Designed conception (20%) Database implementation (20%) Wrote and edited paper (30%)
Michał T. Lorenc	Designed conception (15%) Database implementation (15%) Wrote and edited paper (10%)
Chris Duran	Database implementation (5%)

Lars Smits	Database implementation (15%)
Sahana Manoli	Database implementation (5%)
Jiri Stiller	Database implementation (10%)
David Edwards	Designed conception (40%) Database implementation (10%) Wrote and edited paper (30%)

7. **Kaitao Lai**, David Edwards (2015) WheatGenome.info: A resource for wheat genomics resource. In: Plant Bioinformatics: Methods and Protocols, Methods in Molecular Biology. Ed. Edwards, D. Humana Press (Springer) (USA) In press (accepted). - Incorporated partially in Chapter 4.

Contributor	Statement of contribution
Kaitao Lai (Candidate)	Designed conception (100%) Wrote and edited paper (100%)

Contributions by others to the thesis

Principal supervisor Prof. David Edwards contributed to the conception and design of the project. Prof. David Edwards and co-supervisor Dr Jacqueline Batley respectively contributed to the editing of the sections and critically revising the sections in this thesis. Christopher Duran developed the autoSNPdb method. Michał T. Lorenc implemented SGSAutoSNP. Hong Lee implemented some tools for SNP density analysis. Manuel Zander and Annaliese S. Mason performed the laboratory validation of the varietal SNPs.

Statement of parts of the thesis submitted to qualify for the award of another degree

None.

Acknowledgements

It is a memorable experience during these years to reach the end of my PhD study journey. I suffered many painful and heart-sinking moments, and enjoyed the sheer exhilaration. This PhD thesis is the culmination of my PhD research work with helps from my supervisors and colleagues.

I would like to gratefully and sincerely thank my principal supervisor, Prof. David Edwards, for his guidance, understanding, patience, and most importantly, his recognition of my potential capability in my PhD research. He gave me belief and confidence to cope with the difficulties in research. I would also like to thank my co-supervisor A/Prof. Jacqueline Batley for her assistance and guidance. Jacqueline provided many helps for the assistance in generating research data. In addition, I would like to thank Dr. Jiri Stiller. Jiri supported measured approach to my PhD research with major influence.

I also thank my group colleagues. They have provided help and contributed to my PhD research work. Christopher Duran, Michał T. Lorenc and Hong Lee helped greatly in shaping my thinking about my PhD research.

I acknowledge the funding provided by the Australian Research Council (Projects LP0882095, LP0883462, LP110100200 and DP0985953) and the Grains Research and Development Corporation (Project DAN00117), and the support from the Australian Genome Research Facility (AGRF), the Queensland Cyber Infrastructure Foundation (QCIF), the Australian Partnership for Advanced Computing (APAC), Queensland Facility for Advanced Bioinformatics (QFAB) and Bioplatforms Australia.

To my mother, my father-in-law and mother-in-law, thank you all for your endless love, supporting me and believing in me. To my brothers, Fengkai and Ruokai, thank you all for your encourage, and all for continuously reminding me of the reasons for doing a PhD in the first place.

Finally, and most importantly, I would like to thank my wife Wan Su. Her support, encouragement, quiet patience and unwavering love were undeniably the bedrock upon which the past five years of my life have been built. Her tolerance of my occasional sheer ignorance of her feeling is a testament in itself of her unyielding devotion and love.

Keywords

Triticum aestivum, genetic marker, single nucleotide polymorphism, SNP discovery, transcriptome, SNP density, next-generation sequencing, genomic database, diversity, evolution

Australian and New Zealand Standard Research Classifications (ANZSRC)

ANZSRC code: 060102, Bioinformatics, 50%

ANZSRC code: 060408, Genomics, 40%

ANZSRC code: 060405, Gene Expression, 10%

Fields of Research (FoR) Classification

FoR code: 0601, Biochemistry and Cell Biology, 50%

FoR code: 0604, Genetics, 50%

Table of Contents

Abstract	i
Declaration by author.....	iii
Publications during candidature	iv
Acknowledgements.....	xiii
Keywords	xiv
Australian and New Zealand Standard Research Classifications (ANZSRC).....	xiv
Fields of Research (FoR) Classification.....	xiv
Table of Contents	xv
Chapter 1 Introduction and literature review.....	1
1.1 DNA sequencing technologies.....	3
1.1.1 First-generation DNA sequencing technologies	3
1.1.2 Second and Third-generation DNA sequencing technologies.....	5
1.2 Genetic variation.....	8
1.2.1 Molecular genetic markers.....	8
1.2.1.1 The history of DNA based genetic markers	9
1.2.1.2 Restriction Fragment Length Polymorphism (RFLP)	10
1.2.1.3 Random Amplified Polymorphic DNA (RAPD)	11
1.2.1.4 Amplified fragment length polymorphism (AFLP).....	11
1.2.1.5 Diversity arrays technology (DArT)	12
1.2.1.6 Simple sequence repeats (SSRs).....	12
1.2.1.7 Single nucleotide polymorphisms (SNPs).....	13
1.2.1.8 Discovery of single-nucleotide polymorphisms (SNPs)	14
1.3 The transcriptome	16
1.3.1 Traditional transcriptome analysis	16
1.3.2 Second generation transcriptomics.....	17
1.3.3 Tools for the analysis of second generation transcriptomic data.....	18
1.3.4 Differential homoeologous gene expression.....	19
1.4 Integrated databases for crop plants	21
1.4.1 Molecular databases	22
1.5 <i>Triticum aestivum</i>.....	29
1.5.1 Importance of <i>Triticum aestivum</i>	29
1.5.2 Genome evolution and domestication in <i>Triticum aestivum</i>	33
1.5.3 Approaches to sequence the bread wheat genome	35
1.5.3.1 Transcriptome analysis in wheat	39
1.6 Summary	40
Chapter 2 Single-nucleotide polymorphism (SNP) identification from transcriptome data from Australian bread wheat (<i>Triticum aestivum</i>) cultivars	42
2.1 Introduction	42
2.2 Experimental procedures.....	45
2.2.1 Data generation.....	45
2.2.2 Data processing	45

2.2.2.1	Trimming sequence data	45
2.2.2.2	Assembling sequence data	46
2.2.2.3	SNP discovery and annotation	46
2.2.3	Validation.....	47
2.3	Results and discussion	48
2.3.1	Sequence assembly	48
2.3.2	SNP discovery and abundance	48
2.3.3	Validation.....	52
2.3.4	Analysis of base changes	55
2.3.5	Sequence annotation	56
2.4	Conclusions.....	57
Chapter 3	Identification and characterisation of more than 4 million inter-varietal SNPs across the group 7 chromosomes of bread wheat.....	58
3.1	Introduction	58
3.2	Experimental procedures.....	61
3.2.1	Data selection and generation	61
3.2.2	Data processing	62
3.2.3	SNP prediction.....	63
3.2.4	SNP matrix production and transition/transversion ratio analysis	63
3.2.5	SNP density and gene analysis	64
3.3	Validation	64
3.4	Results	68
3.5	Discussion	76
Chapter 4	WheatGenome.info: An integrated database and portal for wheat genome resources	83
4.1	Introduction	83
4.2	Materials and methods.....	83
4.2.1	Wheat group 7 sequence and 4AL	83
4.2.2	<i>Brachypodium distachyon</i> gene and exon annotation	84
4.2.3	Uniref90 annotation.....	84
4.2.4	Inter-varietal SNPs between 16 Australian wheat varieties	85
4.2.5	Workflow	85
4.3	Database contents.....	86
4.3.1	A wheat genome viewer for annotated chromosome arm assemblies	86
4.3.2	Searching wheat sequence data using TAGdb.....	88
4.3.3	Comparative wheat genome and genetic maps on CMap and CMap3D	90
4.3.4	Annotated wheat EST single nucleotide polymorphisms within autoSNPdb.....	91
4.3.5	Wheat genome Wiki.....	93
4.4	Conclusions and future direction.....	93
Chapter 5	Concluding remarks and future directions.....	94
5.1	Concluding remarks	94
5.2	Future directions	96
References	98	
Appendices	117	

List of Figures

Figure 1-1 Global wheat, area, production and consumption from 1960/1961 to 2010/2011 (Adapted from USDA, PSD Online, January 2011).....	31
Figure 1-2 Global wheat production and consumption from 2009/2010 to 2014/2015 (Adapted from USDA, 2014).....	32
Figure 1-3 Market share of world wheat exports, from 2005/2006 to 2010/2011 (%) (Adapted from USDA, PSD Online, January 2011).....	32
Figure 1-4 Schematic Representation of the Evolutionary History of Wheat Species (<i>Triticum</i> and <i>Aegilops</i>) (Adapted from Chantret et al., 2005).	35
Figure 1-5 Crop genome size of some important crop genomes has been listed in order from smallest to largest.	36
Figure 2-1 Histogram showing the percentage of reads containing SNPs for contigs representing increasing numbers of reads.....	50
Figure 2-2 Histogram presenting the mean number of SNPs for contigs with increasing numbers of reads.....	51
Figure 2-3 Venn diagram displaying the number of reads with combinations of SwissProt, GenBank and <i>Brachypodium</i> orthologue annotation.	56
Figure 3-1 The transition/transversion ratio and standard deviation across chromosomes 7A, 7B and 7D.	70
Figure 3-2 Ts/Tv ratio across the 7A, 7B and 7D syntenic builds. Each red column in the chart represents Ts/Tv ratio of SNPs within the corresponding region. The blue line represents ratio value of 1. Most of Ts/Tv ratios are higher than 1 for the 7A, 7B Syntenic builds while most of Ts/Tv ratios are lower than 1 for the 7D Syntenic build.	71
Figure 3-3 Phylogenetic relationships of 16 Australian wheat varieties based on SNP data obtained in this study. The scalar bar represents the number of base substitutions per 100 positions.	73
Figure 3-4 SNP density across the 7A, 7B and 7D syntenic builds.	75
Figure 4-1 Flowchart demonstrates the general workflow of the WheatGenome.info integrated database.	86
Figure 4-2 Example of the detailed information for the wheat genome 7AS syntenic build from the reference view of GBrowse2. Several tracks of annotation are available, including Uniref90, Genes, Contigs, SNPs and Exons.	88
Figure 4-3 Screenshot of TAGdb showing the alignment of short reads from wheat variety Chinese Spring to a sample query sequence.....	89
Figure 4-4 An inter-species comparison between a physical map of the wheat 7DS syntenic build chromosome, a genetic map of <i>Aegilops tauschii</i> chromosome 7D and a physical map of the <i>Brachypodium distachyon</i> chromosome 3 (between 39 and 45Mbp) using CMap3D.....	91
Figure 4-5 The wheat autoSNPdb web interface displaying the wheat sequence assembly, with predicted SNPs as vertical bars.	92

List of Tables

Table 1-1 The advantages and disadvantages of next-generation sequencing technologies.....	6
Table 1-2 Evolving genotyping methods.....	9
Table 1-3 Examples of molecular marker databases with different types of markers for crop improvement.....	27
Table 2-1 SNP validation.....	53
Table 2-2 Base changes and the number of occurrences within SNPs in wheat autoSNPdb.....	55
Table 3-1 Pairwise inter-varietal SNP matrix for chromosomes 7A, 7B and 7D between 16 Australian wheat varieties.....	66
Table 3-2 Summary of single nucleotide polymorphism validation.....	67
Table 3-3 Subgenomic varietal SNP density for 16 Australian wheat varieties.....	70
Table 3-4 Transition and transversion SNPs for each variety identified within the 7A, 7B and 7D chromosomes.....	72
Table 3-5 Summary of the 16 Australian wheat varieties and associated traits	78

List of Abbreviations used in the thesis

autoSNPdb	automatic annotated Single Nucleotide Polymorphism database
BAC	Bacterial Artificial Chromosome
BLAST	Basic Local Alignment Search Tool
EST	Expressed Sequence Tag
GBrowse2	Generic Genome Browser version 2
NGS	Next-generation Sequencing
PCR	Polymerase Chain Reaction
SNP	Single-Nucleotide Polymorphism
TAGdb	Tag database

Chapter 1 Introduction and literature review

Next-generation high-throughput DNA sequencing (NGS) technologies, also known as second-generation sequencing (SGS) technologies, have opened fascinating opportunities in the life sciences and for the analysis of plants on a genomic scale (Brautigam and Gowik, 2010, Ansorge, 2009). In recent years, NGS methods have become widely available and cost effective. Several kinds of commercial sequencers using NGS technologies have been introduced, including the Roche (454) GS FLX sequencer, Illumina HiSeq sequencing systems, Illumina genome analyser, Applied Biosystems SOLiD sequencer, and the Ion Proton System. NGS can be applied for genome assembly, mutation discovery, enabling metagenomics, defining DNA-protein interactions and regulatory protein binding (Mardis, 2008).

NGS technologies are driving increasingly affordable and high-resolution analyses of plant transcriptomes through sequencing of their associated complementary DNA (cDNA) populations. The related analytical platform is referred to as RNA-sequencing (RNA-seq). RNA-seq has been proven to be a useful tool with a diverse range of applications, from detailed studies of biological processes at the cell type-specific level, to providing insights into fundamental questions in plant biology on an evolutionary time scale. Applications of RNA-seq could help unite the “omics” fields of transcriptomics, proteomics, and metabolomics into a new common systems biology paradigm.

NGS technologies support the completion of reference genome sequences for many important crops, and provide opportunities for improving our understanding of the history of plant domestication and to accelerate crop improvement (Feuillet et al., 2011). The genomes of various important crop species experienced ancestral or recent polyploidisation events (Soltis et al., 2009). Multiple homoeologous gene copies, chromosomal rearrangements and amplification of repetitive DNA within large and complex crop genomes challenge genome analysis and gene discovery. NGS technological advances in molecular genetics and genomics provide unprecedented opportunities to analyse and access ever more recalcitrant genomes (Edwards et al.,

2013). Crop plant comparative genomics is being revolutionised by these reference genome sequences and the new generation of experimental and computational approaches to comparative genomics (Morrell et al., 2011).

NGS technologies and advances in genomic technologies have enabled the discovery of the full spectrum of variants from common to rare alleles in the human population (Indap et al., 2013). The advances in genome sequencing technologies provide unprecedented chances to characterise individual genomic landscapes and identify mutations relevant for diagnosis and therapy. Whole-exome sequencing using NGS technologies is becoming popular in the human genetics community due to the moderate costs and relatively straightforward interpretation and analysis of results (Pabinger et al., 2013).

Chromatin immunoprecipitation followed by sequencing (ChIP-seq) is a technique for genome-wide profiling of DNA-binding proteins, histone modifications or nucleosomes, and is one of the early applications of NGS (Johnson et al., 2007, Barski et al., 2007, Robertson et al., 2007, Mikkelsen et al., 2007). ChIP-seq provides higher resolution, less noise and greater coverage than its array-based predecessor ChIP-chip and therefore provides substantially improved data. ChIP-seq becomes an indispensable tool for studying gene regulation and epigenetic mechanisms. Large quantities of data, and effective computational analysis from ChIP-seq experiments will be regarded as crucial factors for uncovering biological mechanisms (Park, 2009).

Assembling plant genomes *de novo* by NGS technologies remains challenging although genome sequencing is now affordable. The plant kingdom has abundant variation and diversity. Therefore, each plant-sequencing project has unique analysis requirements. Assembling and analysing raw sequence data still requires multiple libraries with different insert sizes using high-coverage sequencing by NGS technologies and substantial bioinformatics effort and expertise (Schatz et al., 2012).

Wheat is a leading example of a major polyploid crop genome with large genome size and great complexity presenting significant challenges (Edwards et al., 2013). Wheat breeding has been a principal contributor to establishing a viable wheat industry in Australia and feeding the world in general. Research has been undertaken through public

and private sector programs and has been instrumental in developing varieties suited to Australia's dry environment. A major objective of these programs is to develop varieties which are resistant to stem rust and leaf rust, both of which prevailed at epidemic levels in the early 20th-century (ABS, 2006). In this thesis, NGS methods have been employed for wheat genome diversity analysis, which provides support for wheat breeding in Australia.

1.1 DNA sequencing technologies

1.1.1 First-generation DNA sequencing technologies

The Sanger enzymatic dideoxy technique was first published in 1977 by Fred Sanger and Alan R. Coulson, through two methodological papers on the rapid determination of DNA sequence (Sanger et al., 1977a, Sanger et al., 1977b). The Sanger sequencing (chain-termination method) method was developed by Fred Sanger (Sanger et al., 1977b). The Maxam-Gilbert sequencing method was developed by Allan Maxam and Walter Gilbert in the same year. This chemical sequencing method allowed purified samples of double-stranded DNA to be used without further cloning (Maxam and Gilbert, 1977). However, Sanger sequencing has the advantages of reduced handling of toxic chemicals and radioisotopes, therefore, Sanger sequencing was the dominant DNA sequencing method used for the next 30 years (Schuster, 2008). Fluorescence detection of the DNA fragments was developed as a method for the partial automation of DNA sequence analysis (Smith et al., 1986). The platforms for massive parallel DNA sequencing read production have been developed and have become widespread (Shendure and Ji, 2008, Mardis, 2008). The first automated DNA sequencers were produced using this method and commercialised by Applied Biosystems (Ansorge, 2009). In addition, Wilhelm Ansorge at the European Molecular Biology Laboratory (EMBL) developed a non-radioactive automated method for DNA sequence determination in 1986 (Ansorge et al., 1986) and an ultrasensitive detection method of fluorescent bands during electrophoresis in 1987 (Ansorge et al., 1987). These methods developed by Ansorge at the EMBL were commercialised by Pharmacia-Amersham, later General Electric (GE) Healthcare (Ansorge, 2009). The Sanger method was adopted for the first sequencing of

a genome region. In this project, sequence determination of the complete gene locus for the hypoxanthine phosphoribosyltransferase (HPRT) gene was performed using above methods developed by Ansorge at the EMBL. In that project, the important concept of paired-end sequencing was also recommended for the first time (Edwards et al., 1990, Ansorge, 2009). The automated Sanger method dominated the market in the DNA sequencing industry for almost two decades between 1986 and 2006 (Metzker, 2010) and was used to achieve a number of significant accomplishments, including the only finished-grade human genome sequence (International Human Genome Sequencing Consortium, 2004).

The demand for the development of techniques for higher sequencing throughput was triggered by the international community in the human genome project (Metzker, 2010). A high-throughput capillary array DNA sequencer was implemented in the Hitachi laboratories by the team of H. Kambara in Japan. Later, two companies, ABI (commercialising the Kambara system) and Amersham (developing further the system set up in the US by the Molecular Dynamics company), commercialised automated sequencing using parallel analysis in systems of up to 384 capillaries. However, the limitation of the Sanger sequencing protocols for larger sequence output were the need for gels or polymers used as sieving separation media and the difficulty of total automation of the sample preparation methods. These limitations initiated efforts to develop techniques without gels, which would allow sequence determination on very large numbers (i.e. millions) of samples in parallel (Ansorge, 2009). Since 2000, the related sequencing methods and equipment developments have continued in several groups in European laboratories and in the US (Ansorge, 2009). A patent application by EMBL presented a large-scale DNA sequencing technique without gels, extending primers in ‘sequencing-by-synthesis, addition and detection of the incorporated base’, proposing the use of the so-called ‘reversible terminators’, entitling array-based DNA sequencing approaches (Turcatti et al., 2008), for speed and efficiency (Ansorge, 1991). The principle described in the patent application is in part similar to some of the second-generation devices that are used today (Ansorge, 2009).

1.1.2 Second and Third-generation DNA sequencing technologies

Since 2006, there has been a fundamental shift from the automated Sanger method to newer methods, termed “second-generation sequencing” or “next-generation sequencing” technologies (Metzker, 2010). Second-generation sequencing had overcome the inertia of a field that relied on Sanger-sequencing for 30 years (Schuster, 2008). One advantage of second-generation sequencing platforms is the determination of the sequence data from amplified DNA fragments, avoiding the requirement for the cloning of DNA fragments (Schuster, 2008).

Second-generation (next-generation) sequencing technologies include Roche/454 pyrosequencing (Margulies et al., 2005), Illumina/Solexa sequencing by synthesis, SOLiD sequencing by ligation, and Ion Torrent sequencing. Each sequencing method has its advantages and disadvantages (Table 1-1). The 454 pyrosequencing technology can deliver long read size and has a fast running time, however, the cost of runs are expensive and the data suffers from homopolymer errors. Sequences of a single repeat unit are known as homopolymers, for example [AAAAAA]. In 454 pyrosequencing technology, the number of bases (the homopolymer length) is determined from the flowgram. As the accuracy of the flowgram decreases with increasing homopolymer length, this causes an increase in the number of insertion/deletion errors around homopolymer tracts (Balzer et al., 2010). In October 2013, Roche announced the discontinuation of the 454 pyrosequencing technology (Nederbragt, 2014). Illumina/Solexa sequencing by synthesis has the potential for high sequence yield as well as relatively high accuracy. It is now the dominant platform in the industry. SOLiD sequencing by ligation delivers short reads for assembly (Liu et al., 2012) but is now being discontinued. Ion Torrent sequencing could deliver cheaper data and a fast running time, however, this method also suffers from homopolymer, as well as substitution errors (Liu et al., 2012, Quail et al., 2012).

Table 1-1 The advantages and disadvantages of next-generation sequencing technologies.

Next-generation sequencing technology	Advantages	Disavantages	References
Roche/454 pyrosequencing	It can deliver long read size and has a fast running time	The cost of runs are expensive and the data suffers from homopolymer errors.	(Balzer et al., 2010)
Illumina/Solexa sequencing by synthesis	It has the potential for high sequence yield as well as relatively high accuracy.		
SOLiD sequencing by ligation		It delivers short reads for assembly.	(Liu et al., 2012)
Ion Torrent sequencing	It could deliver cheaper data and a fast running time	It also suffers from homopolymer, as well as substitution errors.	(Liu et al., 2012, Quail et al., 2012)
Pacific Biosciences	It produces reads with average lengths of 3,000 to 5,000 bp, with the longest reads over 20,000 base pairs.	The accuracy of this sequencing technology remains relatively poor.	www.pacificbiosciences.com

The principle of pyrophosphate detection was proposed in 1985 (Nyren and Lundin, 1985), and adopted as a method for DNA sequencing in 1988 (Hyman, 1988). The technique was further developed into a routinely functioning method (Ronaghi et al., 1996), leading to a technique commercialised for the analysis of samples in parallel in a picotitre plate. The GS instrument was developed by 454 and commercialised by Life Sciences in 2005 as the first second-generation system on the market (Ansorge, 2009). The 454 sequencer dramatically increased DNA sequencing capability and opened up new approaches to identify small RNAs (Rothberg and Leamon, 2008). The recent Roche 454 GS FLX+ Titanium technology is now capable of producing approximately one million reads up to 1,000 bp in read length in a single run of 23 hours (<http://www.454.com>).

The Solexa sequencing platform was commercialised in 2006, and was acquired by Illumina in early 2007. In 2008, Illumina introduced an upgrade with the Genome Analyser II that tripled output compared to the previous Genome Analyser instrument (Ansorge,

2009). Illumina's HiSeq2500 is one of the latest systems and the first Illumina sequencing system to feature two run modes: rapid-run and high-output run mode, using one or two flow cells simultaneously, providing a flexible and scalable platform. In high-output run mode, this sequencing system can deliver 540-600 Gbp read data up to 150 bp in read length from a dual flow cell run of 11 days. Illumina also produce the MiSeq personal sequencer. This sequencer is capable of producing approximately 15 Gbp reads up to 600 bp in read length in a single run of approximately 48 hours (<http://www.illumina.com>). Illumina's HiSeq X Ten is the most powerful sequencing platform to date, which is designed for population-scale human genome sequencing. The HiSeq X Ten is believed to be able to deliver the world's first \$1000 (US dollar) whole human genome (<http://www.illumina.com/systems/hiseq-x-sequencing-system/system.ilmn>).

Ion Torrent sequencing uses complementary metal-oxide semiconductor (CMOS) based high density array technology (Rothberg et al., 2011). The sequencer contains an integrated semiconductor device enabling non-optical DNA sequencing. The Ion PGM sequencer with the Ion 318 Chip v2 can deliver 1.2-2 Gbp 400-base reads in 7.3 hours, while the Ion Proton system with the Ion PI Chip can produce up to 10 Gbp of sequence data with up to 200-base fragment reads in 2-4 hours (<http://www.intorrent.com>).

Following the growth of second-generation sequencing technologies, so called third-generation sequencing technologies are also starting to emerge, predominantly led by single molecule sequencers. Pacific Biosciences established single-molecule, real-time sequencing (SMRT) with a DNA polymerase performing uninterrupted template-directed synthesis. This is considered to be one of the first "third-generation sequencing" sequencing systems to the market (Eid et al., 2009). The PacBio RS II is the latest product from Pacific Biosciences. This apparatus produces reads with average lengths of 3,000 to 5,000 bp, with the longest reads over 20,000 base pairs, enabling *de novo* assembly of complex genomes. However the accuracy of this sequencing technology remains relatively poor (<http://www.pacificbiosciences.com>).

The advantage of second and third-generation sequencing platforms includes greatly reduced cost per nucleotide read. The development of second-generation sequencing platforms has matured to deliver a variety of platforms to help the researchers address

significant biological problems (Robison, 2010). Future third-generation sequencing technologies are expected to produce data which is longer in read length and with a shorter running time than current technologies.

1.2 Genetic variation

1.2.1 Molecular genetic markers

Life's diversity and adaptation are products of evolution. Genetic variation is the foundation of all evolutionary change and is ultimately the basis of all life's diversity. Furthermore, the extent of genetic variation within a population affects its potential to adapt to environmental change (Pierce, 2011).

Molecular genetic markers are based on variation in the genome. They can be used to assess genetic diversity within and between related species, for the production of molecular genetic maps, and to link genotype and phenotype (Edwards and Batley, 2004). The detection and analysis of genetic variation plays an important role in plant breeding and this role is increasing with the continued development of genome technologies. Molecular genetic markers are important tools to characterise genetic variation and assist with genomic breeding. Processing and storing the growing abundance of molecular marker data being produced by advanced genome technologies requires the development of specific bioinformatics tools and advanced databases. Molecular marker databases range from species specific through to organism wide and often host a variety of additional related genetic, genomic or phenotypic information.

The characterisation of genetic variation can provide knowledge to help understand the molecular basis of various biological phenomena in plants. Phenotype-based genetic markers were used in Gregor Mendel's experiments in the nineteenth century. Later, phenotype-based genetic markers helped establish the theory of genetic linkage. More recently, DNA based markers have been developed to overcome the limitations of phenotype-based genetic markers (Agarwal et al., 2008).

The development of molecular marker technology and its application to plants has been extended over the last 30 years (Table 1-2). Molecular marker technology has evolved through several phases. DNA-based methods replaced early methods when the technologies for DNA analysis improved. The most common of early marker technologies was the assay of isozymes. The evolving DNA-based genetic markers include restriction fragment length polymorphism (RFLP) (Tanksley et al., 1989), amplified fragment length polymorphism (AFLP) (Vos et al., 1995), diversity arrays technology (DArT) (Jaccoud et al., 2001), simple sequence repeat (SSR) (Zietkiewicz et al., 1994) and single nucleotide polymorphism (SNP) (Marth et al., 1999).

Table 1-2 Evolving genotyping methods.

Method	Acronym	Technical Details
Restriction fragment length polymorphism	RFLP	It measures a difference in fragment size following digestion of genomic DNA. The limitations include large amounts of DNA, cost, time and low throughput.
Amplified fragment length polymorphism	AFLP	It produces fragments using restriction enzymes followed by the ligation of adaptors and PCR amplification. It is used in linkage studies to generate maps for QTL analysis.
Diversity arrays technology	DArT	DArT provides large number of markers and has been widely applied especially in the construction of genetic maps.
Simple sequence repeat	SSR	It is a highly polymorphic and informative marker and demonstrates a high degree of transferability between closely related species.
Single nucleotide polymorphism	SNP	It is direct marker, providing the exact nature of the allelic variants. It has been used for the high-resolution genetic mapping of traits, association studies, and genome-wide linkage disequilibrium analysis.

1.2.1.1 The history of DNA based genetic markers

The development of DNA analysis methods provided a chance to directly analyse differences in the genome of the organism rather than rely on inference from analysis of expressed proteins (as in isozyme analysis). Early hybridization-based methods employed hybridization of DNA to detect variation in the DNA samples and restriction fragment length polymorphism (RFLP) became the standard approach. Then early

hybridization-based methods were replaced by polymerase chain reaction (PCR) based methods in the 1990s. PCR-based methods significantly increased the feasibility of high-throughput marker screening. Due to a lack of sequence information for species, early PCR-based methods relied upon arbitrary primer sequences. The widespread adoption of more robust microsatellite or simple sequence repeat (SSR) markers overtook early PCR-based methods (Henry, 2012). The two most common modern molecular markers are simple sequence repeats (SSRs), also known as microsatellites, and single nucleotide polymorphisms (SNPs). Second-generation sequencing technologies have significantly accelerated the move to sequence-based markers and single-nucleotide polymorphisms (SNPs) have more recently replaced SSR markers due to available larger volumes of sequence data (Henry, 2008), and automation of SNP genotyping, particular in high-density platforms such as Illumina GoldenGate (Akhunov et al., 2009) and Infinium (Cavanagh et al., 2013, Wang et al., 2014) or Affymetrix Axiom (Hoffmann et al., 2011), and lower density customizable automated methods such as Kompetitive Allele Specific PCR (KASP) from KBioSciences (now part of LGC Genomics) (Semagn et al., 2014), BioMark system from Fluidigm (Wang et al., 2009a), and Sequenom MassARRAY (Gabriel et al., 2001).

Advances in genome sequencing technology and the increasing availability of genome sequences are providing an abundance of dense molecular markers (Lee et al., 2012, Imelfort et al., 2009). For example, sequence polymorphisms developed using the *Brassica rapa* genome sequence (Wang et al., 2011b) have been used to identify and characterise SNPs in agronomically important genes in canola (Hayward et al., 2012a, Hayward et al., 2012b, Tollenrae et al., 2012). In addition, the sequencing of isolated chromosome arms in wheat (Berkman et al., 2011a, Wicker et al., 2011, Berkman et al., 2012b, Hernandez et al., 2012, Berkman et al., 2013, International Wheat Genome Sequencing Consortium, 2014) has led to the identification of large numbers of molecular markers (Nie et al., 2012, Lai et al., 2012b).

1.2.1.2 Restriction Fragment Length Polymorphism (RFLP)

RFLPs are used to measure a difference in fragment size following digestion of genomic DNA by a restriction enzyme and interrogation with a hybridization probe. RFLP analysis can use one or more endonucleases (restriction enzymes) to digest genomic DNA from the target sample (Beckmann and Soller, 1983). The digested DNA is separated by electrophoresis and then transferred to a membrane for analysis of fragments by hybridization with a labeled DNA probe (Orita et al., 1989). The limitations of this method include the need for large amounts of DNA, cost, time and low throughput (Powell et al., 1996).

1.2.1.3 Random Amplified Polymorphic DNA (RAPD)

Random amplified polymorphic DNA (RAPD) techniques were designed for molecular ecology to determine taxonomic identity, assess kinship relationships, analyse mixed genome samples and create specific probes (Hadrys et al., 1992). RAPD analysis provides a cost- and time-effective alternative to restriction fragment-length polymorphism (RFLP) analysis (Penner et al., 1993). They were the first very widely adopted generic PCR marker (Penner et al., 1993) before AFLP technique is popular used.

1.2.1.4 Amplified fragment length polymorphism (AFLP)

The AFLP technique is developed based on the selective PCR amplification of restriction fragments from a total digest of genomic DNA (Vos et al., 1995). The AFLP technique produces fragments using restriction enzymes followed by the ligation of adaptors to the ends of the restriction fragments and PCR amplification from these adapters (Zabeau and Vos, 1993). AFLPs are useful when there is no sequence information available (Mueller and Wolfenbarger, 1999) and they have been widely used to saturate genetic maps. They are highly sensitive and reproducible markers that have been widely used for the identification of genetic variation between varieties or closely related species of plants, fungi and other organisms. AFLP markers have also been used in linkage studies to generate maps for quantitative trait locus (QTL) analysis.

1.2.1.5 Diversity arrays technology (DArT)

Diversity Arrays Technology (DArT) was developed to provide a practical and economical whole-genome profiling capability: typing thousands of markers in a single assay at a fraction of time and cost by comparison with main SNP typing platforms (Kilian et al., 2012). A genomic representation of a species is arrayed for analysis by hybridization in DArT. DArT can provide very large number of markers and has been widely applied especially in the construction of genetic maps (Xia et al., 2005) although this method requires significant development effort for each species. A high-density DArT genome profiling resource has been developed and demonstrated its potential for genome-wide diversity analysis and linkage mapping in several species of *Eucalyptus* (Sansaloni et al., 2010). The first reported genetic linkage maps in any grain legume has been developed for pigeon pea (Yang et al., 2011). One limitation is that the anonymous nature of these markers limits association with reference genomes.

DArTseq is a GBS platform developed by DArT PL in Canberra, Australia. This platform adopted a combination of complexity reduction methods developed initially for array-based DArT and sequencing of resulting representations on next-generation sequencing platforms. DArTseq is essentially the same method as Cornell-style reduced representation sequence-based genotyping by sequencing (GBS) (Li et al., 2015). It supersedes hybridization-based DArT recently because the cost and throughput of sequencers reached a point where any GBS-type approach can compete effectively with hybridisation-based DArT (Li et al., 2015).

1.2.1.6 Simple sequence repeats (SSRs)

Simple sequence repeats (SSRs) (Weber and May, 1989) exist widely in the genome and demonstrate variation in populations. SSRs are highly polymorphic and informative markers for comparative genetic and genomic analysis as PCR primers designed to an SSR from one species frequently amplify a corresponding locus in related species. The mining of SSRs from gene and genome sequence data is now routine (Edwards and Batley, 2010), with large numbers of SSRs identified in a range of plant species including Brassicas (Hong et al., 2007, Burgess et al., 2006), wheat (Nie et al., 2012) and

strawberry (Keniry et al., 2006). SSR loci are also hot spots for SNPs and SSRs may readily be converted to SNP markers (Batley et al., 2003). The limitation of this method is that SSR loci are traditionally expensive to develop and assay (Robinson et al., 2004).

1.2.1.7 Single nucleotide polymorphisms (SNPs)

The highest resolution of genetic variation is the single nucleotide polymorphism. A single nucleotide polymorphism (SNP) is an individual nucleotide base difference between two DNA sequences (Edwards et al., 2007). SNPs are direct markers, providing the exact nature of the allelic variants. SNPs are single base differences between individuals, and represent the most abundant type of sequence variation in plant genomes (Batley and Edwards, 2007). Insertions and deletions (indels) are also often assayed as SNPs. SNPs are evolutionarily conserved markers and have been used as markers for quantitative trait loci (QTL) analysis and in association studies. There are several approaches to identify and genotype SNPs in plants (Chagné et al., 2007), and their diverse applications and high throughput suggest that they will continue to be the dominant DNA molecular marker in the foreseeable future (Batley and Edwards, 2007). Applications of SNPs include their use for the high-resolution genetic mapping of traits, association studies, and genome-wide linkage disequilibrium analysis (Duran et al., 2010b, Rafalski, 2002, Mackay and Powell, 2007, Jannink et al., 2010). Furthermore, SNPs are an invaluable tool for genome mapping, generating very high-density genetic maps and haplotypes around genes or regions of interest (Duran et al., 2009d, Duran et al., 2009b).

Modern methods apply computational algorithms for SNP discovery from abundant DNA sequence data and present the results within searchable databases (Duran et al., 2009a). The main challenge of computational SNP discovery remains the differentiation between true biological variation and the often more abundant errors within the sequence data. This is particularly true for data generated from next generation sequencing platforms, which often sacrifice data quality for quantity.

The application of new sequencing methods is leading to the discovery of large numbers of SNPs in wheat (Allen et al., 2011, Winfield et al., 2012), rice (Kharabian-Masouleh et al.,

2011, Subbaiyan et al., 2012), Brassicas (Trick et al., 2009) and other important crop species (Barker and Edwards, 2009, Bundock et al., 2009).

1.2.1.8 Discovery of single-nucleotide polymorphisms (SNPs)

As a nucleotide base is the smallest unit of inheritance, SNPs are the ultimate form of molecular genetic marker providing the exact nature of the allelic variants. SNPs can be categorised according to nucleotide substitution as either transitions (C/T or G/A) or transversions (C/G, A/T, C/A, or T/G). SNPs provide an important source of useful molecular markers for genetic mapping, map-based positional cloning, detection of marker-trait gene associations through linkage and linkage disequilibrium (LD) mapping and the assessment of genetic relationships between individuals (Edwards et al., 2007). SNPs have a low mutation rate and are useful markers for studying complex genetic traits and understanding genome evolution (Syvanen, 2001).

SNPs are frequently identified in expressed gene sequences. An expressed sequence tag (EST) is a relatively short sub-sequence of a cDNA sequence, resulting from one-shot sequencing of a cloned mRNA. ESTs are used for gene discovery and this sequence data provides a rich source of biologically useful SNPs (Batley et al., 2003) due to high redundancy of sequence reads, high diversity of represented genotypes and that each identified SNP is associated with a functional gene (Picoult-Newberg et al., 1999). Roche 454-based transcriptome sequencing has been demonstrated as an excellent method for the high-throughput acquisition of gene-associated SNPs in maize (Barbazuk et al., 2007).

As with the discovery of many sequence based molecular markers, high initial cost was one of the major limitations of SNP identification (Batley et al., 2003). Traditionally, the identification of SNPs was a laboratory based procedure involving the polymerase chain reaction (PCR)-based amplification of specific genome fragments in individuals of interest and dedicated sequencing of these amplicons. This approach was both costly and laborious. The use of *in silico* methods for SNP discovery supports a much cheaper method of finding SNPs in terms of time and expense (Duran et al., 2009c).

Several approaches have been developed for detecting SNPs to decrease the initial cost and time required. However, *in silico* methods also have limitations, in that true SNP polymorphisms can be difficult to distinguish from abundant sequence errors (0.5-1.0% error per raw base is common) (Pont et al., 2013, Margulies et al., 2005, Bentley et al., 2008). These errors are mostly from the automated base calling of raw data, because of the balance between obtaining the best sequence length and the confidence of accuracy of called bases (Barker et al., 2003). The software Phred is frequently applied to call bases from Sanger sequence chromatogram data (Ewing et al., 1998) and can provide a measure of accuracy of calling of each base by statistical estimation. Therefore true genetic variation can be represented by the sequence difference based on this primary level of confidence. Other methods (Kwok et al., 1994, Garg et al., 1999) also rely on sequence trace file analysis to filter out sequence errors based on trace quality. For example, POLYBAYES (Marth et al., 1999) is one of the methods to differentiate between true SNPs and sequence error, when sequence quality information and trace files are available. The main drawbacks of these methods is that they do not take account of the sequence error from the reverse transcription process required for the generation of cDNA libraries for EST sequencing (Barker et al., 2003).

To overcome these restrictions, the autoSNP software was designed and developed to detect candidate SNPs and small insertions/deletions automatically within EST data, with associated measurements of confidence in the validity of SNP candidates (Barker et al., 2003, Batley et al., 2003). The PERL scripts used a redundancy-based approach to distinguish valid SNPs from erroneous sequence by consensus sequence calling in an alignment of sequence reads. A second measure of validity is the co-segregation score, which is a measure of whether a predicted SNP contributes to define a haplotype. This score is weighted to account for missing data in the assembly (Barker et al., 2003). Subsequently the autoSNPdb system was developed to combine the autoSNP software and sequence annotation within a relational database scheme. This application provides a flexible interface facilitating a variety of queries to identify SNP and indel polymorphisms related to specific genes or traits. AutoSNPdb has been applied in barley, rice and

Brassica species (Duran et al., 2009a). AutoSNPdb is available online (<http://autosnpdb.appliedbioinformatics.com.au>).

1.3 The transcriptome

The transcriptome represents the sequence of RNA molecules transcribed from a genome. The characterisation of the transcriptome is one of the goals of functional genomics (Pierce, 2011). In the process of protein synthesis, genomic DNA is transcribed to produce messenger RNA (mRNA). Proteins are then synthesised by translating this mRNA. The set of all RNA molecules in a cell or tissue is referred to as the 'transcriptome'. Research into the wheat transcriptome has identified expressed genes that control important traits (Coram et al., 2008a). The transcriptome is dynamic and changes rapidly in response to cellular perturbations or during normal developmental events (Lockhart and Winzeler, 2000), and transcriptome sequencing has been used for gene expression profiling, genome annotation, and non-coding RNA discovery (Morozova and Marra, 2008). Gene expression levels can also be regarded as quantitative traits segregating in a population, with transcript levels varying among genetically diverse individuals, and linkage mapping can be used to identify hundreds of expression quantitative trait loci (eQTLs) (Druka et al., 2010).

1.3.1 Traditional transcriptome analysis

Recent studies of high-throughput transcriptome analysis relied on DNA microarray technologies. DNA microarrays can be used to measure gene expression levels, specifically messenger RNA abundance, for tens of thousands of genes simultaneously. High-density arrays of oligonucleotides or cDNAs are attached to a solid surface and hybridised with labelled RNA or cDNA. Hybridisation intensity is then measured and used to calculate gene expression levels. In contrast to microarray methods, sequence-based methods such as the Sanger sequencing of cDNAs, directly determine the mRNA

sequence (Boguski et al., 1994, Gerhard et al., 2004). However, this method is limited by relatively low throughput and high cost, and is generally not considered to provide a basis for quantitative analysis. Tag-based methods such as serial analysis of gene expression (SAGE) (Velculescu et al., 1995) and massively parallel signature sequencing (MPSS) (Brenner et al., 2000) were developed to overcome these limitations to provide more precise ‘digital’ gene expression levels (Wang et al., 2009b). An advantage of SAGE and MPSS over microarrays is that they are sequence based and can identify novel transcripts without requiring *a priori* knowledge of gene sequences (Hu and Polyak, 2006). MPSS technology subsequently evolved into the current Illumina NGS technology through the purchase of MPSS developers Lynx Therapeutics by Solexa in 2005 and subsequent purchase of Solexa by Illumina, and Illumina technology is now applied as a replacement of these previous methods.

1.3.2 Second generation transcriptomics

NGS technologies are now adopted routinely for transcript profiling. These technologies produce sequence tags representing expressed genes without prior knowledge of gene sequence. Second generation transcriptome sequencing can be applied for analysis of gene expression, the structure of genomic loci, and sequence variation present at expressed gene loci (Morozova and Marra, 2008). This can be achieved either by *de novo* assembly of the transcriptome sequence data or by aligning reads to a genome or transcriptome sequence, where these are known.

RNA-Seq is an approach to transcriptome profiling that uses deep-sequencing technologies. In general, a population of RNA (total or fractionated, such as poly(A)) is converted to a library of cDNA fragments with adaptors attached to one or both ends. Each molecule is then sequenced in a high-throughput manner to obtain short sequences from one end (single-end sequencing) or both ends (pair-end sequencing). The reads are generally 30-400 bp, depending on the DNA-sequencing technology used. High-throughput sequencing technologies, such as Illumina, Applied Biosystems SOLiD and Roche 454 Life Science systems have been used for RNA-Seq (Wang et al., 2009b).

Roche 454 technology was the first of the next generation sequencing technologies to become available. The relatively long reads produced by this technology assisted sequence annotation. Illumina technology is becoming increasingly popular for transcriptome research due to the power of vast read depth and read pair technology. Deep sequence coverage, which Illumina's NGS platform provides more readily than 454 sequencing, is important for gene discovery and gene expression analysis (Varshney et al., 2009, Barski et al., 2007, Johnson et al., 2007). In contrast to microarray technology, second generation technologies have the potential to identify all RNA transcripts produced at a specific time, as well as transcript variants caused by differential splicing of genes (Varshney et al., 2009).

1.3.3 Tools for the analysis of second generation transcriptomic data

TopHat (Langmead et al., 2009) is an open-source software developed to align reads from RNA-Seq to a reference genome without relying on known splice sites. TopHat takes a reference genome (as a Bowtie index) and RNA-Seq reads as FASTA or FASTQ and produces alignments in SAM format. With default parameter values TopHat detects junctions even in genes transcribed at very low levels. TopHat version 1.0.7 and later has been extended to browse long paired reads and align reads across splice junctions. It splits a read 75 bp or longer into three or more segments of approximately equal size (25 bp) and maps them independently. It uses Bowtie (<http://bowtie-bio.sourceforge.net>), a short-read mapping program (Langmead et al., 2009), to map non-junction reads (those contained within exons) against the reference genome. All reads that do not map to the genome are set aside as 'initially unmapped reads' (IUM). Tophat searches the IUM reads in order to find reads that span junctions for each splice junction. The MAQ assembly module (Li et al., 2008) is used to build a consensus of the mapped regions. TopHat is implemented in C++ and runs on Linux and Mac OS X. This program uses SeqAn library (Doering et al., 2008) and requires Bowtie and MAQ.

The Cufflinks assembler (Trapnell et al., 2010) is an open-source C++ program and runs on Linux and Mac OS X. It can identify complete novel transcripts and probabilistically assign reads to isoforms. Additionally it contains Cuffcompare and Cuffdiff tools. Cuffcompare validates Cufflinks output, transfrags (assembled transcript fragments), against annotated transcriptomes and also finds transfrags common to multiple assemblies. Cuffdiff then performs differential expression testing. This assembler was designed to investigate transcriptional, splicing regulation and find the minimal number of transcripts that 'explain' the reads (that is, every read should be contained in some transcript). Cufflinks takes as input cDNA fragment sequences that have been aligned to the genome e.g. by TopHat, which can align reads across splice junctions without relying on gene annotation in order to produce spliced alignments.

EST_GENOME (Mott, 1997) is a software tool to facilitate the prediction of genes by sequence homology and to align spliced DNA sequences to unspliced genomic DNA. Standard alignment tools are not ideal for detecting the correct alignment of a spliced product to genomic DNA because large introns can occur in the genomic sequence and software often ignores the conserved sequences found at donor/acceptor splice sites (intron/exon boundaries). The program EST_GENOME overcomes the above limitations. It detects large introns, can recognise splice sites and uses limited memory. The algorithm of this program adopts a modification of the Smith-Waterman algorithm (Smith and Waterman, 1981).

1.3.4 Differential homoeologous gene expression

Polyplodisation frequently occurs during the evolution of flowering plants. This event supplies the raw material for divergence of function in homoeologous genes, leading to phenotypic novelty that can promote the success of polyploids in nature or their selection for use in agriculture (Vidal et al., 2010). Allopolyploidy is believed to be a major factor contributing to speciation, diversification and ecological adaptation in flowering plants (Adams, 2007). The expression of duplicate genes (homoeologues) can be modified leading to functional plasticity and to phenotypic novelty. The absence of signs of

subfunctionalization could indicate whether or not the species have undergone noticeable diploidization (Combes et al., 2012).

Following allopolyploidy, genomic changes including chromosomal rearrangement and changes in gene expression can occur and continue over evolutionary time. Considerable changes in gene expression have been examined in allopolyploids (Adams, 2007). Genes in a variety of functional categories demonstrate altered expression, and the patterns vary considerably by gene. Some changes could be stochastic, whereas others are repeatable. Gene expression changes can be organ specific. Reciprocal silencing of duplicates in different organs has been studied (Adams, 2007). This suggests subfunctionalization and long-term retention of duplicates. It indicates that hybridization has a much greater effect than chromosome doubling on gene expression in allopolyploids. Parent-of-origin effects on gene expression and loss of gene imprinting have been reviewed. Some gene expression changes in polyploids and hybrids can be correlated with phenotypic effects (Adams, 2007). The mechanisms of gene expression changes, including DNA methylation, histone modifications, and antisense RNA, have been demonstrated (Adams, 2007).

The wheat group has evolved through allopolyploidisation among species from the plant genera *Aegilops* and *Triticum* followed by genome doubling (Feldman and Levy, 2012). Allopolyploidy speeds up genome evolution in wheat in two ways, including revolutionary changes and evolutionary changes: (1) allopolyploidisation triggers rapid genome alterations (revolutionary changes) through the instantaneous generation of a variety of cardinal genetic and epigenetic changes; (2) the allopolyploid condition assists sporadic genomic changes during the life of the species (evolutionary changes) that are not attainable at the diploid level (Feldman and Levy, 2009). The revolutionary alterations occurred during the formation of the allopolyploid and lead to rapid cytological and genetic diploidization and facilitated the successful establishment of the newly formed allopolyploid in nature. Meanwhile, evolutionary changes occurred during the life of the allopolyploids and increased the intra-specific genetic diversity, and consequently, increased fitness, adaptability and competitiveness. These phenomena emphasised the dynamic plasticity of both structure and function for the allopolyploid wheat genome

(Feldman and Levy, 2009). The associations between homoeologous gene expression changes and polyhaploidization has been addressed in bread wheat (Wang et al., 2011a).

1.4 Integrated databases for crop plants

In recent years, significant changes have occurred in the genome bioinformatics community: from studies of a single genome or chromosome to many thousands of genomes; using different sequencing platforms with various data types and error models. The growth in genome information has led to a challenge for bioinformatics researchers to transform the vast quantities of data being produced into collective knowledge. As sequence availability has increased, data access, representation, analysis and visualisation present significant challenges (Ning and Montgomery, 2010, Duran et al., 2009a). In this context, online databases for genome and genomic data are very much in demand.

Molecular markers are used in plant breeding and genetic research, including mapping of genes and quantitative trait loci (QTL), phylogenetic studies, comparative genomics, and marker assisted breeding (Prasad et al., 2000, Stein and Graner, 2005, Varshney et al., 2005). While several diverse DNA based marker types have been developed, single nucleotide polymorphisms (SNPs) now dominate and are widely used in plant breeding, genomic research and modern genetic analysis (Edwards and Batley, 2008, Duran et al., 2009a).

Many molecular marker databases host SNP markers (Batley and Edwards, 2009). Some databases also include other types of marker that are not commonly used. These markers include simple sequence repeats (SSRs), restriction fragment length polymorphism (RFLP), amplified fragment length polymorphism (AFLP), random amplification of polymorphic DNA (RAPD), short tandem repeat (STR) and diversity arrays technology (DArT). In this section, I present some of the features of plant molecular genetic marker databases, highlight the various types of marker resources and predict the potential future direction of crop marker databases.

1.4.1 Molecular databases

With the ever increasing amount of genetic and genomic information, there is a requirement to manage the data to make it available and accessible to researchers (Duran et al., 2009a, Duran et al., 2009b). This includes the development of custom visualisation tools (Lim et al., 2007, Duran et al., 2010a, Duran et al., 2009b) and bioinformatics systems to traverse the genome to phenotype divide (Duran et al., 2010b, Edwards and Batley, 2004). Many molecular marker databases provide various types of markers for a range of species while some databases provide information on a single type of marker (Lai et al., 2012c). The largest single marker database is GeneBank dbSNP (<http://www.ncbi.nlm.nih.gov/projects/SNP/>) (Wheeler et al., 2008, Benson et al., 2009, Karsch-Mizrachi et al., 2012), which provides SNP data mostly for human and other vertebrates, although it also includes some plant data.

There are several databases for the grasses. The Gramene database (<http://gramene.org>) hosts many types of markers based on the genomes of rice, and maize (Monaco et al., 2014). This website provides a search engine, and users can search for specific markers. Marker details are displayed in text format, including database cross-references and map positions linked to chromosomes in CMap (Youens-Clark et al., 2009). The sources of SSR markers include the International Rice Genome Sequencing Project, IRMI (International Rice Microsatellite Initiative), MaizeGDB, the Cornell SSR library, and the Indian Agricultural Research Institute.

GrainGenes (<http://wheat.pw.usda.gov>) hosts multiple types of markers for Triticeae and Avena (Matthews et al., 2003, Carollo et al., 2005, O'Sullivan, 2007). The website also provides comparative map views for wheat, barley, rye and oats using CMap. Marker types include SSR, RFLP and SNP. Most of the SNP markers are from two publications (Close et al., 2009, Szűcs et al., 2009). An improved SNP-based consensus genetic map has been developed from 1,133 individuals from 10 mapping populations. This database provides a search panel with query name or a list of marker names as input.

MaizeGDB (<http://www.maizegdb.org>) (Lawrence, 2007, Lawrence et al., 2007, Schaeffer et al., 2011) provides a search engine to identify ESTs, AFLPs, RAPD probes and sequence data for Maize. The legume information system (LIS) (<http://legumeinfo.org/>) (Gonzales et al., 2005, Gonzales et al., 2007) provides access to markers such as SNP, SSR, RFLP and RAPDs for diverse legumes, including peanut, soybean, alfalfa and common bean.

The Panzea (<http://panzea.org>) database (Zhao et al., 2006, Canaran et al., 2006, Canaran et al., 2008) describes the genetic architecture of complex traits in maize and Teosinte. This database also provides a marker search interface. Two common types of marker, SNP and SSR, can be searched for. The search results display a list of markers with position details related to different chromosomes. When the marker is selected, the web site can display this marker in pre-computed multiple sequence alignments using the Look-Align viewer (Canaran et al., 2006).

TriMEDB (Triticeae mapped EST database) (Mochida et al., 2008) provides information on mapped cDNA markers that are related between barley and wheat. The current version of TriMEDB provides map-location data for barley and wheat. These data were retrieved from three published barley linkage maps: the barley SNP database of SCRI (http://bioinf.scri.ac.uk/barley_snpdb/), the barley transcript map of IPK (http://pgrc.ipk-gatersleben.de/transcript_map/), HarvEST barley versions 1.63 and 1.68 (<http://harvest.ucr.edu/>) and one diploid wheat (Hori et al., 2007). Users can search the database from the search markers page using marker and chromosome names. The search results include the name of any retrieved marker, related linkage maps, chromosome number, map positions, primer pairs for PCR, EST contigs for each sequence resource, a link to the cDNA assembly and comparative maps for the rice genome. The database can be accessed at <http://trimedb.psc.riken.jp>.

The database of the Rice Genome Annotation Project (Ouyang et al., 2007) hosts SSR markers, ESTs and SNPs in the rice genome pseudomolecules (<http://rice.plantbiology.msu.edu/>). The rice genome annotation project pseudomolecules (Release 7) and the unified Os-Nipponbare-Reference-IRGSP-1.0 pseudomolecules (Kawahara et al., 2013) were used for marker annotation. This database provides a web

interface and displays predicted SSR markers, ESTs and SNPs filtered by type and/or chromosome, as well as a Gbrowse view to display the related sequences.

With the exception of some important species, databases for non-grass species tend to be more limited in scope. There are a large number of Brassica molecular markers developed together with bioinformatics resources (Lorenc et al., 2012a, Duran et al., 2011). The central Brassica portal for all things Brassica (<http://www.brassica.info>) provides a link to access a range of Brassica molecular markers, including SNP/InDel, SSR, RFLP, AFLP, and RAPD. This website provides a summary of available information for Brassica SSRs, and provides a means to exchange and distribute these markers at the Brassica microsatellite information exchange (Choi et al., 2007).

The Sol Genomics Network database (SGN; <http://solgenomics.net>) is a clade-oriented database (COD) hosting biological data for species in the Solanaceae and their close relatives. The data types range from chromosomes and genes to phenotypes and accessions. SGN hosts more than 20 genetic and physical maps for tomato, potato, pepper and tobacco with thousands of markers. Genetic marker types in the database include SNP, SSR, AFLP, PCR, and RFLP (Bombarely et al., 2011).

SoyBase (<http://soybase.org/>) (Grant et al., 2010) hosts genome and genetic data for soybean. The markers include SNP, SSR, RFLP, RAPD and AFLP. The markers can be viewed in CMap and have also been linked to corresponding location in a Gbrowse2 genome viewer. Each marker comes with the genomic sequence, detection method, and information source.

VegMarks (<http://vegmarks.nivot.affrc.go.jp>) is a database for vegetable genetic markers developed by the National Institute of Vegetable and Tea Science (NIVTS) in Japan (Fukuoka et al., 2012). This database provides various marker characteristics, including ID number, map position, nucleotide sequence of the clones/PCR primers and polymorphism data among varieties/accessions for Chinese cabbage, bunching onion, cucumber, eggplant, melon and tomato. The markers hosted in this database include SNP, SSR and RFLPs. Some marker data is restricted for registered users only. This database provides a single map for each chromosome together with marker position information.

MoccaDB (<http://moccadb.mpl.ird.fr>) is an integrative database for functional, comparative and diversity studies in the Rubiaceae family which includes coffee (Plechakova et al., 2009). It provides an easy access to markers, such as SSR, SNP and RFLP and related information data such as PCR assay conditions, cross amplification within related species, locus position on different linkage maps and diversity parameters. It also provides a search engine for searching related markers by keywords and downloads of related data in Microsoft Office Excel format.

The Cotton Microsatellite Database (CMD) (<http://www.cottonmarker.org>) is a curated and integrated web-based relational database providing centralised access to publicly available cotton SSRs. CMD contains publication, sequence, primer, mapping and homology data for nine major cotton SSR projects, collectively representing 5,484 SSR markers (Blenda et al., 2006).

In addition to species specific databases, other databases focus on specific marker types. The autoSNPdb database (Duran et al., 2009a) is based on an early pipeline for SNP discovery from EST sequence data (Barker et al., 2003, Batley et al., 2003). It provides an interface facilitating a variety of queries to search for SNPs within known genes from a range of species including Brassica, rice, barley (Duran et al., 2009c) and wheat (Lai et al., 2012b). The SNP search method was developed based on polymorphisms related to specific genes identified through keyword, sequence similarity or comparative genomics approaches and the results provide sequence annotation and SNP information in tabular and graphical format.

There are an increasing number of bioinformatics resources available for wheat (Duran et al., 2011). WheatGenome.info is an integrated database resource which supplies a variety of web-based systems hosting wheat genetic and genomic data. Wheatgenome.info (Lai et al., 2012a) provides a GBrowse2 based wheat genome viewer, CMap and CMap3D comparative genetic map viewers (Duran et al., 2010a, Youens-Clark et al., 2009). From the GBrowse2-based wheat genome viewer, wheat reference genomic sequences are currently only available for wheat group 7 chromosomes (Berkman et al., 2011a, Berkman et al., 2012b, Berkman et al., 2013). SGSAutoSNP (Second Generation Sequencing autoSNP) software was used to identify more than 4 000 000 intervarietal

SNPs between 16 Australian wheat varieties along this chromosome group (Lai et al., 2014).

SSR Primer 2 (<http://www.appliedbioinformatics.com.au/projects/ssrPrimer>) (Jewell et al., 2006) provides the real time discovery of SSRs within submitted DNA sequences, with the concomitant design of PCR primers for SSR amplification (Robinson et al., 2004). The success of this system has been demonstrated in Brassica (Batley et al., 2007, Hopkins et al., 2007, Ling et al., 2007) and strawberry (Keniry et al., 2006).

Molecular marker databases are expanding rapidly as increasing numbers of markers are developed from the latest high throughput DNA sequencing technologies. There is an increasing challenge to manage and maintain this expanding data as well as integrate marker data with the growth of available genome sequences. Finally, the greatest challenge will be to fully integrate genetic diversity information with heritable trait information, bridging the genome to phenotype divide and providing the tools for more advanced breeding and crop improvement.

A list of molecular marker databases with web links and references is presented in Table 1-3.

Table 1-3 Examples of molecular marker databases with different types of markers for crop improvement.

Database Name	Viewer	SNPs	SSRs	RFLPs	RAPDs	AFLPs	ESTs	BACs	DArTs	DNA Probes	PCR Primers	Web Link	References
autoSNPdb	*	+										http://autosnpdb.appliedbioinformatics.com.au/	(Duran et al., 2009a, Duran et al., 2009c, Lai et al., 2012b)
Brassica.info		+	+	+	+	+						http://www.brassica.info/resource/markers.php	(Choi et al., 2007)
<i>Brassica rapa</i> genome database	*		+									http://brassicadb.org/brad/geneticMarker.php	(Cheng et al., 2011)
Cotton Marker Database (CMD)	*		+				+				+	http://www.cottonmarker.org/cgi-bin/cmd_search_marker_result.cgi	(Blenda et al., 2006)
GenBank dbSNP	*	+										http://www.ncbi.nlm.nih.gov/projects/SNP/	(Wheeler et al., 2008, Benson et al., 2009, Karsch-Mizrachi et al., 2012)
GrainGenes	*	+	+	+		+			+	+	+	http://wheat.pw.usda.gov	(Matthews et al., 2003, Carollo et al., 2005, O'Sullivan, 2007)
Gramene	*	+	+	+	+	+	+		+	+	+	http://www.gramene.org/db/markers/marker_view	(Monaco et al., 2014)
ICRISAT chickpea root EST database		+										http://www.icrisat.org	(Azam et al., 2012)
Legume Information System (LIS)	*	+	+	+	+	+						http://legumeinfo.org/	(Gonzales et al., 2005, Gonzales et al., 2007)

MaizeGDB			+		+	+	+			+		http://www.maizegdb.org/probe.php	(Lawrence, 2007, Lawrence et al., 2007, Schaeffer et al., 2011)
MoccaDB		+	+	+								http://moccadb.mpl.ird.fr/index.php?cat=1	(Plechakova et al., 2009)
Panzea	*	+	+									http://www.panzea.org/db/searches/webform.marker_search	(Zhao et al., 2006, Canaran et al., 2006, Canaran et al., 2008)
Rice Genome Annotation Project	*	+	+				+					http://rice.plantbiology.msu.edu	(Ouyang et al., 2007)
SSR Primer 2			+									http://www.appliedbioinformatics.com.au/projects/ssrPrimer	(Jewell et al., 2006)
SOL Genomics Network (SGN)	*	+	+	+		+				+		http://solgenomics.net	(Bombarely et al., 2011)
SoyBase	*	+	+	+	+	+				+		http://soybase.org	(Grant et al., 2010)
TreeGenes	*		+									https://dendrome.ucdavis.edu/treegenes	(Wegrzyn et al., 2012)
Triticeae Mapped EST DataBase ver.2.0 (TriMEDB)	*	+					+					http://trimedb.psc.riken.jp	(Mochida et al., 2008)
VegMarks	*	+	+	+								http://vegmarks.nivot.affrc.go.jp/VegMarks/jsp/page.do?transition=marker	(Fukuoka et al., 2012)
Wheat genome information		+	+									http://www.wheatgenome.info	(Lai et al., 2012a, Edwards et al., 2012b)

(* indicates that this database provides viewer. + indicates that this database supplies this type of marker)

1.5 *Triticum aestivum*

1.5.1 Importance of *Triticum aestivum*

Wheat is a major food crop, ranked within the top four agricultural commodities globally by production and value according to the Food and Agricultural Organization of the United Nations (FAO; <http://www.fao.org/home/en/>), and used widely for making products including breads, pastries, noodles and dumplings. Substantial cultivation of tetraploid wheat occurs, primarily ‘durum’ wheat (*Triticum durum*). However, greater than 90% of the world’s cultivated wheat is the hexaploid species *Triticum aestivum* (Shewry, 2009), known as ‘common’ or ‘bread’ wheat.

Bread wheat was domesticated at the dawn of agriculture about 10,000 years ago and has since been adapted to grow in a broad range of climates throughout the world to become one of the major crops (Dubcovsky and Dvorak, 2007, Krasileva et al., 2013). Bread wheat (*Triticum aestivum* L.) is one of the most important crop plants and most widely grown cereal grains worldwide, occupying 17% (one sixth) of crop acreage of the world (Gupta et al., 2008). It is a staple food, feeding about 35% of the world’s population, providing more calories and protein in the global diet than any other crop (www.idrc.ca/en/ev-31631-201-1-DO_TOPIC.html). Annual global wheat consumption has exceeded 600 Mt (<http://www.fao.org>), of approximately 620 million tons annually, provides about one-fifth of the calories consumed by humans (FAO, 2006, Pfeifer et al., 2014), which means about 100 kg per capita, and the requirement for wheat production is predicted to grow by over 40% by 2020 (Bhalla, 2006). Global wheat production has recently declined by 5.5% due to climatic trends (Lobell et al., 2011). A doubling in global food demand projected for the next 50 years poses challenges for the sustainability of food production (Tilman et al., 2002), including wheat production.

Wheat is one of the most important commodities produced by the Australian agriculture industry. In 2003-2004, almost 30,000 farmers in Australia grew wheat using half of the agricultural land dedicated to cropping. In addition, the gross value of production of wheat

in 2003-04 in Australia was AUD\$5.6 billion, which represented 15% of the total value of farm production (ABS, 2006). The gross value of production of wheat in 2008-2009 and 2009-2010 in Australia was AUD\$6.0 billion and AUD\$4.8 billion respectively (ABS, 2012). Wheat was Australia's largest production volume (21.8 million tones with 13.9 million hectares) of grain crop in 2009-2010, being approximately three times Australia's second largest production volume of grain crop, barley, in terms of land area usage and production volume (ABS, 2012).

Northern Asian countries including China have had historically high wheat consumption, whereas rice has been staple in the southern parts of Asia (and southern China) because of a more suitable climate for rice production. Wheat has been consumed in China for about 3000 years and spread to other Asian countries about 1000 years ago. The consumption of wheat based products has increased in recent years (Miskelly, 2005). In addition, the demand for better quality food products is also increasing (DFAT, 2004). However, climate change has impacted wheat production in China and severe drought prevailed in Northern and Western China, the major winter wheat producing areas in the 2008/2009 winter wheat season (FAO, 2009).

Wheat plant breeding is limited by the lack of knowledge of the genetic architecture of most complex traits, such as yield, and resistance to abiotic stress. Marker-assisted selection (MAS) is being practised for improvement of traits in wheat in several countries including Australia, USA and Canada. The traits being analysed include disease/pest resistance, and quality traits such as grain protein and bread making quality. Further traits including tolerance to drought, heat, salinity, waterlogging, and metal toxicity are being targeted for wheat breeding programs in the future (Gupta et al., 2010).

The total production and consumption of wheat have increased steadily from 1960 to 2011 (Figure 1-1) (USDA, 2011), however, the total size of area harvested didn't increase over this period. In addition, the production and consumption of wheat have continued to increase from 2009/2010 to 2014/2015 (Figure 1-2) (USDA, 2014). Wheat yield improvement is a high priority. The market share of world wheat exports from Australia occupied around 9.43 per cent as the 5th largest exporter globally from 2005 to 2011 (Figure 1-3) (USDA, 2011). In addition, the expansion of the global population is expected

to reach 9 billion by 2050 (Godfray et al., 2010). Therefore, identifying genetic variation, markers and genes to assist breeders design high-yielding wheat is crucial. The research of this PhD thesis is focusing on identifying genetic variation, such as single nucleotide polymorphisms (SNPs), homoeologous gene expression, and SNP density analysis in wheat based on NGS technologies and bioinformatics. In addition, related databases have been developed to make the results accessible.

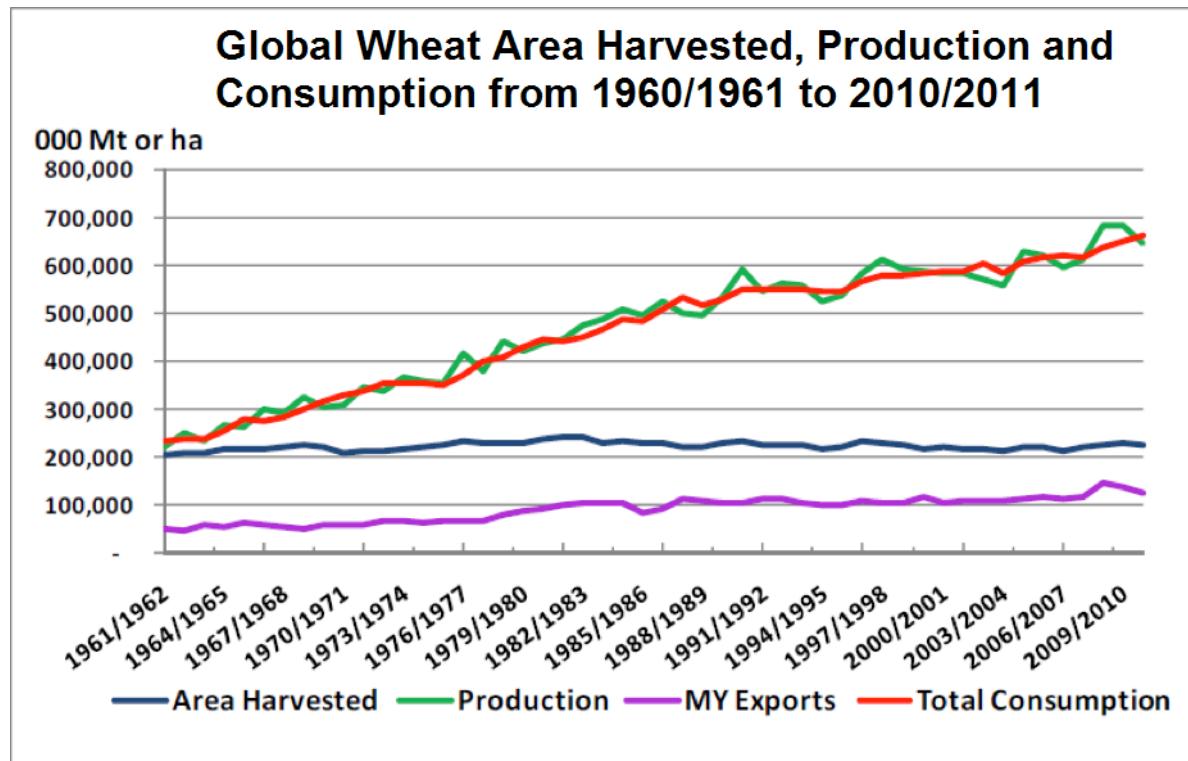


Figure 1-1 Global wheat, area, production and consumption from 1960/1961 to 2010/2011 (Adapted from USDA, PSD Online, January 2011).

Global Wheat Production and Consumption from 2009/2010 to 2014/2015

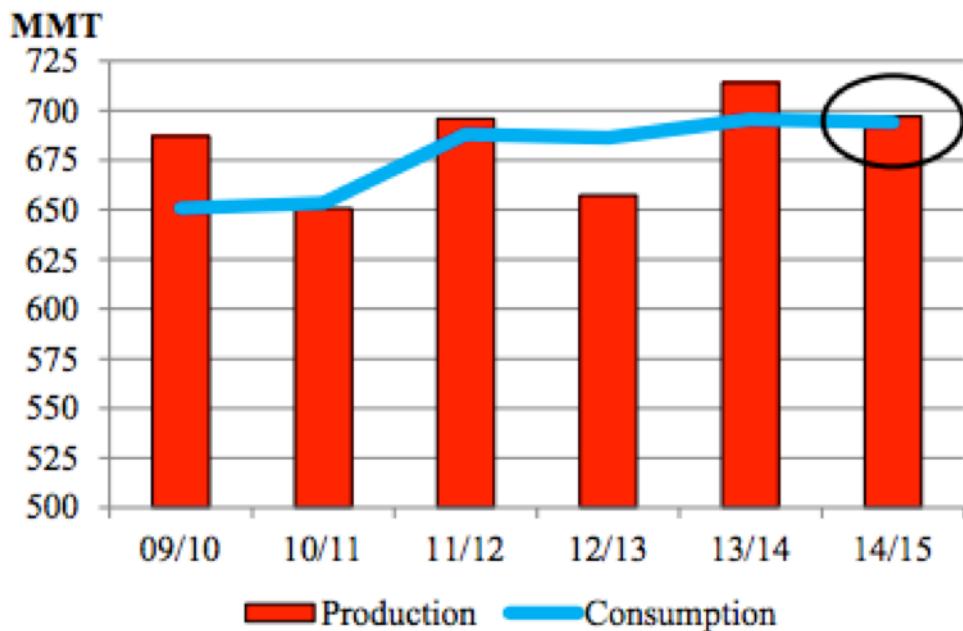


Figure 1-2 Global wheat production and consumption from 2009/2010 to 2014/2015
(Adapted from USDA, 2014).

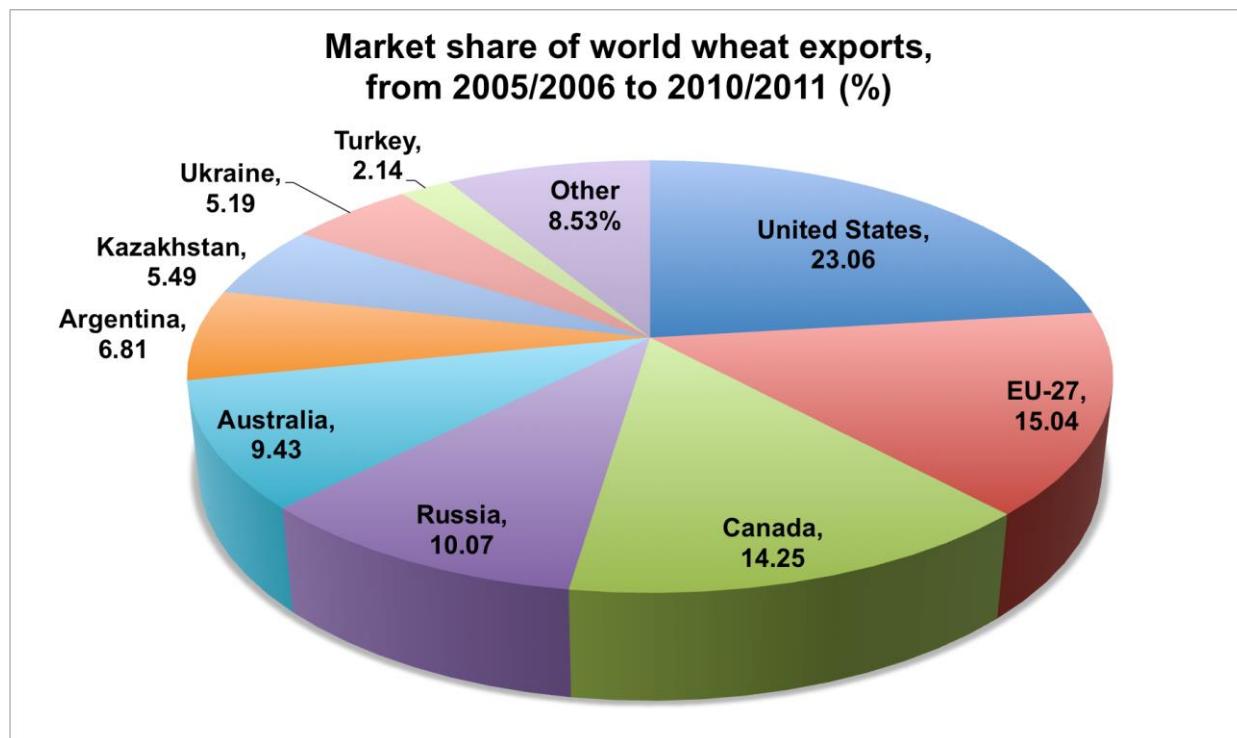


Figure 1-3 Market share of world wheat exports, from 2005/2006 to 2010/2011 (%)
(Adapted from USDA, PSD Online, January 2011).

1.5.2 Genome evolution and domestication in *Triticum aestivum*

The most important wheat species grown today are bread wheat (*Triticum aestivum*) and durum wheat (*Triticum turgidum*) (Feuillet et al., 2008). Bread wheat and durum wheat are different from one another in genomic make-up, in grain composition and in food end-use quality attributes. Bread wheat is an allohexaploid (AABBDD), and is known as hard wheat or soft wheat, depending on grain hardness (Peña, 2002). Durum wheat is a cultivated tetraploid wheat with the AABB genome composition, and its endosperm has the hardest texture of all wheats (Liu et al., 1996). Durum wheat is used for the production of pasta, couscous, and various types of bread in some areas (Fabriani et al., 1988).

Genomic duplications have been proposed as the main force for genome evolution and domestication of plant species (Ohno, 1970, Chantret et al., 2005). Domestication involving few individuals leads to reduced genetic diversity of cultivated plant species compared with their wild progenitors (Tanksley and McCouch, 1997). In addition, domestication leads to the selection and spreading of specific genes and alleles that control traits of agronomic importance and distinguish crops from their wild relatives (Clark et al., 2004).

The hexaploid genome of bread wheat formed through two allopolyploidisation events. In a divergence event between 2.5 and 6 million years ago ancestral wheat split into at least 3 different diploid species. In a hybridization event that happened between 0.5 and 3 MYA, the diploid genomes of *Triticum urartu* (AuAu) and an unidentified species (BB), similar to *Aegilops speltoides*, combined to produce the allotetraploid genome of wild emmer wheat or *Triticum turgidum* (AuAuBB) (Chantret et al., 2005, Eckardt, 2001, Huang et al., 2002). A small group of humans made the shift from hunter-gathering to cultivating plants for sustained survival around 12,000 years ago. However, the research community is still debating about where, how and why agriculture originated (Salamini et al., 2002). The fertile tetraploid (AABB) was domesticated by early farmers around 10 thousand years ago and became known as Emmer wheat. Approximately 8000 years ago, probably in a region close to the Caspian sea, a second event combined the genomes of tetraploid *T. turgidum* (AuAuBB) and *Aegilops tauschii* (DD), producing the allohexaploid *T. aestivum*

(bread wheat) genome (AuAuBBDD) (McFadden and Sears, 1946). This cross normally yields sterile hybrids; however, a doubling of the chromosomes in gametes or progeny gave rise to a fertile, hexaploid species, *T. aestivum* (Feuillet et al., 2008, Chantret et al., 2005). Figure 1-4 (Chantret et al., 2005) summarizes the above evolution events: AABB was domesticated, then grown with wild DD species, formed a hexaploid which crossed back to the tetraploid. Genetic exchange between AABB and hexaploid increased diversity in the AABB genomes of hexaploid, while there was little or no genetic exchange between the wild DD species and the new hexaploid (Berkman et al., 2013). An AABB x AABBDD cross will produce AABBD pentaploid progeny. An initial hybrid that is monosomic for every D chromosome. The low fertility of the pentaploid (AABBD) plants presented a common problem in direct cross of tetraploid (AABB) to hexaploid (AABBDD). The highly sterile (fertility of 1%-2%) was reported by Grama and Gerechter-Amitai from the pentaploid interspecific hybrids between the wild emmer accession G-25 and common wheat cultivars (Grama and Gerechte.Zk, 1974). The sterility of the pentaploid hybrids is generally conquered by further crossing or backcrossing the hybrids with common wheat cultivars (AABBDD) (Liu et al., 2002, Brown-Guedira et al., 2005, Hua et al., 2009, Ogbonnaya and Abdalla, 2013).

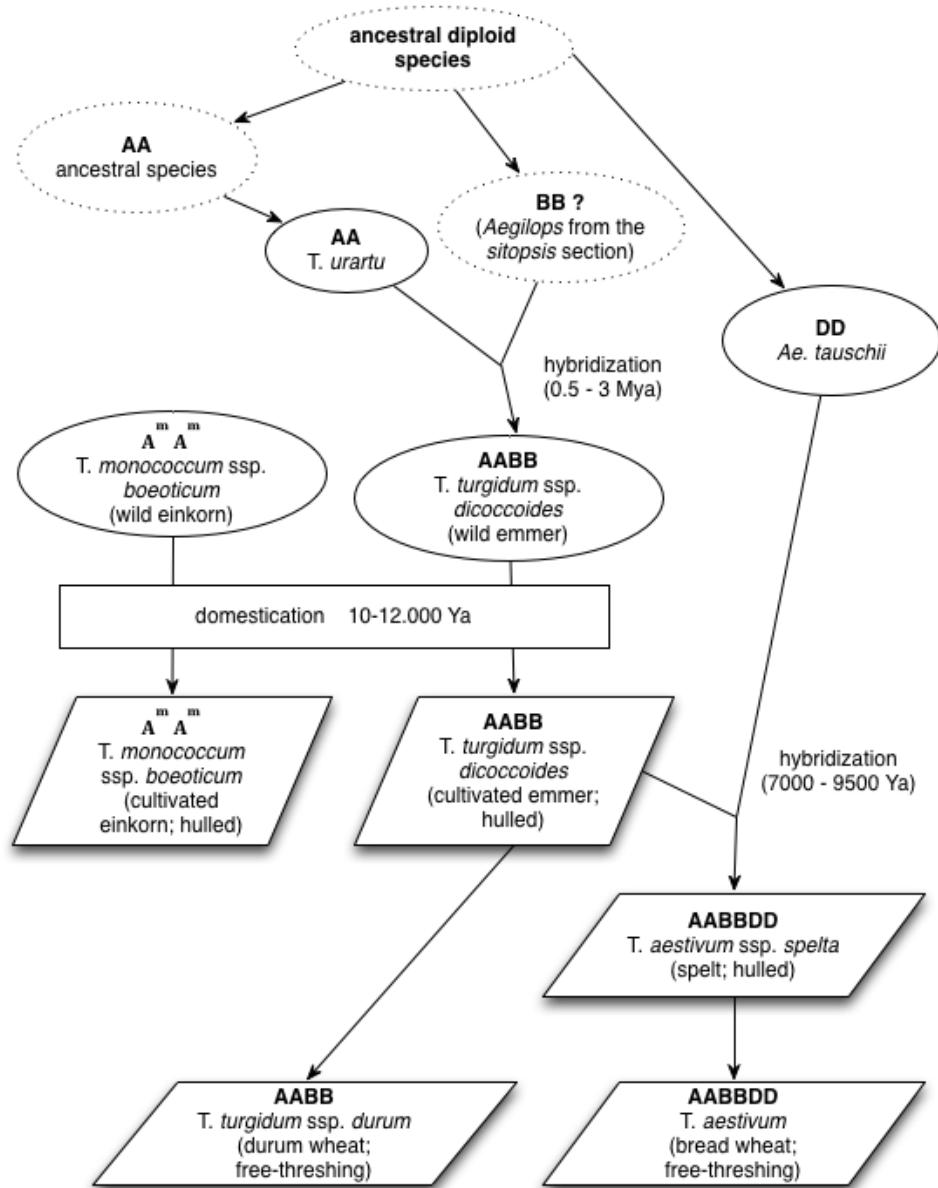


Figure 1-4 Schematic Representation of the Evolutionary History of Wheat Species (*Triticum* and *Aegilops*) (Adapted from Chantret et al., 2005).

1.5.3 Approaches to sequence the bread wheat genome

The completion of reference genome sequences for many important crops and the expansion of NGS technologies provide opportunities for accelerating crop improvement. The list of available or published reference genome for several important crops (Morrell et al., 2011, Bolger et al., 2014) is presented in Figure 1-5.

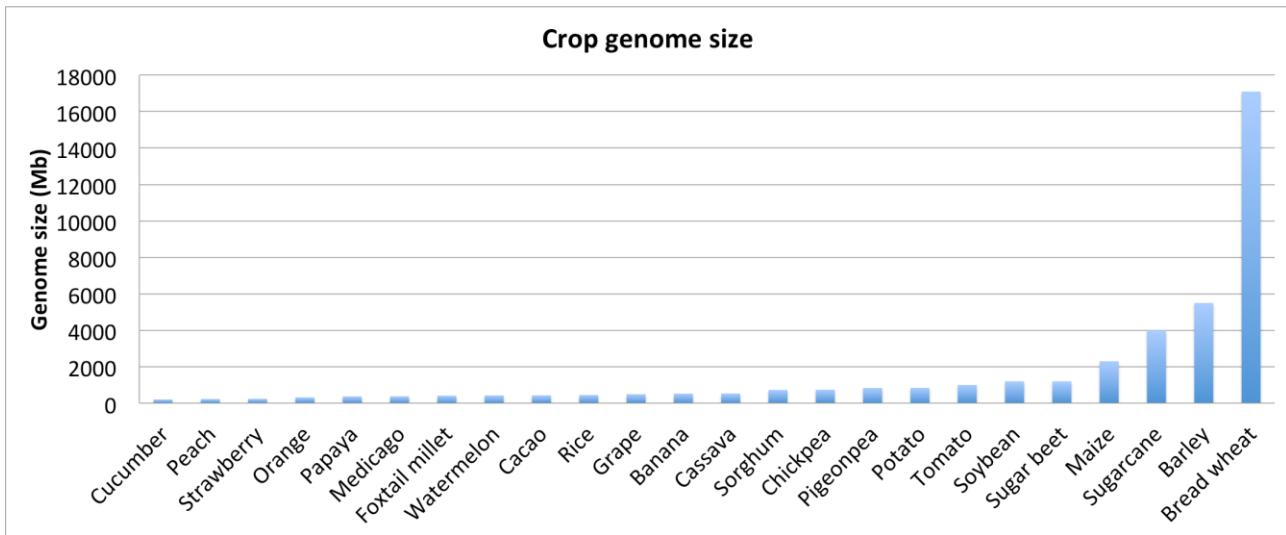


Figure 1-5 Crop genome size of some important crop genomes has been listed in order from smallest to largest.

Wheat possesses a large (17 Gb) genome, which is complex due to being hexaploid and containing a high proportion (~80%) of repetitive DNA (Chantret et al., 2005, Paux et al., 2008, Wanjigi et al., 2009, Gupta et al., 2008, Morrell et al., 2011), making genomic analysis significantly challenging. Wheat genome sequencing is ongoing.

The IWGSC (International Wheat Genome Sequencing Consortium) (<http://www.wheatgenome.org/>) was formed following a wheat genome sequencing workshop in November 2003 (Gill et al., 2004). This consortium was focusing on sequencing wheat to enhance the knowledge of the structure and function of the genome for wheat improvement (www.wheatgenome.org) (Eversole, 2012). This consortium was adopting a BAC (bacterial artificial chromosome) by BAC approach, which aimed to deliver a complete high quality genome sequence by 2016. Hexaploid bread wheat was selected rather than the individual, ancestral diploid genomes because it is the species grown in 95% of the wheat-growing areas, and the A, B and D genomes of bread wheat do not correspond physically and functionally to the sum of the ancestral A (*Triticum urartu*), B (unknown species that is likely to be related to *Aegilops speltoides*) and D (*Aegilops tauschii*) genomes (Feuillet et al., 2011). The consortium has chosen a chromosome-based strategy to construct physical BAC clone maps and subsequently to sequence each of the 21 individual chromosomes (Dolezel et al., 2007).

A Czech laboratory has developed a strategy to produce BAC libraries from single chromosomes or chromosome arms for reduction of the large wheat genome into more manageable pieces. The chromosome-based approach was expected to deliver a complete, finished reference genome sequence of sufficient quality to enable gene isolation, functional analyses, new allele discovery for pre-breeding and polymorphism discovery for marker-assisted selection (Feuillet et al., 2011). The IWGSC has completed construction of chromosome- and chromosome arm-specific BAC libraries, derived from wheat chromosomes (Šafář et al., 2010). The first physical map of the largest wheat chromosome 3B has been produced (Paux et al., 2008) and its sequencing is complete (<http://urgi.versailles.inra.fr/index.php/urgi/Projects/3BSeq>) (International Wheat Genome Sequencing Consortium, 2014). The IWGSC previously expected to deliver the complete genome sequence by 2014 (Eversole, 2012), but this has been extended to 2016 and is dependent on funding.

Other groups are applying next-generation sequencing technologies to gain insights into this complex genome to complement the BAC by BAC approach. A consortium from the UK produced 5x coverage of the bread wheat genome using Roche 454 technology (<http://www.cerealsdb.uk.net/>). Although this is insufficient to produce a finished genome assembly, the data is a valuable resource for gene discovery and genetic variation analysis. This data has contributed to our understanding of the bread wheat genome. For example, Brenchley *et al.* (Brenchley et al., 2012) compared this with the sequences of diploid ancestral and progenitor genomes recently. They have identified between 94,000 and 96,000 genes, and assigned two-thirds to the three component genomes (A, B and D) of hexaploid wheat (Brenchley et al., 2012). They demonstrate that the hexaploid genome is highly dynamic, with significant loss of gene family members on polyploidisation and domestication (Brenchley et al., 2012). To identify exome-based co-dominant SNP-based assays, which have capability of distinguishing between heterozygotes and homozygotes, Allen *et al.* used targeted re-sequencing of the wheat exome to generate large amounts of genomic sequences from 8 varieties (Allen et al., 2013).

Even with the advantage of next-generation sequencing technology, it is still relatively expensive to generate whole genome sequences for more than a few wheat genomes at

any one time. To resolve this problem, Winfield *et al.* have developed a targeted-capture re-sequencing protocol based upon NimbleGen array technology to capture and characterise 56.5 Mb of genomic DNA with sequence similarity to over 100,000 transcripts from 8 UK allohexaploid wheat varieties. A total of 500,000 putative single-nucleotide polymorphisms (SNPs) were then identified, with 80% representing variants between the homoeologous genomes, while a relatively small number (20%) were putative varietal SNPs between the 8 varieties studied (Winfield *et al.*, 2012). Cereals DB 2.0, an online resource containing databases for DArT markers and EST sequences, and in excess of 100,000 putative varietal SNPs, has been developed by the functional genomics group at the University of Bristol (Wilkinson *et al.*, 2012). A draft wheat genome assembly has also been produced from the donor species (Chantret *et al.*, 2005) of the wheat A-genome, *Triticum Monococcum*; the wheat B-genome from tetraploid wheat *Triticum Durum* (AABB); the wheat D-genome, *Aegilops tauschii* (Jia *et al.*, 2013).

The sequencing of isolated chromosome arms provides valuable information of the gene content of wheat. This was first demonstrated by isolating chromosome arm 7DS and applying second-generation sequencing technology and appropriate algorithmic analysis to sequence and assemble low copy and genic regions of this chromosome arm (Berkman *et al.*, 2011a). The subsequent comparison of wheat homoeologues revealed that ~84% of genes previously identified in 7DS have a homoeologue on 7BS or 4AL (Berkman *et al.*, 2012b). The value of applying second generation sequencing technologies to assemble gene-rich regions of complex genomes and investigate polyploid genome evolution has demonstrated that selection and dispersion of wheat has shaped the modern wheat genome (Berkman *et al.*, 2013) and also defined a translocation between chromosome arms 7BS and 4AL to gene level resolution (Berkman *et al.*, 2012b). A recent endeavor to shotgun sequence and characterise all the chromosome arms of wheat has been published (International Wheat Genome Sequencing Consortium, 2014).

With the increasing volume of wheat genome data becoming available through such efforts, it is essential to provide resources that can integrate wheat-specific sequence

information in a manner accessible to crop improvement researchers (Edwards and Batley, 2010).

1.5.3.1 Transcriptome analysis in wheat

High quality transcriptomics of bread wheat is a powerful method to increase molecular understanding of the biology of bread wheat. However, the application of NGS technologies are not yet widely applied in transcriptome analysis in wheat (Edwards et al., 2013).

The transcriptome of hexaploid wheat has been analysed using microarray methods. In 2004, high-density microarrays of a publicly available wheat EST resource containing 26,382 sequences were produced, based on 35 individual cDNA libraries representing specific developmental stages of different tissues of both grains and plant development (Wilson et al., 2004). A wheat unigene cDNA microarray containing 9,155 features was developed from this resource and used to investigate changes in the wheat embryo transcriptome during late grain development and maturation and during the first 48-hours of post imbibition germination (Wilson et al., 2005). In addition, a 9K wheat unigene cDNA microarray has been produced from cDNA libraries prepared mainly from developing wheat seed (Gregersen et al., 2005).

An Affymetrix wheat GeneChip oligonucleotide array has been developed, with over 61,127 probe sets representing 55,052 transcripts (<http://www.affymetrix.com>), and this has enabled the generation of numerous high-quality gene expression datasets (Coram et al., 2008a). This GeneChip was used to assay the transcriptome of developing grains (Coram et al., 2008b) and to identify the genes involved in the development of low temperature tolerance (Laudencia-Chingcuanco et al., 2011). The wheat GeneChip has further been applied to profile the changes occurring after inoculation with *Puccinia striiformis*, the causal agent of stripe rust in wheat lines that differed for the presence of the Yr5 gene (Coram et al., 2008b). A 90k SNP genotyping array chip has been developed and used to characterise genetic variation in allohexaploid and allotetraploid wheat populations. A significant fraction of common genome-wide distributed SNPs are

included in this array chip and are represented in populations of diverse geographical origin. Low-intensity clusters that are detected in assays are expected to provide insight into the distribution of presence-absence variation (PAV) in wheat populations (Wang et al., 2014).

Despite the demonstrated value of NGS technology in transcriptome analysis in other plant species, these have not been widely applied to transcriptome analysis in wheat. A number of research groups have recently analysed wheat transcriptome NGS data and several publications have been published (International Wheat Genome Sequencing Consortium, 2014, Pfeifer et al., 2014). The combination of wheat transcriptome data and the availability of isolated wheat chromosome arm assemblies will likely lead to a greater understanding of the structure, expression and evolution of the wheat genome. Such a study will have implications not just in wheat, but also more broadly in polyploid plant genomics.

1.6 Summary

Bread wheat is an extremely important crop plant widely grown worldwide feeding a significant fraction of the world population. Wheat improvement is required to meet the demands of population growth in the face of climate change. Next-generation high-throughput sequencing (NGS) technologies support to the completion of reference genome sequences for many important crops, and offer approaches for the profiling of mRNAs and small RNAs. In addition, NGS technologies enable the discovery of genome diversity. Single nucleotide polymorphisms (SNPs) are regarded as the highest resolution of genetic variation in comparison with other kind of markers. SNPs provide the exact nature of the allelic variants.

The purpose of this thesis is to establish and apply an approach that utilises NGS data and genome diversity identification and characterisation methods, and apply this to identify inter-varietal SNPs from wheat transcriptome data from 3 wheat varieties and Illumina paired read genomic data from 16 Australian diverse bread wheat varieties.

AutoSNPdb method has been successfully applied to identify SNPs from wheat transcriptome sequence data that have been generated using 454 Life Sciences technology. Each SNP is within a consensus sequence contig with annotation, providing resource for the selection of polymorphism within genes of interest.

SGSautoSNP method has been adopted for identifying large number of genomic SNPs across 16 Australian diverse bread wheat varieties. The processes of SNP density analysis and SNP transition/transversion ratio analysis have been developed for investigating the evolution and breeding history of this important crop.

An online integrated database and portal for wheat genomics has been established and released. This research aims to understand the effect of sequence variation on the form and function of wheat for improving the wheat production.

Chapter 2 Single-nucleotide polymorphism (SNP) identification from transcriptome data from Australian bread wheat (*Triticum aestivum*) cultivars

2.1 Introduction

The phenotype of an organism is largely determined by the genes expressed within it. These expressed genes can be described by a “transcriptome” indicating the identity of each expressed gene and its level of expression for a defined population of cells. The transcriptome can be regarded as a dynamic link between an organism’s genome and its physical characteristics (Velculescu et al., 1997).

Transcriptome sequencing using next-generation sequencing (NGS) technology provides high-resolution transcriptome data and is a useful tool for studying global transcriptional networks (Oono et al., 2013). Wheat transcriptome sequencing can be applied to identify candidate genes associated with trait expression and to develop SNP (Single Nucleotide Polymorphism) markers for tracking favourable alleles in breeding programs (Schreiber et al., 2012). The analysis of the transcriptome in bread wheat would also prove to be an excellent tool to unravel regulatory networks associated with important traits. The aim of this chapter is identifying biological information related to specific genes or traits from transcriptome analysis. SNP identification and characterisation from transcriptome data, and differential homoeologous gene expression identification and characterisation are approaches for transcriptome analysis. The 454 transcriptome reads used in this chapter were produced from Excalibur and RAC875 and Kukri. A total of 4,694,141 transcriptome sequence reads from the three bread wheat varieties were assembled to identify a total of 38,928 candidate SNPs. Each candidate SNP is within an assembled consensus contig with annotation of genes.

Next generation whole genome re-sequencing has been used for the identification of large numbers of SNPs in plant genomes (Edwards and Batley, 2010, Berkman et al., 2012a), including Arabidopsis (Ossowski et al., 2008), rice (Yamamoto et al., 2010, Deschamps et al., 2010, Subbaiyan et al., 2012), soybean (Lam et al., 2010, Deschamps et al., 2010, Song et al., 2013), maize (Lai et al., 2010), common bean (Blair et al., 2013), and *Brassica napus* (Dalton-Morgan et al., 2014). Where a suitable reference is not available, SNPs can be identified using transcriptome data or reduced representation sequence libraries. The discovery of SNPs from next generation transcriptome data was pioneered in maize, where 454-based sequencing was demonstrated as an excellent method for the high-throughput acquisition of gene-associated SNPs (Barbazuk et al., 2007). More recently this approach has been applied in hexaploid oat to identify 9,448 candidate SNPs with a validation accuracy of 54% (Oliver et al., 2011), and a combination of Illumina and 454 transcriptome sequencing was used to identify SNPs in chickpea (Hiremath et al., 2011, Azam et al., 2012). More than 20,000 SNPs were identified within Illumina transcriptome data from Brassica (Trick et al., 2009), and 454 sequencing of amplicons in polyploid sugarcane was used to identify more than 2,000 SNPs (Bundock et al., 2009). In addition, 454 technology in Melon (*Cucumis melo L.*) was used to identify about 40,000 SNPs (Blanca et al., 2012).

SNP discovery in bread wheat has been hampered by the combination of the large size and hexaploid nature of the wheat genome, which often confound the differentiation of homoeologous and inter varietal SNPs, together with relatively low levels of polymorphism within this highly inbred crop (Barker and Edwards, 2009, Kaur et al., 2012, Lai et al., 2014). Genomic approaches have been successfully applied in diploid and tetraploid wheat. A combination of Roche 454 and AB SOLiD sequencing was used to identify SNPs across the diploid *Aegilops tauschii* genome (You et al., 2011), and a total of 2,659 putatively homozygous SNPs were identified within 1,206 consensus sequences from four tetraploid durum wheat cultivars following 454 sequencing of CRoPS reduced representation libraries (Treibbi et al., 2011). Next generation SNP discovery in hexaploid bread wheat has been shown to be a significant challenge which is slowly being addressed. Illumina sequence data has been applied to discover 14,000 SNPs in the

wheat transcriptome (Allen et al., 2011), while in a separate study, bulk segregant analysis was used to identify more than 4,000 SNPs from Illumina RNA-Seq data with a validation rate of 56-58% (Trick et al., 2012). A 9k Illumina Infinium SNP array was developed and used to genotype a total of 8,630 SNPs, with a validation accuracy of 90% (Cavanagh et al., 2013), with many of these SNPs contributing to a new developed 90K Illumina Infinium array (Wang et al., 2014).

With the increasing ability to identify large numbers of SNPs from next generation sequence data (Imelfort et al., 2009), one of the remaining challenges is to maintain information about these SNPs within an annotated database (Duran et al., 2009b, Duran et al., 2009a, Lee et al., 2012). AutoSNP software was developed to automatically identify candidate SNPs and small insertions/deletions from Sanger EST data, with associated measurements of confidence for the validity of SNP candidates (Barker et al., 2003, Batley et al., 2003). The script uses a redundancy-based approach to distinguish valid SNPs from erroneous sequence. AutoSNPdb was later developed to combine autoSNP with sequence annotation, and maintain the results within a web searchable relational database (Duran et al., 2009a). This application provides a flexible interface facilitating a variety of queries to identify SNP and indel polymorphisms related to specific genes or traits, and has been applied to successfully identify SNPs in barley, rice and *Brassica* species from Sanger EST data (Duran et al., 2009a).

I have modified and applied the autoSNPdb method to identify SNPs between Australian wheat varieties from Roche 454 transcriptome data. These varieties include drought tolerant Excalibur and RAC875 and the drought sensitive Kukri variety. A total of 4,694,141 wheat sequence reads from the three varieties were assembled to identify 38,928 candidate SNPs. Each SNP within an assembly is annotated, enabling the selection of polymorphisms within genes of interest or within regions syntenic with the related model plant *Brachypodium distachyon*. SNPs within annotated genes are maintained within a public database at <http://autosnpdb.appliedbioinformatics.com.au/>.

2.2 Experimental procedures

2.2.1 Data generation

An average of 3.21 million reads per variety were generated by the Victorian State Government Department of Primary Industries (DPI), with plant growth, RNA extraction and sequencing performed by Dr Matthew Hayden. The DSN-normalised mSEQ cDNA libraries correspond to approximately 500 Mb of paired-end sequence data. Preliminary data analysis from DPI indicated that DSN normalisation of transcript abundance was efficient.

Seeds for Excalibur, RAC875 and Kukri were germinated and seedlings were grown for 8-10 days. In addition, florets of different developmental stages, ranging from pre-meiosis to just prior to anthesis, were dissected from heads and snap frozen into liquid nitrogen prior to RNA extraction. Total RNA was isolated from seedling leaves, roots and florets using TRIzol (Invitrogen, Carlsbad, USA) according to the manufacturer's instructions and further purified on RNAeasy Spin columns (Qiagen, Germany).

Wheat 454 transcriptome data was produced using the method of Meyer *et al.* (Meyer et al., 2009) and included a step transcript normalization using Kamchatka crab duplex specific nuclease (Zhulidov et al., 2004).

2.2.2 Data processing

2.2.2.1 Trimming sequence data

Three cDNA fragment types were present in the wheat 454 transcriptome data: 5'-terminal fragments, 3'-terminal fragments and internal fragments. The residual SMART oligo signature linked to the cap core and Btitn suppressor tag was used to identify 5'-terminal fragments. Atitn and Btitn suppressor tag sequences are suppression tags. During PCR amplification, suppression tags invoke PCR suppression for the fragments that end up flanked by the same kind of adapter. In these fragments, the Atitn primer is found on the inside of the original cDNA sequence while the Btitn primer can be either inside or outside (Matz et al., 2009). Internal fragments were identified by SMART, BC

signature, cap core, and Btitn suppressor tag sequences. All reads were trimmed of these sequences using a custom Perl script prior to sequence assembly. The sequences of the above cDNA fragments were sorted and stored as a text file. This custom Perl script could match 5'-terminal fragments or 3'-terminal fragments with each end of each cDNA sequence (wheat 454 transcriptome read). Meanwhile, the custom Perl script could also match other primers with each whole cDNA sequence.

2.2.2.2 Assembling sequence data

MIRA (Chevreux, 2005) was designed and developed as a *de novo* transcriptome assembler for 454 transcriptome reads. The trimmed sequences with retained quality information were assembled with MIRA using the parameters 'denovo, est, normal, 454 – GE:not=30'. The *de novo* transcriptome consensus assemblies and related constituent reads were processed and parsed into a custom MySQL database using standard autoSNPdb scripts (Duran et al., 2009a).

2.2.2.3 SNP discovery and annotation

The autoSNPdb method (Duran et al., 2009a) was applied to identify candidate SNPs. Indel calling within assemblies was disabled. Assemblies containing four or more sequences were examined for polymorphisms. To enrich for intervarietal SNPs over the more frequent homoeologous SNPs, all bases within the same cultivar were required to be the same at each SNP position to prevent the false calling of SNPs from misassembled reads. The minimum redundancy score associated with a polymorphism was varied in proportion to the number of sequences within the assembly at the SNP position. A minimum redundancy score of 2 was required for predicting SNPs represented by 7 sequences; 3 for between 8 and 11 sequences; 4 for between 12 and 19 sequences, and a minimum redundancy of 5 was required for 20 or more sequence reads (Duran et al., 2009a). A SNP co-segregation score was calculated using each SNP compared to all SNPs in an assembly. The weighted co-segregation score was calculated according to the proportion of missing data at that position in the assembly (Duran et al., 2009a). All of the unique singletons and consensus assemblies were compared with GenBank (Release 182) and SwissProt (Release 2011_02) using BLAST (Altschul et al., 1997). Annotations with an e-value of >10⁻⁵ were parsed to the wheat autoSNPdb database.

Consensus sequences were also compared with the *Brachypodium distachyon* reference genome to identify candidate gene orthologs.

2.2.3 Validation

Thirty-eight contigs were selected randomly for validation of 50 SNPs, which was performed by Manuel Zander and Annaliese S. Mason. These had a range of redundancy and co-segregation scores. Genomic DNA was isolated from the three cultivars Excalibur, RAC875 and Kukri, according to a protocol adapted from (Fulton et al., 1995). PCR amplification of the 38 loci was performed using primers designed to conserved sequence surrounding the SNPs (Table 2-1) in a 20 µL reaction volume containing 1 × iTaq PCR buffer (100 mM Tris-HCl and 500 mM KCl, pH 8.3) (Scientifix), 200 µM each dNTP (Scientifix), 0.5 µM each primer, 1.5U iTaq DNA polymerase (Scientifix), RNase and DNase free water (Gibco) and 60 ng DNA. Thermocycling conditions for the reaction were 94 °C for 2 min, followed by 35 cycles of: 94 °C for 30 sec, annealing for 1 min at 60 °C and extension for 1 min at 72 °C. Final extension was performed at 72 °C for 10 min. The PCR products were electrophoresed on a 1% (w/v) agarose gel in 1 × TAE buffer (Sambrook and Russel, 2001) containing ethidium bromide resolved products, which were excised and purified using a silica method based on (Boyle and Lew, 1995).

The purified PCR products were Sanger sequenced using BigDye 3.1, using forward PCR primers, and electrophoresed using an ABI3730xl. The sequences for each locus and line were aligned and compared using Geneious Pro v5.4.6 (Drummond et al., 2011) with a cost matrix of 65%, a gap open penalty of 6, and a gap extension penalty of 3, and each of the SNPs assessed manually.

2.3 Results and discussion

2.3.1 Sequence assembly

A total of 5,407,382 wheat transcriptome reads were produced using a Roche 454 DNA sequencer by Matthew J. Hayden from Department of Primary Industries, Victorian AgriBiosciences Centre. After trimming and filtering, 4,694,141 reads representing the three cultivars remained: Excalibur (1,484,565 reads); RAC875 (1,596,757 reads); and Kukri (1,612,819 reads). The filtered reads represented 1,336,401,100 nucleotides and had an average read length of 285 bp. MIRA assembly produced a total of 202,405 contigs and 222,298 singletons. AutoSNPdb requires a minimum redundancy score of two (that is the polymorphic base should be represented by at least two reads from two or more varieties). Of the assembled contigs, 113,747 (56.2%) contained four or more sequences and could therefore be used for redundancy based SNP discovery. There was an average of 7.4 cultivar specific reads per contig which is substantially greater than the 0.12 cultivar specific reads per contig within the barley autoSNPdb assembly (Duran et al., 2009c). This largely reflects the smaller number of varieties sequenced in this study and greater read depth obtained by next-generation sequencing (three wheat varieties compared to 128 varieties and Sanger sequencing in the barley study). The greater sequencing depth for wheat was used to help differentiate between inter-varietal and homoeologous SNPs between related sub-genomes.

2.3.2 SNP discovery and abundance

The AutoSNPdb method has previously been applied to barley, *Brassica* and rice sequence data. Several changes were made to AutoSNPdb in this project: a SNP conflict represents the nucleotide differences that exist within same variety. SNPs that conflict within a variety have been eliminated. This change can improve the confidence of identifying real inter-varietal SNPs and reduce the false SNP identification. In addition, the original autoSNPdb method used CAP3 for sequence assembly. The wheat NGS transcriptome data was assembled using MIRA (Chevreux, 2005) which can generate more accurate assemblies from 454 data.

A total of 38,928 candidate intervarietal SNPs were identified within 15,032 contigs from the wheat transcriptome. The percentage of contigs containing SNPs increased with the increasing number of reads until around 35 reads per contig, and then levelled off (Figure 2-1). The number of SNPs identified per contig also leveled off for contigs containing more than 35 sequence reads (Figure 2-2). The majority of contigs contained more than 35 reads which suggests that additional data would not substantially increase the total number of SNPs predicted.

The 38,928 SNPs were identified over a total region of 22,154,849 bp, suggesting an average SNP density of one SNP per 569 bp. This is less than the density observed in barley (one SNP in every 240 bp) using a similar method (Duran et al., 2009c) and reflects the relatively few wheat varieties sequenced, the low level of diversity in hexaploid wheat and the high stringency of the autoSNPdb pipeline.

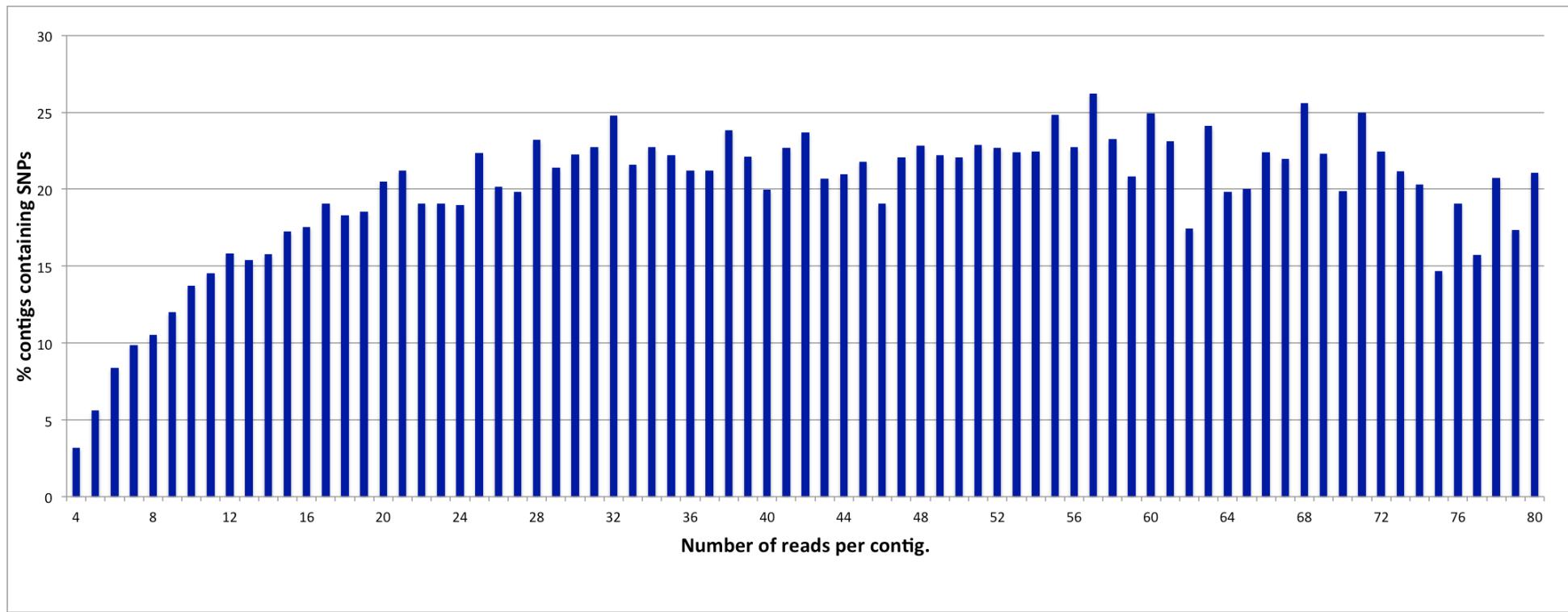


Figure 2-1 Histogram showing the percentage of reads containing SNPs for contigs representing increasing numbers of reads.

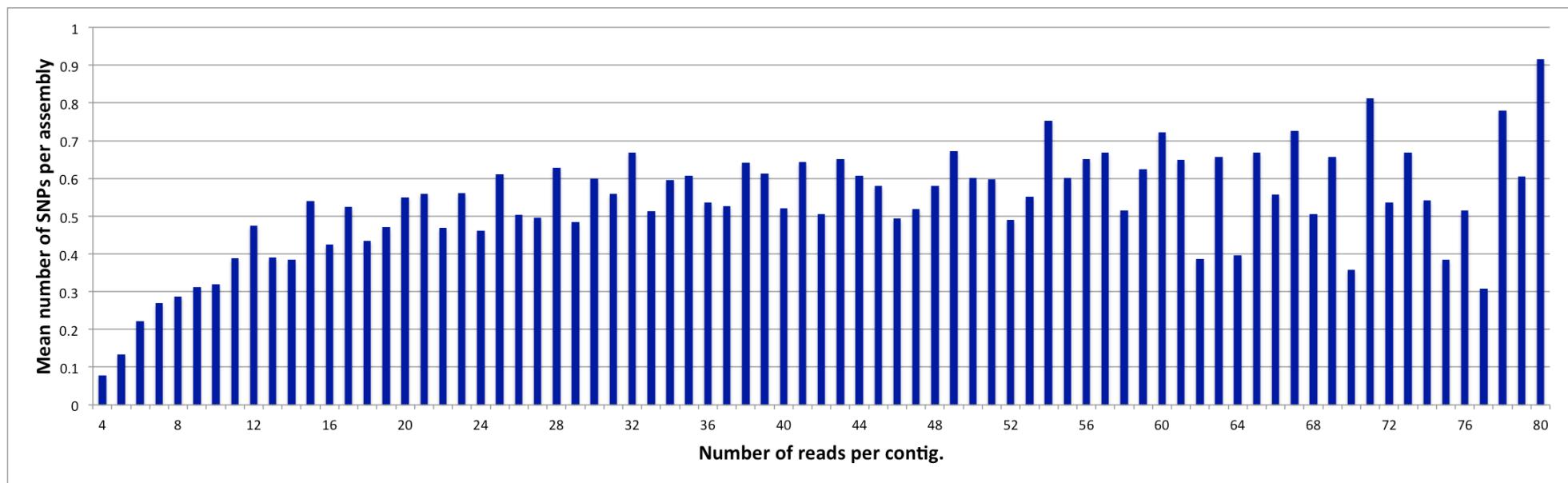


Figure 2-2 Histogram presenting the mean number of SNPs for contigs with increasing numbers of reads.

2.3.3 Validation

Fifty candidate SNPs, within 38 contigs, were investigated by direct sequencing of PCR products to assess the false discovery rate. The SNPs were randomly chosen to have a range of redundancy and co-segregation scores. Of the 50 candidate SNPs, 39 (78%) were shown to be true polymorphisms (Table 2-1), from 31 (81.5%) of the 38 contigs, while 11 (22%) were mono-allelic SNPs. Of these validated SNPs, 26 (66%) were sub-genome specific and 13 were coincident with sub-genomic variation between homoeologous gene copies (ie. detected as heterozygous in one or more variety). Of the candidate SNPs which were assessed to be mono-allelic, they appeared to come from multi-gene families or gene homoeologues, where the primers had been designed to be specific to only one gene copy. Overall, the validated SNPs had an average SNP redundancy score of 3.4 (range 2-7), this is higher than the mono-allelic SNPs, which had an average SNP score of 2.3 (range 2-3). This suggests that selecting SNPs with higher SNP scores would increase the accuracy of true polymorphism identification. You *et al.* utilised next generation sequencing for the identification of SNPs in the diploid D genome of wheat, *Ae. Tauschii*, and validated over 80% of predicted SNPs (You *et al.*, 2011). In contrast, Trebbi *et al.* had 36% SNPs matching the expected genotypes from the SNP discovery phase when they applied CRoPS technology for the identification of SNPs in tetraploid durum wheat (Trebbi *et al.*, 2011), which is lower than the 54% single locus intervarietal SNPs observed in this study, and demonstrates the difficulty of accurate prediction of inter-varietal SNPs in polyploid species. The SNP validation rate in our study is similar to that reported for SNP discovery from 454 cDNA sequencing in hexaploid oat (Oliver *et al.*, 2011), in which 54% of tested SNPs were validated as polymorphic in the parents of a mapping population and slightly less than validated in a previous study in hexaploid wheat (67%) with a SNP score of 2 (Allen *et al.*, 2011).

Table 2-1 SNP validation.

Contig	Forward primer	Reverse Primer	SNP position	Excalibur alleles	Kukri alleles	RAC875 alleles	SNP score	Result
2184	GCTCAAAGGACCTGGTCTG	TTATTGTGCAGTCGCCATC	1229, 1232 &1235	T, A & T	T, A & T	G, G & G	3,2,2	FALSE
2281	CTAATCCAAGGCCGTTCAAG	ACCTCGTGGTGAAGGAAGTG	1986	T	G	G	4	intervarietal
4628	CTCAACACTGCAGCATATCCA	GCCATGAACAGACCCAAAAAA	2266	A	G	A	5	heterozygous
4943	CAGAAAGGGAGTCGAAACAGT	GTGCAAAGCTCGTTGATATG	465	T	T	C	2	intervarietal
4989	ATGTGTGTGCAGGTACAAGG	ATGTGTGTGCAGGTACAAGG	1846	T	C	C	5	intervarietal
6456	TTATGCCAGTAGGGTTGGT	AACAGAAAGGATGGTTGGA	855	C	No reads	T	2	intervarietal
7931	GCACCTCTGTTGACTGCTCT	CAAGGGAAACCAAGCAGCATCAT	2837	A	G	G	5	intervarietal
8145	AGCAGACAAAGCAATGTGAA	CCATAGACTGTTCATGCCA	1019, 1084	G, A	T, G	G,A	3,5	intervarietal
9444	GTCCTTGGCTTGCTGGTT	CAGTTATCAGCTGCGCAAAA	314	G	G	C	7	intervarietal
10770	TAGGCTTGAGCCCCATCAAT	TGCCCCTTCCATGAGTTTAT	1087	G	A	no reads	2	heterozygous
10153	CGACTCTCCCTCTATCTGGA	ACACCTGACGGAACCTGAAT	1263	No reads	A	G	2	FALSE
11551	TGGCCTTCTCGGAGAGAGTA	AACCGTGAAAGGGAGGAGTT	2490	C	C	T	5	heterozygous in RAC only
12063	CACAGTTGGGTACGCAGTG	GGACACGAGGAAACCTTGTGTC	898 &908	T, A	T, A	C, G	3 & 3	FALSE
12307	ATTTGGGAGAAATATGTGCG	CCTTTATCAGCAAGGAGGAA	2480	T	T	C	2	heterozygous
12551	CATTACCACGTCGATCAGAA	TAATCCAGATTGCAACCAAA	1890	T	C	C	4	intervarietal
13284	CCTTGCGCCAAAGTATTTC	ACGTTACCAAGGTCGAGAGG	611	T	C	no reads	2	FALSE
14357	ATCAGCAAGATCCTACAGCC	GGTATCACCAAAATCCTCCA	391	A	A	G	4	intervarietal
15072	GGTCTCATTGCTGCAAGG	GGCTACAGAGAGATGCATGTA	527	T	T	C	3	intervarietal
15553	TAAGTTGATAAGGCCAGGG	ATTGCAGCGTATGTACCAAA	749, 769	T, T	T, T	C, C	2, 2	FALSE
16531	CTCGGGAAACCTATCCCTA	TAAGGCCACCCACATCTTCGT	783, 837	C, C	T,T	C,C	2,3	intervarietal
17468	TCGGTTCACTGACTGAAGAA	GTGCTGATAAGGTCTTGCT	1081	C	T	T	3	intervarietal
18396	ACAGAAGCTTGTGAAATCC	CACATCCATGGTGAAGAGAA	1269, 1284, 1290, 1317	G, A, A, T	A, no reads	A, C, G, A	2, 2, 2, 2	heterozygous
18835	AGATAGAAGTGGTGGTGGC	CAATGGCAGGGACATTATT	379	C	No reads	T	2	heterozygous
193961	CCGTCAGTGTCAATTGTGG	GCCAAGGAAACTGGTGTATGT	318	no reads	C	G	2	heterozygous in RAC only

19691	GATCGTCAGTGTACACCAT	TTTGTGACAACGTATGGGAG	421	no reads	C	T	2	heterozygous
19817	GACCAAAGGCTTGGAAATA	GTAATTGACCTGGAGGGAGC	990	No reads	C	T	2	intervarietal
20034	ATGGGGGATAAAATGTGCTA	AGTTGGTCTGTTGGGTCACT	737	C	C	G	5	intervarietal
20795	TTTGTTCAGATGTTGTCGG	GATTGGCAACTACCAAGGAC	1013	C	G	G	2	heterozygous
22002	GGCAGCTGGGTTAGTTGATT	CATTCCGTTGCTCATCCTTC	699, 724	G,T	no reads	A,C	3,4	intervarietal
22419	TTAAGTGGTTGGGGAGGTAA	GCAATAGTTGCCTCCAGTTT	1524	C	G	G	5	intervarietal
25034	GATTTCACTCGTCCTGATCC	TATTCTCCTGGATCTTCGC	1161	A	C	A	3	heterozygous
31151	TAAGAGTAAACCCCTCACCG	ACGCCATGGAAGTAATCTGT	1040	T	C	T	2	intervarietal
33054	AGTTCCAGAAAAGGGTGGTA	CTCTGCCTCAGTTCCAATCT	330	T	T	G	2	intervarietal
42926	AACGTCTGTTCTGGGTGCTT	TGCTGGAAGAACACCTGATCG	1197	G	C	G	2	intervarietal
44073	GGGTGAGGAGGTCACTTGAA	CTTCTCACCTCCGGGTTATC	469, 472, 527	A, C, A	G, T, G	A, C, A	3, 3 & 3	intervarietal
44973	TGTTCTCCGTAGGAATAGCC	AGCTCCAAGTCCAGAACATGTC	1369	T	T	C	2	FALSE
45851	AGCAGGAAGATGATACCGAG	ATACCGGTCGATGAAGTGTT	545	A	No reads	C	2	FALSE
50002	TGCTTAATCACATGCCTTC	GATTAGCTCAGAACGCCGAT	953	C	A	A	2	intervarietal

2.3.4 Analysis of base changes

The frequency of nucleotide substitutions for the putative SNPs is summarised in Table 2-2. Transitions were far more abundant than transversions. A transition abundance bias is commonly observed for true SNPs and reflects the high frequency of C to T mutation following methylation (Coulondre et al., 1978). The bias observed in wheat is greater than that observed in barley using a similar method (Duran et al., 2009c) and may reflect a higher level of methylation in the wheat genome. DNA methylation leading to gene silencing represents revolutionary changes at a functional (genetic diplodisation) level (Feldman and Levy, 2009). An increased abundance of C/G transversions was observed compared with A/T, A/C and G/T transversions. A similar unexpected bias was also observed in barley (Duran et al., 2009c).

Table 2-2 Base changes and the number of occurrences within SNPs in wheat autoSNPdb.

Base Change	Type	Number
A-G	Transition	12323
C-T	Transition	12954
A-C	Transversion	3324
A-T	Transversion	3345
C-G	Transversion	3625
G-T	Transversion	3354

2.3.5 Sequence annotation

Of the 202,405 assemblies, 74,177 (36.6%) had SwissProt annotations, 145,504 (71.9%) had GenBank annotations, 111,714 (55.2%) showed a significant match with the *Brachypodium distachyon* genome, and 48,303 had no annotation (Figure 2-3). In total, 61,724 assemblies (30.5%) were annotated by all methods. Sequences with no annotation may be novel genes in wheat or may be too short for accurate annotation. The annotations can assist the wheat research community to search for SNPs in genes of predicted function or in a reference genome location of interest.

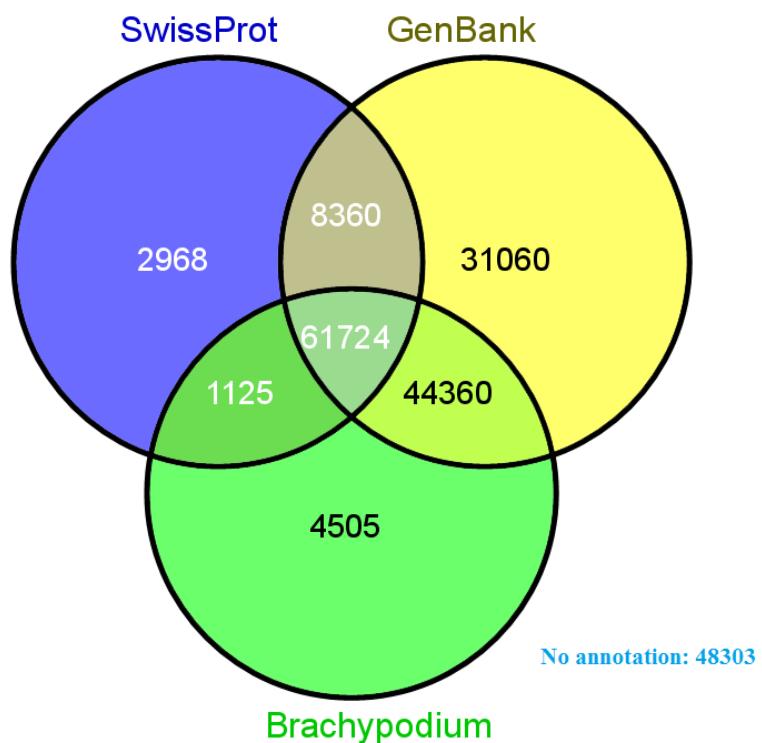


Figure 2-3 Venn diagram displaying the number of reads with combinations of SwissProt, GenBank and *Brachypodium* orthologue annotation.

2.4 Conclusions

I have modified and applied the autoSNPdb SNP discovery and annotation pipeline to identify 38,928 candidate SNPs between three hexaploid wheat varieties from Roche 454 NGS data. These SNPs and annotations are maintained within a customised web searchable database at <http://autosnpdb.appliedbioinformatics.com.au/> and provide a valuable source of annotated genetic markers in wheat that can be applied for high-resolution genetic map construction and trait association. These markers may be used to validate and assist the assembly of the wheat genome.

Chapter 3 Identification and characterisation of more than 4 million inter-varietal SNPs across the group 7 chromosomes of bread wheat

3.1 Introduction

Genome analysis in bread wheat poses substantial challenges; in addition to the complexity associated with its hexaploid structure, the bread wheat genome is very large (~ 17 Gb; around 40 times the size of rice, or nearly six times larger than the human genome) and consists of between 80-90% repetitive sequence (Šafář et al., 2010, Wanjigi et al., 2009).

Genome mapping using molecular markers has played a central role in genetics since the 1980s (Schlotterer, 2004) revolutionising fundamental research approaches such as the definition of haplotypes, the discovery of genomic regions associated with specific traits, and the assessment of evolutionary relationships between organisms. In addition to being critical for research in cereal crops such as wheat, molecular markers play a crucial role in modern cereal breeding (Duran et al., 2009b, Rafalski, 2002). For example, genotyping using molecular markers facilitates accurate identification and maintenance of genetic stocks, and guides the development of genetically diverse populations for selection programs. In some instances traditional marker-assisted selection, wherein selection for a specific trait is guided using a marker or markers that accurately predict the inheritance of that trait, has enabled rapid incorporation of favourable alleles into elite cereal cultivars (Xu and Crouch, 2008). Furthermore, an increase in the number of markers available for cereal crops and a decrease in the cost of genotyping is beginning to enable new approaches including genome wide association studies (GWAS) (Rosenberg et al., 2010, Schlotterer, 2004, Tian et al., 2011) and genomic selection (GS) in cereals (Heffner et al., 2009, Poland et al., 2012). Single nucleotide polymorphisms (SNPs) represent the most

frequent type of genetic polymorphism and can therefore allow the development of the highest density of molecular markers. (Batley and Edwards, 2007). Powerful next-generation sequencing (NGS) technologies provide the possibility of large-scale SNP discovery by comparing whole-genome shotgun sequences of individuals with high-quality reference genome sequences (Edwards et al., 2013, Edwards et al., 2012a, Imelfort et al., 2009).

Expressed genes have been a traditional source of data for SNP discovery. AutoSNP (Barker et al., 2003, Batley et al., 2003) and the associated autoSNPdb (Duran et al., 2009a) are tools for this purpose, and use redundancy and haplotype co-segregation to distinguish true polymorphism from sequence error. The autoSNPdb pipeline described above has recently been applied for the discovery of 38,928 candidate SNPs from 4,694,141 reads of wheat 454 transcriptome data (Lai et al., 2012b). SGSSautoSNP (Second Generation Sequencing autoSNP) is an additional SNP discovery pipeline designed specifically to predict SNPs from whole genome Illumina shotgun sequence data. SGSSautoSNP has recently been applied to identify more than 800,000 SNPs between four varieties of bread wheat with accuracy greater than 93% (Lorenc et al., 2012b), using the wheat group 7 isolated chromosome arm assemblies as a reference (Berkman et al., 2011b, Berkman et al., 2012b, Berkman et al., 2013).

A fundamental research problem in biology is understanding how natural selection has shaped the evolution of gene regulation. SNP density is sensitive to mutation rate biases and mutation rate heterogeneity (Chen and Rajewsky, 2006). Distribution of SNPs, in indica rice inbreds, was non-random across the chromosomes, with high and low SNP density regions (Subbaiyan et al., 2012).

The large data volumes from NGS platforms provide the potential to discover very large numbers of SNPs, both in expressed sequences and elsewhere throughout the genome (Visendi et al., 2013). For example Lai *et al.* identified more than 1 million SNPs between six inbred maize lines; furthermore the authors were able to detect a large number of presence/absence variations (PAVs) and suggested that this phenomenon may contribute to heterosis in this species (Lai et al., 2010). High-throughput SNP discovery from NGS data has recently been applied to identify SNPs between two accessions of the diploid

wheat genome progenitor *Ae. tauschii*. For this purpose, an ‘Annotation-based genome-wide SNP discovery pipeline’ (AGSNP) (You et al., 2011) was developed to facilitate SNP discovery from species with large and complex genomes. Using this pipeline, the authors combined data from Roche 454 sequencing of *Ae. tauschii* accession AL8/78, with Applied Biosystems SOLiD and Illumina sequencing of genomic DNA and cDNA from *Ae. tauschii* accession AS75, to identify a total of 497,118 candidate SNPs (You et al., 2011). In hexaploid wheat, Allen *et al.* identified 14,078 putative SNPs in 6,255 distinct reference sequences via *de novo* assembly of Illumina GAIIx cDNA sequence data from wheat lines Avalon, Cadenza, Rialto, Savannah and Recital, supplemented with publically available EST sequences. The authors obtained a validation rate of 67% for a subset of 1,659 of these markers (Allen et al., 2011).

More recently, 4,694,141 wheat transcriptome sequence reads have been produced, using 454 sequencing technology, and assembled from three bread wheat varieties. As described in Chapter 2, a total of 38,928 candidate SNPs, with a validation rate of 78% of predicted SNPs, were identified from three wheat transcriptome assemblies using the autoSNPdb method (Lai et al., 2012b). Allen *et al.* used targeted re-sequencing of the wheat exome to generate large amounts of genomic sequence from 8 bread wheat varieties and identified 95,266 putative SNPs (Allen et al., 2013). Of these, 10,251 (10.76%) were predicted to be genome specific, putative co-dominant SNP markers with a validation accuracy of 96%.

High-density SNP genotyping arrays are a useful tool for discovery of genomic patterns of diversity, inferring ancestral relationships between individuals in populations and for studying marker-trait associations in genetic mapping experiments (Wang et al., 2014). A 9K Illumina Infinium SNP array was recently constructed and used to genotype a total of 8,630 SNPs, with a validation accuracy of 90% (Cavanagh et al., 2013), with many of these SNPs contributing to a new 90K Illumina Infinium array (Wang et al., 2014). SNPs on the 90K Illumina Infinium array were demonstrated to be polymorphic across multiple populations of different geographical origin, implying that the array can be applied as a genotyping platform in various wheat genetic studies. From the 90K Illumina Infinium SNP

array, 69% of the 81,587 functional iSelect bead chip assays revealed polymorphism among the unrelated wheat accessions (Wang et al., 2014).

A large national initiative was established in Australia in 2010 to coordinate diverse wheat genetic and genomic activities and establish a resource for Australian crop improvement (Edwards et al., 2012b).

In this study, I discovered more than 4 million candidate inter-varietal SNPs across the wheat group 7 chromosomes from 16 Australian wheat varieties using the SGSAutoSNP pipeline (Lorenc et al., 2012b). This abundance of SNPs has permitted an assessment of SNP density variation across the length of these chromosomes and a comparison of variation in homoeologous chromosomes representing the A, B and D genomes of wheat. Our results demonstrate the impact of evolution and breeding on bread wheat genome diversity and provide a valuable resource for the further characterisation and improvement of this important crop.

3.2 Experimental procedures

3.2.1 Data selection and generation

Despite being a major international crop, our understanding of the wheat genome is relatively poor due to its large size and complexity. To gain a greater understanding of wheat genome diversity, I have identified single nucleotide polymorphisms between 16 Australian bread wheat varieties. The Illumina whole genome re-sequencing paired-end reads were sequenced from 16 Australian wheat varieties. The cost of Illumina whole genome re-sequencing remains high. Therefore, only a limited number of wheat varieties were selected to represent the genetic diversity in Australia.

The data selection and generation was performed and supplied to our research team by Bioplatforms Australia. The 16 wheat cultivars were chosen based on three criteria: genetic diversity, economic impact and key varieties. Five to ten plants of each line were grown and DNA from leaf samples of each plant was extracted using a standard phenol/chloroform method. A NanoDrop ND-1000 spectrophotometer (ThermoScientific,

Willmington, DE) and standard agarose gel electrophoresis were used for quality assessment of the wheat genomic DNA. Illumina TruSeq DNA Library Preparation Kits (Cat. No. FC-390-1021, Illumina Inc., San Diego, CA) and associated recommended protocol were applied for DNA libraries and sequencing (Edwards et al., 2012b).

3.2.2 Data processing

Whole genome Illumina PE data for 16 Australian bread wheat varieties were downloaded from the website of Bioplatforms Australia (https://downloads.bioplatforms.com/wheat_cultivars/) and clonal reads were removed using a custom Perl script: remove_possible_clones.pl. The remaining sequence data for each of the 16 Australian wheat cultivars were mapped to the three group 7 wheat chromosome assemblies (Berkman et al., 2011a, Berkman et al., 2012b, Berkman et al., 2013), as well as an assembly of chromosome arm 4AL (Hernandez et al., 2012), using the alignment tool SOAP v2.21 (Li et al., 2009b) with default parameters, allowing up to 2 mismatches per read and only retaining read pairs mapping uniquely to the reference with parameter “-r 0”. The chromosome arm 4AL was included to prevent reads from the translocated 7BS/4AL region mapping to homoeologous locations on 7AS or 7DS.

The resulting SOAP files (text files) from each library were concatenated as one SOAP file for each variety. These files were converted into BAM files (binary files). Each BAM file was processed using FixMateInformation executable jar file from Picard command-line tools. FixMateInformation ensures that all mate-pair information is in synchronisation between each read and its mate pair within the lane-specific BAM file and simultaneously sort the BAM file by genomic coordinate. The resulting BAM files were merged using samtools v0.1.17 (r973:277) (Li et al., 2009a).

3.2.3 SNP prediction

A SNP discovery pipeline called SGSautoSNP (Second-Generation Sequencing AutoSNP) has been developed to identify SNPs from SGS sequence data. The SGSautoSNP pipeline calls SNPs between different individuals using Illumina paired read data aligned to references. SGSautoSNP uses BAM (Binary Alignment/Map) format to save memory and storage space (Lorenc et al., 2012b). SNP predication between 16 Australian wheat varieties was performed using SGSautoSNP, with output in.snp format for subsequent analysis, vcf format for multiple SNP analysis software applications, map and genotype format for the graphical genotype visualisation viewer Flapjack (Milne et al., 2010), and gff format for presentation on a GBrowse genome viewer at www.wheatgenome.info (Lai et al., 2012a). The variant call format (VCF) is a generic format for storing DNA polymorphism data, for example SNPs (Danecek et al., 2011).

3.2.4 SNP matrix production and transition/transversion ratio analysis

The.snp files generated by SGSautoSNP were parsed using a custom Python script filter_snps_matrix.py to generate the SNP matrix file (Table 3-1). The SNP matrix was subsequently converted to phylip format and I applied neighbor joining using RapidNJ (<http://birc.au.dk/Software/RapidNJ/>) to convert to Newick format (Simonsen et al., 2008). The phylogenetic tree was constructed using the Philodendron web-based application (<http://iubio.bio.indiana.edu/treeapp/>) (Gilbert, 1999).

The SnpEff variant annotation tool (Cingolani et al., 2012), for predicting the effect of the identified SNPs, was used to calculate transition/transversion ratios for SNPs between varieties. VCFtools is a software suite that implements various utilities for processing VCF files, including validation, merging, comparing DNA polymorphism data (Danecek et al., 2011). In this study, the transition/transversion ratio for each chromosome was calculated based on bins of 500 SNPs using VCFtools.

3.2.5 SNP density and gene analysis

The SNP density plots for each chromosome were generated using a custom Python script that calculates relative density based on a window size of 50,000 bp. Subsequent analysis was performed to identify genes in low SNP density regions, defined as those for which SNP density in the regions 2 kbp upstream and downstream of a gene was significantly lower than the mean for all genes on the chromosome.

Genes identified as being in low SNP density regions were annotated by comparison with the Swissprot database (Release 2013_06) using BLASTX (BLASTALL 2.2.6) (Altschul et al., 1990) with an E-value cut-off 1e-5.

3.3 Validation

SNP validation was performed by Manuel Zander. A total of 22 SNPs were selected from the three group 7 reference genomes for validation. These SNPs had a range of redundancy scores. Genomic DNA was isolated from 11 cultivars, Alsen, Chara, Drysdale, Excalibur, Gladius, H45, Kukri, RAC875, VolcaniDDI, Xiaoyan 54 and Yitpi, according to a protocol adapted from Fulton *et al.* (Fulton et al., 1995). PCR amplification of the 22 loci was performed using primers designed to bind to conserved sequence surrounding the SNPs (Table 3-2) in a 25 µL reaction volume containing 1 x iTaq PCR buffer (containing 100 mM Tris–HCl and 500 mM KCl, pH 8.3) (Bio-Rad, Hercules, CA), 200 µM each dNTP (Bio-Rad), 0.5 µM each primer, 1.5 U iTaq DNA polymerase (Bio-Rad), RNase- and DNase-free water (Gibco; Life Technologies, Carlsbad, CA) and 5 ng DNA. Thermocycling conditions for the reaction were 94°C for 2 min, followed by 35 cycles of 94°C for 30 s, annealing for 1 min at 59-62°C and extension for 1 min at 72°C. Final extension was performed at 72°C for 10 min. Gel electrophoresis was performed on a 1% (w/v) agarose gel in 1 x TAE buffer (Sambrook and Russel, 2001) containing ethidium bromide-resolved products. The Australian Genome Research Facility's (AGRF) PD+ service was used to purify and subsequently sequence the PCR products. The purified PCR products were Sanger-sequenced using Big-Dye 3.1 (PerkinElmer, Waltham, MA), using forward and reverse PCR primers, and analysed using an ABI3730xl. The

sequences for each locus and cultivar were aligned and compared using Geneious Pro v5.4.6 (Drummond et al., 2011) with a cost matrix of 65%, a gap open penalty of 6, and a gap extension penalty of 3, and each of the alignments assessed visually to determine the SNP.

In addition to the 22 resequenced SNP loci, flanking sequences for predicted SNPs from the recently published 90K Infinium array (Wang et al., 2014) were compared to the group 7 assemblies using WU-BLAST 2.0 (Gish, 1996-2006, <http://blast.wustl.edu/>) and an e value of 1e-10. SNP positions of flanking sequences matching the group 7 assemblies were extracted in tabular format and compared to predicted SNP locations using R scripts. Where Infinium SNP positions matched predicted SNP positions from SGSAutoSNP, the Infinium SNP call was compared with the SGSAutoSNP prediction.

Table 3-1 Pairwise inter-varietal SNP matrix for chromosomes 7A, 7B and 7D between 16 Australian wheat varieties.

AC Barrie	0															
Alsen	194,725	0														
Baxter	328,294	246,218	0													
Chara	592,193	438,075	146,171	0												
Drysdale	429,530	319,401	392,632	730,606	0											
Excalibur	346,557	273,217	324,087	567,179	367,279	0										
Gladius	529,898	327,659	472,457	906,611	616,253	491,885	0									
H45	385,753	265,113	339,227	627,589	298,414	280,576	519,690	0								
Kukri	245,356	208,666	290,506	541,524	428,134	318,029	480,575	345,358	0							
Pastor	302,731	289,053	340,269	603,323	336,029	284,559	552,119	309,025	302,231	0						
RAC875	412,818	257,630	390,967	722,089	429,038	368,152	158,973	386,145	418,037	375,137	0					
VolcaniD DI	508,175	413,676	412,553	808,658	696,467	600,478	813,067	633,916	498,017	586,694	643,205	0				
Westonia	354,599	276,490	310,192	623,591	500,461	362,800	557,464	405,842	346,683	349,542	403,411	678,631	0			
Wyalkatc hem	525,289	341,043	433,228	800,300	560,759	327,888	386,213	449,614	436,777	442,941	235,924	800,137	505,345	0		
Xiaoyan 54	458,214	332,986	368,604	761,864	540,264	324,881	696,677	377,053	401,191	413,462	522,021	897,807	622,449	569,223	0	
Yitpi	544,440	328,216	468,743	968,088	690,017	548,694	233,539	587,310	530,687	580,060	287,648	951,537	654,967	444,084	844,785	0
	AC Barrie	Alsen	Baxter	Chara	Drysdale	Excalibur	Gladius	H45	Kukri	Pastor	RAC875	VolcaniD DI	Westonia	Wyalkatc hem	Xiaoyan 54	Yitpi

Table 3-2 Summary of single nucleotide polymorphism validation.

SNP ID	Forward Primer	Reverse Primer	Redundancy score	Chr.	SNP	Validation
UQ01TA7A495714	AGGTGTTGTTCTTCACCGT	AAGGATCTTGTGAAGTGGC	5	7A	A/G	True SNP
UQ01TA7A1381138	CATAACCGCTTCCTTGTT	ATCGGTAGACCTGCTTTG	36	7A	A/G	True SNP
UQ01TA7A781968	CAGATGAAGGCAGCAGTATG	TTTCGTGACTACATCCGTG	23	7A	G/A	True SNP
UQ01TA7A14292	TTCTTATGTCGTGTTGTGCC	AAAAAGGACACGAAGAGGAA	28	7A	C/T	True SNP
UQ01TA7A19199	CCTACGCTTAGCCACTGAT	CTCCCTTACAATGAACCAGC	4	7A	T/C	True SNP
UQ01TA7A19247	TAGGGATTTGCATGGATT	CCAACTTGTCGTCGTCATT	16	7A	A/G	True SNP
UQ01TA7A04421	GGCGAGCTGACAATAAGTT	TGTTTGCAAATGAATGCTT	13	7A	T/C	True SNP
UQ01TA7A16656	ACAACCTCAGGTGAGAGAGC	TTGCCTGTCATGTCGATTAC	5	7A	C/T	True SNP
UQ01TA7A30454	CCATCATCATTGGAACAGAA	GATCAGATGTGGAAGAACGCC	84	7A	C/T	True SNP
UQ01TA7B299842	TTTTATCAGGCTAGTGGGGA	TGTCGTTGTGTAGGTATCCG	2	7B	C/T	Failed
UQ01TA7B64149	GTTGCATTCATCTTCGACAA	TGCTGGATCTTGACTTGAA	25	7B	C/T	True SNP
UQ01TA7B588806	GTGCCAGTTTCCATAAC	GTGACGGACTTGGAGAACAC	12	7B	C/T	True SNP
UQ01TA7B1552504	GATGATCCTCGAAAAGGAAA	AAATAGTGGCCTTCATTCCA	4	7B	C/A	True SNP
UQ01TA7B1734345	TGCAAATGACATGCACATAA	TGCTAATGAGATGAAGAGCG	29	7B	G/A	True SNP
UQ01TA7B03509	AATGGGGATATTGTTCGTG	ATGTCCTGGAGCTTTTCAG	51	7B	A/G	True SNP
UQ01TA7B03075	TGGAATCATGTGATGTTGGT	GATATCCGTCCTCCATTCTG	50	7B	C/T	True SNP
UQ01TA7D349608	GAAAGAAGCGAATACCCAGA	GTCAAAC TGATCCCAAGGAG	14	7D	T/C	True SNP
UQ01TA7D523654	GGGCTAAAGAAATGGTCAA	CGAGATAATGCCAGAGGGA	20	7D	A/C	True SNP
UQ01TA7D19459	GCCAGTGGAGAAGAGTCAT	ACTTCCAGGTGTGTTTGGT	25	7D	T/C	True SNP
UQ01TA7D09646	CGTGCTGGATAACTGTCTTG	GATCCC GTTACCAAATGAC	22	7D	C/G	True SNP
UQ01TA7D05992	AGGGCAACATTGTCTTCAT	GCAAGCTACGACATTTGA	4	7D	G/T	True SNP
UQ01TA7D12121	GGTCAGTTCTTGATGGCT	CGAAGAGAGTATTTCCGC	25	7D	C/T	True SNP

3.4 Results

Whole genome Illumina paired read sequence data was generated from 16 Australian bread wheat varieties (Edwards et al., 2012b). After filtering to remove poor quality and clonal reads, a total of 13,642 million read pairs remained. Alignment of these read pairs to the wheat group 7 and 4AL chromosome assemblies (Berkman et al., 2011a, Berkman et al., 2012b, Hernandez et al., 2012, Berkman et al., 2013) using strict parameters, resulted in 3.05%, 3.76% and 3.43% of read pairs mapping uniquely to chromosomes 7A, 7B and 7D respectively. Each syntenic builds representing the majority of genes from the chromosome arm was produced by assembling chromosome arm-specific sequence data generated by Illumina short-read sequencing technology (Berkman et al., 2011a). SNP calling using the SGSautoSNP pipeline (Lorenc et al., 2012b) predicted a total of 4,018,311 inter-varietal SNPs.

The majority of SNPs were identified on contigs which do not form part of the syntenic builds and are predominantly within intergenic regions, and a substantially greater number of SNPs were predicted on chromosomes 7A and 7B, compared to 7D (Table 3-3). The SNP transition/transversion ratio (Tr/Tv) was determined for each of the three chromosomes. As can be seen in Figure 3-1, the Tr/Tv ratio for chromosomes 7A and 7B are both above 1.6; meanwhile, the Tr/Tv ratio for chromosomes 7D is approximately 1.0. In Figure 3-2, it can be seen that most of the Ts/Tv ratio bars are above 1 (indicated as blue line in the plots) for the 7A syntenic build and the 7B syntenic build; meanwhile, most of the Ts/Tv ratio bars are below 1 for the 7D syntenic build. In Table 3-4, it is shown that most of the Ts/Tv ratio values are greater than 1 for chromosome 7A and 7B; meanwhile, most of the Ts/Tv ratio values are less than 1 for chromosome 7D. Overall, the average Tr/Tv ratios within the A and the B genomes were found to be significantly higher than observed for the D genome (Figure 3-1; Figure 3-2; Table 3-4).

An inter-varietal SNP matrix was constructed representing SNPs between each pair of the 16 Australian wheat varieties (Table 3-1). SNPs between varieties varied from 146,171, between Chara and Baxter, to 968,088, between Chara and Yitpi. The average

number of SNPs between varieties was 465,278, and the majority of pairwise wheat combinations (117 out of 120) featured more than 200,000 SNPs. This matrix was used to produce a phylogenetic tree representing similarity between the 16 varieties (Figure 3-3). A total of 6 clades have been identified from this phylogenetic tree. Each clade represents a group. The 16 varieties can be grouped based on these 6 clades: group 1 consists of Alsen, AC Barrie and Kukri; group 2 consists of Chara, Baxter, and VolcaniDDI; group 3 consists of Westonia and Pastor; group 4 consists of Xiaoyan 54 and Excalibur; group 5 consists of Yitpi, Gladius, RAC875 and Wyalkatchem; group 6 consists of H45 and Drysdale.

Table 3-3 Subgenomic varietal SNP density for 16 Australian wheat varieties.

	Total		Syntenic Build	
	# SNPs	SNPs/Mb	# SNPs	SNPs/Mb
7A	1,486,040	4077	42,041	3212
7B	1,860,295	4737	38,508	3384
7D	671,976	1939	20,563	1088

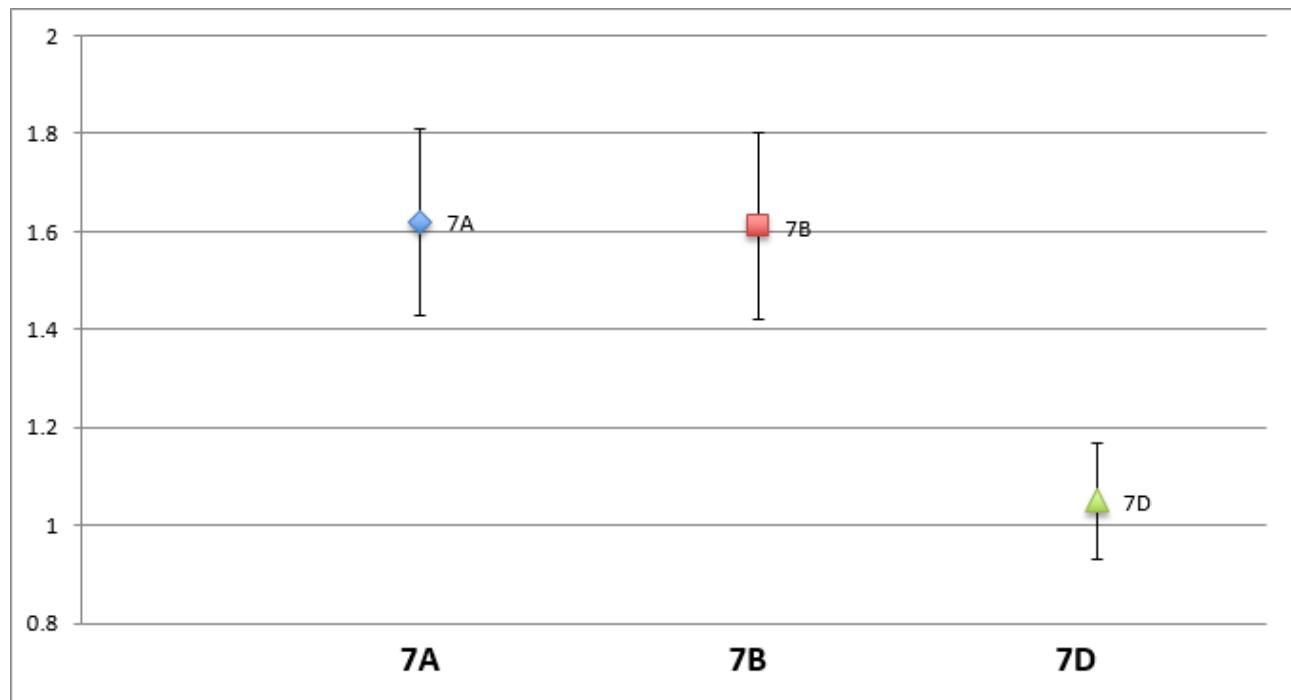


Figure 3-1 The transition/transversion ratio and standard deviation across chromosomes 7A, 7B and 7D.

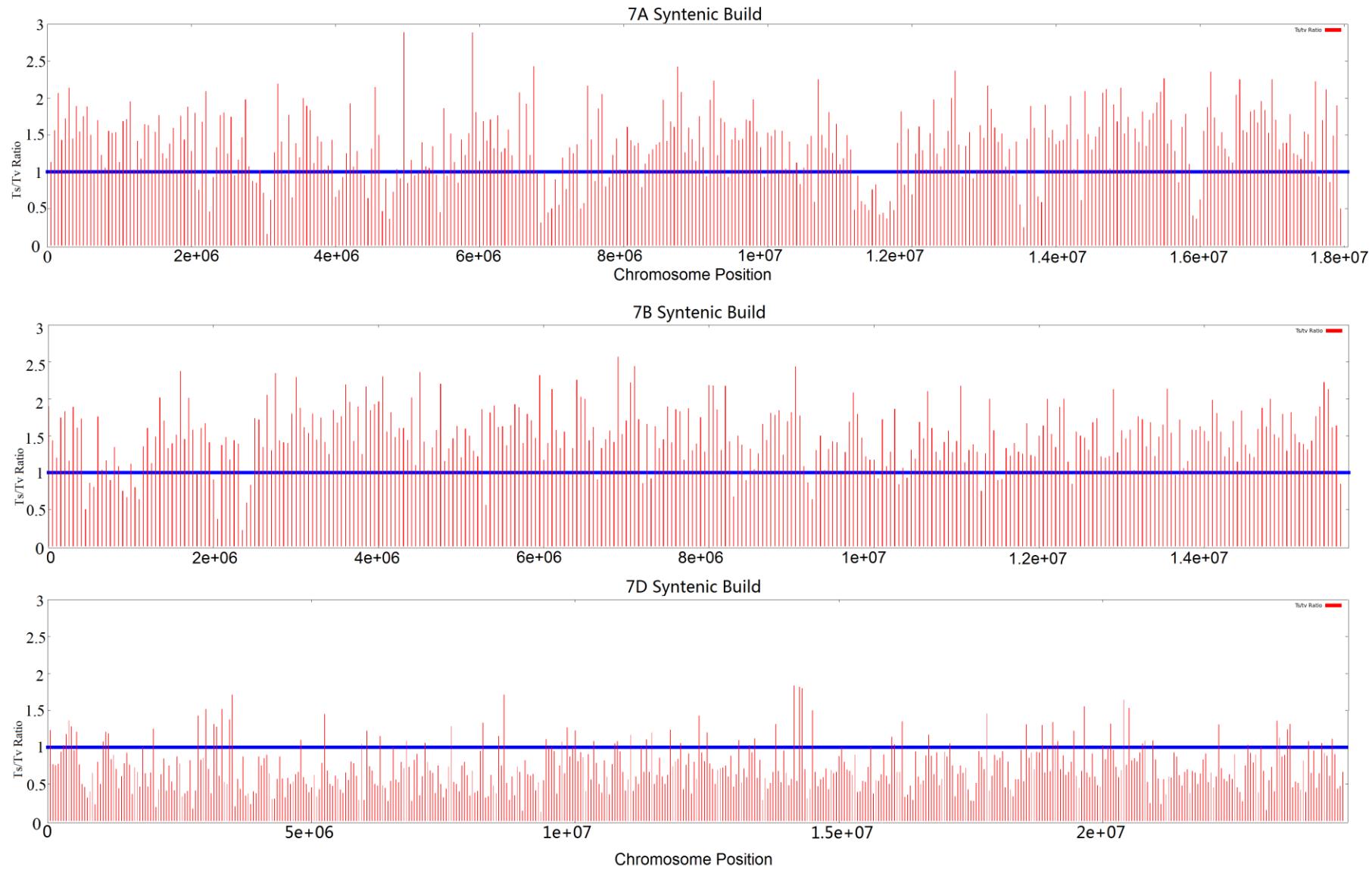


Figure 3-2 Ts/Tv ratio across the 7A, 7B and 7D syntenic builds. Each red column in the chart represents Ts/Tv ratio of SNPs within the corresponding region. The blue line represents ratio value of 1. Most of Ts/Tv ratios are higher than 1 for the 7A, 7B Syntenic builds while most of Ts/Tv ratios are lower than 1 for the 7D Syntenic build.

Table 3-4 Transition and transversion SNPs for each variety identified within the 7A, 7B and 7D chromosomes.

	7A chromosome			7B chromosome			7D chromosome		
Wheat Variety	Ts	Tv	Ts/Tv	Ts	Tv	Ts/Tv	Ts	Tv	Ts/Tv
AC Barrie	413,822	263,086	1.57	577,788	361,380	1.60	132,959	144,560	0.92
Alsen	290,867	175,106	1.66	407,378	244,576	1.67	92,245	99,470	0.93
Baxter	271,785	168,837	1.61	369,655	224,003	1.65	87,188	98,739	0.88
Chara	617,192	381,594	1.62	828,588	501,770	1.65	199,201	207,601	0.96
Drysdale	630,520	387,535	1.63	874,217	532,535	1.64	205,147	212,349	0.97
Excalibur	547,495	312,203	1.75	744,284	424,592	1.75	174,467	174,104	1.00
Gladius	622,903	379,410	1.64	872,024	530,411	1.64	195,376	203,878	0.96
H45	563,223	346,230	1.63	769,624	468,867	1.64	172,117	181,714	0.95
Kukri	440,364	250,765	1.76	601,675	340,593	1.77	146,880	143,791	1.02
Pastor	477,143	294,948	1.62	669,038	411,183	1.63	155,085	164,307	0.94
RAC875	480,457	273,311	1.76	701,676	397,320	1.77	164,922	162,738	1.01
VolcaniDDI	624,919	387,402	1.61	818,139	496,611	1.65	248,290	239,878	1.04
Westonia	432,855	288,880	1.50	523,797	349,238	1.50	198,943	192,550	1.03
Wyalkatchem	620,035	382,581	1.62	839,238	517,444	1.62	194,306	203,492	0.95
Xiaoyan 54	788,251	488,038	1.62	1,006,491	621,479	1.62	301,027	287,724	1.05
Yitpi	708,345	434,410	1.63	974,774	596,274	1.63	264,411	258,115	1.02

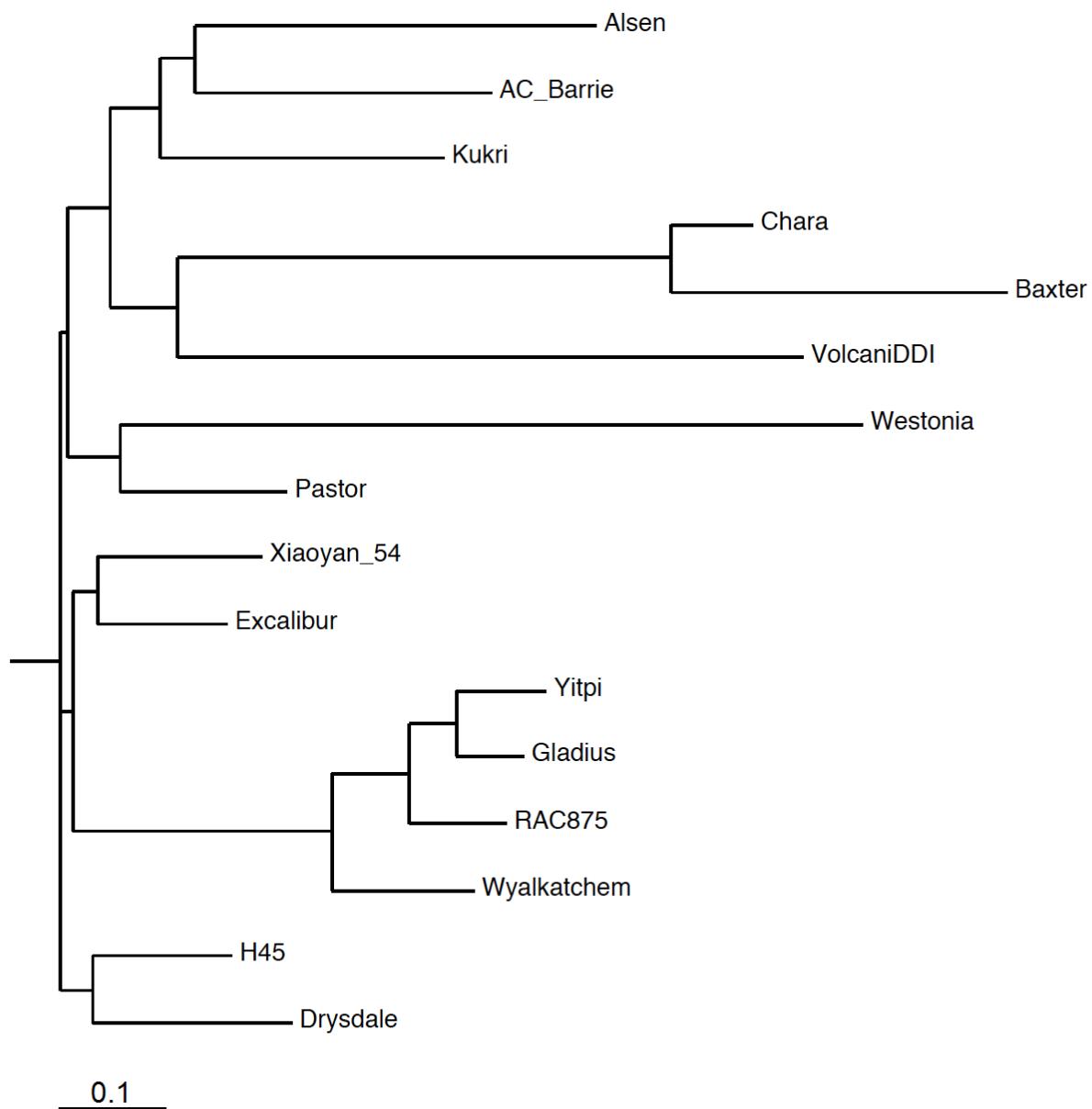


Figure 3-3 Phylogenetic relationships of 16 Australian wheat varieties based on SNP data obtained in this study. The scalar bar represents the number of base substitutions per 100 positions.

In addition to SNP density variation between the chromosomes, SNP density also varied along the lengths of chromosome syntenic builds (Figure 3-4). Normalised number of SNPs represents the raw number of SNPs divided by the coverage using a sliding window. To assess whether this variation is associated with selection for genes exhibiting specific characteristics, SNP density was calculated in regions 2 Kbp upstream and downstream of each predicted gene. A total of 146 genes were predicted to be in low SNP density regions, representing 40, 27 and 79 genes on the A, B and D genome respectively (Appendix 1). These genes include MADS box and Myb transcription factors, signal transduction pathway

genes, a sodium transporter, candidate disease resistance genes, an iron responsive transcription factor, hormone response genes, a potassium transporter, callose synthase, sucrose synthase and sugar transporters. These genes may reflect reduced diversity and selection for agronomic traits. In contrast, a total of 14 genes were predicted to be in high SNP density regions, representing 10, 3 and 1 gene(s) on the A, B and D genome respectively (Appendix 2). These genes include cellulose synthase, argonaute and ethylene response factors.

Twenty-two candidate SNPs were amplified by PCR, and Sanger sequenced to assess the false discovery rate associated with the approach used in this study. SNPs were chosen to represent all three of the group 7 chromosomes, including syntenic builds and unplaced contigs, and reflected a range of redundancy scores. Of the 22 SNPs, one assay failed to amplify a PCR product; of the 21 which amplified successfully, all were shown to be true inter-varietal polymorphisms (Table 3-2). The validated SNPs had an average redundancy score of 23.6 (range 4 - 84); in contrast, the SNP which failed to amplify had a redundancy score of 2.

The SNPs from the recently published wheat 90K infinium array (Wang et al., 2014) were compared with those predicted by SGSAutoSNP. A total of 850 SNPs were identified as having a match on the group 7 chromosomes at the same position as a predicted SNP in our study (Appendix 3). Of these, 482 (57%) were classified as polymorphic single locus, 316 (37%) as being polymorphic multi locus, while only 52 (6%) were monomorphic.

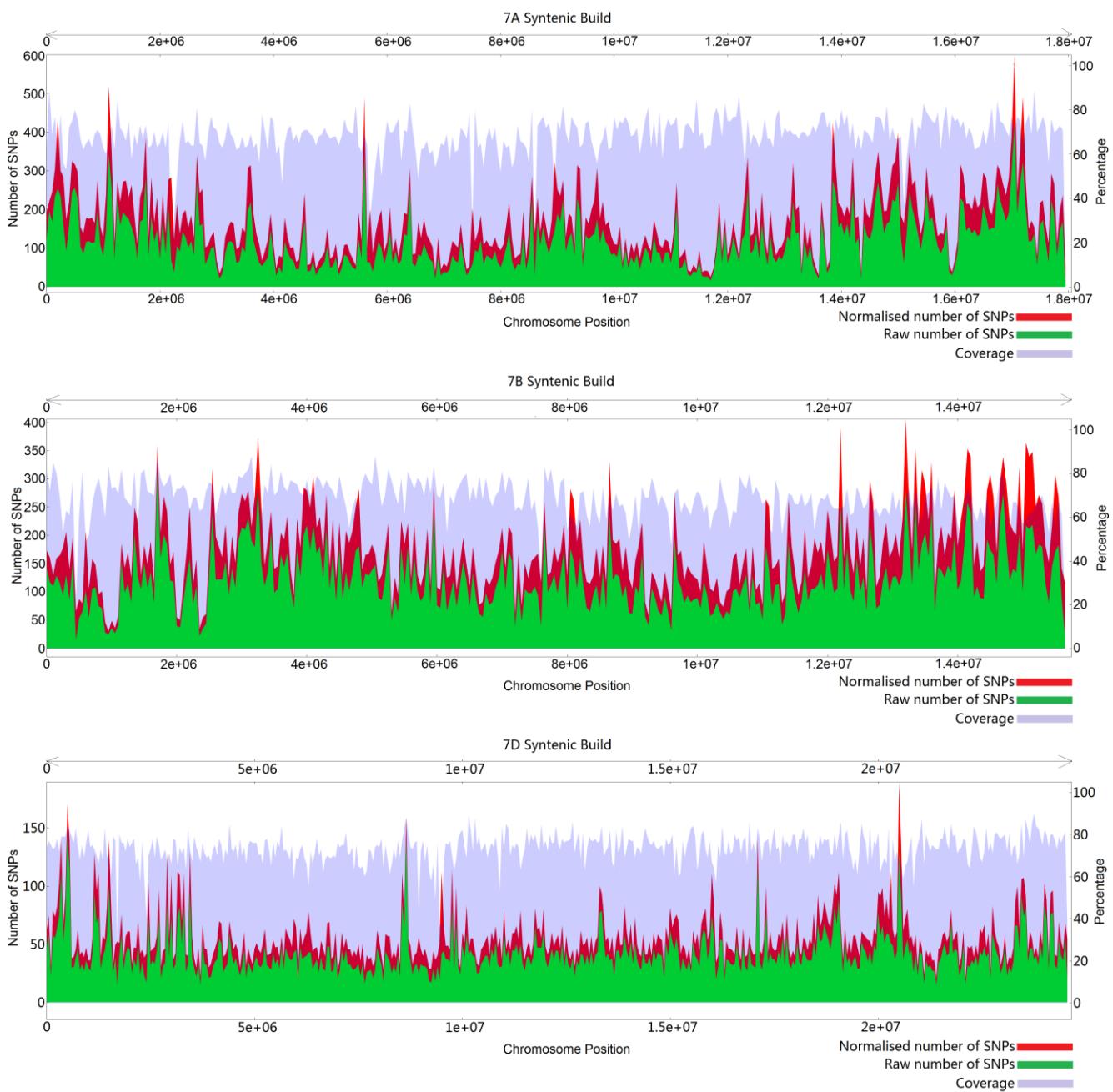


Figure 3-4 SNP density across the 7A, 7B and 7D syntenic builds.

3.5 Discussion

Domesticated crops are subject to human-mediated selection, aimed at developing high-yielding varieties adapted to local conditions. Genome-wide comparative diversity analysis has been performed in hexaploid bread wheat. Since its origin near the Caspian Sea around 8000 years ago (Nesbitt and Samuel, 1998), hexaploid bread wheat has demonstrated its adaptability to a wider range of environmental conditions and agricultural practices aimed at developing improved, high-yielding varieties (Dubcovsky and Dvorak, 2007). In order to identify regions of the wheat genome subject to selection during improvement, Cavanagh *et al.* developed a high-throughput array to interrogate 9,000 gene-associated SNPs in a worldwide sample of 2,994 accessions of hexaploid wheat, including landraces and modern cultivars. The majority of the selected alleles were identified at low frequency in local populations, suggesting either weak selection pressure or temporal variation in the targets of directional selection during breeding probably associated with changing agricultural practices or environmental conditions. Low genetic differentiation between landraces and modern cultivars suggests that selection during wheat breeding has not dramatically altered allele frequency genome-wide, but may have been accomplished by selection of a relatively limited number of loci (Cavanagh *et al.*, 2013).

I have identified more than four million candidate inter-varietal SNPs, across the group 7 chromosomes, between 16 Australian bread wheat varieties. This represents the greatest number of SNPs identified to date for this important crop. By re-sequencing 22 loci in different varieties we obtained a SNP validation rate of 95%, and comparison of SNPs with results from the recently published wheat infinium study (Wang *et al.*, 2014) show that 94% of SNPs identified in both studies were polymorphic. This compares to an overall polymorphism rate across the infinium assay of only 69% (Wang *et al.*, 2014). Our results are similar to the 93% we observed in a previous study examining four varieties (Lorenc *et al.*, 2012b). This is also similar to a study in the diploid D genome species of *Ae. tauschii*, which validated over 80% of predicted SNPs (You *et al.*, 2011), and the recent 96% validation of 10,251 putative co-dominant SNP markers in bread wheat (Allen *et al.*, 2013). This is significantly higher than the validation of SNPs on the recently produced 9k Illumina infinium array where 65% of

SNPs demonstrated accurate genotype calling (Wurschum et al., 2013), though this rises to 90% across more diverse germplasm and following manual data curation (Cavanagh et al., 2013). SGSAutoSNP identified SNPs with high average redundancy score from SOAP alignment assemblies, which improved the validation rate. In addition, the SNP which failed to amplify had a low redundancy score.

Phylogenetic analysis has been applied to many species, for example, *Brassica oleracea* and *Arabidopsis thaliana* (Town et al., 2006), lettuce lines (Kwon et al., 2013), peach varieties (Cao et al., 2014), wheat (Petersen et al., 2006) and many more. Substantial variation in pairwise SNP numbers between varieties was observed in this study (Table 3-1), with the greatest polymorphism identified between Chara and Yitpi, and least polymorphism identified between Chara and Baxter. Understanding the level of genomic diversity in populations can facilitate breeding and selection, ensuring that crosses lead to progeny with high levels of sequence diversity for the mapping of segregating traits. The phylogenetic tree produced based on pairwise SNP similarity (Figure 3-3) reflects the known breeding history of these varieties (Edwards et al., 2012b). For example, one clade has been demonstrated in previous studies, which is consistent with the fact that RAC-875 is present in Gladius pedigree (breeding) history (Edwards et al., 2012b). A total of 6 clades represent 6 groups of wheat varieties in this phylogenetic tree. The groups are associated with phenotypes. From group 1, Alsen and AC Barrie are both classified as Hard Red Spring Wheat (Alsen in US and AC Barrie in Canada) and high protein. In addition, Kukri has high protein. From group 2, Chara and Baxter are both classified as Australian Prime Hard (Chara in Southern NSW and Baxter in Queensland and Northern NSW) and high yielding. From group 5, Yitpi and Gladius are both classified as Australian Hard, mid season maturity and moderately susceptible to stripe rust and leaf rust; Yitpi and RAC875 are both good grain size; Gladius and Wyalkatchem are both high yielding. From group 6, H45 and Drysdale are both classified Australian Hard (NSW) and Australian Premium White (H45 in Western Australia and Victoria, Drysdale in Victoria). The associated traits for 16 Australian wheat varieties have been summarised in Table 3-5.

Table 3-5 Summary of the 16 Australian wheat varieties and associated traits

Variety	Variety Code	Classification
DRYSDALE	DRY	Classified as Australian Hard (NSW), Australian Premium White (Victoria)
GLADIUS	GLA	Classified as Australian Premium White (Western Australia), Australian Hard (South Australia, Victoria, NSW)
RAC-875	RAC	
EXCALIBUR	EXC	Classified as Australian Standard White
KUKRI	KUK	Classified as Australian Hard, premium quality white wheat
AC-BARRIE	ACB	Classified as Hard Red Spring Wheat (Canada)
BAXTER	BAX	Classified as Australian Premium Hard (Queensland and Northern NSW)
CHARA, AUS	CH7	Classified as Australian Prime Hard (Southern NSW), Australian Hard (South Australia and Victoria)
VOLCANI-DD-1	VOL	Breeding line from Israel
WESTONIA	WES	Classified as Australian Premium White
PASTOR	PAS	CIMMYT (International Maize and Wheat Improvement Centre, Mexico) cultivar
XIAOYAN-54	XIA	Chinese Winter wheat cultivar; Derived from the very successful cultivar Xiaoyan6
YITPI	YIT	Classified as Australian Hard; Mid to long season maturity
ALSEN, USA	ALS	Classified as Hard Red Spring Wheat (US)
WYALKATCHEM	WYA	Classified as Australian Premium White (Western Australia, South Australia and Southern NSW) and Australian Standard White (Victoria)
H-45	H45	Classified Australian Hard (NSW) and Australian Premium White (Western Australia and Victoria)

(Bioplatforms Australia Metadata, Wheat Cultivar Samples: https://ccgapps.com.au/bpa-metadata/wheat_cultivars/samples)

The majority of SNPs were identified outside of the syntenic builds. The syntenic builds reflect gene containing contigs which display similarity with genes from syntenic regions of related species and represent only 4% of the total assembly. SNP densities on chromosomes 7A, 7B, and 7D had 4077, 4737 and 1939 SNPs/Mb respectively (Table 3-3). This difference in SNP density is consistent with previous observations (Berkman et al., 2013, Chao et al., 2009) and

reflects the early evolutionary history of this crop. In an evolutionary event believed to have occurred near the Caspian Sea around 8000 years ago, tetraploid emmer wheat crossed with wild D genome progenitor *Ae. tauschii*, to form the hexaploid species *T. aestivum*, which became common wheat (bread wheat) (Giles and Brown, 2006, Nesbitt and Samuel, 1998, Salamini et al., 2002); a greater number of genes for domestication traits are found on the A and B genomes (Gegas et al., 2010), consistent with domestication of emmer wheat prior to the formation of the hexaploid. During the subsequent evolution of bread wheat, gene flow is predicted to have occurred between *T. aestivum* and its tetraploid progenitor *T. turgidum* (AuAuBB), however, no substantial gene flow is predicted to have occurred between the hexaploid and *Ae. tauschii* (DD) (Berkman et al., 2013, Caldwell et al., 2004, Dvorak et al., 2006, Talbert et al., 1998). This would be expected to result in a substantial increase in polymorphism on the A and B genomes relative to the D genome in modern cultivated wheat, consistent with patterns of SNP diversity identified in this study.

In addition to the variation detected between chromosomes, SNP density also varied across the lengths of the individual syntenic builds. Regions of low SNP density may reflect selection at loci associated with domestication or important agronomic traits, with a loss of diversity in and around genes which display favourable alleles. In contrast, genes in high SNP density regions may be associated with regions introgressed from related species. To assess this, genes within low and high SNP density regions were identified and analysed (Appendix 1; Appendix 2). Similar evidence of selection by Jordan et al. 2015 also was observed later after this finding as low polymorphism in certain regions of certain linkage groups. Allopolyploidy may have increased the possibility of beneficial allele recovery by spreading the set of potential selection targets (Jordan et al., 2015).

In addition, other potential possibilities for above variation in polymorphism we observed need to be considered. The distal regions of chromosomes have high levels of recombination and this is directly connected to elevated levels of polymorphism. This observation has been hypothesized to be associated with the need to generate increased genetic variation required for adaptation, for example, to disease resistance. Alternatively proximal regions indicate much lower rates of recombination and hence lower genetic polymorphism.

Recently, Cavanagh *et al.* found evidence for selection around a major “green revolution” dwarfing gene Rht-B1 (Cavanagh et al., 2013). The genes identified here in low SNP density regions are good candidates for further assessment to explore possible contributions to desirable characteristics of cultivated wheat. It appears likely that assessment of SNP density around genes, as performed in this study, will identify alleles selected during breeding, some of which could be targets for further crop improvement. In contrast to the 146 predicted genes identified in low SNP density regions, 14 genes were identified in high SNP density regions. These may reflect natural variation in SNP density across the genome or may have been introgressed from other diverse lines or species leading to regions of high polymorphism in this population. In low SNP density regions, important genes, which may reflect reduced diversity and selection for agronomic traits, are expected to be identified. These genes include MADS box and Myb transcription factors, signal transduction pathway genes, a sodium transporter, candidate disease resistance genes, an iron responsive transcription factor, hormone response genes, a potassium transporter, callose synthase, sucrose synthase and sugar transporters. MADS-box genes are recognised as important transcription factors for plant development, especially floral organogenesis (Alvarez-Buylla et al., 2000, Causier et al., 2002). The Myb family is very important in transcriptional control in higher plants because of the number of genes involved and because of their roles in the control of plant-specific processes (Martin and Paz-Ares, 1997). Sugar-transport proteins play an important role in the cell-to-cell and long-distance distribution of sugars throughout the plant (Williams et al., 2000). In contrast to low SNP density regions, in high SNP density regions, important genes, which may reflect reduced diversity and selection for agronomic traits, are rarely identified. For example, cellulose synthase-like protein D2 is product of cellulose synthase-like genes, whose amino acid sequences are related to the cellulose synthase (CesA) genes (Hazen et al., 2002). Cellulose is a main component of plant and most algal cell walls (Roberts et al., 2002).

Performing genomic selection (GS) in many breeding programs, especially for non-major crops, requires large-scale genotyping (Nakaya and Isobe, 2012). Habier *et al.* proposed to use a panel of evenly spaced low-density SNPs across the genome to estimate genome-assisted breeding values of selection candidates in pedigree populations by

tracking the effects of high-density SNP alleles within families based on the utilisation of cosegregation information (Habier et al., 2009). Iwata and Jannink imputed missing marker scores in a low-density genotyped panel by referencing a high-density panel in barley (Iwata and Jannink, 2011). Both approaches were based on a common concept of predicting the interval genotypes of a population using low-density allelic data. GS uses all genome-wide molecular marker data as predictors of performance and delivers more accurate predictions rather than searching to identify individual loci significantly associated with a trait (Jannink et al., 2010). Our analysis of genes within low and high SNP density regions could provide a resource for GS in wheat breeding programs, and also for GWAS and subsequent candidate gene identification. GS adopting SNP arrays is reasonably less expensive with a typical array (Poland et al., 2012). The high-throughput SNP array and diversity map will can provide a valuable resource for improving analysis of identification of genes targeted by selection, designing of high-power genome-wide association studies (GWAS) experiments, marker-assisted breeding, and genomic selection (GS) (Cavanagh et al., 2013).

The SNP transition/transversion ratio (Tr/Tv) has been proposed in a framework for variation discovery and genotyping using next-generation DNA sequencing data (DePristo et al., 2011). This ratio has been observed in species such as hexaploid wheat (Winfield et al., 2012), *Arabidopsis thaliana* (Belfield et al., 2012), *Brassica napus* (Bus et al., 2012) and *Setaria viridis* (Huang et al., 2014). The SNP transition/transversion ratio (Tr/Tv) was identified as 1.81 for a total of 511,439 putative SNPs from over 100,000 transcripts from eight different UK allohexaploid wheat varieties (Winfield et al., 2012). This transition/transversion ratio (1.81) from all genomes (Winfield et al., 2012) was slightly greater than the transition/transversion ratios (1.62, 1.61 and 1.05 on chromosome 7A, 7B and 7D respectively) from our study (Figure 3-1). An additional observation made in this study was that chromosomes 7A and 7B feature a higher SNP transition/transversion ratio (Tr/Tv) than chromosome 7D (Figure 3-1; Figure 3-2; Table 3-4). A relatively high frequency of transitions has been observed in many species and is thought to be predominantly due to the tendency of methyl cytosine to mutate to uracil, which is then corrected to thymine (Coulondre et al., 1978); transitions can thus be considered an ‘evolutionary footprint’ of methylation (Buckler and Holtsford, 1996). It has also previously been demonstrated that gene loss is greater in the A and B genomes than the D

genome (Berkman et al., 2013, Pont et al., 2013). Genome-wide methylation and associated gene silencing (Bottley et al., 2006, Charmet, 2011) is an immediate result of polyploidisation (Feldman and Levy, 2009). It may be that the higher Tr/Tv ratio and frequency of gene loss observed in the A and B genomes is a result of the additional polyploidy event involving these genomes compared to the D genome during the formation of hexaploid wheat.

Overall, this study has identified a large number of polymorphisms across the chromosome 7 homoeologues of hexaploid wheat amongst elite Australian varieties. This resource is publically available to assist additional genetic analysis and breeding. Furthermore, observed patterns of SNPs across the homoeologous group 7 chromosomes has provided insight into the molecular consequences of the evolution and selection that resulted in modern hexaploid wheat.

Chapter 4 WheatGenome.info: An integrated database and portal for wheat genome resources

4.1 Introduction

In recent years, the growth in genome information has led to a challenge for bioinformatics researchers to transform the vast quantities of data being produced into collective knowledge. As sequence availability has increased, data access, representation, analysis and visualisation present significant challenges (Ning and Montgomery, 2010). In this context, online databases for genetic and genomic data are very much in demand.

This chapter describes an integrated wheat genome data resource, WheatGenome.info, which provides a variety of web-based systems for access to wheat genetic and genomic data to support crop research and applied crop improvement. Moreover, this interface includes links to wheat related web-based data hosted at other research organisations. WheatGenome.info is available at <http://www.wheatgenome.info/>.

4.2 Materials and methods

4.2.1 Wheat group 7 sequence and 4AL

Despite the significance of wheat for resolving food shortage globally, its complex and large genome impedes efforts in genome sequencing. Berkman *et al.* have assembled genomic regions representing unique and low copy regions for isolated chromosome arms. These genomic regions include syntetic builds for chromosomes 7A, 7B and 7D (Berkman *et al.*, 2011a, Berkman *et al.*, 2012b, Berkman *et al.*, 2013). The syntetic builds represent gene containing contigs which demonstrate similarity with genes from related species, and represent around 4% of the total assembly. These genomic regions, containing all, or nearly all, genes for these chromosomes, have been assembled, and the majority of these genes have been ordered and aligned based on synteny with *B. distachyon*, *O. sativa*, and *S. bicolor*

(Berkman et al., 2013). The majority of wheat contigs, which are outside of the syntenic builds, are also included in the WheatGenome.info database.

4.2.2 *Brachypodium distachyon* gene and exon annotation

Grasses not only provide the bulk of nutrition for humans, but also produce a source of sustainable energy (Somerville, 2006). *Brachypodium* is a member of the Pooideae subfamily. The diploid ecotype of *Brachypodium distachyon* has the smallest reported genome size in the Poaceae (Draper et al., 2001) and this was the first member of the Pooideae subfamily to be sequenced (published by The International Brachypodium Initiative). *B. distachyon* is an important model for developing new energy and food crops (International Brachypodium Initiative, 2010). *Brachypodium distachyon* gene and exon annotation were completed by Paul Berkman.

4.2.3 Uniref90 annotation

Uniref90 annotation was completed by other group members. Clustering of protein sequence space based on sequence similarity provides information for reducing overrepresentation of sequences. The UniRef (UniProt Reference Clusters) provide clustered sets of sequences from UniProt Knowledgebase (UniprotKB) and selected UniProt Archive records. UniRef90 are constructed by clustering UniRef100 sequences at the 90% sequence identity levels (Suzek et al., 2007). The UniRef90 database was downloaded from the UniProt website, and the wheat chromosomes and extra contigs were compared with the UniRef90 database using BLAST. Uniref90 protein annotation can provide biological information related to specific genes or traits. The Uniref90 annotation was performed by Berkman *et al.* (Berkman et al., 2013).

4.2.4 Inter-varietal SNPs between 16 Australian wheat varieties

To obtain a greater understanding of wheat genome diversity, I have identified inter-varietal single nucleotide polymorphisms (SNPs) between 16 Australian bread wheat varieties, which include AC Barrie, Alsen, Baxter, Chara, Drysdale, Excalibur, Gladius, H45, Kukri, Pastor, RAC875, VolcaniDDI, Westonia, Wyalkatchem, Xiaoyan 54 and Yitpi (Chapter 3). These 16 Australian bread wheat varieties were sequenced using whole genome shotgun Illumina paired read sequencing by Bioplatform Australia. I have aligned these paired reads to the draft assemblies of chromosomes 7A, 7B and 7D. As described in the previous chapter, total of over 4,018,311 inter-varietal SNPs have been identified between these 16 Australian wheat varieties (Lai et al., 2014).

4.2.5 Workflow

The WheatGenome.info integrated database and portal can be split into distinct sections: The generative pipeline, the data storage component and the visual interface (Figure 4-1). In pipeline processing, BLAST tools were used for comparison with genome references and annotation. Annotated genetic maps were generated in CMAP. All of the processed data is stored in a MySQL relational database using customised schema for data storage. Users interact with each web-based system through their portals.

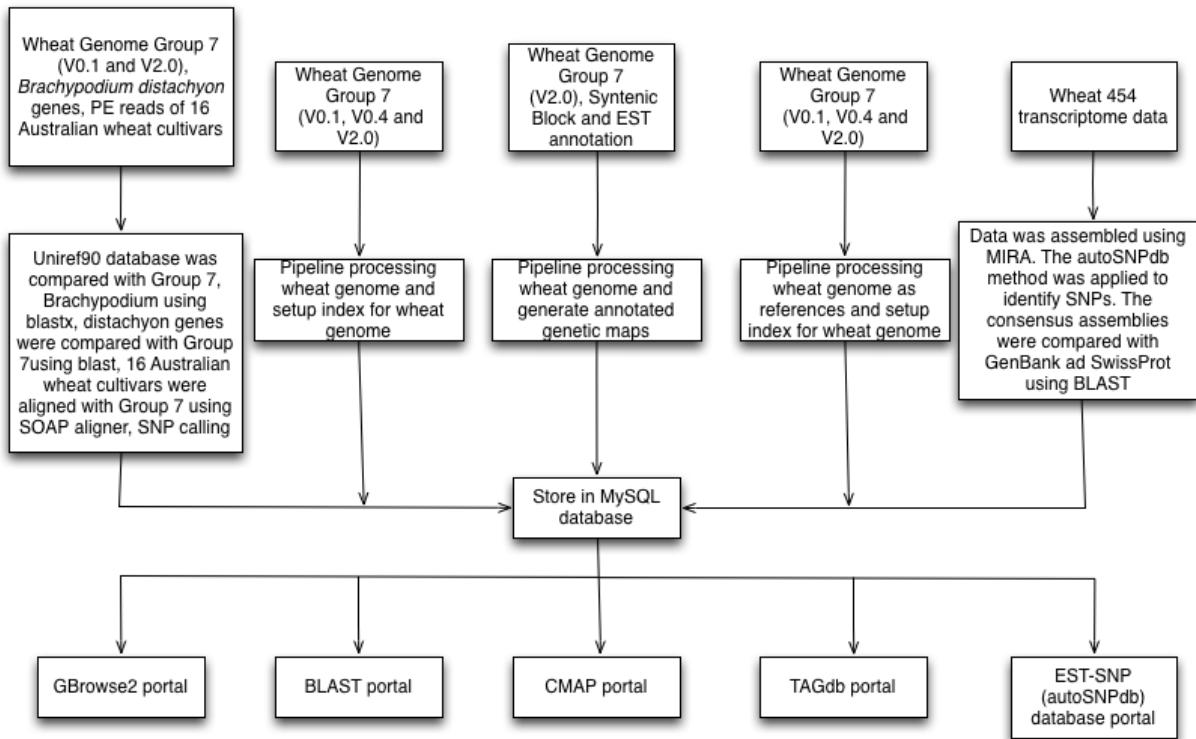


Figure 4-1 Flowchart demonstrates the general workflow of the WheatGenome.info integrated database.

4.3 Database contents

The wheatGenome.info database integrates several main web-based systems. These include an annotated wheat genome viewer based on GBrowse2, searchable using keywords, genome location or by sequence similarity using the BLAST portal; a CMap genetic and physical mapping database; TAGdb for searching wheat short read sequences; an annotated wheat EST-SNP (autoSNPdb) database; and a wheat genome Wiki.

4.3.1 A wheat genome viewer for annotated chromosome arm assemblies

The application of next generation sequencing technology and advanced bioinformatics tools has enabled the assembly and annotation of the genes and low copy regions of isolated wheat chromosome arms, producing syntenic builds containing the majority of wheat genes (Berkman et al., 2011a, Berkman et al., 2012b, Berkman et al., 2013). Assemblies and syntenic builds for each of the group seven chromosome arms are hosted in a Gbrowse2 database. Gbrowse2 is a user-friendly generic genome browser for genome sequence data

and annotation (Arnaoudova et al., 2009, Donlin, 2007). Each wheat chromosome arm has been annotated with predicted genes, Uniref90 gene similarities, as well as homoeologue specific single nucleotide polymorphisms (SNPs).

As well as annotation keyword searches, a BLAST portal enables sequence similarity searches of assembled wheat chromosome arm data, with results displayed in the GBrowse2 viewer. DNA or protein query sequence can be uploaded or pasted in the web based form in FASTA format. The results are displayed in three sliding windows: the Overview window, Region window and Details window. The reference view can be dragged and zoomed. Several tracks of annotation are available, including Uniref90, Genes, Contigs, SNPs and Exons. All of these features can be expanded by clicking the associated plus button, and each feature provides a link to show the feature details (Figure 4-2).

The GBrowse database enables the rapid dissemination of wheat chromosome arm sequence information prior to publication. In the absence of a finished wheat reference genome upon which to base crop improvement efforts, this tool represents the first opportunity for wheat researchers to interact with chromosome-scale gene-based sequence scaffolds in an intuitive and user-friendly manner. It allows for a more rigorous interrogation of genes surrounding a locus of interest than was previously possible in wheat, to assist the identification of the genomic basis of important traits. With the expansion of wheat genome sequencing activities by several groups internationally, this resource will increasingly provide access to wheat genome information for crop improvement research. Group 7 references are currently available online. Recently, IWGSC has published the draft sequence of the hexaploid bread wheat genome. This draft sequence would be used as reference for annotation and SNP identification, and results would be released online once completed in the future.

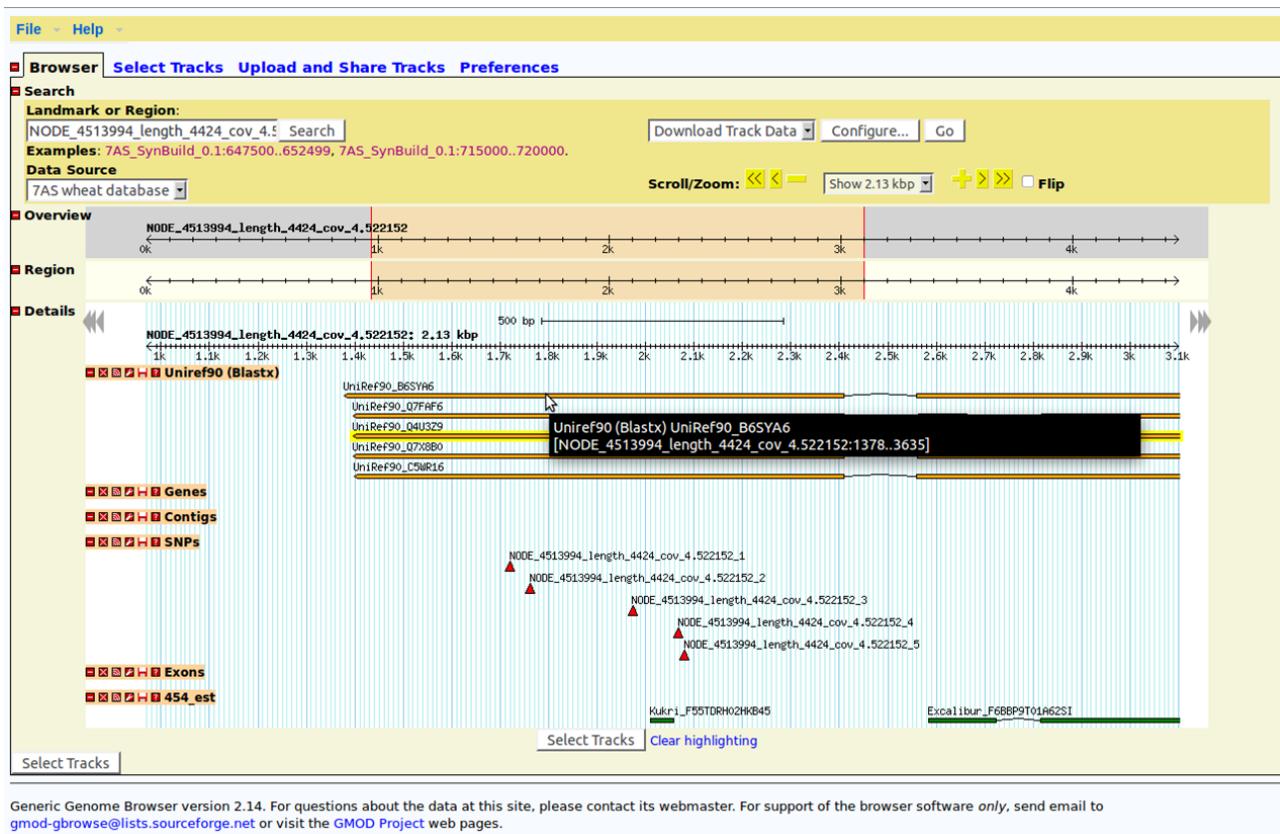


Figure 4-2 Example of the detailed information for the wheat genome 7AS syntetic build from the reference view of GBrowse2. Several tracks of annotation are available, including UniRef90, Genes, Contigs, SNPs and Exons.

4.3.2 Searching wheat sequence data using TAGdb

TAGdb is an online database system designed to identify and visualise next generation paired sequence tags that share identity with a submitted query sequence (Marshall et al., 2010). The TAGdb interface requests a FASTA format query sequence, of up to 5000 bp, as well as a contact email address, so users can retrieve previous query searches. Users can then select a variety of wheat short read data libraries. After starting the process, TAGdb sends an email to the user acknowledging that the job started successfully and provides a link to the results web page. Once the search is complete, TAGdb sends a second email to confirm completion, together with a link to the results. Two windows display an overview and zoomed region of the read alignments (Figure 4-3), paired reads are connected by a line, with a blue rectangle confirming that the result conforms to the expected orientation and paired read distance.

Matching reads, together with their matching or non-matching read pairs are viewed as a table or can be downloaded as a multi-FASTA format file for further analysis.

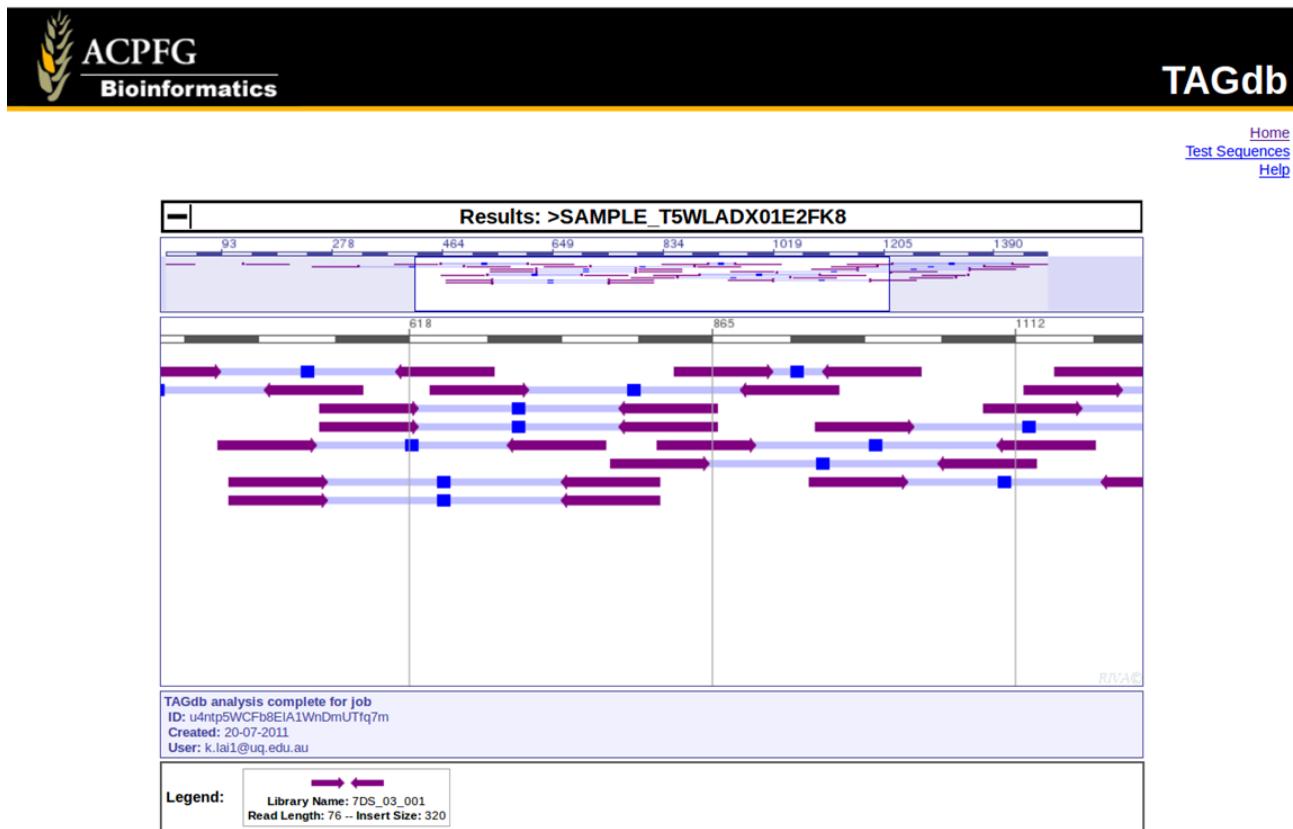


Figure 4-3 Screenshot of TAGdb showing the alignment of short reads from wheat variety Chinese Spring to a sample query sequence.

The key value of this tool is that it provides researchers with rapid, yet simple, access to the wheat genome sequence data being produced by new sequencing technologies. The identification of a large number of matching reads may enable the local assembly of the wheat genomic region. Where fewer reads are identified, read pairs may be used to design PCR primers to amplify and sequence the gene, as well as genomic sequence flanking the matching query. Wheat TAGdb currently hosts whole genome paired-read libraries of wheat varieties, AC Barrie, Alsen, Baxter, Chara, Chinese Spring, Drysdale, Excalibur, Gladius, H45, Kukri, Pastor, RAC875, VolcaniDDI, Westonia, Wyalkatchem, Xiaoyan 54, Yitpi, including specific data for the long and short arms of isolated chromosomes. Read lengths vary between 35 bp and 100 bp, with a range of insert sizes from 300 bp to 3700 bp. Additional wheat short read data for different wheat varieties will be hosted on TAGdb as it becomes publicly available in the future.

4.3.3 Comparative wheat genome and genetic maps on CMap and CMap3D

CMap is a generic, extensible web-based comparative map viewer for displaying and comparing genetic and physical maps from any species (Youens-Clark et al., 2009). There are two main CMap databases of interest to wheat researchers. The most comprehensive is hosted within GrainGenes (Carollo et al., 2005, Matthews et al., 2003) and is linked from the wheatgenome.info front page. The wheatgenome.info installation of the CMap system aims to specifically link the assembled wheat chromosome arm information with the sequenced genomes of *Brachypodium distachyon* and rice, as well as a genetic map of the D genome donor of hexaploid wheat, *Aegilops tauschii*. Bread wheat genome data includes syntetic builds for chromosome arms 7DS and 7BS, with other chromosomes being added as an ongoing process.

A CMap summary interface provides links to the CMap viewer, administration, tutorial document, map search and feature search functionalities. When a main reference sequence is selected, users can add a second physical sequence or genetic map. As genetic and physical maps become more abundant, their effective visualisation becomes a challenge. CMap3D is a tool developed based on CMap for the visualisation and comparison of multiple genome or genetic maps. This software is a stand-alone client and available for Windows, OSX and Linux (Duran et al., 2010a). The comparative maps present each corresponding marker and the links between maps as a three-dimensional view (Figure 4-4). CMap3D overcomes the limitation of comparing multiple adjacent aligned maps and provides a more user friendly comparison of multiple genomes or genetic maps in three-dimensional space.

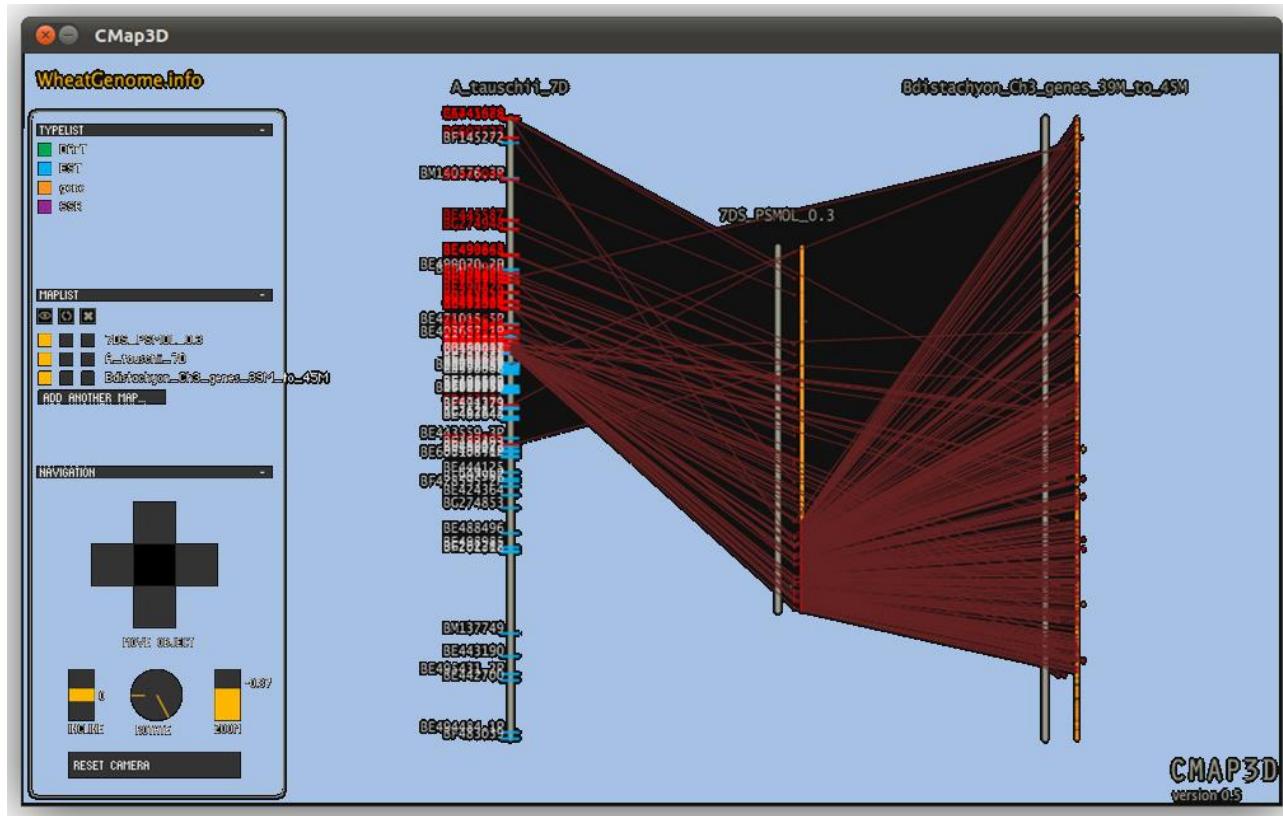


Figure 4-4 An inter-species comparison between a physical map of the wheat 7DS syntentic build chromosome, a genetic map of *Aegilops tauschii* chromosome 7D and a physical map of the *Brachypodium distachyon* chromosome 3 (between 39 and 45Mbp) using CMap3D.

4.3.4 Annotated wheat EST single nucleotide polymorphisms within autoSNPdb

Advances in second generation sequencing technologies have greatly increased the scale and scope to interrogate genomes and uncover genetic variation. However, differentiating between sequence errors and true SNPs remains a challenge, particularly for large and complex genomes such as wheat (Imelfort et al., 2009, Duran et al., 2009d). An approach to improve polymorphism prediction accuracy includes deep sequencing and multiple measures of prediction confidence.

AutoSNPdb (Duran et al., 2009a, Duran et al., 2009c) is a version of SNP discovery software which started with autoSNP (Barker et al., 2003, Batley et al., 2003) and includes SNPServer (Savage et al., 2005) and SGSAutoSNP (Lorenc et al., 2012b). It provides an extensible and user-friendly graphical interface facilitating a variety of queries to identify SNP polymorphisms related to specific genes or traits. This application processes multiple consensus sequences

from Expressed Sequence Tag (EST) reads to identify candidate SNPs using a series of Perl scripts.

The current application hosts data for important crops including rice, barley, Brassica and wheat (Duran et al., 2010b). Within wheat autoSNPdb, the accuracy of polymorphism detection has been improved by adopting the strategy of deep coverage sequencing of specific wheat cultivars. Wheat ESTs generated by Roche 454 second generation sequencing have been assembled using MIRA, with the resulting assembly processed using autoSNPdb Perl scripts to identify SNPs. Wheat autoSNPdb provides a valuable resource of annotated genetic markers of wheat, which can be used for genetic diversity analysis, cultivar identification, and high-resolution genetic map construction (Lai et al., 2012b) (Chapter 2).

Wheat autoSNPdb can be searched using keywords, similarity to a query sequence or by selecting SNPs which differentiate between varieties. A list of consensus contigs is displayed which includes the consensus sequence with aligned reads and highlighted SNPs (Figure 4-5). Full annotation of potential gene function is also displayed, and SNPs can also be searched based on homologous locations in the rice genome. AutoSNPdb is recommended to be viewed by using Mozilla Firefox as Internet Explorer may not provide full functionality.

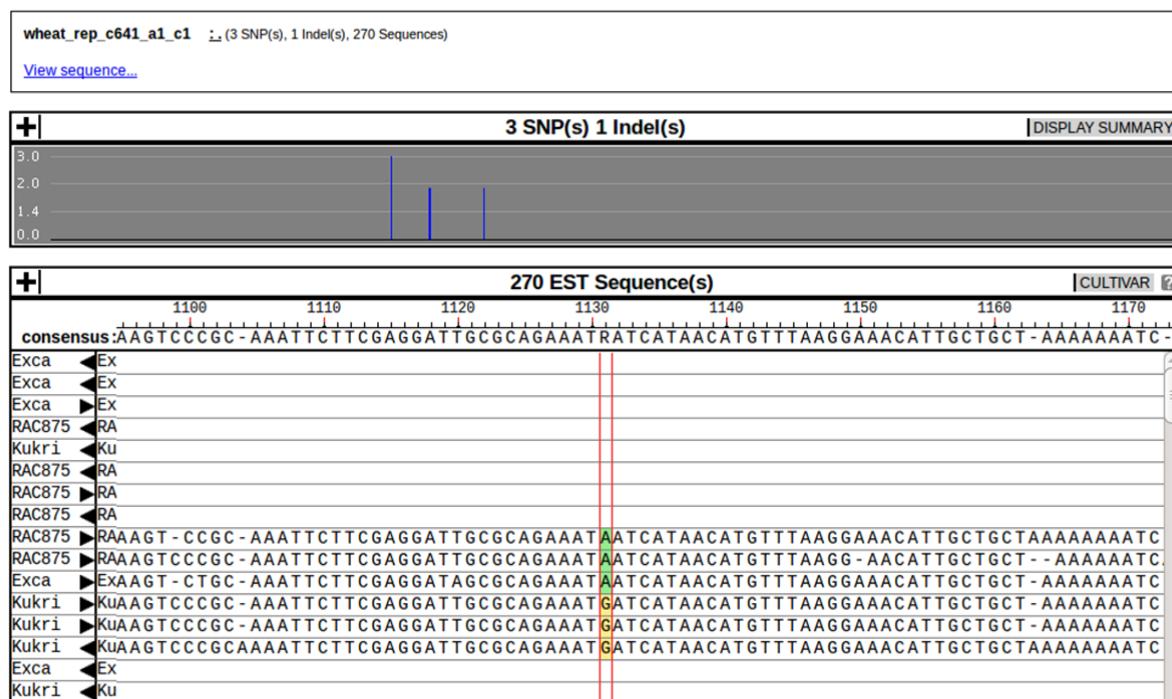


Figure 4-5 The wheat autoSNPdb web interface displaying the wheat sequence assembly, with predicted SNPs as vertical bars.

4.3.5 Wheat genome Wiki

The Wiki hosted at wheatgenome.info aims to assist communication between international groups undertaking diverse wheat sequencing activities. The wiki is based on the popular free web-based wiki software application from MediaWiki (<http://www.mediawiki.org>) which is also used by Wikipedia. This Wiki can provide an economic and efficient way to communicate and collaborate, and any research group which is undertaking wheat genome sequencing is welcome to describe their activities on the wiki, with secure access provided on request.

4.4 Conclusions and future direction

The wheatgenome.info system hosts a range of wheat genome information with unrestricted public access. Wheat genome sequencing is still in its infancy, and a complete high quality genome sequence is not expected until 2017, at the earliest. This system is maintained by David Edwards research group. This group would continue seek funding from the Australian Research Council (ARC), with server hosting at the University of Western Australia (UWA). Meanwhile, the number and quality of draft genome assemblies is likely to increase, together with an increasing amount of genome information relating to different wheat cultivars and wild relatives. The wheatgenome.info resource provides researchers with early access to this genetic and genomic data allowing them to compare query sequences with genomic data, identify genes at loci of interest, extract new genetic marker information, distinguish between homoeologous and varietal SNP markers, and access a hub for discussion on wheat genome sequencing activities beyond the current scope of the international consortium. The collation of this information within one place, together with links to external wheat genome resources greatly facilitates researchers who wish to use this information to improve this valuable crop.

Chapter 5 Concluding remarks and future directions

5.1 Concluding remarks

Sequencing data, including 454 transcriptome reads and Illumina whole genome re-sequencing paired-end reads from Australian wheat varieties have been produced and offered as a valuable resource for genetic variation characterisation.

Single nucleotide polymorphisms (SNPs) are the most abundant type of molecular genetic marker and can be used for producing high-resolution genetic maps, marker-trait association studies and marker assisted breeding. Large polyploid genomes such as wheat present a challenge for SNP discovery due to the potential presence of multiple homoeologues for each gene. AutoSNPdb has been successfully applied to identify SNPs from Sanger sequence data for several species, including barley, rice and Brassica, but the volume of data required to accurately call SNPs in the complex genome of wheat has prevented its application to this important crop. DNA sequencing has been revolutionised by the introduction of next generation sequencing technology, and it is now possible to generate several million sequence reads in a timely and cost effective manner. Chapter 2 describes the result of SNP discovery from the wheat transcriptome. Matthew J. Hayden has produced wheat transcriptome sequence data using Roche 454 technology. I applied these wheat transcriptome sequence data for SNP discovery using a modified autoSNPdb method, which integrates SNP and gene annotation information with a graphical viewer. A total of 4,694,141 sequence reads from these three bread wheat varieties were assembled to identify a total of 38,928 candidate SNPs. Each SNP is within an assembly complete with annotation, enabling the selection of polymorphism within genes of interest (Lai et al., 2012b).

Chapter 3 describes the result of SNP discovery using Illumina paired read sequence data from 16 Australian bread wheat varieties. Whole genome shotgun Illumina paired read

sequence data was mapped to the draft assemblies of chromosomes 7A, 7B and 7D to identify more than 4 million inter-varietal SNPs. SNP density varied between the three genomes, with much greater density observed on the A and B genomes than the D genome. This variation may be a result of substantial gene flow from the tetraploid *Triticum turgidum*, which possesses A and B genomes, to the ABD hexaploid *Triticum aestivum* during early co-cultivation of tetraploid and hexaploid wheat. In addition, I examined SNP density variation along the chromosome syntenic builds and identified genes in low SNP density regions which may have been selected during domestication and breeding. This study highlights the impact of evolution and breeding on the bread wheat genome and provides a substantial resource for trait association and crop improvement (Lai et al., 2014).

The SNPs identified from both 454 transcriptome reads, from 3 varieties, and the Illumina whole genome re-sequencing paired-end reads from 16 Australian wheat varieties provides a valuable source of annotated genetic markers in wheat for high-resolution genetic map development and trait association. The number of SNPs from 16 Australian wheat varieties are showing “B > A > D” pattern, which matches the result from the International Wheat Genome Sequencing (IWGSC) (International Wheat Genome Sequencing Consortium, 2014). The genetic map and trait association information will help researchers to identify and implement genotyping solutions, and help the breeders to implement novel breeding strategies and breed advanced wheat cultivars.

The International Wheat Genome Sequencing Consortium (IWGSC) has produced an ordered draft sequence of the 17-gigabase hexaploid bread wheat genome by sequencing isolated chromosome arms. Gene loci have been annotated, and comparative gene analysis of wheat subgenomes and extant diploid and tetraploid wheat relatives has been completed (International Wheat Genome Sequencing Consortium, 2014).

Wheat crop improvement has been limited due to its large and complex genome. Advances in genomics are supporting this improvement. We provide a variety of web-based systems hosting wheat genome and genomic data to support wheat research and crop improvement. Chapter 4 describes the integrated database WheatGenome.info and the data available within this system. WheatGenome.info hosts several web-based applications. These include a GBrowse2 based wheat genome viewer with a BLAST sequence search portal, TAGdb for

searching wheat next generation genome sequence data, wheat autoSNPdb, links to wheat genetic maps using CMap and CMap3D, and a wheat genome Wiki to allow interaction between diverse wheat genome sequencing activities. This system also includes links to a variety of wheat genome resources hosted at other research organisations. This integrated database aims to support wheat genome research and is freely accessible via the web interface at <http://www.wheatgenome.info/>.

5.2 Future directions

Wheat improvement partially depends on the use of molecular markers to improve selection efficiencies and to support the precise transfer of genes and QTL between different genetic backgrounds (International Wheat Genome Sequencing Consortium, 2014). The predicted SNPs between 16 Australian wheat varieties provide a resource for genotyping and marker trait association studies. Breeders could use this information to design and implement novel breeding strategies.

Low coverage genome wide Illumina paired-end reads could be mapped to the references to genotype the predicted SNP positions. Meanwhile, missing positions could be imputed using known haplotype block information.

The shotgun sequencing of the remaining chromosome arms 1-6 has recently been completed (International Wheat Genome Sequencing Consortium, 2014) and provides the latest complete reference for mapping of current Illumina paired read sequence data to identify new SNPs. These Illumina paired read sequence data from 16 varieties could be mapped onto these wheat references to identify genome wide SNPs.

The NCBI ESTs and 454 transcriptomes of bread wheat, from our homoeologous gene expression analysis, could be mapped on the whole ordered draft sequence of bread wheat genome that was produced by IWGSC (International Wheat Genome Sequencing Consortium, 2014). The identification of differential homoeologous gene expression from this analysis could be used to search for evidence for genome-wide transcriptional dominance of

an individual subgenome, which was not identified by IWGSC (International Wheat Genome Sequencing Consortium, 2014).

The analyses from my PhD research deals with the identification of extensive genetic variation. The increased number and quality of draft genome assemblies, combined with increasing amounts of genome information, including newly identified SNPs and more homoeologous gene expression, relating to different wheat cultivars and wild relatives could be loaded into WheatGenome.info resources (Lai et al., 2012a). These resources could support future population genomic and genome-wide association studies, rapid genetic marker development, identification of genes underlying important agronomic traits, improvement of agronomical and industrial traits, and support wheat breeding to satisfy the increasing demand of food worldwide.

When the BAC by BAC data of hexaploid bread wheat becomes available, an ultra high-resolution genetic map can be developed. These BAC libraries of bread wheat should be suitable for map-based cloning of wheat genes, physical mapping of the wheat genome, and cloning genes for important traits of interest in hexaploid wheat.

References

- ABS 2006. The Australian Wheat Industry. *Year of Australia, Australian Bureau of Statistics*.
- ABS 2012. Agricultural Production. *Year Book Australia, Australian Bureau of Statistics*.
- ADAMS, K. L. 2007. Evolution of Duplicate Gene Expression in Polyploid and Hybrid Plants. *Journal of Heredity*, 98, 136-141.
- AGARWAL, M., SHRIVASTAVA, N. & PADH, H. 2008. Advances in molecular marker techniques and their applications in plant sciences. *Plant Cell Reports*, 27, 617-631.
- AKHUNOV, E., NICOLET, C. & DVORAK, J. 2009. Single nucleotide polymorphism genotyping in polyploid wheat with the Illumina GoldenGate assay. *Theoretical and Applied Genetics*, 119, 507-17.
- ALLEN, A. M., BARKER, G. L., BERRY, S. T., COGHILL, J. A., GWILLIAM, R., KIRBY, S., ROBINSON, P., BRENCHLEY, R. C., D'AMORE, R., MCKENZIE, N., WAITE, D., HALL, A., BEVAN, M., HALL, N. & EDWARDS, K. J. 2011. Transcript-specific, single-nucleotide polymorphism discovery and linkage analysis in hexaploid bread wheat (*Triticum aestivum* L.). *Plant Biotechnology Journal*, 9, 1086-99.
- ALLEN, A. M., BARKER, G. L., WILKINSON, P., BURRIDGE, A., WINFIELD, M., COGHILL, J., UAUY, C., GRIFFITHS, S., JACK, P., BERRY, S., WERNER, P., MELICHAR, J. P., MCDOUGALL, J., GWILLIAM, R., ROBINSON, P. & EDWARDS, K. J. 2013. Discovery and development of exome-based, co-dominant single nucleotide polymorphism markers in hexaploid wheat (*Triticum aestivum* L.). *Plant Biotechnology Journal*, 11, 279-95.
- ALTSCHUL, S., GISH, W., MILLER, W., MYERS, E. & LIPMAN, D. 1990. Basic local alignment search tool. *Journal of Molecular Biology*, 215, 403 - 410.
- ALTSCHUL, S. F., MADDEN, T. L., SCHÄFFER, A. A., ZHANG, J., ZHANG, Z., MILLER, W. & LIPMAN, D. J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, 25, 3389-3402.
- ALVAREZ-BUYLLA, E. R., PELAZ, S., LILJEGREN, S. J., GOLD, S. E., BURGEFF, C., DITTA, G. S., DE POUPLANA, L. R., MARTINEZ-CASTILLA, L. & YANOFSKY, M. F. 2000. An ancestral MADS-box gene duplication occurred before the divergence of plants and animals. *Proceedings of the National Academy of Sciences of the United States of America*, 97, 5328-5333.
- ANSORGE, W. 1991. *Process for sequencing nucleic acids without gel sieving media on solid support and DNA chips*. German patent application.
- ANSORGE, W., SPROAT, B., STEGEMANN, J., SCHWAGER, C. & ZENKE, M. 1987. Automated DNA sequencing: ultrasensitive detection of fluorescent bands during electrophoresis. *Nucleic Acids Research*, 15, 4593-602.
- ANSORGE, W., SPROAT, B. S., STEGEMANN, J. & SCHWAGER, C. 1986. A non-radioactive automated method for DNA sequence determination. *Journal of Biochemical and Biophysical Methods*, 13, 315-23.
- ANSORGE, W. J. 2009. Next-generation DNA sequencing techniques. *New Biotechnology*, 25, 195-203.
- ARNAOUDOVA, E. G., BOWENS, P. J., CHUI, R. G., DINKINS, R. D., HESSE, U., JAROMCZYK, J. W., MARTIN, M., MAYNARD, P., MOORE, N. & SCHARDL, C. L. 2009. Visualizing and sharing results in bioinformatics projects: GBrowse and GenBank exports. *BMC Bioinformatics*, 10.
- AZAM, S., THAKUR, V., RUPEROAO, P., SHAH, T., BALAJI, J., AMINDALA, B., FARMER, A. D., STUDHOLME, D. J., MAY, G. D., EDWARDS, D., JONES, J. D. & VARSHNEY, R. K. 2012. Coverage-based consensus calling (CbCC) of short sequence reads and comparison of CbCC results to identify SNPs in chickpea (*Cicer arietinum*; Fabaceae), a crop species without a reference genome. *American Journal of Botany*, 99, 186-92.
- BALZER, S., MALDE, K., LANZEN, A., SHARMA, A. & JONASSEN, I. 2010. Characteristics of 454 pyrosequencing data--enabling realistic simulation with flowsim. *Bioinformatics*, 26, i420-5.
- BARBAZUK, W. B., EMRICH, S. J., CHEN, H. D., LI, L. & SCHNABLE, P. S. 2007. SNP discovery via 454 transcriptome sequencing. *Plant Journal*, 51, 910-918.
- BARKER, G., BATLEY, J., O'SULLIVAN, H., EDWARDS, K. J. & EDWARDS, D. 2003. Redundancy based detection of sequence polymorphisms in expressed sequence tag data using autoSNP. *Bioinformatics*, 19, 421-422.

- BARKER, G. L. A. & EDWARDS, K. J. 2009. A genome-wide analysis of single nucleotide polymorphism diversity in the world's major cereal crops. *Plant Biotechnology Journal*, 7, 318-325.
- BARSKI, A., CUDDAPAH, S., CUI, K., ROH, T. Y., SCHONES, D. E., WANG, Z., WEI, G., CHEPELEV, I. & ZHAO, K. 2007. High-resolution profiling of histone methylations in the human genome. *Cell*, 129, 823-37.
- BATLEY, J., BARKER, G., O'SULLIVAN, H., EDWARDS, K. J. & EDWARDS, D. 2003. Mining for single nucleotide polymorphisms and insertions/deletions in maize expressed sequence tag data. *Plant Physiology*, 132, 84-91.
- BATLEY, J. & EDWARDS, D. 2007. SNP applications in plants. In: ORAGUZIE, N., RIKKERINK, E., GARDINER, S. & DE SILVA, H. (eds.) *Association Mapping in Plants*. New York: Springer
- BATLEY, J. & EDWARDS, D. 2009. Mining for Single Nucleotide Polymorphism (SNP) and Simple Sequence Repeat (SSR) molecular genetic markers. In: POSADA, D. (ed.) *Bioinformatics for DNA Sequence Analysis*. USA: Humana Press.
- BATLEY, J., HOPKINS, C. J., COGAN, N. O. I., HAND, M., JEWELL, E., KAUR, J., KAUR, S., LI, X., LING, A. E., LOVE, C., MOUNTFORD, H., TODOROVIC, M., VARDY, M., WALKIEWICZ, M., SPANGENBERG, G. C. & EDWARDS, D. 2007. Identification and characterization of simple sequence repeat markers from *Brassica napus* expressed sequences. *Molecular Ecology Notes*, 7, 886-889.
- BECKMANN, J. S. & SOLLER, M. 1983. Restriction fragment length polymorphisms in genetic improvement: methodologies, mapping and costs. *Theoretical and Applied Genetics*, 67, 35-43.
- BELFIELD, E. J., GAN, X., MITHANI, A., BROWN, C., JIANG, C., FRANKLIN, K., ALVEY, E., WIBOWO, A., JUNG, M., BAILEY, K., KALWANI, S., RAGOUESSIS, J., MOTT, R. & HARBERD, N. P. 2012. Genome-wide analysis of mutations in mutant lineages selected following fast-neutron irradiation mutagenesis of *Arabidopsis thaliana*. *Genome Research*, 22, 1306-15.
- BENSON, D. A., KARSCH-MIZRACHI, I., LIPMAN, D. J., OSTELL, J. & SAYERS, E. W. 2009. GenBank. *Nucleic Acids Research*, 37, 26-31.
- BENTLEY, D. R., BALASUBRAMANIAN, S., SWERDLOW, H. P., SMITH, G. P., MILTON, J., BROWN, C. G., HALL, K. P., EVERS, D. J., BARNES, C. L., BIGNELL, H. R., BOUTELL, J. M., BRYANT, J., CARTER, R. J., CHEETHAM, R. K., COX, A. J., ELLIS, D. J., FLATBUSH, M. R., GORMLEY, N. A., HUMPHRAY, S. J., IRVING, L. J., KARBELASHVILI, M. S., KIRK, S. M., LI, H., LIU, X. H., MAISINGER, K. S., MURRAY, L. J., OBRADOVIC, B., OST, T., PARKINSON, M. L., PRATT, M. R., RASOLONJATOVO, I. M. J., REED, M. T., RIGATTI, R., RODIGHIERO, C., ROSS, M. T., SABOT, A., SANKAR, S. V., SCALLY, A., SCHROTH, G. P., SMITH, M. E., SMITH, V. P., SPIRIDOU, A., TORRANCE, P. E., TZONEV, S. S., VERMAAS, E. H., WALTER, K., WU, X. L., ZHANG, L., ALAM, M. D., ANASTASI, C., ANIEBO, I. C., BAILEY, D. M. D., BANCARZ, I. R., BANERJEE, S., BARBOUR, S. G., BAYBAYAN, P. A., BENOIT, V. A., BENSON, K. F., BEVIS, C., BLACK, P. J., BOODHUN, A., BRENNAN, J. S., BRIDGHAM, J. A., BROWN, R. C., BROWN, A. A., BUERMANN, D. H., BUNDU, A. A., BURROWS, J. C., CARTER, N. P., CASTILLO, N., CATENAZZI, M. C. E., CHANG, S., COOLEY, R. N., CRAKE, N. R., DADA, O. O., DIAKOUUMAKOS, K. D., DOMINGUEZ-FERNANDEZ, B., EARNSHAW, D. J., EGBUJOR, U. C., ELMORE, D. W., ETCHIN, S. S., EWAN, M. R., FEDURCO, M., FRASER, L. J., FAJARDO, K. V. F., FUREY, W. S., GEORGE, D., GIETZEN, K. J., GODDARD, C. P., GOLDA, G. S., GRANIERI, P. A., GREEN, D. E., GUSTAFSON, D. L., HANSEN, N. F., HARNISH, K., HAUDENSCHILD, C. D., HEYER, N. I., HIMS, M. M., HO, J. T., HORGAN, A. M., et al. 2008. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*, 456, 53-59.
- BERKMAN, B. J., SKARSHEWSKI, A., LORENC, M. T., LAI, K., DURAN, C., LING, E. Y. S., STILLER, J., SMITS, L., IMELFORT, M., MANOLI, S., MCKENZIE, M., KUBALAKOVA, M., SIMKOVA, H., BATLEY, J., FLEURY, D., DOLEŽEL, J. & EDWARDS, D. 2011a. Sequencing and assembly of low copy and genic regions of isolated *Triticum aestivum* chromosome arm 7DS. *Plant Biotechnology Journal*, 9, 768-775.
- BERKMAN, P. J., LAI, K., LORENC, M. T. & EDWARDS, D. 2012a. Next generation sequencing applications for wheat crop improvement. *American Journal of Botany*, 99, 365-371.
- BERKMAN, P. J., MANOLI, S., MCKENZIE, M., KUBALÁKOVÁ, M., ŠIMKOVÁ, H., BATLEY, J., FLEURY, D., DOLEŽEL, J., EDWARDS, D., SKARSHEWSKI, A., LORENC, M. T., LAI, K., DURAN, C., LING, E. Y. S., STILLER, J., SMITS, L. & IMELFORT, M. 2011b. Sequencing and assembly of low copy and genic regions of isolated *Triticum aestivum* chromosome arm

- 7DS. *Plant Biotechnology Journal*, 9, 768-775.
- BERKMAN, P. J., SKARSHEWSKI, A., MANOLI, S., LORENC, M. T., STILLER, J., LARS, SMITS, L., LAI, K., CAMPBELL, E., KUBALAKOVA, M., SIMKOVA, H., BATLEY, J., DOLEZEL, J., HERNANDEZ, P. & EDWARDS, D. 2012b. Sequencing wheat chromosome arm 7BS delimits the 7BS/4AL translocation and reveals homoeologous gene conservation. *Theoretical and Applied Genetics*, 124, 423-432.
- BERKMAN, P. J., VISENDI, P., LEE, H. C., STILLER, J., MANOLI, S., LORENC, M. T., LAI, K., BATLEY, J., FLEURY, D., SIMKOVA, H., KUBALAKOVA, M., WEINING, S., DOLEZEL, J. & EDWARDS, D. 2013. Dispersion and domestication shaped the genome of bread wheat. *Plant Biotechnology Journal*, 11, 564-571.
- BHALA, P. L. 2006. Genetic engineering of wheat--current challenges and opportunities. *Trends in Biotechnology*, 24, 305-11.
- BLAIR, M. W., CORTES, A. J., PENMETSA, R. V., FARMER, A., CARRASQUILLA-GARCIA, N. & COOK, D. R. 2013. A high-throughput SNP marker system for parental polymorphism screening, and diversity analysis in common bean (*Phaseolus vulgaris* L.). *Theoretical and Applied Genetics*, 126, 535-48.
- BLANCA, J., ESTERAS, C., ZIARSOLO, P., PEREZ, D., FERNA NDEZ-PEDROSA, V., COLLADO, C., RODRA GUEZ DE PABLOS, R., BALLESTER, A., ROIG, C., CANIZARES, J. & PICO, B. 2012. Transcriptome sequencing for SNP discovery across *Cucumis melo*. *BMC Genomics*, 13, 280.
- BLEENDA, A., SCHEFFLER, J., SCHEFFLER, B., PALMER, M., LACAPE, J. M., YU, J. Z., JESUDURAI, C., JUNG, S., MUTHUKUMAR, S., YELLAMBALASE, P., FICKLIN, S., STATON, M., ESHELMAN, R., ULLOA, M., SAHA, S., BURR, B., LIU, S., ZHANG, T., FANG, D., PEPPER, A., KUMPATLA, S., JACOBS, J., TOMKINS, J., CANTRELL, R. & MAIN, D. 2006. CMD: a Cotton Microsatellite Database resource for *Gossypium* genomics. *BMC Genomics*, 7, 132.
- BOGUSKI, M. S., TOLSTOSHEV, C. M. & BASSETT, D. E., JR. 1994. Gene discovery in dbEST. *Science*, 265, 1993-4.
- BOLGER, M. E., WEISSHAAR, B., SCHOLZ, U., STEIN, N., USADEL, B. & MAYER, K. F. 2014. Plant genome sequencing - applications for crop improvement. *Curr Opin Biotechnol*, 26, 31-7.
- BOMBARELY, A., MENDA, N., TECLE, I. Y., BUELS, R. M., STRICKLER, S., FISCHER-YORK, T., PUJAR, A., LETO, J., GOSSELIN, J. & MUELLER, L. A. 2011. The Sol Genomics Network (solgenomics.net): growing tomatoes using Perl. *Nucleic Acids Research*, 39, D1149-D1155.
- BOTTLEY, A., XIA, G. M. & KOEBNER, R. M. D. 2006. Homoeologous gene silencing in hexaploid wheat. *The Plant Journal*, 47, 897-906.
- BOYLE, J. S. & LEW, A. M. 1995. An inexpensive alternative to glassmilk for DNA purification. *Trends in Genetics*, 11, 8.
- BRAUTIGAM, A. & GOWIK, U. 2010. What can next generation sequencing do for you? Next generation sequencing as a valuable tool in plant research. *Plant Biology*, 12, 831-41.
- BRENCHLEY, R., SPANNAGL, M., PFEIFER, M., BARKER, G. L. A., D'AMORE, R., ALLEN, A. M., MCKENZIE, N., KRAMER, M., KERHORNOU, A., BOLSER, D., KAY, S., WAITE, D., TRICK, M., BANCROFT, I., GU, Y., HUO, N., LUO, M.-C., SEHGAL, S., GILL, B., KIANIAN, S., ANDERSON, O., KERSEY, P., DVORAK, J., MCCOMBIE, W. R., HALL, A., MAYER, K. F. X., EDWARDS, K. J., BEVAN, M. W. & HALL, N. 2012. Analysis of the bread wheat genome using whole-genome shotgun sequencing. *Nature*, 491, 705-710.
- BRENNER, S., JOHNSON, M., BRIDGHAM, J., GOLDA, G., LLOYD, D. H., JOHNSON, D., LUO, S., MCCURDY, S., FOY, M., EWAN, M., ROTH, R., GEORGE, D., ELETR, S., ALBRECHT, G., VERMAAS, E., WILLIAMS, S. R., MOON, K., BURCHAM, T., PALLAS, M., DUBRIDGE, R. B., KIRCHNER, J., FEARON, K., MAO, J. & CORCORAN, K. 2000. Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nature Biotechnology*, 18, 630-4.
- BROWN-GUEDIRA, G. L., HATCHETT, J. H., LIU, X. M., FRITZ, A. K., OWUOCHE, J. O., GILL, B. S., SEARS, R. G., COX, T. S. & CHEN, M. S. 2005. Registration of KS99WGRC42 Hessian fly resistant hard red winter wheat germplasm. *Crop Science*, 45, 804-805.
- BUCKLER, E. S. & HOLTSFORD, T. P. 1996. Zea ribosomal repeat evolution and substitution patterns. *Molecular Biology and Evolution*, 13, 623-632.

- BUNDOCK, P. C., ELIOTT, F. G., ABLETT, G., BENSON, A. D., CASU, R. E.,AITKEN, K. S. & HENRY, R. J. 2009. Targeted single nucleotide polymorphism (SNP) discovery in a highly polyploid plant species using 454 sequencing. *Plant Biotechnology Journal*, 7, 347-354.
- BURGESS, B., MOUNTFORD, H., HOPKINS, C. J., LOVE, C., LING, A. E., SPANGENBERG, G. C., EDWARDS, D. & BATLEY, J. 2006. Identification and characterization of simple sequence repeat (SSR) markers derived in silico from *Brassica oleracea* genome shotgun sequences. *Molecular Ecology Notes*, 6, 1191-1194.
- BUS, A., HECHT, J., HUETTEL, B., REINHARDT, R. & STICH, B. 2012. High-throughput polymorphism detection and genotyping in *Brassica napus* using next-generation RAD sequencing. *Bmc Genomics*, 13.
- CALDWELL, K. S., DVORAK, J., LAGUDAH, E. S., AKHUNOV, E., LUO, M. C., WOLTERS, P. & POWELL, W. 2004. Sequence polymorphism in polyploid wheat and their D-genome diploid ancestor. *Genetics*, 167, 941-7.
- CANARAN, P., BUCKLER, E. S., GLAUBITZ, J. C., STEIN, L., SUN, Q., ZHAO, W. & WARE, D. 2008. Panzea: an update on new content and features. *Nucleic Acids Research*, 36, D1041-3.
- CANARAN, P., STEIN, L. & WARE, D. 2006. Look-Align: an interactive web-based multiple sequence alignment viewer with polymorphism analysis support. *Bioinformatics*, 22, 885-886.
- CAO, K., ZHENG, Z., WANG, L., LIU, X., ZHU, G., FANG, W., CHENG, S., ZENG, P., CHEN, C., WANG, X., XIE, M., ZHONG, X., WANG, X., ZHAO, P., BIAN, C., ZHU, Y., ZHANG, J., MA, G., CHEN, C., LI, Y., HAO, F., LI, Y., HUANG, G., LI, Y., LI, H., GUO, J., XU, X. & WANG, J. 2014. Comparative population genomics reveals the domestication history of the peach, *prunus persica*, and human influences on perennial fruit crops. *Genome Biology*, 15, 415.
- CAROLLO, V., MATTHEWS, D. E., LAZO, G. R., BLAKE, T. K., HUMMEL, D. D., LUI, N., HANE, D. L. & ANDERSON, O. D. 2005. GrainGenes 2.0. An improved resource for the small-grains community. *Plant Physiology*, 139, 643-651.
- CAUSIER, B., KIEFFER, M. & DAVIES, B. 2002. Plant biology. MADS-box genes reach maturity. *Science*, 296, 275-6.
- CAVANAGH, C. R., CHAO, S., WANG, S., HUANG, B. E., STEPHEN, S., KIANI, S., FORREST, K., SAINTENAC, C., BROWN-GUEDIRA, G. L., AKHUNOVA, A., SEE, D., BAI, G., PUMPHREY, M., TOMAR, L., WONG, D., KONG, S., REYNOLDS, M., DA SILVA, M. L., BOCKELMAN, H., TALBERT, L., ANDERSON, J. A., DREISIGACKER, S., BAENZIGER, S., CARTER, A., KORZUN, V., MORRELL, P. L., DUBCOVSKY, J., MORELL, M. K., SORRELLS, M. E., HAYDEN, M. J. & AKHUNOV, E. 2013. Genome-wide comparative diversity uncovers multiple targets of selection for improvement in hexaploid wheat landraces and cultivars. *Proceedings of the National Academy of Sciences of the United States of America*, 110, 8057-62.
- CHAGNÉ, D., BATLEY, J., EDWARDS, D. & FORSTER, J. W. 2007. Single Nucleotide Polymorphism genotyping in plants. In: ORAGUZIE, N., RIKKERINK, E., GARDINER, S. & DE SILVA, H. (eds.) *Association Mapping in Plants*. New York: Springer
- CHANTRET, N., SALSE, J., SABOT, F., RAHMAN, S., BELLEC, A., LAUBIN, B., DUBOIS, I., DOSSAT, C., SOURDILLE, P., JOUDRIER, P., GAUTIER, M. F., CATTOLICO, L., BECKERT, M., AUBOURG, S., WEISSENBACH, J., CABOCHE, M., BERNARD, M., LEROY, P. & CHALHOUB, B. 2005. Molecular basis of evolutionary events that shaped the hardness locus in diploid and polyploid wheat species (*Triticum* and *aegilops*). *Plant Cell*, 17, 1033-1045.
- CHAO, S., ZHANG, W., AKHUNOV, E., SHERMAN, J., MA, Y., LUO, M.-C. & DUBCOVSKY, J. 2009. Analysis of gene-derived SNP marker polymorphism in US wheat (*Triticum aestivum* L.) cultivars. *Molecular Breeding*, 23, 23-33.
- CHARMET, G. 2011. Wheat domestication: Lessons for the future. *Comptes Rendus Biologies*, 334, 212-220.
- CHEN, K. & RAJEWSKY, N. 2006. Natural selection on human microRNA binding sites inferred from SNP data. *Nature Genetics*, 38, 1452-6.
- CHENG, F., LIU, S., WU, J., FANG, L., SUN, S., LIU, B., LI, P., HUA, W., WANG, X., CHENG, F., LIU, S. Y., WU, J., FANG, L., SUN, S. L., LIU, B., LI, P. X., HUA, W. & WANG, X. W. 2011. BRAD, the genetics and genomics database for *Brassica* plants. *BMC Plant Biology*, 11.
- CHEVREUX, B. 2005. MIRA: an automated genome and EST assembler. *Ruprecht-Karls University, Heidelberg, Germany*.
- CHOI, S. R., TEAKLE, G. R., PLAHA, P., KIM, J. H., ALLENDER, C. J., BEYNON, E., PIAO, Z. Y., SOENGAS, P., HAN, T. H., KING, G. J., BARKER, G. C., HAND, P., LYDIATE, D. J., BATLEY, J., EDWARDS, D., KOO, D. H., BANG, J. W., PARK, B. S. & LIM, Y. P.

2007. The reference genetic linkage map for the multinational *Brassica rapa* genome sequencing project. *Theoretical and Applied Genetics*, 115, 777-92.
- CINGOLANI, P., PLATTS, A., WANG LE, L., COON, M., NGUYEN, T., WANG, L., LAND, S. J., LU, X. & RUDEN, D. M. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)*, 6, 80-92.
- CLARK, R. M., LINTON, E., MESSING, J. & DOEBLEY, J. F. 2004. Pattern of diversity in the genomic region near the maize domestication gene tb1. *Proceedings of the National Academy of Sciences of the United States of America*, 101, 700-7.
- CLOSE, T. J., BHAT, P. R., LONARDI, S., WU, Y., ROSTOKS, N., RAMSAY, L., DRUKA, A., STEIN, N., SVENSSON, J. T., WANAMAKER, S., BOZDAG, S., ROOSE, M. L., MOSCOU, M. J., CHAO, S., VARSHNEY, R. K., SZUCS, P., SATO, K., HAYES, P. M., MATTHEWS, D. E., KLEINHOFS, A., MUEHLBAUER, G. J., DEYOUNG, J., MARSHALL, D. F., MADISHETTY, K., FENTON, R. D., CONDAMINE, P., GRANER, A. & WAUGH, R. 2009. Development and implementation of high-throughput SNP genotyping in barley. *BMC Genomics*, 10, 582.
- COMBES, M.-C., CENCI, A., BARAILLE, H., BERTRAND, B. & LASHERMES, P. 2012. Homeologous Gene Expression in Response to Growing Temperature in a Recent Allopolyploid (*Coffea arabica* L.). *Journal of Heredity*, 103, 36-46.
- CORAM, T. E., BROWN-GUEDIRA, G. & CHEN, X. 2008a. Using transcriptomics to understand the wheat genome. *CAB Reviews: Perspectives in Agriculture, Veterinary Science, Nutrition and Natural Resources*, 083, 1-9.
- CORAM, T. E., WANG, M. & CHEN, X. 2008b. Transcriptome analysis of the wheat–*Puccinia striiformis* f. sp. *tritici* interaction. *Molecular Plant Pathology*, 9, 157-169.
- COULONDRE, C., MILLER, J. H., FARABAUGH, P. J. & GILBERT, W. 1978. Molecular-Basis of Base Substitution Hotspots in *Escherichia coli*. *Nature*, 274, 775-780.
- DALTON-MORGAN, J., HAYWARD, A., ALAMERY, S., TOLLENAERE, R., MASON, A., CAMPBELL, E., PATEL, D., LORENC, M., YI, B., LONG, Y., MENG, J., RAMAN, R., RAMAN, H., LAWLEY, C., EDWARDS, D. & BATLEY, J. 2014. A high-throughput SNP array in the amphidiploid species *Brassica napus* shows diversity in resistance genes. *Functional & Integrative Genomics*, 1-13.
- DANECEK, P., AUTON, A., ABECASIS, G., ALBERS, C. A., BANKS, E., DEPRISTO, M. A., HANDSAKER, R. E., LUNTER, G., MARTH, G. T., SHERRY, S. T., MCVEAN, G., DURBIN, R. & GENOMES PROJECT ANALYSIS, G. 2011. The variant call format and VCFtools. *Bioinformatics*, 27, 2156-8.
- DEPRISTO, M. A., BANKS, E., POPLIN, R., GARIMELLA, K. V., MAGUIRE, J. R., HARTL, C., PHILIPPakis, A. A., DEL ANGEL, G., RIVAS, M. A., HANNA, M., MCKENNA, A., FENNELL, T. J., KERNYTSKY, A. M., SIVACHENKO, A. Y., CIBULSKIS, K., GABRIEL, S. B., ALTSHULER, D. & DALY, M. J. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, 43, 491-+.
- DESCHAMPS, S., ROTA, M., RATASHAK, J., BIDDLE, P., THUREEN, D., FARMER, A., LUCK, S., BEATTY, M., NAGASAWA, N. & MICHAEL, L. 2010. Rapid genome-wide single nucleotide polymorphism discovery in soybean and rice via deep resequencing of reduced representation libraries with the Illumina genome analyzer. *The Plant Genome*, 3, 53 - 68.
- DFAT 2004. Asian Agrifood Demands Trends and Outlook to 2010.
- DOERING, A., WEESE, D., RAUSCH, T. & REINERT, K. 2008. SeqAn An efficient, generic C++ library for sequence analysis. *BMC Bioinformatics*, 9.
- DOLEZEL, J., KUBALAKOVA, M., PAUX, E., BARTOS, J. & FEUILLET, C. 2007. Chromosome-based genomics in the cereals. *Chromosome Research : an International Journal on the Molecular, Supramolecular and Evolutionary Aspects of Chromosome Biology*, 15, 51-66.
- DONLIN, M. 2007. Using the Generic Genome Browser (GBrowse). *Current Protocols in Bioinformatics*, Chapter 9, 9.
- DRAPER, J., MUR, L. A., JENKINS, G., GHOSH-BISWAS, G. C., BABLAK, P., HASTEROK, R. & ROUTLEDGE, A. P. 2001. Brachypodium distachyon. A new model system for functional genomics in grasses. *Plant Physiol*, 127, 1539-55.
- DRUKA, A., POTOKINA, E., LUO, Z., JIANG, N., CHEN, X., KEARSEY, M. & WAUGH, R. 2010. Expression quantitative trait loci

- analysis in plants. *Plant Biotechnology Journal*, 8, 10-27.
- DRUMMOND, A. J., ASHTON, B., S. B., CHEUNG, M., COOPER, A., DURAN, C., FIELD, M., HELED, J., KEARSE, M., MARKOWITZ, S., MOIR, R., STONES-HAVAS, S., STURROCK, S., THIERER, T. & WILSON, A. 2011. *Geneious v5.4*. <http://www.geneious.com>. [Online]. Available: <http://www.geneious.com>.
- DUBCOVSKY, J. & DVORAK, J. 2007. Genome plasticity a key factor in the success of polyploid wheat under domestication. *Science*, 316, 1862-6.
- DURAN, C., APPLEBY, N., CLARK, T., WOOD, D., IMELFORT, M., BATLEY, J. & EDWARDS, D. 2009a. AutoSNPdb: an annotated single nucleotide polymorphism database for crop plants. *Nucleic Acids Research*, 37, D951-D953.
- DURAN, C., APPLEBY, N., EDWARDS, D. & BATLEY, J. 2009b. Molecular Genetic Markers: Discovery, Applications, Data Storage and Visualisation. *Current Bioinformatics*, 4, 16-27.
- DURAN, C., APPLEBY, N., VARDY, M., IMELFORT, M., EDWARDS, D. & BATLEY, J. 2009c. Single nucleotide polymorphism discovery in barley using autoSNPdb. *Plant Biotechnology Journal*, 7, 326-333.
- DURAN, C., BOSKOVIC, Z., BATLEY, J. & EDWARDS, D. 2011. Role of bioinformatics as a tool for vegetable Brassica species. In: J. S. (ed.) *Vegetable Brassicas*. New Hampshire: Science Publishers, Inc.
- DURAN, C., BOSKOVIC, Z., IMELFORT, M., BATLEY, J., HAMILTON, N. A. & EDWARDS, D. 2010a. CMap3D: A 3D visualisation tool for comparative genetic maps. *Bioinformatics*, 26, 273-274.
- DURAN, C., EALES, D., MARSHALL, D., IMELFORT, M., STILLER, J., BERKMAN, P. J., CLARK, T., MCKENZIE, M., APPLEBY, N., BATLEY, J., BASFORD, K. & EDWARDS, D. 2010b. Future tools for association mapping in crop plants. *Genome*, 53, 1017-1023.
- DURAN, C., EDWARDS, D. & BATLEY, J. 2009d. Genetic maps and the use of synteny. In: SOMERS, D., LANGRIDGE, P. & GUSTAFSON, J. (eds.) *Plant Genomics*. Humana Press (USA).
- DVORAK, J., AKHUNOV, E. D., AKHUNOV, A. R., DEAL, K. R. & LUO, M. C. 2006. Molecular characterization of a diagnostic DNA marker for domesticated tetraploid wheat provides evidence for gene flow from wild tetraploid wheat to hexaploid wheat. *Molecular Biology and Evolution*, 23, 1386-96.
- ECKARDT, N. A. 2001. A sense of self: The role of DNA sequence elimination in allopolyploidization. *Plant Cell*, 13, 1699-1704.
- EDWARDS, A., VOSS, H., RICE, P., CIVITELLO, A., STEGEMANN, J., SCHWAGER, C., ZIMMERMANN, J., ERFLE, H., CASKEY, C. T. & ANSORGE, W. 1990. Automated DNA sequencing of the human HPRT locus. *Genomics*, 6, 593-608.
- EDWARDS, D. & BATLEY, J. 2004. Plant Bioinformatics: From Genome to Phenome. *Trends in Biotechnology*, 22, 232-237.
- EDWARDS, D. & BATLEY, J. 2008. Bioinformatics: fundamentals and applications in plant genetics, mapping and breeding. In: KOLE, C. & ABBOTT, A. G. (eds.) *Principles and Practices of Plant Genomics*. USA: Science Publishers, Inc.
- EDWARDS, D. & BATLEY, J. 2010. Plant genome sequencing: applications for crop improvement. *Plant Biotechnology Journal*, 7, 1-8.
- EDWARDS, D., BATLEY, J. & SNOWDON, R. J. 2013. Accessing complex crop genomes with next-generation sequencing. *Theoretical and applied genetics*, 126, 1-11.
- EDWARDS, D., FORESTER, J. W., CHAGNÉ, D. & BATLEY, J. 2007. What are SNPs? In: ORAGUZIE, N. C., RIKKERINK, E. H. A., GARDINER, S. E. & SILVA, H. N. D. (eds.) *Association Mapping in Plants*. Springer NY.
- EDWARDS, D., HENRY, R. J. & EDWARDS, K. J. 2012a. Preface: advances in DNA sequencing accelerating plant biotechnology. *Plant Biotechnology Journal*, 10, 621-622.
- EDWARDS, D., WILCOX, S., BARRERO, R. A., FLEURY, D., CAVANAGH, C. R., FORREST, K. L., HAYDEN, M. J., MOOLHUIJSEN, P., KEEBLE-GAGNÈRE, G., BELLGARD, M. I., LORENC, M. T., SHANG, C. A., BAUMANN, U., TAYLOR, J. M., MORELL, M. K., LANGRIDGE, P., APPELS, R. & FITZGERALD, A. 2012b. Bread matters: a national initiative to profile the genetic diversity of Australian wheat. *Plant Biotechnology Journal*, 10, 703-708.
- EID, J., FEHR, A., GRAY, J., LUONG, K., LYLE, J., OTTO, G., PELUSO, P., RANK, D., BAYBAYAN, P., BETTMAN, B., BIBILLO, A., BJORNSON, K., CHAUDHURI, B., CHRISTIANS, F., CICERO, R., CLARK, S., DALAL, R., DEWINTER, A., DIXON, J., FOQUET, M., GAERTNER, A., HARDENBOL, P., HEINER, C., HESTER, K., HOLDEN, D., KEARNS, G., KONG, X., KUSE, R.,

- LACROIX, Y., LIN, S., LUNDQUIST, P., MA, C., MARKS, P., MAXHAM, M., MURPHY, D., PARK, I., PHAM, T., PHILLIPS, M., ROY, J., SEBRA, R., SHEN, G., SORENSEN, J., TOMANEY, A., TRAVERS, K., TRULSON, M., VIECELJ, J., WEGENER, J., WU, D., YANG, A., ZACCARIN, D., ZHAO, P., ZHONG, F., KORLACH, J. & TURNER, S. 2009. Real-time DNA sequencing from single polymerase molecules. *Science*, 323, 133-8.
- EVERSOLE, K. The International Wheat Genome Sequencing Consortium (IWGSC): Building the foundation for a paradigm shift in wheat breeding. The 15th International EWAC Conference, 2012 Novi Sad, Serbia.
- EWING, B., HILLIER, L., WENDL, M. C. & GREEN, P. 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Research*, 8, 175-185.
- FABRIANI, G., LINTAS, C. & CHEMISTS, A. A. O. C. 1988. *Durum wheat: chemistry and technology*, American Association of Cereal Chemists.
- FAO 2006. Statistical Yearbook 2005–2006.
- FAO 2009. GIEWS Global Watch.
- FELDMAN, M. & LEVY, A. A. 2009. Genome evolution in allopolyploid wheat--a revolutionary reprogramming followed by gradual changes. *Journal of Genetics and Genomics*, 36, 511-8.
- FELDMAN, M. & LEVY, A. A. 2012. Genome Evolution Due to Allopolyploidization in Wheat. *Genetics*, 192, 763-774.
- FEUILLET, C., LANGRIDGE, P. & WAUGH, R. 2008. Cereal breeding takes a walk on the wild side. *Trends in Genetics*, 24, 24-32.
- FEUILLET, C., LEACH, J. E., ROGERS, J., SCHNABLE, P. S. & EVERSOLE, K. 2011. Crop genome sequencing: lessons and rationales. *Trends in Plant Science*, 16, 77-88.
- FUKUOKA, H., MIYATAKE, K., NUNOME, T., NEGORO, S., SHIRASAWA, K., ISOBE, S., ASAMIZU, E., YAMAGUCHI, H. & OHYAMA, A. 2012. Development of gene-based markers and construction of an integrated linkage map in eggplant by using Solanum orthologous (SOL) gene sets. *Theoretical and Applied Genetics*, 125, 47-56.
- FULTON, T., CHUNWONGSE, J. & TANKSLEY, S. 1995. Microprep protocol for extraction of DNA from tomato and other herbaceous plants. *Plant Molecular Biology Reporter*, 13, 207-209.
- GABRIEL, S., ZIAUGRA, L. & TABBA, D. 2001. SNP Genotyping Using the Sequenom MassARRAY iPLEX Platform. *Current Protocols in Human Genetics*. John Wiley & Sons, Inc.
- GARG, K., GREEN, P. & NICKERSON, D. A. 1999. Identification of candidate coding region single nucleotide polymorphisms in 165 human genes using assembled expressed sequence tags. *Genome Research*, 9, 1087-1092.
- GEGAS, V. C., NAZARI, A., GRIFFITHS, S., SIMMONDS, J., FISH, L., ORFORD, S., SAYERS, L., DOONAN, J. H. & SNAPE, J. W. 2010. A genetic framework for grain size and shape variation in wheat. *Plant Cell*, 22, 1046-56.
- GERHARD, D. S., WAGNER, L., FEINGOLD, E. A., SHENMEN, C. M., GROUSE, L. H., SCHULER, G., KLEIN, S. L., OLD, S., RASOOLY, R., GOOD, P., GUYER, M., PECK, A. M., DERGE, J. G., LIPMAN, D., COLLINS, F. S., JANG, W., SHERRY, S., FEOLO, M., MISQUITTA, L., LEE, E., ROTMISTROVSKY, K., GREENHUT, S. F., SCHAEFER, C. F., BUETOW, K., BONNER, T. I., HAUSSLER, D., KENT, J., KIEKHAUS, M., FUREY, T., BRENT, M., PRANGE, C., SCHREIBER, K., SHAPIRO, N., BHAT, N. K., HOPKINS, R. F., HSIE, F., DRISCOLL, T., SOARES, M. B., CASAVANT, T. L., SCHEETZ, T. E., BROWN-STEIN, M. J., USDIN, T. B., TOSHIYUKI, S., CARNINCI, P., PIAO, Y., DUDEKULA, D. B., KO, M. S., KAWAKAMI, K., SUZUKI, Y., SUGANO, S., GRUBER, C. E., SMITH, M. R., SIMMONDS, B., MOORE, T., WATERMAN, R., JOHNSON, S. L., RUAN, Y., WEI, C. L., MATHAVAN, S., GUNARATNE, P. H., WU, J., GARCIA, A. M., HULYK, S. W., FUH, E., YUAN, Y., SNEED, A., KOWIS, C., HODGSON, A., MUZNY, D. M., MCPHERSON, J., GIBBS, R. A., FAHEY, J., HELTON, E., KETTEMAN, M., MADAN, A., RODRIGUES, S., SANCHEZ, A., WHITING, M., MADARI, A., YOUNG, A. C., WETHERBY, K. D., GRANITE, S. J., KWONG, P. N., BRINKLEY, C. P., PEARSON, R. L., BOUFFARD, G. G., BLAKESLY, R. W., GREEN, E. D., DICKSON, M. C., RODRIGUEZ, A. C., GRIMWOOD, J., SCHMUTZ, J., MYERS, R. M., BUTTERFIELD, Y. S., GRIFFITH, M., GRIFFITH, O. L., KRZYWINSKI, M. I., LIAO, N., MORIN, R., PALMQUIST, D., et al. 2004. The status, quality, and expansion of the NIH full-length cDNA project: the Mammalian Gene Collection (MGC). *Genome Research*, 14, 2121-7.
- GILBERT, D. G. 1999. *Phylogenodendron, an application for drawing phylogenetic trees* [Online]. Available: <http://iubio.bio.indiana.edu/treeapp/>.

- GILES, R. & BROWN, T. 2006. GluD_y allele variations in *Aegilops tauschii* and *Triticum aestivum*: implications for the origins of hexaploid wheats. *Theoretical and Applied Genetics*, 112, 1563-1572.
- GILL, B. S., APPELS, R., BOTHA-OBERHOLSTER, A. M., BUELL, C. R., BENNETZEN, J. L., CHALHOUB, B., CHUMLEY, F., DVORAK, J., IWANAGA, M., KELLER, B., LI, W., MCCOMBIE, W. R., OGIHARA, Y., QUETIER, F. & SASAKI, T. 2004. A workshop report on wheat genome sequencing: International Genome Research on Wheat Consortium. *Genetics*, 168, 1087-96.
- GODFRAY, H. C., BEDDINGTON, J. R., CRUTE, I. R., HADDAD, L., LAWRENCE, D., MUIR, J. F., PRETTY, J., ROBINSON, S., THOMAS, S. M. & TOULMIN, C. 2010. Food security: the challenge of feeding 9 billion people. *Science*, 327, 812-8.
- GONZALES, M. D., ARCHULETA, E., FARMER, A., GAJENDRAN, K., GRANT, D., SHOEMAKER, R., BEAVIS, W. D. & WAUGH, M. E. 2005. The Legume Information System (LIS): an integrated information resource for comparative legume biology. *Nucleic Acids Research*, 33, D660-D665.
- GONZALES, M. D., GAJENDRAN, K., FARMER, A. D., ARCHULETA, E. & BEAVIS, W. D. 2007. Leveraging model legume information to find candidate genes for soybean sudden death syndrome using the legume information system. In: EDWARDS, D. (ed.) *Methods in Molecular Biology*. Totowa, NJ, USA: Humana Press.
- GRAMA, A. & GERECHTE.ZK 1974. Inheritance of Resistance to Stripe Rust (*Puccinia-Striiformis*) in Crosses between Wild Emmer (*Triticum-Dicoccoides*) and Cultivated Tetraploid and Hexaploid Wheats .2. *Triticum-Aestivum*. *Euphytica*, 23, 393-398.
- GRANT, D., NELSON, R. T., CANNON, S. B. & SHOEMAKER, R. C. 2010. SoyBase, the USDA-ARS soybean genetics and genomics database. *Nucleic Acids Research*, 38, D843-6.
- GREGERSEN, P. L., BRINCH-PEDERSEN, H. & HOLM, P. B. 2005. A microarray-based comparative analysis of gene expression profiles during grain development in transgenic and wild type wheat. *Transgenic research*, 14, 887-905.
- GUPTA, P. K., LANGRIDGE, P. & MIR, R. R. 2010. Marker-assisted wheat breeding: present status and future possibilities. *Molecular Breeding*, 26, 145-161.
- GUPTA, P. K., MIR, R. R., MOHAN, A. & KUMAR, J. 2008. Wheat genomics: present status and future prospects. *International Journal of Plant Genomics*, 2008, 896451.
- HABIER, D., FERNANDO, R. L. & DEKKERS, J. C. M. 2009. Genomic Selection Using Low-Density Marker Panels. *Genetics*, 182, 343-353.
- HADRYS, H., BALICK, M. & SCHIERWATER, B. 1992. Applications of random amplified polymorphic DNA (RAPD) in molecular ecology. *Molecular Ecology*, 1, 55-63.
- HAYWARD, A., DALTON-MORGAN, J., MASON, A., ZANDER, M., EDWARDS, D. & BATLEY, J. 2012a. SNP discovery and applications in *Brassica napus*. *Journal of Plant Biotechnology*, 39, 49-61.
- HAYWARD, A., VIGHNES, G., DELAY, C., SAMIAN, M. R., MANOLI, S., STILLER, J., MCKENZIE, M., EDWARDS, D. & BATLEY, J. 2012b. Second-generation sequencing for gene discovery in the Brassicaceae. *Plant biotechnology journal*, 10, 750-9.
- HAZEN, S. P., SCOTT-CRAIG, J. S. & WALTON, J. D. 2002. Cellulose synthase-like genes of rice. *Plant Physiology*, 128, 336-40.
- HEFFNER, E. L., SORRELLS, M. E. & JANNINK, J.-L. 2009. Genomic Selection for Crop Improvement. *Crop Science*, 49, 1-12.
- HENRY, R. J. 2008. *Plant genotyping II: SNP technology*, CAB International.
- HENRY, R. J. 2012. Evolution of DNA Marker Technology in Plants. *Molecular Markers in Plants*. Blackwell Publishing Ltd.
- HERNANDEZ, P., MARTIS, M., DORADO, G., PFEIFER, M., GALVEZ, S., SCHAAF, S., JOUVE, N., SIMKOVA, H., VALARIK, M., DOLEZEL, J. & MAYER, K. F. 2012. Next-generation sequencing and syntenic integration of flow-sorted arms of wheat chromosome 4A exposes the chromosome structure and gene content. *The Plant Journal*, 69, 377-86.
- HIREMATH, P. J., FARMER, A., CANNON, S. B., WOODWARD, J., KUDAPA, H., TUTEJA, R., KUMAR, A., BHANUPRAKASH, A., MULAOSMANOVIC, B., GUJARIA, N., KRISHNAMURTHY, L., GAUR, P. M., KAVIKISHOR, P. B., SHAH, T., SRINIVASAN,

- R., LOHSE, M., XIAO, Y., TOWN, C. D., COOK, D. R., MAY, G. D. & VARSHNEY, R. K. 2011. Large-scale transcriptome analysis in chickpea (*Cicer arietinum* L.), an orphan legume crop of the semi-arid tropics of Asia and Africa. *Plant Biotechnology Journal*, 9, 922-931.
- HOFFMANN, T. J., KVALE, M. N., HESSELSON, S. E., ZHAN, Y., AQUINO, C., CAO, Y., CAWLEY, S., CHUNG, E., CONNELL, S., ESHRAGH, J., EWING, M., GOLLUB, J., HENDERSON, M., HUBBELL, E., IRIBARREN, C., KAUFMAN, J., LAO, R. Z., LU, Y., LUDWIG, D., MATHAUDA, G. K., MCGUIRE, W., MEI, G., MILES, S., PURDY, M. M., QUESENERRY, C., RANATUNGA, D., ROWELL, S., SADLER, M., SHAPERO, M. H., SHEN, L., SHENOY, T. R., SMETHURST, D., VAN DEN EEDEN, S. K., WALTER, L., WAN, E., WEARLEY, R., WEBSTER, T., WEN, C. C., WENG, L., WHITMER, R. A., WILLIAMS, A., WONG, S. C., ZAU, C., FINN, A., SCHAEFER, C., KWOK, P. Y. & RISCH, N. 2011. Next generation genome-wide association tool: design and coverage of a high-throughput European-optimized SNP array. *Genomics*, 98, 79-89.
- HONG, C. P., PIAO, Z. Y., KANG, T. W., BATLEY, J., YANG, T. J., HUR, Y. K., BHAK, J., PARK, B. S., EDWARDS, D. & LIM, Y. P. 2007. Genomic distribution of simple sequence repeats in *Brassica rapa*. *Molecules and Cells*, 23, 349-356.
- HOPKINS, C. J., COGAN, N. O. I., HAND, M., JEWELL, E., KAUR, J., LI, X., LIM, G. A. C., LING, A. E., LOVE, C., MOUNTFORD, H., TODOROVIC, M., VARDY, M., SPANGENBERG, G. C., EDWARDS, D. & BATLEY, J. 2007. Sixteen new simple sequence repeat markers from *Brassica juncea* expressed sequences and their cross-species amplification. *Molecular Ecology Notes*, 7, 697-700.
- HORI, K., TAKEHARA, S., NANKAKU, N., SATO, K., SASAKUMA, T. & TAKEDA, K. 2007. Barley EST Markers Enhance Map Saturation and QTL Mapping in Diploid Wheat. *Breeding Science*, 57, 39-45.
- HU, M. & POLYAK, K. 2006. Serial analysis of gene expression. *Nature Protocols*, 1, 1743-60.
- HUA, W., LIU, Z., ZHU, J., XIE, C., YANG, T., ZHOU, Y., DUAN, X., SUN, Q. & LIU, Z. 2009. Identification and genetic mapping of pm42, a new recessive wheat powdery mildew resistance gene derived from wild emmer (*Triticum turgidum* var. *dicoccoides*). *Theoretical and Applied Genetics*, 119, 223-230.
- HUANG, P., FELDMAN, M., SCHRODER, S., BAHRI, B. A., DIAO, X., ZHI, H., ESTEP, M., BAXTER, I., DEVOS, K. M. & KELLOGG, E. A. 2014. Population genetics of *Setaria viridis*, a new model system. *Molecular Ecology*.
- HUANG, S., SIRIKHACHORNKIT, A., SU, X., FARIS, J., GILL, B., HASELKORN, R. & GORNICKI, P. 2002. Genes encoding plastid acetyl-CoA carboxylase and 3-phosphoglycerate kinase of the *Triticum/Aegilops* complex and the evolutionary history of polyploid wheat. *Proceedings of the National Academy of Sciences of the United States of America*, 99, 8133-8138.
- HYMAN, E. D. 1988. A new method of sequencing DNA. *Analytical biochemistry*, 174, 423-36.
- IMELFORT, M., DURAN, C., BATLEY, J. & EDWARDS, D. 2009. Discovering genetic polymorphisms in next-generation sequencing data. *Plant Biotechnology Journal*, 7, 312-317.
- INDAP, A. R., COLE, R., RUNGE, C. L., MARTH, G. T. & OLIVIER, M. 2013. Variant discovery in targeted resequencing using whole genome amplified DNA. *BMC Genomics*, 14, 468.
- INTERNATIONAL BRACHYPODIUM INITIATIVE 2010. Genome sequencing and analysis of the model grass *Brachypodium distachyon*. *Nature*, 463, 763-8.
- INTERNATIONAL HUMAN GENOME SEQUENCING CONSORTIUM 2004. Finishing the euchromatic sequence of the human genome. *Nature*, 431, 931-45.
- INTERNATIONAL WHEAT GENOME SEQUENCING CONSORTIUM 2014. A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. *Science*, 345, 1251788.
- IWATA, H. & JANNINK, J. L. 2011. Accuracy of Genomic Selection Prediction in Barley Breeding Programs: A Simulation Study Based On the Real Single Nucleotide Polymorphism Data of Barley Breeding Lines. *Crop Science*, 51, 1915-1927.
- JACCOUD, D., PENG, K., FEINSTEIN, D. & KILIAN, A. 2001. Diversity arrays: a solid state technology for sequence information independent genotyping. *Nucleic Acids Research*, 29, E25.
- JANNINK, J.-L., LORENZ, A. J. & IWATA, H. 2010. Genomic selection in plant breeding: from theory to practice. *Briefings in Functional Genomics*, 9, 166-177.

- JEWELL, E., ROBINSON, A., SAVAGE, D., ERWIN, T., LOVE, C. G., LIM, G. A. C., LI, X., BATLEY, J., SPANGENBERG, G. C. & EDWARDS, D. 2006. SSRPrimer and SSR Taxonomy Tree: Biome SSR discovery. *Nucleic Acids Research*, 34, W656-W659.
- JIA, J., ZHAO, S., KONG, X., LI, Y., ZHAO, G., HE, W., APPELS, R., PFEIFER, M., TAO, Y., ZHANG, X., JING, R., ZHANG, C., MA, Y., GAO, L., GAO, C., SPANNAGL, M., MAYER, K. F., LI, D., PAN, S., ZHENG, F., HU, Q., XIA, X., LI, J., LIANG, Q., CHEN, J., WICKER, T., GOU, C., KUANG, H., HE, G., LUO, Y., KELLER, B., XIA, Q., LU, P., WANG, J., ZOU, H., ZHANG, R., XU, J., GAO, J., MIDDLETON, C., QUAN, Z., LIU, G., WANG, J., INTERNATIONAL WHEAT GENOME SEQUENCING, C., YANG, H., LIU, X., HE, Z., MAO, L. & WANG, J. 2013. Aegilops tauschii draft genome sequence reveals a gene repertoire for wheat adaptation. *Nature*, 496, 91-5.
- JOHNSON, D. S., MORTAZAVI, A., MYERS, R. M. & WOLD, B. 2007. Genome-wide mapping of in vivo protein-DNA interactions. *Science*, 316, 1497-502.
- JORDAN, K. W., WANG, S., LUN, Y., GARDINER, L. J., MACLACHLAN, R., HUCL, P., WIEBE, K., WONG, D., FORREST, K. L., CONSORTIUM, I., SHARPE, A. G., SIDEBOTTOM, C. H., HALL, N., TOOMAJIAN, C., CLOSE, T., DUBCOVSKY, J., AKHUNOVA, A., TALBERT, L., BANSAL, U. K., BARIANA, H. S., HAYDEN, M. J., POZNIAK, C., JEDDELOH, J. A., HALL, A. & AKHUNOV, E. 2015. A haplotype map of allohexaploid wheat reveals distinct patterns of selection on homoeologous genomes. *Genome Biology*, 16, 48.
- KARSCH-MIZRACHI, I., NAKAMURA, Y. & COCHRANE, G. 2012. The International Nucleotide Sequence Database Collaboration. *Nucleic Acids Research*, 40, D33-7.
- KAUR, S., FRANCKI, M. G. & FORSTER, J. W. 2012. Identification, characterization and interpretation of single-nucleotide sequence variation in allopolyploid crop species. *Plant Biotechnology Journal*, 10, 125-138.
- KAWAHARA, Y., DE LA BASTIDE, M., HAMILTON, J., KANAMORI, H., MCCOMBIE, W. R., OUYANG, S., SCHWARTZ, D., TANAKA, T., WU, J., ZHOU, S., CHILDS, K., DAVIDSON, R., LIN, H., QUESADA-OCAMPO, L., VAILLANCOURT, B., SAKAI, H., LEE, S. S., KIM, J., NUMA, H., ITOH, T., BUELL, C. R. & MATSUMOTO, T. 2013. Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data. *Rice*, 6, 4.
- KENIRY, A., HOPKINS, C. J., JEWELL, E., MORRISON, B., SPANGENBERG, G. C., EDWARDS, D. & BATLEY, J. 2006. Identification and characterization of simple sequence repeat (SSR) markers from *Fragaria x ananassa* expressed sequences. *Molecular Ecology Notes*, 6, 319-322.
- KHARABIAN-MASOULEH, A., WATERS, D. L. E., REINKE, R. F. & HENRY, R. J. 2011. Discovery of polymorphisms in starch-related genes in rice germplasm by amplification of pooled DNA and deeply parallel sequencing†. *Plant Biotechnology Journal*, 9, 1074-1085.
- KILIAN, A., WENZL, P., HUTTNER, E., CARLING, J., XIA, L., BLOIS, H., CAIG, V., HELLER-USZYNSKA, K., JACCOUD, D., HOPPER, C., ASCHENBRENNER-KILIAN, M., EVERE, M., PENG, K., CAYLA, C., HOK, P. & USZYNSKI, G. 2012. Diversity Arrays Technology: A Generic Genome Profiling Technology on Open Platforms. In: POMPANON, F. & BONIN, A. (eds.) *Data Production and Analysis in Population Genomics*. Humana Press.
- KRASILEVA, K. V., BUFFALO, V., BAILEY, P., PEARCE, S., AYLING, S., TABBITA, F., SORIA, M., WANG, S., CONSORTIUM, I., AKHUNOV, E., UAUY, C. & DUBCOVSKY, J. 2013. Separating homeologs by phasing in the tetraploid wheat transcriptome. *Genome Biology*, 14, R66.
- KWOK, P. Y., CARLSON, C., YAGER, T. D., ANKENER, W. & NICKERSON, D. A. 1994. Comparative-Analysis of Human DNA Variations by Fluorescence-Based Sequencing of Pcr Products. *Genomics*, 23, 138-144.
- KWON, S., SIMKO, I., HELLIER, B., MOU, B. & HU, J. 2013. Genome-wide association of 10 horticultural traits with expressed sequence tag-derived SNP markers in a collection of lettuce lines. *The Crop Journal*, 1, 25-33.
- LAI, J. S., LI, R. Q., XU, X., JIN, W. W., XU, M. L., ZHAO, H. N., XIANG, Z. K., SONG, W. B., YING, K., ZHANG, M., JIAO, Y. P., NI, P. X., ZHANG, J. G., LI, D., GUO, X. S., YE, K. X., JIAN, M., WANG, B., ZHENG, H. S., LIANG, H. Q., ZHANG, X. Q., WANG, S. C., CHEN, S. J., LI, J. S., FU, Y., SPRINGER, N. M., YANG, H. M., WANG, J. A., DAI, J. R., SCHNABLE, P. S. & WANG, J. 2010. Genome-wide patterns of genetic variation among elite maize inbred lines. *Nature Genetics*, 42, 1027-U158.

- LAI, K., BERKMAN, P. J., LORENC, M. T., DURAN, C., SMITS, L., MANOLI, S., STILLER, J. & EDWARDS, D. 2012a. WheatGenome.info: An Integrated Database and Portal for Wheat Genome Information. *Plant and Cell Physiology*, 53, e2.
- LAI, K., DURAN, C., BERKMAN, P. J., LORENC, M. T., STILLER, J., MANOLI, S., HAYDEN, M. J., FORREST, K. L., FLEURY, D., BAUMANN, U., ZANDER, M., MASON, A. S., BATLEY, J. & EDWARDS, D. 2012b. Single nucleotide polymorphism discovery from wheat next-generation sequence data. *Plant Biotechnology Journal*, 10, 743-749.
- LAI, K., LORENC, M. T. & EDWARDS, D. 2012c. Genomic databases for crop improvement. *Agronomy*, 2, 62-73.
- LAI, K., LORENC, M. T., LEE, H., BERKMAN, P. J., BAYER, P. E., MUHINDIRA, P. V., RUPEROAO, P., FITZGERALD, T. L., ZANDER, M., CHAN, C.-K. K., MANOLI, S., STILLER, J., BATLEY, J. & EDWARDS, D. 2014. Identification and characterisation of more than 4 million inter-varietal SNPs across the group 7 chromosomes of bread wheat. *Plant Biotechnology Journal*.
- LAM, H. M., XU, X., LIU, X., CHEN, W. B., YANG, G. H., WONG, F. L., LI, M. W., HE, W. M., QIN, N., WANG, B., LI, J., JIAN, M., WANG, J. A., SHAO, G. H., WANG, J., SUN, S. S. M. & ZHANG, G. Y. 2010. Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection. *Nature Genetics*, 42, 1053-U41.
- LANGMEAD, B., TRAPNELL, C., POP, M. & SALZBERG, S. L. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*, 10.
- LAUDENCIA-CHINGCUANCO, D., GANESHAN, S., YOU, F., FOWLER, B., CHIBBAR, R. & ANDERSON, O. 2011. Genome-wide gene expression analysis supports a developmental model of low temperature tolerance gene regulation in wheat (*Triticum aestivum* L.). *BMC Genomics*, 12, 299.
- LAWRENCE, C. J. 2007. MaizeGDB - The maize genetics and genomics database. In: EDWARDS, D. (ed.) *Methods in Molecular Biology*. Totowa, NJ, USA: Humana Press.
- LAWRENCE, C. J., SCHAEFFER, M. L., SEIGFRIED, T. E., CAMPBELL, D. A. & HARPER, L. C. 2007. MaizeGDB's new data types, resources and activities. *Nucleic Acids Research*, 35, D895-D900.
- LEE, H., LAI, K., LORENC, M. T., IMELFORT, M., DURAN, C. & EDWARDS, D. 2012. Bioinformatics tools and databases for analysis of next generation sequence data. *Briefings in Functional Genomics*, 2, 12-24.
- LI, H., HANDSAKER, B., WYSOKER, A., FENNELL, T., RUAN, J., HOMER, N., MARTH, G., ABECASIS, G., DURBIN, R. & GENOME PROJECT DATA PROCESSING, S. 2009a. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25, 2078-9.
- LI, H., RUAN, J. & DURBIN, R. 2008. Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Research*, 18, 1851-1858.
- LI, H., VIKRAM, P., SINGH, R. P., KILIAN, A., CARLING, J., SONG, J., BURGUENO-FERREIRA, J. A., BHAVANI, S., HUERTA-ESPINO, J., PAYNE, T., SEHGAL, D., WENZL, P. & SINGH, S. 2015. A high density GBS map of bread wheat and its application for dissecting complex disease resistance traits. *BMC Genomics*, 16, 216.
- LI, R., YU, C., LI, Y., LAM, T. W., YIU, S. M., KRISTIANSEN, K. & WANG, J. 2009b. SOAP2: an improved ultrafast tool for short read alignment. *Bioinformatics*, 25, 1966-7.
- LIM, G. A. C., JEWELL, E. G., XI, L., ERWIN, T. A., LOVE, C., BATLEY, J., SPANGENBERG, G. & EDWARDS, D. 2007. A comparative map viewer integrating genetic maps for *Brassica* and *Arabidopsis*. *BMC Plant Biology*, 7.
- LING, A. E., KAUR, J., BURGESS, B., HAND, M., HOPKINS, C. J., LI, X., LOVE, C. G., VARDY, M., WALKIEWICZ, M., SPANGENBERG, G., EDWARDS, D. & BATLEY, J. 2007. Characterization of simple sequence repeat markers derived in silico from *Brassica rapa* bacterial artificial chromosome sequences and their application in *Brassica napus*. *Molecular Ecology Notes*, 7, 273-277.
- LIU, C. Y., SHEPHERD, K. W. & RATHJEN, A. J. 1996. Improvement of durum wheat pastamaking and breadmaking qualities. *Cereal Chemistry*, 73, 155-166.
- LIU, L., LI, Y., LI, S., HU, N., HE, Y., PONG, R., LIN, D., LU, L. & LAW, M. 2012. Comparison of next-generation sequencing systems. *Journal of Biomedicine & Biotechnology*, 2012, 251364.
- LIU, Z. Y., SUN, Q. X., NI, Z. F., NEVO, E. & YANG, T. M. 2002. Molecular characterization of a novel powdery mildew

- resistance gene Pm30 in wheat originating from wild emmer. *Euphytica*, 123, 21-29.
- LOBELL, D. B., SCHLENKER, W. & COSTA-ROBERTS, J. 2011. Climate Trends and Global Crop Production Since 1980. *Science*, 333, 616-620.
- LOCKHART, D. J. & WINZELER, E. A. 2000. Genomics, gene expression and DNA arrays. *Nature*, 405, 827-836.
- LORENC, M. T., BOSKOVIC, Z., STILLER, J., DURAN, C. & EDWARDS, D. 2012a. Role of Bioinformatics as a Tool for Oilseed Brassica Species. In: EDWARDS, D., PARKIN, I. A. P. & BATLEY, J. (eds.) *Genetics, Genomics and Breeding of Oilseed Brassicas*. New Hampshire: Science Publishers Inc.
- LORENC, M. T., HAYASHI, S., STILLER, J., LEE, H., MANOLI, S., RUPERO, P., VISENDI, P., BERKMAN, P. J., LAI, K., BATLEY, J. & EDWARDS, D. 2012b. Discovery of Single Nucleotide Polymorphisms in Complex Genomes Using SGSAutoSNP. *Biology*, 1, 370-382.
- MACKAY, I. & POWELL, W. 2007. Methods for linkage disequilibrium mapping in crops. *Trends in Plant Science*, 12, 57-63.
- MARDIS, E. R. 2008. The impact of next-generation sequencing technology on genetics. *Trends in Genetics*, 24, 133-41.
- MARGULIES, M., EGHLOM, M., ALTMAN, W. E., ATTILA, S., BADER, J. S., BEMBEN, L. A., BERKA, J., BRAVERMAN, M. S., CHEN, Y. J., CHEN, Z., DEWELL, S. B., DU, L., FIERRO, J. M., GOMES, X. V., GODWIN, B. C., HE, W., HELGESEN, S., HO, C. H., IRZYK, G. P., JANDO, S. C., ALENQUER, M. L., JARVIE, T. P., JIRAGE, K. B., KIM, J. B., KNIGHT, J. R., LANZA, J. R., LEAMON, J. H., LEFKOWITZ, S. M., LEI, M., LI, J., LOHMAN, K. L., LU, H., MAKHIJANI, V. B., MCDADE, K. E., MCKENNA, M. P., MYERS, E. W., NICKERSON, E., NOBILE, J. R., PLANT, R., PUC, B. P., RONAN, M. T., ROTH, G. T., SARKIS, G. J., SIMONS, J. F., SIMPSON, J. W., SRINIVASAN, M., TARTARO, K. R., TOMASZ, A., VOGT, K. A., VOLKMER, G. A., WANG, S. H., WANG, Y., WEINER, M. P., YU, P., BEGLEY, R. F. & ROTHBERG, J. M. 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, 437, 376-80.
- MARSHALL, D. J., HAYWARD, A., EALES, D., IMELFORT, M., STILLER, J., BERKMAN, P. J., CLARK, T., MCKENZIE, M., LAI, K., DURAN, C., BATLEY, J. & EDWARDS, D. 2010. Targeted identification of genomic regions using TAGdb. *Plant Methods*, 6, 19.
- MARTH, G. T., KORF, I., YANDELL, M. D., YEH, R. T., GU, Z. J., ZAKERI, H., STITZEL, N. O., HILLIER, L., KWOK, P. Y. & GISH, W. R. 1999. A general approach to single-nucleotide polymorphism discovery. *Nature Genetics*, 23, 452-456.
- MARTIN, C. & PAZ-ARES, J. 1997. MYB transcription factors in plants. *Trends in Genetics*, 13, 67-73.
- MATTHEWS, D. E., CAROLLO, V. L., LAZO, G. R. & ANDERSON, O. D. 2003. GrainGenes, the genome database for small-grain crops. *Nucleic Acids Research*, 31, 183-6.
- MATZ, M. V., MEYER, E. & AGLYAMOVA, G. 2009. Methods and compositions for nucleic acid sequencing. Google Patents.
- MAXAM, A. M. & GILBERT, W. 1977. A new method for sequencing DNA. *Proceedings of the National Academy of Sciences of the United States of America*, 74, 560-4.
- MCFADDEN, E. & SEARS, E. 1946. The origin of *Triticum spelta* and its free-threshing hexaploid relatives. *Journal of Heredity*, 37, 81 - 89.
- METZKER, M. L. 2010. Sequencing technologies - the next generation. *Nature Reviews Genetics*, 11, 31-46.
- MEYER, E., AGLYAMOVA, G., WANG, S., BUCHANAN-CARTER, J., ABREGO, D., COLBOURNE, J., WILLIS, B. & MATZ, M. 2009. Sequencing and de novo analysis of a coral larval transcriptome using 454 GSFlx. *BMC Genomics*, 10, 219.
- MIKKELSEN, T. S., KU, M., JAFFE, D. B., ISSAC, B., LIEBERMAN, E., GIANNOUKOS, G., ALVAREZ, P., BROCKMAN, W., KIM, T. K., KOCH, R. P., LEE, W., MENDENHALL, E., O'DONOVAN, A., PRESSER, A., RUSS, C., XIE, X., MEISSNER, A., WERNIG, M., JAENISCH, R., NUSBAUM, C., LANDER, E. S. & BERNSTEIN, B. E. 2007. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature*, 448, 553-60.
- MILNE, I., SHAW, P., STEPHEN, G., BAYER, M., CARDLE, L., THOMAS, W. T., FLAVELL, A. J. & MARSHALL, D. 2010. Flapjack--graphical genotype visualization. *Bioinformatics*, 26, 3133-4.
- MISKELLY, D. M. Asian foods - research, product range and quality attributes. 55h Australian Cereal Chemistry Conference & Pacific Rim Symposium, 2005 Wagga Wagga, NSW Australia. The Regional Institute Ltd.
- MOCHIDA, K., SAISHO, D., YOSHIDA, T., SAKURAI, T. & SHINOZAKI, K. 2008. TriMEDB: A database to integrate transcribed markers and facilitate genetic studies of the tribe Triticeae. *BMC Plant Biology*, 8, 72.

- MONACO, M. K., STEIN, J., NAITHANI, S., WEI, S., DHARMAWARDHANA, P., KUMARI, S., AMARASINGHE, V., YOUENS-CLARK, K., THOMASON, J., PREECE, J., PASTERNAK, S., OLSON, A., JIAO, Y., LU, Z., BOLSER, D., KERHORNOU, A., STAINES, D., WALTS, B., WU, G., D'EUSTACHIO, P., HAW, R., CROFT, D., KERSEY, P. J., STEIN, L., JAISWAL, P. & WARE, D. 2014. Gramene 2013: comparative plant genomics resources. *Nucleic Acids Research*, 42, D1193-9.
- MOROZOVA, O. & MARRA, M. A. 2008. Applications of next-generation sequencing technologies in functional genomics. *Genomics*, 92, 255-264.
- MORRELL, P. L., BUCKLER, E. S. & ROSS-IBARRA, J. 2011. Crop genomics: advances and applications. *Nature Reviews Genetics*, 13, 85-96.
- MOTT, R. 1997. EST_GENOME: a program to align spliced DNA sequences to unspliced genomic DNA. *Computer Applications in the Biosciences*, 13, 477-8.
- MUELLER, U. G. & WOLFENBARGER, L. L. 1999. AFLP genotyping and fingerprinting. *Trends Ecol Evol*, 14, 389-394.
- NAKAYA, A. & ISOBE, S. N. 2012. Will genomic selection be a practical method for plant breeding? *Annals of Botany*, 110, 1303-1316.
- NEDERBRAGT, A. J. 2014. On the middle ground between open source and commercial software-the case of the Newbler program. *Genome Biology*, 15, 113.
- NESBITT, M. & SAMUEL, D. 1998. Wheat domestication: Archaeobotanical evidence. *Science*, 279, 1431.
- NIE, X., LI, B., WANG, L., LIU, P., BIRADAR, S. S., LI, T., DOLEZEL, J., EDWARDS, D., LUO, M. & WEINING, S. 2012. Development of chromosome-arm-specific microsatellite markers in *Triticum aestivum* (Poaceae) using NGS technology. *American Journal of Botany*, 99, e369-71.
- NING, Z. & MONTGOMERY, S. 2010. Out of the sequencer and into the wiki as we face new challenges in genome informatics. *Genome Biology*, 11, 308.
- NYREN, P. & LUNDIN, A. 1985. Enzymatic method for continuous monitoring of inorganic pyrophosphate synthesis. *Analytical Biochemistry*, 151, 504-9.
- O'SULLIVAN, H. 2007. GrainGenes - A genomic database for Triticeae and Avena. In: EDWARDS, D. (ed.) *Methods in Molecular Biology*. Totowa, NJ, USA: Humana Press.
- OGBONNAYA, F. C. & ABDALLA, O. 2013. Synthetic Hexaploids: Harnessing Species of the Primary Gene Pool for Wheat Improvement. In: JANICK, J. (ed.) *Plant Breeding Reviews*. 1 ed. Somerset: Wiley.
- OHNO, S. 1970. *Evolution by gene duplication*, Berlin ; New York :, Springer-Verlag.
- OLIVER, R., LAZO, G., LUTZ, J., RUBENFIELD, M., TINKER, N., ANDERSON, J., WISNIEWSKI MOREHEAD, N., ADHIKARY, D., JELLEN, E., MAUGHAN, P. J., BROWN GUEDIRA, G., CHAO, S., BEATTIE, A., CARSON, M., RINES, H., OBERT, D., BONMAN, J. M. & JACKSON, E. 2011. Model SNP development for complex genomes based on hexaploid oat using high-throughput 454 sequencing technology. *BMC Genomics*, 12, 77.
- OONO, Y., KOBAYASHI, F., KAWAHARA, Y., YAZAWA, T., HANDA, H., ITOH, T. & MATSUMOTO, T. 2013. Characterisation of the wheat (*Triticum aestivum* L.) transcriptome by de novo assembly for the discovery of phosphate starvation-responsive genes: gene expression in Pi-stressed wheat. *BMC Genomics*, 14, 77.
- ORITA, M., IWAHANA, H., KANAZAWA, H., HAYASHI, K. & SEKIYA, T. 1989. Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms. *Proceedings of the National Academy of Sciences*, 86, 2766-2770.
- OSSOWSKI, S., SCHNEEBERGER, K., CLARK, R. M., LANZ, C., WARTHMANN, N. & WEIGEL, D. 2008. Sequencing of natural strains of *Arabidopsis thaliana* with short reads. *Genome Research*, 18, 2024-2033.
- OUYANG, S., ZHU, W., HAMILTON, J., LIN, H., CAMPBELL, M., CHILDS, K., THIBAUD-NISSEN, F., MALEK, R. L., LEE, Y., ZHENG, L., ORVIS, J., HAAS, B., WORTMAN, J. & BUELL, C. R. 2007. The TIGR Rice Genome Annotation Resource: Improvements and new features. *Nucleic Acids Research*, 35, D883-D887.
- PABINGER, S., DANDER, A., FISCHER, M., SNAJDER, R., SPERK, M., EFREMOVA, M., KRABICHLER, B., SPEICHER, M. R., ZSCHOCKE, J. & TRAJANOSKI, Z. 2013. A survey of tools for variant analysis of next-generation genome

- sequencing data. *Briefings in Bioinformatics*.
- PARK, P. J. 2009. ChIP-seq: advantages and challenges of a maturing technology. *Nature reviews. Genetics*, 10, 669-80.
- PAUX, E., SOURDILLE, P., SALSE, J., SAINTENAC, C., CHOULET, F., LEROY, P., KOROL, A., MICHALAK, M., KIANIAN, S., SPIELMEYER, W., LAGUDAH, E., SOMERS, D., KILIAN, A., ALAUX, M., VAUTRIN, S., BERGES, H., EVERSOLE, K., APPELS, R., SAFAR, J., SIMKOVA, H., DOLEZEL, J., BERNARD, M. & FEUILLET, C. 2008. A physical map of the 1-gigabase bread wheat chromosome 3B. *Science*, 322, 101-4.
- PEÑA, R. 2002. Wheat for bread and other foods. *bread wheat improvement and production. Curtis, BC; Rajaram, SS and Gómez, M. H (eds). Food and Agriculture Organization of the United Nations. Rome*, 483-542.
- PENNER, G., BUSH, A., WISE, R., KIM, W., DOMIER, L., KASHA, K., LAROCHE, A., SCOLES, G., MOLNAR, S. & FEDAK, G. 1993. Reproducibility of random amplified polymorphic DNA (RAPD) analysis among laboratories. *Genome Research*, 2, 341-345.
- PETERSEN, G., SEBERG, O., YDE, M. & BERTHELSEN, K. 2006. Phylogenetic relationships of Triticum and Aegilops and evidence for the origin of the A, B, and D genomes of common wheat (*Triticum aestivum*). *Molecular Phylogenetics and Evolution*, 39, 70-82.
- PFEIFER, M., KUGLER, K. G., SANDVE, S. R., ZHAN, B., RUDI, H., HVIDSTEN, T. R., INTERNATIONAL WHEAT GENOME SEQUENCING, C., MAYER, K. F. & OLSEN, O. A. 2014. Genome interplay in the grain transcriptome of hexaploid bread wheat. *Science*, 345, 1250091.
- PICOULT-NEWBERG, L., IDEKER, T. E., POHL, M. G., TAYLOR, S. L., DONALDSON, M. A., NICKERSON, D. A. & BOYCE-JACINO, M. 1999. Mining SNPs From EST Databases. *Genome Research*, 9, 167-174.
- PIERCE, B. A. 2011. *Genetics: A Conceptual Approach*, W. H. Freeman.
- PLECHAKOVA, O., TRANCHANT-DUBREUIL, C., BENEDET, F., COUDERC, M., TINAUT, A., VIADER, V., DE BLOCK, P., HAMON, P., CAMPA, C., DE KOCHKO, A., HAMON, S. & PONCET, V. 2009. MoccaDB - an integrative database for functional, comparative and diversity studies in the Rubiaceae family. *Bmc Plant Biology*, 9, 123.
- POLAND, J., ENDELMAN, J., DAWSON, J., RUTKOSKI, J., WU, S., MANES, Y., DREISIGACKER, S., CROSSA, J., SÁNCHEZ-VILLEDA, H., SORRELLS, M. & JANNINK, J.-L. 2012. Genomic Selection in Wheat Breeding using Genotyping-by-Sequencing. *The Plant Genome*, 5, 103-113.
- PONT, C., MURAT, F., GUIZARD, S., FLORES, R., FOUCRIER, S., BIDET, Y., QURAISHI, U. M., ALAUX, M., DOLEŽEL, J., FAHIMA, T., BUDAK, H., KELLER, B., SALVI, S., MACCAFERRI, M., STEINBACH, D., FEUILLET, C., QUESNEVILLE, H. & SALSE, J. 2013. Wheat syntenome unveils new evidences of contrasted evolutionary plasticity between paleo- and neoduplicated subgenomes. *The Plant Journal*, 76, 1030-1044.
- POWELL, W., MORGANTE, M., ANDRE, C., HANAFAY, M., VOGEL, J., TINGEY, S. & RAFALSKI, A. 1996. The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. *Molecular Breeding*, 2, 225-238.
- PRASAD, M., VARSHNEY, R. K., ROY, J. K., BALYAN, H. S. & GUPTA, P. K. 2000. *The use of microsatellites for detecting DNA polymorphism, genotype identification and genetic diversity in wheat*, Heidelberg, Allemagne, Springer.
- QUAIL, M. A., SMITH, M., COUPLAND, P., OTTO, T. D., HARRIS, S. R., CONNOR, T. R., BERTONI, A., SWERDLOW, H. P. & GU, Y. 2012. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics*, 13, 341.
- RAFALSKI, A. 2002. Applications of single nucleotide polymorphisms in crop genetics. *Current Opinion in Plant Biology*, 5, 94-100.
- ROBERTS, A. W., ROBERTS, E. M. & DELMER, D. P. 2002. Cellulose synthase (CesA) genes in the green alga *Mesotaenium caldariorum*. *Eukaryotic Cell*, 1, 847-55.
- ROBERTSON, G., HIRST, M., BAINBRIDGE, M., BILENKY, M., ZHAO, Y. J., ZENG, T., EUSKIRCHEN, G., BERNIER, B., VARHOL, R., DELANEY, A., THIESSEN, N., GRIFFITH, O. L., HE, A., MARRA, M., SNYDER, M. & JONES, S. 2007. Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. *Nature Methods*, 4, 651-657.
- ROBINSON, A. J., LOVE, C. G., BATLEY, J., BARKER, G. & EDWARDS, D. 2004. Simple sequence repeat marker loci discovery

- using SSR primer. *Bioinformatics*, 20, 1475-1476.
- ROBISON, K. 2010. Editorial: Second-generation sequencing. *Briefings in Bioinformatics*, 11, 455-6.
- RONAGHI, M., KARAMOHAMED, S., PETTERSSON, B., UHLÉN, M. & NYRÉN, P. 1996. Real-Time DNA Sequencing Using Detection of Pyrophosphate Release. *Analytical Biochemistry*, 242, 84-89.
- ROSENBERG, N. A., HUANG, L., JEWETT, E. M., SZPIECH, Z. A., JANKOVIC, I. & BOEHNKE, M. 2010. Genome-wide association studies in diverse populations. *Nature Reviews Genetics*, 11, 356-366.
- ROTHBERG, J. M., HINZ, W., REARICK, T. M., SCHULTZ, J., MILESKI, W., DAVEY, M., LEAMON, J. H., JOHNSON, K., MILGREW, M. J., EDWARDS, M., HOON, J., SIMONS, J. F., MARRAN, D., MYERS, J. W., DAVIDSON, J. F., BRANTING, A., NOBLE, J. R., PUC, B. P., LIGHT, D., CLARK, T. A., HUBER, M., BRANCIFORTE, J. T., STONER, I. B., CAWLEY, S. E., LYONS, M., FU, Y., HOMER, N., SEDOVA, M., MIAO, X., REED, B., SABINA, J., FEIERSTEIN, E., SCHORN, M., ALANJARY, M., DIMALANTA, E., DRESSMAN, D., KASINSKAS, R., SOKOLSKY, T., FIDANZA, J. A., NAMSARAEV, E., MCKERNAN, K. J., WILLIAMS, A., ROTH, G. T. & BUSTILLO, J. 2011. An integrated semiconductor device enabling non-optical genome sequencing. *Nature*, 475, 348-52.
- ROTHBERG, J. M. & LEAMON, J. H. 2008. The development and impact of 454 sequencing. *Nature Biotechnology*, 26, 1117-24.
- ŠAFÁŘ, J., ŠIMKOVÁ, H., KUBALÁKOÁ, M., ČÍHALÍKOVÁ, J., SUCHÁNKOVÁ, P., BARTOŠ, J. & DOLEŽEL, J. 2010. Development of chromosome-specific BAC resources for genomics of bread wheat. *Cytogenetic and Genome Research*, 129, 211-223.
- SALAMINI, F., OZKAN, H., BRANDOLINI, A., SCHAFER-PREGL, R. & MARTIN, W. 2002. Genetics and geography of wild cereal domestication in the near east. *Nature Reviews Genetics*, 3, 429-441.
- SAMBROOK, J. & RUSSEL, D. W. 2001. *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor, NY, Cold Spring Harbor Laboratory Press. .
- SANGER, F., AIR, G. M., BARRELL, B. G., BROWN, N. L., COULSON, A. R., FIDDES, C. A., HUTCHISON, C. A., SLOCOMBE, P. M. & SMITH, M. 1977a. Nucleotide sequence of bacteriophage phi X174 DNA. *Nature*, 265, 687-95.
- SANGER, F., NICKLEN, S. & COULSON, A. R. 1977b. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*, 74, 5463-7.
- SANSALONI, C. P., PETROLI, C. D., CARLING, J., HUDSON, C. J., STEANE, D. A., MYBURG, A. A., GRATTAPAGLIA, D., VAILLANCOURT, R. E. & KILIAN, A. 2010. A high-density Diversity Arrays Technology (DArT) microarray for genome-wide genotyping in Eucalyptus. *Plant Methods*, 6, 16.
- SAVAGE, D., BATLEY, J., ERWIN, T., LOGAN, E., LOVE, C. G., LIM, G. A. C., MONGIN, E., BARKER, G., SPANGENBERG, G. C. & EDWARDS, D. 2005. SNPServer: a real-time SNP discovery tool. *Nucleic Acids Research*, 33, W493-W495.
- SCHAEFFER, M. L., HARPER, L. C., GARDINER, J. M., ANDORF, C. M., CAMPBELL, D. A., CANNON, E. K., SEN, T. Z. & LAWRENCE, C. J. 2011. MaizeGDB: curation and outreach go hand-in-hand. *Database : The Journal of Biological Databases and Curation*, 2011, bar022.
- SCHATZ, M. C., WITKOWSKI, J. & MCCOMBIE, W. R. 2012. Current challenges in de novo plant genome sequencing and assembly. *Genome Biology*, 13, 243.
- SCHLOTTERER, C. 2004. The evolution of molecular markers [mdash] just a matter of fashion? *Nature Reviews Genetics*, 5, 63-69.
- SCHREIBER, A. W., HAYDEN, M. J., FORREST, K. L., KONG, S. L., LANGRIDGE, P. & BAUMANN, U. 2012. Transcriptome-scale homoeolog-specific transcript assemblies of bread wheat. *BMC Genomics*, 13, 492.
- SCHUSTER, S. C. 2008. Next-generation sequencing transforms today's biology. *Nature Methods*, 5, 16-8.
- SEMAGN, K., BABU, R., HEARNE, S. & OLSEN, M. 2014. Single nucleotide polymorphism genotyping using Kompetitive Allele Specific PCR (KASP): overview of the technology and its application in crop improvement. *Molecular Breeding*, 33, 1-14.
- SHENDURE, J. & JI, H. 2008. Next-generation DNA sequencing. *Nat Biotechnol*, 26, 1135-45.
- SHEWRY, P. R. 2009. Wheat. *Journal of Experimental Botany*, 60, 1537-1553.

- SIMONSEN, M., MAILUND, T. & PEDERSEN, C. N. S. 2008. Rapid Neighbour-Joining. *Algorithms in Bioinformatics*, Wabi 2008, 5251, 113-122.
- SMITH, L. M., SANDERS, J. Z., KAISER, R. J., HUGHES, P., DODD, C., CONNELL, C. R., HEINER, C., KENT, S. B. & HOOD, L. E. 1986. Fluorescence detection in automated DNA sequence analysis. *Nature*, 321, 674-9.
- SMITH, T. F. & WATERMAN, M. S. 1981. Identification of common molecular subsequences. *Journal of molecular biology*, 147, 195-7.
- SOLTIS, D. E., ALBERT, V. A., LEEBENS-MACK, J., BELL, C. D., PATERSON, A. H., ZHENG, C., SANKOFF, D., DEPAMPHILIS, C. W., WALL, P. K. & SOLTIS, P. S. 2009. Polyploidy and angiosperm diversification. *American Journal of Botany*, 96, 336-48.
- SOMERVILLE, C. 2006. The billion-ton biofuels vision. *Science*, 312, 1277.
- SONG, Q., HYTEN, D. L., JIA, G., QUIGLEY, C. V., FICKUS, E. W., NELSON, R. L. & CREGAN, P. B. 2013. Development and evaluation of SoySNP50K, a high-density genotyping array for soybean. *PLoS One*, 8, e54985.
- STEIN, N. & GRANER, A. 2005. Map-Based Gene Isolation in Cereal Genomes. In: GUPTA, P. & VARSHNEY, R. (eds.) *Cereal Genomics*. Springer Netherlands.
- SUBBAIYAN, G. K., WATERS, D. L., KATIYAR, S. K., SADANANDA, A. R., VADDADI, S. & HENRY, R. J. 2012. Genome-wide DNA polymorphisms in elite indica rice inbreds discovered by whole-genome sequencing. *Plant Biotechnology Journal*, 10, 623-34.
- SUZEK, B. E., HUANG, H., MCGARVEY, P., MAZUMDER, R. & WU, C. H. 2007. UniRef: comprehensive and non-redundant UniProt reference clusters. *Bioinformatics*, 23, 1282-8.
- SYVANEN, A. C. 2001. Accessing genetic variation: genotyping single nucleotide polymorphisms. *Nature Reviews Genetics*, 2, 930-42.
- SZŰCS, P., BLAKE, V. C., BHAT, P. R., CHAO, S., CLOSE, T. J., CUESTA-MARCOS, A., MUEHLBAUER, G. J., RAMSAY, L., WAUGH, R. & HAYES, P. M. 2009. An Integrated Resource for Barley Linkage Map and Malting Quality QTL Alignment. *The Plant Genome*, 2, 134-140.
- TALBERT, L. E., SMITH, L. Y. & BLAKE, M. K. 1998. More than one origin of hexaploid wheat is indicated by sequence comparison of low-copy DNA. *Genome*, 41, 402-407.
- TANKSLEY, S. D. & MCCOUCH, S. R. 1997. Seed banks and molecular maps: unlocking genetic potential from the wild. *Science*, 277, 1063-6.
- TANKSLEY, S. D., YOUNG, N. D., PATERSON, A. H. & BONIERBALE, M. W. 1989. RFLP Mapping in Plant-Breeding - New Tools for an Old Science. *Nature Biotechnology*, 7, 257-264.
- TIAN, F., BRADBURY, P. J., BROWN, P. J., HUNG, H., SUN, Q., FLINT-GARCIA, S., ROCHEFORD, T. R., MCMULLEN, M. D., HOLLAND, J. B. & BUCKLER, E. S. 2011. Genome-wide association study of leaf architecture in the maize nested association mapping population. *Nature Genetics*, 43, 159-162.
- TILMAN, D., CASSMAN, K. G., MATSON, P. A., NAYLOR, R. & POLASKY, S. 2002. Agricultural sustainability and intensive production practices. *Nature*, 418, 671-7.
- TOLLENAERE, R., HAYWARD, A., DALTON-MORGAN, J., CAMPBELL, E., MCCLANDERS, J., LORENC, M., MANOLI, S., STILLER, J., RAMAN, R., RAMAN, H., EDWARDS, D. & BATLEY, J. 2012. Identification and characterisation of candidate Rlm4 blackleg resistance genes in *Brassica napus* using next generation sequencing. *Plant Biotechnology Journal*, 10, 709-715.
- TOWN, C. D., CHEUNG, F., MAITI, R., CRABTREE, J., HAAS, B. J., WORTMAN, J. R., HINE, E. E., ALTHOFF, R., ARBOGAST, T. S., TALLON, L. J., VIGOUROUX, M., TRICK, M. & BANCROFT, I. 2006. Comparative genomics of *Brassica oleracea* and *Arabidopsis thaliana* reveal gene loss, fragmentation, and dispersal after polyploidy. *The Plant Cell*, 18, 1348-59.
- TRAPNELL, C., WILLIAMS, B. A., PERTEA, G., MORTAZAVI, A., KWAN, G., VAN BAREN, M. J., SALZBERG, S. L., WOLD, B. J. & PACHTER, L. 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnology*, 28, 511-515.
- TREBBI, D., MACCAFERRI, M., DE HEER, P., SØRENSEN, A., GIULIANI, S., SALVI, S., SANGUINETI, M., MASSI, A., VAN DER

- VOSSEN, E. & TUBEROZA, R. 2011. High-throughput SNP discovery and genotyping in durum wheat (*Triticum durum*). *Theoretical and Applied Genetics*, 123, 555-569.
- TRICK, M., ADAMSKI, N., MUGFORD, S., JIANG, C.-C., FEBRER, M. & UAUY, C. 2012. Combining SNP discovery from next-generation sequencing data with bulked segregant analysis (BSA) to fine-map genes in polyploid wheat. *BMC Plant Biology*, 12, 14.
- TRICK, M., LONG, Y., MENG, J. L. & BANCROFT, I. 2009. Single nucleotide polymorphism (SNP) discovery in the polyploid *Brassica napus* using Solexa transcriptome sequencing. *Plant Biotechnology Journal*, 7, 334-346.
- TURCATTI, G., ROMIEU, A., FEDURCO, M. & TAIRI, A. P. 2008. A new class of cleavable fluorescent nucleotides: synthesis and optimization as reversible terminators for DNA sequencing by synthesis. *Nucleic Acids Research*, 36.
- USDA 2011. Production, Supply and Distribution Online. January 2011 ed.: USDA.
- USDA 2014. Grain: World Markets and Trade.
- VARSHNEY, R. K., NAYAK, S. N., MAY, G. D. & JACKSON, S. A. 2009. Next-generation sequencing technologies and their implications for crop genetics and breeding. *Trends in Biotechnology*, 27, 522-530.
- VARSHNEY, R. K., SIGMUND, R., BÖRNER, A., KORZUN, V., STEIN, N., SORRELLS, M. E., LANGRIDGE, P. & GRANER, A. 2005. Interspecific transferability and comparative mapping of barley EST-SSR markers in wheat, rye and rice. *Plant Science*, 168, 195-202.
- VELCULESCU, V. E., ZHANG, L., VOGELSTEIN, B. & KINZLER, K. W. 1995. Serial analysis of gene expression. *Science*, 270, 484-7.
- VELCULESCU, V. E., ZHANG, L., ZHOU, W., VOGELSTEIN, J., BASRAI, M. A., BASSETT, D. E., JR., HIETER, P., VOGELSTEIN, B. & KINZLER, K. W. 1997. Characterization of the yeast transcriptome. *Cell*, 88, 243-51.
- VIDAL, R. O., MONDEGO, J. M., POT, D., AMBROSIO, A. B., ANDRADE, A. C., PEREIRA, L. F., COLOMBO, C. A., VIEIRA, L. G., CARAZZOLLE, M. F. & PEREIRA, G. A. 2010. A high-throughput data mining of single nucleotide polymorphisms in Coffea species expressed sequence tags suggests differential homeologous gene expression in the allotetraploid *Coffea arabica*. *Plant Physiology*, 154, 1053-66.
- VISENDI, P., BATLEY, J. & EDWARDS, D. 2013. Next Generation Characterisation of Cereal Genomes for Marker Discovery. *Biology*, 2, 1357-1377.
- VOS, P., HOGERS, R., BLEEKER, M., REIJANS, M., VAN DE LEE, T., HORNES, M., FRIJTERS, A., POT, J., PELEMAN, J., KUIPER, M. & ET AL. 1995. AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research*, 23, 4407-14.
- WANG, J., LIN, M., CRENSHAW, A., HUTCHINSON, A., HICKS, B., YEAGER, M., BERNDT, S., HUANG, W. Y., HAYES, R. B., CHANOCK, S. J., JONES, R. C. & RAMAKRISHNAN, R. 2009a. High-throughput single nucleotide polymorphism genotyping using nanofluidic Dynamic Arrays. *BMC Genomics*, 10.
- WANG, J., LIU, D., GUO, X., YANG, W., WANG, X., ZHAN, K. & ZHANG, A. 2011a. Variability of Gene Expression After Polyhaploidization in Wheat (*Triticum aestivum* L.). *G3: Genes, Genomes, Genetics*, 1, 27-33.
- WANG, S., WONG, D., FORREST, K., ALLEN, A., CHAO, S., HUANG, B. E., MACCAFERRI, M., SALVI, S., MILNER, S. G., CATTIVELLI, L., MASTRANGELO, A. M., WHAN, A., STEPHEN, S., BARKER, G., WIESEKE, R., PLIESKE, J., INTERNATIONAL WHEAT GENOME SEQUENCING, C., LILLEMO, M., MATHER, D., APPELS, R., DOLFERUS, R., BROWN-GUEDIRA, G., KOROL, A., AKHUNOVA, A. R., FEUILLET, C., SALSE, J., MORGANTE, M., POZNIAK, C., LUO, M.-C., DVORAK, J., MORELL, M., DUBCOVSKY, J., GANAL, M., TUBEROZA, R., LAWLEY, C., MIKOULITCH, I., CAVANAGH, C., EDWARDS, K. J., HAYDEN, M. & AKHUNOV, E. 2014. Characterization of polyploid wheat genomic diversity using a high-density 90 000 single nucleotide polymorphism array. *Plant Biotechnology Journal*, n/a-n/a.
- WANG, X., WANG, H., WANG, J., SUN, R., WU, J., LIU, S., BAI, Y., MUN, J.-H., BANCROFT, I., CHENG, F., HUANG, S., LI, X., HUA, W., WANG, J., WANG, X., FREELING, M., PIRES, J. C., PATERSON, A. H., CHALHOUB, B., WANG, B., HAYWARD, A., SHARPE, A. G., PARK, B.-S., WEISCHAAR, B., LIU, B., LI, B., LIU, B., TONG, C., SONG, C., DURAN, C., PENG, C., GENG, C., KOH, C., LIN, C., EDWARDS, D., MU, D., SHEN, D., SOUMPOUROU, E., LI, F., FRASER, F., CONANT, G., LASSALLE, G., KING, G. J., BONNEMA, G., TANG, H., WANG, H., BELCRAM, H., ZHOU, H., HIRAKAWA, H., ABE, H., GUO, H., WANG, H., JIN, H., PARKIN, I. A. P., BATLEY, J., KIM, J.-S., JUST, J., LI, J., XU, J., DENG, J., KIM, J. A., LI, J., YU,

- J., MENG, J., WANG, J., MIN, J., POULAIN, J., HATAKEYAMA, K., WU, K., WANG, L., FANG, L., TRICK, M., LINKS, M. G., ZHAO, M., JIN, M., RAMCHIARY, N., DROU, N., BERKMAN, P. J., CAI, Q., HUANG, Q., LI, R., TABATA, S., CHENG, S., ZHANG, S., ZHANG, S., HUANG, S., SATO, S., SUN, S., KWON, S.-J., CHOI, S.-R., LEE, T.-H., FAN, W., ZHAO, X., TAN, X., XU, X., WANG, Y., QIU, Y., YIN, Y., LI, Y., DU, Y., et al. 2011b. The genome of the mesopolyploid crop species *Brassica rapa*. *Nature Genetics*, 43, 1035-U157.
- WANG, Z., GERSTEIN, M. & SNYDER, M. 2009b. RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics*, 10, 57-63.
- WANJUGI, H., COLEMAN-DERR, D., HUO, N., KIANIAN, S. F., LUO, M. C., WU, J., ANDERSON, O. & GU, Y. Q. 2009. Rapid development of PCR-based genome-specific repetitive DNA junction markers in wheat. *Genome*, 52, 576-87.
- WEBER, J. L. & MAY, P. E. 1989. Abundant Class of Human DNA Polymorphisms Which Can Be Typed Using the Polymerase Chain-Reaction. *American Journal of Human Genetics*, 44, 388-396.
- WEGRZYN, J., MAIN, D., FIGUEROA, B., CHOI, M., YU, J., NEALE, D., JUNG, S., LEE, T., STANTON, M., ZHENG, P., FICKLIN, S., CHO, I., PEACE, C., EVANS, K., VOLK, G., ORAGUZIE, N., CHEN, C., OLMSTEAD, M., GMITTER, G. & ABBOTT, A. 2012. Uniform standards for genome databases in forest and fruit trees. *Tree Genetics & Genomes*, 1-9.
- WHEELER, D. L., BARRETT, T., BENSON, D. A., BRYANT, S. H., CANESE, K., CHETVERNIN, V., CHURCH, D. M., DICUCCIO, M., EDGAR, R., FEDERHEN, S., FEOLO, M., GEER, L. Y., HELMBERG, W., KAPUSTIN, Y., KHOBAYKO, O., LANDSMAN, D., LIPMAN, D. J., MADDEN, T. L., MAGLOTT, D. R., MILLER, V., OSTELL, J., PRUITT, K. D., SCHULER, G. D., SHUMWAY, M., SEQUEIRA, E., SHERRY, S. T., SIROTKIN, K., SOUVOROV, A., STARICHENKO, G., TATUSOV, R. L., TATUSOVA, T. A., WAGNER, L. & YASCHENKO, E. 2008. Database resources of the national center for biotechnology information. *Nucleic Acids Research*, 36, D13-D21.
- WICKER, T., MAYER, K. F. X., GUNDLACH, H., MARTIS, M., STEUERNAGEL, B., SCHOLZ, U., SIMKOVA, H., KUBALAKOVA, M., CHOULET, F., TAUDIEN, S., PLATZER, M., FEUILLET, C., FAHIMA, T., BUDAK, H., DOLEZEL, J., KELLER, B. & STEIN, N. 2011. Frequent Gene Movement and Pseudogene Evolution Is Common to the Large and Complex Genomes of Wheat, Barley, and Their Relatives. *Plant Cell*, 23, 1706-1718.
- WILKINSON, P. A., WINFIELD, M. O., BARKER, G. L., ALLEN, A. M., BURRIDGE, A., COGHILL, J. A. & EDWARDS, K. J. 2012. CerealsDB 2.0: an integrated resource for plant breeders and scientists. *BMC Bioinformatics*, 13, 219.
- WILLIAMS, L. E., LEMOINE, R. & SAUER, N. 2000. Sugar transporters in higher plants--a diversity of roles and complex regulation. *Trends in Plant Science*, 5, 283-90.
- WILSON, I. D., BARKER, G. L., BESWICK, R. W., SHEPHERD, S. K., LU, C., COGHILL, J. A., EDWARDS, D., OWEN, P., LYONS, R., PARKER, J. S., LENTON, J. R., HOLDSWORTH, M. J., SHEWRY, P. R. & EDWARDS, K. J. 2004. A transcriptomics resource for wheat functional genomics. *Plant Biotechnology Journal*, 2, 495-506.
- WILSON, I. D., BARKER, G. L., LU, C., COGHILL, J. A., BESWICK, R. W., LENTON, J. R. & EDWARDS, K. J. 2005. Alteration of the embryo transcriptome of hexaploid winter wheat (*Triticum aestivum* cv. Mercia) during maturation and germination. *Functional & Integrative Genomics*, 5, 144-54.
- WINFIELD, M. O., WILKINSON, P. A., ALLEN, A. M., BARKER, G. L., COGHILL, J. A., BURRIDGE, A., HALL, A., BRENCHLEY, R. C., D'AMORE, R., HALL, N., BEVAN, M. W., RICHMOND, T., GERHARDT, D. J., JEDDELOH, J. A. & EDWARDS, K. J. 2012. Targeted re-sequencing of the allohexaploid wheat exome. *Plant Biotechnology Journal*, 10, 733-42.
- WURSCHUM, T., LANGER, S. M., LONGIN, C. F., KORZUN, V., AKHUNOV, E., EBMEYER, E., SCHACHSCHNEIDER, R., SCHACHT, J., KAZMAN, E. & REIF, J. C. 2013. Population structure, genetic diversity and linkage disequilibrium in elite winter wheat assessed with SNP and SSR markers. *Theoretical and Applied Genetics*, 126, 1477-86.
- XIA, L., PENG, K., YANG, S., WENZL, P., DE VICENTE, M. C., FREGENE, M. & KILIAN, A. 2005. DArT for high-throughput genotyping of Cassava (*Manihot esculenta*) and its wild relatives. *Theoretical and Applied Genetics*, 110, 1092-8.
- XU, Y. & CROUCH, J. H. 2008. Marker-Assisted Selection in Plant Breeding: From Publications to Practice. *Crop Science*, 48, 391-407.
- YAMAMOTO, T., NAGASAKI, H., YONEMARU, J.-I., EBANA, K., NAKAJIMA, M., SHIBAYA, T. & YANO, M. 2010. Fine definition of the pedigree haplotypes of closely related rice cultivars by means of genome-wide discovery of

- single-nucleotide polymorphisms. *BMC Genomics*, 11, 267.
- YANG, S. Y., SAXENA, R. K., KULWAL, P. L., ASH, G. J., DUBEY, A., HARPER, J. D., UPADHYAYA, H. D., GOTHALWAL, R., KILIAN, A. & VARSHNEY, R. K. 2011. The first genetic map of pigeon pea based on diversity arrays technology (DArT) markers. *Journal of Genetics*, 90, 103-9.
- YOU, F., HUO, N., DEAL, K., GU, Y., LUO, M.-C., MCGUIRE, P., DVORAK, J. & ANDERSON, O. 2011. Annotation-based genome-wide SNP discovery in the large and complex *Aegilops tauschii* genome using next-generation sequencing without a reference genome sequence. *BMC Genomics*, 12, 59.
- YOUENS-CLARK, K., FAGA, B., YAP, I. V., STEIN, L. & WARE, D. 2009. CMap 1.01: a comparative mapping application for the Internet. *Bioinformatics*, 25, 3040-3042.
- ZABEAU, M. & VOS, P. 1993. *Selective restriction fragment amplification : a general method for DNA fingerprinting*. EP19920402629 19920924.
- ZHAO, W., CANARAN, P., JURKUTA, R., FULTON, T., GLAUBITZ, J., BUCKLER, E., DOEBLEY, J., GAUT, B., GOODMAN, M., HOLLAND, J., KRESOVICH, S., McMULLEN, M., STEIN, L. & WARE, D. 2006. Panzea: a database and resource for molecular and functional diversity in the maize genome. *Nucleic Acids Research*, 34, D752-7.
- ZHULIDOV, P. A., BOGDANOVA, E. A., SHCHEGLOV, A. S., VAGNER, L. L., KHASPEKOV, G. L., KOZHEMYAKO, V. B., MATZ, M. V., MELESHKEVITCH, E., MOROZ, L. L., LUKYANOV, S. A. & SHAGIN, D. A. 2004. Simple cDNA normalization using kamchatka crab duplex-specific nuclease. *Nucleic Acids Research*, 32,
- ZIETKIEWICZ, E., RAFALSKI, A. & LABUDA, D. 1994. Genome fingerprinting by simple sequence repeat (SSR)-anchored polymerase chain reaction amplification. *Genomics*, 20, 176-83.

Appendices

Appendix 1 Genes identified in low SNP density regions on chromosomes 7A, 7B and 7D.

UniProtKB Entry	Query Seq ID	Protein names
	7A	
A2YQK9	7A_SynBuild_v2.0_11335166_11337705	MADS-box transcription factor 26 (FDRMADS3) (OsMADS26) (RMADS220)
E2IUB0	7A_SynBuild_v2.0_5448868_5460813	Cycloartenol synthase (KdCAS) (EC 5.4.99.8)
F4JC97	7A_SynBuild_v2.0_5028864_5044067	Proteasome activator subunit 4 (Proteasome activator PA200)
H2KWF1	7A_SynBuild_v2.0_5448868_5460813	Parkeol synthase (EC 5.4.99.47)
O49397	7A_SynBuild_v2.0_2917217_2925239	Two-component response regulator ARR10 (Receiver-like protein 4)
O64477	7A_SynBuild_v2.0_15132848_15149802	G-type lectin S-receptor-like serine/threonine-protein kinase At2g19130 (EC 2.7.11.1)
P0C2J7	7AL_v2_extra_contigs_NODE_1373857_21991_23665	Transposon Ty4-H Gag-Pol polyprotein
P20025	7A_SynBuild_v2.0_370868_399976	Myb-related protein Zm38
P20025	7A_SynBuild_v2.0_390441_428999	Myb-related protein Zm38
P31922	7A_SynBuild_v2.0_3235981_3244137	Sucrose synthase 1 (EC 2.4.1.13) (Sucrose-UDP glucosyltransferase 1)
P43188	7A_SynBuild_v2.0_11494407_11498586	Adenylate kinase, chloroplastic (AK) (EC 2.7.4.3) (ATP-AMP transphosphorylase) (ATP:AMP phosphotransferase) (Adenylate monophosphate kinase)
P59230	7A_SynBuild_v2.0_8413567_8415224	60S ribosomal protein L10a-2
Q2R712	7A_SynBuild_v2.0_5448868_5460813	Achilleol B synthase (EC 5.4.99.48)
Q2R712	7A_SynBuild_v2.0_5466001_5466082	Achilleol B synthase (EC 5.4.99.48)
Q501H5	7AL_v2_extra_contigs_NODE_1198305_4015_4881	Phosphatidylinositol/phosphatidylcholine transfer protein SFH13
Q56XP4	7A_SynBuild_v2.0_5142750_5146058	Sodium/hydrogen exchanger 2 (Na(+)/H(+) exchanger 2) (NHE-2)
Q5XVJ4	7A_SynBuild_v2.0_2613170_2614566	Fanconi-associated nuclelease 1 homolog (EC 3.1.-.-)
Q5Z9Z0	7A_SynBuild_v2.0_5185271_5190644	Beta-glucosidase 24 (Os6bglu24) (EC 3.2.1.21)
Q6BE24	7A_SynBuild_v2.0_5448868_5460813	Cucurbitadienol synthase (EC 5.4.99.33)
Q6L400	7AS_v2_extra_contigs_NODE_309985_3977_5592	Putative late blight resistance protein homolog R1B-16
Q6NQK2	7A_SynBuild_v2.0_12738157_12808393	NAC domain-containing protein 8 (ANAC008)
Q6R0H1	7A_SynBuild_v2.0_10359753_10360637	Protein LHY (MYB-related transcription factor LHY) (Protein LATE ELONGATED HYPOCOTYL)

Q84UI5	7A_SynBuild_v2.0_2022901_2026218	Mitogen-activated protein kinase 1 (MAP kinase 1) (EC 2.7.11.24)
Q8LPQ6	7A_SynBuild_v2.0_17842889_17845184	ABC transporter B family member 28 (ABC transporter ABCB.28) (AtABCB28) (Non-intrinsic ABC protein 8) (TAP-related protein 1)
Q93XX5	7A_SynBuild_v2.0_11632886_11634020	PI-PLC X domain-containing protein At5g67130
Q9FDX8	7A_SynBuild_v2.0_4694623_4697169	Zinc finger protein HD1 (Protein CONSTANS-like) (Protein HEADING DATE 1) (OsHd1) (Protein PHOTOPERIOD SENSITIVITY 1)
Q9FG77	7A_SynBuild_v2.0_6916713_6919471	Probable WRKY transcription factor 2 (WRKY DNA-binding protein 2)
Q9FKS8	7A_SynBuild_v2.0_11061372_11063328	Lysine histidine transporter 1
Q9FNP1	7A_SynBuild_v2.0_8325824_8334284	Peroxisome biogenesis protein 1 (Peroxin-1) (AtPEX1)
Q9LP24	7AS_v2_extra_contigs_NODE_721474_2997_3944	Probable leucine-rich repeat receptor-like protein kinase At1g35710 (EC 2.7.11.1)
Q9LUM0	7A_SynBuild_v2.0_11591113_11597844	1-phosphatidylinositol-3-phosphate 5-kinase FAB1B
Q9LXZ5	7A_SynBuild_v2.0_9458695_9462663	Alpha-soluble NSF attachment protein 1 (Alpha-SNAP1) (N-ethylmaleimide-sensitive factor attachment protein alpha 1)
Q9LYS2	7A_SynBuild_v2.0_2284187_2284370	ABC transporter C family member 10
Q9LYS2	7A_SynBuild_v2.0_2288812_2289089	ABC transporter C family member 10
Q9SKR2	7A_SynBuild_v2.0_15392057_15396878	Synaptotagmin-1 (NTMC2T1.1) (Synaptotagmin A)
Q9SMU7	7A_SynBuild_v2.0_4445035_4445421	Signal recognition particle 9 kDa protein (SRP9)
Q9SQK3	7A_SynBuild_v2.0_4104385_4105575	Ankyrin repeat domain-containing protein EMB506, chloroplastic (Protein EMBRYO DEFECTIVE 506)
Q9SS90	7A_SynBuild_v2.0_13784420_13784561	E3 ubiquitin-protein ligase RGLG1 (EC 6.3.2.-) (RING domain ligase 1)
Q9XFH4	7AS_v2_extra_contigs_NODE_317854_5583_7567	ATP-dependent DNA helicase DDM1 (EC 3.6.4.12)
Q9ZUZ2	7A_SynBuild_v2.0_14346914_14352565	CDPK-related kinase 3 (AtCRK3) (EC 2.7.11.1) (Calcium/calmodulin-dependent protein kinase 4) (AtCK)
	7B	
F4IV66	7B_SynBuild_v2.0_5148560_5150292	Ribonuclease E/G-like protein, chloroplastic (RNase E/G-like protein) (EC 3.1.26.-)
O23372	7B_SynBuild_v2.0_9733736_9740633	Histone-lysine N-methyltransferase MEDEA (EC 2.1.1.43)
P0CH33	7B_SynBuild_v2.0_13598538_13599045	Polyubiquitin 11 [Cleaved into: Ubiquitin]
P0DI10	7B_SynBuild_v2.0_15557632_15557974	Ubiquitin-like protein-NEDD8-like protein RUB3
P22843	7B_SynBuild_v2.0_725038_725101	ABC transporter C family member 8 (ABC transporter ABCC.8) (AtABCC8) (EC 3.6.3.44)
P93831	7B_SynBuild_v2.0_5300114_5303606	Histone-lysine N-methyltransferase EZA1 (EC 2.1.1.43)

Q10NB9	7BL_v2_extra_contigs_NODE_1571469_2980_8103	Probable protein phosphatase 2C 31 (OsPP2C31) (EC 3.1.3.16)
Q500U8	7B_SynBuild_v2.0_6317126_6318886	Tetraketide alpha-pyrone reductase 1 (EC 1.1.1.-)
Q5SNL7	7B_SynBuild_v2.0_1094794_1095277	Fanconi-associated nuclease 1 homolog (EC 3.1.-.-)
Q6R0H1	7B_SynBuild_v2.0_9473563_9474447	Protein LHY (MYB-related transcription factor LHY) (Protein LATE ELONGATED HYPOCOTYL)
Q6YZX6	7BS_v2_extra_contigs_NODE_342101_6265_7978	Putative aconitate hydratase, cytoplasmic (Aconitase) (EC 4.2.1.3)
Q7Y1I7	7BL_v2_extra_contigs_NODE_555852_12671_13612	DNA (cytosine-5)-methyltransferase 1A (OsMET1a) (EC 2.1.1.37)
Q8S4P6	7B_SynBuild_v2.0_949893_955206	Protein LHY (MYB-related transcription factor LHY) (Protein LATE ELONGATED HYPOCOTYL)
Q9FKC3	7B_SynBuild_v2.0_1121282_1123834	Sec-independent protein translocase protein TATB, chloroplastic (Protein HIGH CHLOROPHYLL FLUORESCENCE 106)
Q9LYS2	7B_SynBuild_v2.0_489128_576877	ABC transporter C family member 10
Q9LYS2	7B_SynBuild_v2.0_515804_548218	ABC transporter C family member 10
Q9LYS2	7B_SynBuild_v2.0_544578_644934	ABC transporter C family member 10
Q9LYS2	7B_SynBuild_v2.0_548805_548872	ABC transporter C family member 10
Q9LYS2	7B_SynBuild_v2.0_549377_601130	ABC transporter C family member 10
Q9LYS2	7B_SynBuild_v2.0_552131_552419	Multidrug resistance-associated protein 1 (ATP-binding cassette sub-family C member 1)
Q9LYS2	7B_SynBuild_v2.0_569708_569815	ABC transporter C family member 10
Q9LYS2	7B_SynBuild_v2.0_575449_575764	ABC transporter C family member 10
Q9LYS2	7B_SynBuild_v2.0_597833_655253	ABC transporter C family member 10
Q9LYS2	7B_SynBuild_v2.0_603167_603319	ABC transporter C family member 10
Q9LYS2	7B_SynBuild_v2.0_641528_641794	Tetraketide alpha-pyrone reductase 1 (EC 1.1.1.-)
Q9LYS2	7B_SynBuild_v2.0_647770_648044	ABC transporter C family member 10 (
Q9XH75	7B_SynBuild_v2.0_11022783_11024138	Sec-independent protein translocase protein TATB, chloroplastic (Protein HIGH CHLOROPHYLL FLUORESCENCE 106)
	7D	
Q8GYB1	7D_SynBuild_v2.0_10031060_10031201	Nudix hydrolase 15, mitochondrial (AtNUDT15) (EC 3.6.1.-)
Q6YYA3	7D_SynBuild_v2.0_10470480_10489093	Chloroplastic group IIA intron splicing facilitator CRS1, chloroplastic (Chloroplastic RNA splicing factor 1)
Q6ET36	7D_SynBuild_v2.0_10570860_10604462	Ent-copalyl diphosphate synthase 1, chloroplastic
P42730	7D_SynBuild_v2.0_11098568_11099087	Chaperone protein ClpB1

Q0J6T3	7D_SynBuild_v2.0_11149787_11152642	Putative cinnamyl alcohol dehydrogenase 5 (OsCAD5) (EC 1.1.1.195)
Q38707	7D_SynBuild_v2.0_11149787_11152642	Mannitol dehydrogenase (EC 1.1.1.255)
Q9FWX7	7D_SynBuild_v2.0_11638052_11642375	ABC transporter B family member 11
O62518	7D_SynBuild_v2.0_11844056_11848181	Brix domain-containing protein ZK795.3
Q9C5L3	7D_SynBuild_v2.0_11929267_11932658	Putative glycerol-3-phosphate transporter 1
Q9FE65	7D_SynBuild_v2.0_11954300_11957005	60S ribosomal protein L34-2
P40071	7D_SynBuild_v2.0_12264360_12267854	Transmembrane 9 superfamily member 3
Q9BRS2	7D_SynBuild_v2.0_13340317_13341646	Serine/threonine-protein kinase RIO1 (EC 2.7.11.1)
P43188	7D_SynBuild_v2.0_13967380_13971677	Adenylate kinase, chloroplastic (AK) (EC 2.7.4.3)
C0LGE0	7D_SynBuild_v2.0_14311414_14311717	Probable LRR receptor-like serine/threonine-protein kinase At1g07650 (EC 2.7.11.1)
Q6Z1Z3	7D_SynBuild_v2.0_14363778_14367332	B3 domain-containing protein IDEF1 (Protein IRON DEFICIENCY-RESPONSIVE ELEMENT FACTOR 1)
Q9FJH6	7D_SynBuild_v2.0_14538593_14544571	ABC transporter F family member 1 (ABC transporter ABCF.1)
O82089	7D_SynBuild_v2.0_14704464_14707665	Copper transport protein CCH (Copper chaperone CCH)
A1EA00	7D_SynBuild_v2.0_15175370_15179623	DNA-directed RNA polymerase subunit beta" (EC 2.7.7.6)
Q94BN2	7D_SynBuild_v2.0_15411876_15415836	Spermine synthase (SPMSY) (EC 2.5.1.22)
Q6YW46	7D_SynBuild_v2.0_16457911_16460371	Elongation factor 1-gamma 2 (EF-1-gamma 2)
Q8VYE4	7D_SynBuild_v2.0_16568623_16570048	Protein NRT1/ PTR FAMILY 4.5 (AtNPF4.5) (Protein ABA-IMPORTING TRANSPORTER 2)
O64791	7D_SynBuild_v2.0_16748882_16750011	Syntaxin-124 (AtSYP124)
Q69VE0	7D_SynBuild_v2.0_16905756_16906842	Auxin-responsive protein IAA23 (Indoleacetic acid-induced protein 23)
Q9C5U3	7D_SynBuild_v2.0_17017430_17020286	26S protease regulatory subunit 8 homolog A
Q94B74	7D_SynBuild_v2.0_18174687_18178506	Nudix hydrolase 2 (AtNUDT2) (EC 3.6.1.-)
O49339	7D_SynBuild_v2.0_18208508_18212385	PTI1-like tyrosine-protein kinase 2 (PTI1-2) (EC 2.7.10.2)
Q6R2J8	7D_SynBuild_v2.0_18208508_18212385	Protein STRUBBELIG-RECEPTOR FAMILY 8 (Leucine-rich repeat receptor kinase-like protein SRF8)
Q8VYP9	7D_SynBuild_v2.0_18755222_18758985	Probable isoprenylcysteine alpha-carbonyl methylesterase ICME1 (EC 3.1.1.n2)
Q43135	7D_SynBuild_v2.0_18885489_18885715	Tyrosine N-monoxygenase (EC 1.14.13.41)
A2WQG7	7D_SynBuild_v2.0_21151793_21162781	Probable histone H2A.5
A2YG67	7D_SynBuild_v2.0_21151793_21162781	Auxin response factor 17
Q652J4	7D_SynBuild_v2.0_21367853_21370095	Probable potassium transporter 13 (OsHAK13)

Q9FEF8	7D_SynBuild_v2.0_21701203_21704604	Probable mediator of RNA polymerase II transcription subunit 36b (Histone-glutamine methyltransferase) (EC 2.1.1.-)
Q0DAE4	7D_SynBuild_v2.0_22117858_22121645	Glutaredoxin-C8 (Glutaredoxin-C4 homolog)
O65570	7D_SynBuild_v2.0_22134888_22143027	Villin-4
P38385	7D_SynBuild_v2.0_22317396_22318909	Protein transport protein Sec61 subunit gamma
P42785	7D_SynBuild_v2.0_22681297_22700793	Lysosomal Pro-X carboxypeptidase (EC 3.4.16.2)
Q8LCL3	7D_SynBuild_v2.0_23209963_23210264	60S ribosomal protein L27-2
Q9LYS2	7D_SynBuild_v2.0_2332353_2379080	ABC transporter C family member 10
Q9LYS2	7D_SynBuild_v2.0_2361880_2361984	ABC transporter C family member 10
Q9SL03	7D_SynBuild_v2.0_23832584_23849235	Callose synthase 2 (EC 2.4.1.34) (1,3-beta-glucan synthase) (Protein GLUCAN SYNTHASE-LIKE 3)
Q96290	7D_SynBuild_v2.0_23889788_23891374	Monosaccharide-sensing protein 1 (Monosaccharide transporter 1) (Sugar transporter MSSP1) (Sugar transporter MT1)
Q9LYS2	7D_SynBuild_v2.0_2390790_2415142	ABC transporter C family member 10
O23290	7D_SynBuild_v2.0_24406175_24406788	60S ribosomal protein L36a
Q5SNL7	7D_SynBuild_v2.0_2787706_2789934	Fanconi-associated nuclease 1 homolog (EC 3.1.-.)
Q9S775	7D_SynBuild_v2.0_3109988_3114490	CHD3-type chromatin-remodeling factor PICKLE (EC 3.6.4.-) (Protein GYMNO)
Q69TG5	7D_SynBuild_v2.0_3859442_3861913	MADS-box transcription factor 55 (OsMADS55)
F4HXP9	7D_SynBuild_v2.0_4184247_4191295	Myosin-9 (Myosin XI C) (AtXIC)
P37834	7D_SynBuild_v2.0_4322805_4323862	Peroxidase 1 (EC 1.11.1.7)
Q9SSJ8	7D_SynBuild_v2.0_4675693_4681811	Putative 1-phosphatidylinositol-3-phosphate 5-kinase FAB1C (Phosphatidylinositol 3-phosphate 5-kinase) (EC 2.7.1.150)
Q2PS26	7D_SynBuild_v2.0_5623133_5632633	Ubiquitin-like-specific protease 1D (EC 3.4.22.68) (Protein OVERLY TOLERANT TO SALT 1)
Q5T440	7D_SynBuild_v2.0_698075_709386	Putative transferase CAF17, mitochondrial (EC 2.1.-.) (Iron-sulfur cluster assembly factor homolog)
Q9T029	7D_SynBuild_v2.0_9338787_9338984	40S ribosomal protein S25-4
O81906	7D_SynBuild_v2.0_9933280_9934574	G-type lectin S-receptor-like serine/threonine-protein kinase B120 (EC 2.7.11.1)
Q5Z749	7DL_v2_extra_contigs_NODE_1136104_21658_22778	Auxin-responsive protein IAA21 (Indoleacetic acid-induced protein 21)
Q6NWL4	7DL_v2_extra_contigs_NODE_1279139_2537_3330	Cell differentiation protein RCD1 homolog (Rcd-1)
P17801	7DL_v2_extra_contigs_NODE_1537192_2603_3527	Putative receptor protein kinase ZmPK1 (EC 2.7.11.1)
A3CCP9	7DL_v2_extra_contigs_NODE_1540539_14888_28761	Putative protein phosphatase 2C 76 (OsPP2C76) (EC 3.1.3.16)

P34788	7DL_v2_extra_contigs_NODE_1715594_12707_13719	40S ribosomal protein S18
Q08480	7DL_v2_extra_contigs_NODE_239023_6223_11775	Adenylate kinase B (AK B) (EC 2.7.4.3)
A2XLI0	7DL_v2_extra_contigs_NODE_294551_3705_4264	Probable histone H2AXa
Q8S920	7DL_v2_extra_contigs_NODE_319602_9668_11787	Ubiquitin-conjugating enzyme E2 5A (EC 6.3.2.19)
P86207	7DL_v2_extra_contigs_NODE_344662_2082_2643	Ras-related protein Rab-2A (Fragments)
Q7G6K7	7DL_v2_extra_contigs_NODE_377658_9210_9711	Formin-like protein 3 (OsFH3)
A1XBB7	7DL_v2_extra_contigs_NODE_615413_2015_2562	Protein IN2-1 homolog B (Glutathione S-transferase GSTZ5)
Q9SIA3	7DL_v2_extra_contigs_NODE_644910_20039_22224	MATE efflux family protein 6 (Protein DETOXIFICATION 1-like 2)
Q7Y1I7	7DL_v2_extra_contigs_NODE_646167_8276_11167	DNA (cytosine-5)-methyltransferase 1A (OsMET1a) (EC 2.1.1.37) (
Q65XK0	7DL_v2_extra_contigs_NODE_722669_6481_8410	Ketol-acid reductoisomerase, chloroplastic (EC 1.1.1.86)
Q6ZF89	7DL_v2_extra_contigs_NODE_781813_2047_2612	Putative mixed-linked glucan synthase 1 (EC 2.4.1.-)
Q6ATG6	7DL_v2_extra_contigs_NODE_833934_2655_4203	Ribonuclease 3-like protein 2 (EC 3.1.26.-)
Q0JF58	7DL_v2_extra_contigs_NODE_837727_3292_8437	Protein argonaute 4B (OsAGO4b)
Q0D7E4	7DL_v2_extra_contigs_NODE_848417_3344_5969	Metal transporter Nramp1 (OsNramp1)
Q9ZPW2	7DL_v2_extra_contigs_NODE_919525_26759_28090	Anaphase-promoting complex subunit 10 (APC10) (Cyclosome subunit 10)
P92516	7DL_v2_extra_contigs_NODE_926099_2910_7024	Uncharacterized mitochondrial protein AtMg00750 (ORF119)
Q5W274	7DS_v2_extra_contigs_NODE_1403520_2594_14383	Pleiotropic drug resistance protein 3 (NtPDR3)
Q9SZAT7	7DS_v2_extra_contigs_NODE_1414556_5302_7681	Probable disease resistance protein At4g33300
Q8VYI1	7DS_v2_extra_contigs_NODE_1695704_9211_9778	Sphinganine C(4)-monooxygenase 1 (EC 1.14.13.169)
O65857	7DS_v2_extra_contigs_NODE_260032_7519_8579	Probable glutathione S-transferase GSTF1 (EC 2.5.1.18) (GST-I)
A2WSD3	7DS_v2_extra_contigs_NODE_289317_6528_8598	Bidirectional sugar transporter SWEET6b (OsSWEET6b)

Appendix 2 Genes identified in high SNP density regions on chromosomes 7A, 7B and 7D.

UniProtKB Entry	Query Seq ID	Protein names
	7A	
A2YU42	7A_SynBuild_v2.0_1433345_1433508	Cellulose synthase-like protein D2 (EC 2.4.1.-) (OsCslD2)
A9LYH7	7AS_v2_extra_contigs_NODE_1358090_3699_5299	DNA-directed RNA polymerase subunit beta (EC 2.7.7.6)
B8AUI3	7AL_v2_extra_contigs_NODE_783049_7780_9224	Peroxisomal (S)-2-hydroxy-acid oxidase GLO3 (EC 1.1.3.15)
O22259	7AL_v2_extra_contigs_NODE_696888_5263_7578	Ethylene-responsive transcription factor ERF071
Q0JF58	7AL_v2_extra_contigs_NODE_839047_5148_10124	Protein argonaute 4B (OsAGO4b)
Q6ZI17	7A_SynBuild_v2.0_124830_128903	Protein MEI2-like 2 (OML2) (MEI2-like protein 2)
Q75QN6	7AL_v2_extra_contigs_NODE_255280_2817_7995	Dual specificity protein phosphatase PHS1 (EC 3.1.3.16) (EC 3.1.3.48) (Protein PROPYZAMIDE-HYPERSENSITIVE 1)
Q7WHB6	7AS_v2_extra_contigs_NODE_54339_5381_7009	Chaperone protein ClpB
Q8H199	7AS_v2_extra_contigs_NODE_614155_8921_11139	Cysteine-rich receptor-like protein kinase 14 (Cysteine-rich RLK14) (EC 2.7.11.-)
Q9SX8	7AL_v2_extra_contigs_NODE_696888_5263_7578	Ethylene-responsive transcription factor 3
	7B	
Q6YZY5	7B_SynBuild_v2.0_9761044_9761980	Germin-like protein 8-11
Q9SIA4	7B_SynBuild_v2.0_15563397_15564454	MATE efflux family protein 5 (Protein DETOXIFICATION 1-like 1) (Protein DTX3)
Q9SSJ8	7B_SynBuild_v2.0_3256714_3263835	Putative 1-phosphatidylinositol-3-phosphate 5-kinase FAB1C
	7D	
Q9FKS8	7D_SynBuild_v2.0_13316857_13321111	Lysine histidine transporter 1

Appendix 3 SNPs from 90K SNP array matched SNPs from 16 Australian wheat.

SNP ID	SNP Name from 90K SNP array	Genotype	SNP Annotation	SNP Comment	Assay Type	Amino Add	Change Type I	Change Type II	Mapping Refence ID	Position	SNP name from 16	SNP	SNP
IWB67146	Tdurum contig11542_548	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7A_SvnBuild_v2.0	10134891	UQ01TA7A22004	35	175
IWA2381	wsnp_Ex_c19198_28115812	CT	Codominant	detects poly single locus	InfiniumII	K->R	Transition	nonsynonymous	7A_SvnBuild_v2.0	10284395	UQ01TA7A22223	29	132
IWA788	wsnp_CAP11_c592_400447	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7A_SvnBuild_v2.0	10354512	UQ01TA7A22330	32	91
IWB5246	BobWhite rep_c60436_231	CT	Codominant	detects poly single locus	InfiniumII	M->V	Transition	nonsynonymous	7A_SvnBuild_v2.0	10585686	UQ01TA7A22671	16	150
IWB69456	Tdurum contig28195_198	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7A_SvnBuild_v2.0	1072903	UQ01TA7A03429	12	64
IWB1318	BobWhite_c19875_271	AG	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7A_SvnBuild_v2.0	10785828	UQ01TA7A23020	58	148
IWB61248	RAC875_c98675_226	AG	AA Multiallelic	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7A_SvnBuild_v2.0	1092936	UQ01TA7A03474	38	114
IWB12840	CAP11_c3214_133	CT	Recode AB to BB	detects poly single locus	InfiniumII	no hit	Transition	.	7A_SvnBuild_v2.0	1093101	UQ01TA7A03475	38	126
IWB41953	Kukri_c18365_943	AG	Codominant	detects poly single locus	InfiniumII	N->N	Transition	synonymous	7A_SvnBuild_v2.0	10932284	UQ01TA7A23269	55	157
IWB43661	Kukri_c29154_977	AG	MSV	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7A_SvnBuild_v2.0	1098620	UQ01TA7A03513	17	95
IWB3612	BobWhite_c47716_503	AG	MSV	detects poly duplicated loci	InfiniumII	I->V	Transition	nonsynonymous	7A_SvnBuild_v2.0	11044371	UQ01TA7A23430	79	165
IWB1816	BobWhite_c24259_117	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7A_SvnBuild_v2.0	11137633	UQ01TA7A23731	12	34
IWB71784	Tdurum contig47309_609	GT	Codominant	detects poly single locus	InfiniumII	M->L	Transversion	nonsynonymous	7A_SvnBuild_v2.0	1118937	UQ01TA7A03643	2	21
IWB26795	Excalibur_c46601_212	AC	Codominant	detects poly single locus	InfiniumII	L->L	Transversion	synonymous	7A_SvnBuild_v2.0	1119338	UQ01TA7A03648	31	99
IWB26796	Excalibur_c46601_265	CT	BB Multiallelic	detects poly duplicated loci	InfiniumII	S->G	Transition	nonsynonymous	7A_SvnBuild_v2.0	1119391	UQ01TA7A03649	40	101
IWB54679	RAC875_c18446_521	AG	MSV	detects poly duplicated loci	InfiniumII	G->G	Transition	synonymous	7A_SvnBuild_v2.0	1125996	UQ01TA7A03702	5	19
IWB73326	Tdurum contig75819_3136	AG	Codominant	detects poly single locus	InfiniumII	V->V	Transition	synonymous	7A_SvnBuild_v2.0	1164653	UQ01TA7A03952	78	178
IWB69508	Tdurum contig28368_89	CT	MSV	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7A_SvnBuild_v2.0	11870965	UQ01TA7A24382	2	59
IWA4173	wsnp_Ex_c53387_56639804	CT	Codominant Null	detects poly single locus	InfiniumII	T->A	Transition	nonsynonymous	7A_SvnBuild_v2.0	11871706	UQ01TA7A24393	15	72
IWA4174	wsnp_Ex_c53387_56639835	AG	MSV	detects poly duplicated loci	InfiniumII	I->I	Transition	synonymous	7A_SvnBuild_v2.0	11871737	UQ01TA7A24394	27	93
IWB73997	Tdurum contig97505_172	AC	Codominant	detects poly single locus	InfiniumII	no hit	Transversion	.	7A_SvnBuild_v2.0	11873424	UQ01TA7A24402	21	175
IWA4176	wsnp_Ex_c53387_56640789	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7A_SvnBuild_v2.0	11873433	UQ01TA7A24403	49	160
IWA4177	wsnp_Ex_c53387_56641291	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7A_SvnBuild_v2.0	11873935	UQ01TA7A24407	38	85
IWA1554	wsnp_Ex_c11928_19131833	AC	Codominant	detects poly single locus	InfiniumII	S->A	Transversion	nonsynonymous	7A_SvnBuild_v2.0	11966816	UQ01TA7A24674	11	30
IWB56709	RAC875_c32212_84	CT	Recode AB to AA	detects poly single locus	InfiniumII	N->N	Transition	synonymous	7A_SvnBuild_v2.0	12162736	UQ01TA7A25023	18	152
IWB51440	Ra_c2171_1347	CT	Codominant Null	detects poly single locus	InfiniumII	no hit	Transition	.	7A_SvnBuild_v2.0	124004	UQ01TA7A00359	51	112
IWB8231	BS00040590_51	AG	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7A_SvnBuild_v2.0	12441874	UQ01TA7A25734	23	202
IWA4277	wsnp_Ex_c558_1105911	AG	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7A_SvnBuild_v2.0	12517418	UQ01TA7A25885	60	128
IWB7063	BS00022406_51	AG	BB Multiallelic Null	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7A_SvnBuild_v2.0	1251969	UQ01TA7A04204	54	161
IWB38834	Ku_c22875_172	AG	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7A_SvnBuild_v2.0	12558100	UQ01TA7A25992	20	137
IWB59123	RAC875_c5549_149	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7A_SvnBuild_v2.0	12610438	UQ01TA7A26084	71	169
IWA7293	wsnp_Ku_c7593_13054436	AG	Codominant	detects poly single locus	InfiniumII	H->R	Transition	nonsynonymous	7A_SvnBuild_v2.0	12611542	UQ01TA7A26087	48	141
IWB34725	IAAV3506	AT	Codominant	detects poly single locus	InfiniumII	G->G	Transversion	synonymous	7A_SvnBuild_v2.0	12611603	UQ01TA7A26088	16	93
IWA1832	wsnp_Ex_c1395_2672002	AG	Codominant	detects poly single locus	InfiniumII	H->R	Transition	nonsynonymous	7A_SvnBuild_v2.0	12611856	UQ01TA7A26090	55	134
IWB10876	BS00076379_51	AC	Codominant	detects poly single locus	InfiniumII	no hit	Transversion	.	7A_SvnBuild_v2.0	12613042	UQ01TA7A26094	65	163
IWA1834	wsnp_Ex_c1395_2672221	AG	Codominant	detects poly single locus	InfiniumII	K->R	Transition	nonsynonymous	7A_SvnBuild_v2.0	12613846	UQ01TA7A26096	55	146
IWA4637	wsnp_Ex_c7216_12390889	AG	BB Multiallelic	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7A_SvnBuild_v2.0	12652201	UQ01TA7A26241	59	122
IWA4638	wsnp_Ex_c7216_12391182	AG	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7A_SvnBuild_v2.0	12652494	UQ01TA7A26242	99	202
IWA2301	wsnp_Ex_c18352_27178687	AC	Codominant	detects poly single locus	InfiniumII	no hit	Transversion	.	7A_SvnBuild_v2.0	12926775	UQ01TA7A26695	10	103
IWA1491	wsnp_Ex_c11417_18429357	AG	Codominant	detects poly single locus	InfiniumII	Y->Y	Transition	synonymous	7A_SvnBuild_v2.0	13152424	UQ01TA7A27149	49	138
IWB51323	Ra_c18741_604	AG	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7A_SvnBuild_v2.0	13190902	UQ01TA7A27342	4	16
IWB12565	BS00110561_51	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7A_SvnBuild_v2.0	13364421	UQ01TA7A27698	67	143
IWB29223	Excalibur_c8801_1154	AG	BB Multiallelic	detects poly duplicated loci	InfiniumII	C->R	Transition	nonsynonymous	7A_SvnBuild_v2.0	13617817	UQ01TA7A28065	26	128

IWA8115	w.snp_Ra_rep_c105182_89171305	AG	Codominant	detects poly single locus	InfiniumII	M->T	Transition	nonsynonymous	7A SvnBuild v2.0	13891917	UQ01TA7A28725	4	13
IWA6376	w.snp_Ku_c10202_16937059	CT	BB Multiallelic	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7A SvnBuild v2.0	13969369	UQ01TA7A29037	32	138
IWB34697	IAAV3305	CT	MSV	detects poly duplicated loci	InfiniumII	T->A	Transition	nonsynonymous	7A SvnBuild v2.0	13993198	UQ01TA7A29110	3	114
IWA7770	w.snp_Ra_c23285_32795964	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7A SvnBuild v2.0	14067175	UQ01TA7A29338	62	142
IWA3701	w.snp_Ex_c40011_47158369	CT	Codominant	detects poly single locus	InfiniumII	R->G	Transition	nonsynonymous	7A SvnBuild v2.0	14069478	UQ01TA7A29345	59	124
IWB34076	GENE-4644_70	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7A SvnBuild v2.0	140713	UQ01TA7A00427	2	62
IWB58819	RAC875_c52318_188	CT	MSV	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7A SvnBuild v2.0	14084300	UQ01TA7A29385	38	175
IWA3082	w.snp_Ex_c26923_36144147	CT	Codominant	detects poly single locus	InfiniumII	N->D	Transition	nonsynonymous	7A SvnBuild v2.0	14164300	UQ01TA7A29635	40	101
IWB13779	CAP7_c12333_392	CT	MSV	detects poly duplicated loci	InfiniumII	T->A	Transition	nonsynonymous	7A SvnBuild v2.0	14322834	UQ01TA7A30152	32	174
IWB33880	GENE-4270_242	AG	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7A SvnBuild v2.0	14418588	UQ01TA7A30243	73	149
IWB25388	Excalibur_c33866_404	GT	Codominant	detects poly single locus	InfiniumII	no hit	Transversion	.	7A SvnBuild v2.0	14419922	UQ01TA7A30247	57	127
IWB58209	RAC875_c46055_710	GT	Codominant	detects poly single locus	InfiniumII	D->A	Transversion	nonsynonymous	7A SvnBuild v2.0	14423791	UQ01TA7A30262	8	138
IWB34290	GENE-4981_53	AC	Codominant	detects poly single locus	InfiniumII	no hit	Transversion	.	7A SvnBuild v2.0	14498530	UQ01TA7A30516	66	169
IWB11559	BS00088374_51	AG	MSV	detects poly duplicated loci	InfiniumII	V->V	Transition	synonymous	7A SvnBuild v2.0	14529947	UQ01TA7A30600	50	153
IWA3662	w.snp_Ex_c38981_46383475	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7A SvnBuild v2.0	14685878	UQ01TA7A31253	41	125
IWB3343	BobWhite_c4399_447	AG	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7A SvnBuild v2.0	14699919	UQ01TA7A31289	12	141
IWB53764	RAC875_c13696_226	AG	Codominant	detects poly single locus	InfiniumII	G->G	Transition	synonymous	7A SvnBuild v2.0	14761212	UQ01TA7A31540	28	121
IWA808	w.snp_CAP11_c78_114341	AG	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7A SvnBuild v2.0	14769961	UQ01TA7A31561	26	113
IWB47336	Kukri_c6676_172	AG	MSV	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7A SvnBuild v2.0	14780758	UQ01TA7A31598	45	133
IWB72890	Tdurum_contig62357_527	AG	AA Multiallelic	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7A SvnBuild v2.0	14794950	UQ01TA7A31627	12	141
IWB12865	CAP11_c3781_95	AC	Recode AB to AA	detects poly single locus	InfiniumII	I->I	Transversion	synonymous	7A SvnBuild v2.0	15053308	UQ01TA7A32686	12	55
IWB20855	Ex_c5177_1092	AC	MSV	detects poly duplicated loci	InfiniumII	T->T	Transversion	synonymous	7A SvnBuild v2.0	15174157	UQ01TA7A32924	3	22
IWB3124	BobWhite_c40535_218	AG	BB Multiallelic	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7A SvnBuild v2.0	1518305	UQ01TA7A05218	11	41
IWB41762	Kukri_c17417_571	CT	BB Multiallelic	detects poly duplicated loci	InfiniumII	T->A	Transition	nonsynonymous	7A SvnBuild v2.0	1520077	UQ01TA7A05224	2	23
IWB4276	BobWhite_c7082_184	AG	AA Multiallelic	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7A SvnBuild v2.0	15273076	UQ01TA7A33375	37	158
IWB4277	BobWhite_c7082_196	AG	BB Multiallelic	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7A SvnBuild v2.0	15273088	UQ01TA7A33376	2	163
IWB5501	BobWhite_rep_c64772_309	AG	MSV	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7A SvnBuild v2.0	15306748	UQ01TA7A33485	57	146
IWB29719	Excalibur_rep_c101421_191	AG	Codominant	detects poly single locus	InfiniumII	N->N	Transition	synonymous	7A SvnBuild v2.0	153082	UQ01TA7A00510	35	72
IWB25077	Excalibur_c31216_744	AG	Codominant	detects poly single locus	InfiniumII	G->G	Transition	synonymous	7A SvnBuild v2.0	15417900	UQ01TA7A33820	55	137
IWB74590	tolb0036a12_207	AG	MSV	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7A SvnBuild v2.0	15433465	UQ01TA7A33847	65	172
IWB3767	BobWhite_c5235_710	AG	Recode AB to BB	detects poly single locus	InfiniumII	H->H	Transition	synonymous	7A SvnBuild v2.0	15458607	UQ01TA7A33891	16	140
IWA8076	w.snp_Ra_c8394_14242358	AG	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7A SvnBuild v2.0	15471336	UQ01TA7A33936	42	148
IWB10482	BS00070626_51	AG	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7A SvnBuild v2.0	15481169	UQ01TA7A33967	42	130
IWB42182	Kukri_c19696_60	CT	Codominant	detects poly single locus	InfiniumII	E->E	Transition	synonymous	7A SvnBuild v2.0	15719111	UQ01TA7A34775	30	115
IWB68176	Tdurum_contig14265_689	AG	Codominant	detects poly single locus	InfiniumII	I->I	Transition	synonymous	7A SvnBuild v2.0	15788176	UQ01TA7A34889	5	171
IWB69997	Tdurum_contig30290_64	CT	AA Multiallelic	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7A SvnBuild v2.0	15833898	UQ01TA7A35008	9	192
IWB69996	Tdurum_contig30290_175	AG	Codominant	detects poly single locus	InfiniumII	S->S	Transition	synonymous	7A SvnBuild v2.0	15834009	UQ01TA7A35009	5	196
IWB56839	RAC875_c33451_241	GT	Mono	Mono	InfiniumII	O->P	Transversion	nonsynonymous	7A SvnBuild v2.0	16033320	UQ01TA7A35303	4	83
IWB29518	Excalibur_c95707_285	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7A SvnBuild v2.0	16151530	UQ01TA7A35635	68	147
IWA5912	w.snp_JD_c20555_18262260	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7A SvnBuild v2.0	16256864	UQ01TA7A36032	80	163
IWA5913	w.snp_JD_c20555_18262317	AG	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7A SvnBuild v2.0	16256921	UQ01TA7A36033	76	164
IWB53096	RAC875_c105310_155	AG	Codominant	detects poly single locus	InfiniumII	M->T	Transition	nonsynonymous	7A SvnBuild v2.0	16355221	UQ01TA7A36305	42	98
IWB35597	IAAV9161	CT	Codominant	detects poly single locus	InfiniumII	D->G	Transition	nonsynonymous	7A SvnBuild v2.0	16355710	UQ01TA7A36309	51	181
IWB50776	Ra_c105310_660	GT	Recode AB to AA	detects poly single locus	InfiniumII	K->Q	Transversion	nonsynonymous	7A SvnBuild v2.0	16356095	UQ01TA7A36311	36	107
IWA4620	w.snp_Ex_c7071_12171222	CT	Codominant	detects poly single locus	InfiniumII	T->A	Transition	nonsynonymous	7A SvnBuild v2.0	16368531	UQ01TA7A36348	21	82
IWA4621	w.snp_Ex_c7071_12171619	AG	AA Multiallelic	detects poly duplicated loci	InfiniumII	P->P	Transition	synonymous	7A SvnBuild v2.0	16369817	UQ01TA7A36352	11	38
IWA2270	w.snp_Ex_c17899_26666328	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7A SvnBuild v2.0	16377249	UQ01TA7A36375	53	107

IWB23034	Excalibur c17899_352	CT	Codominant	detects poly single locus	InfiniumII	T->A	Transition	nonsynonymous	7A SvnBuild v2.0	16377418	UQ01TA7A36377	11	130
IWB8059	BS00036553_51	AG	MSV	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7A SvnBuild v2.0	1646372	UQ01TA7A05481	3	42
IWB50931	Ra_c114158_328	GT	MSV	detects poly duplicated loci	InfiniumII	no hit	Transversion	.	7A SvnBuild v2.0	16476943	UQ01TA7A36626	31	130
IWB38907	Ku_c24644_599	AG	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7A SvnBuild v2.0	1651515	UQ01TA7A05501	26	63
IWA4187	wsnp_Ex_c53442_56678505	AG	Codominant	detects poly single locus	InfiniumII	A->A	Transition	synonymous	7A SvnBuild v2.0	16542739	UQ01TA7A36881	45	133
IWA7642	wsnp_Ra_c16020_24570294	CT	Mono	Mono	InfiniumII	Q->R	Transition	nonsynonymous	7A SvnBuild v2.0	16717759	UQ01TA7A37451	2	22
IWB43786	Kukri_c2990_129	AG	MSV	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7A SvnBuild v2.0	16757924	UQ01TA7A37588	60	167
IWB39103	Ku_c2990_604	GT	Codominant	detects poly single locus	InfiniumII	D->A	Transversion	nonsynonymous	7A SvnBuild v2.0	16758399	UQ01TA7A37592	49	152
IWB39102	Ku_c2990_1997	AC	Codominant	detects poly single locus	InfiniumII	no hit	Transversion	.	7A SvnBuild v2.0	16759852	UQ01TA7A37595	48	161
IWB28064	Excalibur_c61603_1209	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7A SvnBuild v2.0	16769734	UQ01TA7A37632	36	115
IWA4437	wsnp_Ex_c61603_61581218	AG	MSV	detects poly duplicated loci	InfiniumII	I->I	Transition	synonymous	7A SvnBuild v2.0	16773443	UQ01TA7A37649	21	147
IWB34981	IAAV5167	AG	MSV	detects poly duplicated loci	InfiniumII	W->R	Transition	nonsynonymous	7A SvnBuild v2.0	16780690	UQ01TA7A37728	2	129
IWB50566	Kukri_rep_c98227_230	AG	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7A SvnBuild v2.0	16853577	UQ01TA7A38057	37	99
IWA4925	wsnp_Ex_c9476_15710162	AG	BB Multiallelic	detects poly duplicated loci	InfiniumII	V->V	Transition	synonymous	7A SvnBuild v2.0	16855978	UQ01TA7A38065	30	152
IWB69742	Tdurum_contig29240_206	AG	MSV	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7A SvnBuild v2.0	16901655	UQ01TA7A38241	45	147
IWB62310	RAC875_rep_c15178_103	AG	BB Multiallelic	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7A SvnBuild v2.0	16985772	UQ01TA7A38538	24	67
IWB12588	BS00110683_51	CT	MSV	detects poly duplicated loci	InfiniumII	P->P	Transition	synonymous	7A SvnBuild v2.0	16985973	UQ01TA7A38542	33	74
IWB12587	BS00110681_51	CT	Codominant	detects poly single locus	InfiniumII	M->V	Transition	nonsynonymous	7A SvnBuild v2.0	16986005	UQ01TA7A38543	41	82
IWA7184	wsnp_Ku_c5693_10079278	AC	Codominant	detects poly single locus	InfiniumII	L->R	Transversion	nonsynonymous	7A SvnBuild v2.0	16990081	UQ01TA7A38561	21	54
IWB46622	Kukri_c5693_1983	AG	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7A SvnBuild v2.0	16991337	UQ01TA7A38570	37	110
IWA2905	wsnp_Ex_c24486_33732900	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7A SvnBuild v2.0	16992052	UQ01TA7A38572	8	35
IWB69347	Tdurum_contig27856_230	AG	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7A SvnBuild v2.0	17077446	UQ01TA7A39228	17	81
IWB1874	BobWhite_c25105_507	CT	Codominant	detects poly single locus	InfiniumII	I->M	Transition	nonsynonymous	7A SvnBuild v2.0	17086886	UQ01TA7A39287	10	60
IWB72673	Tdurum_contig59633_56	AG	AA Multiallelic	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7A SvnBuild v2.0	17204522	UQ01TA7A39791	14	36
IWB38552	Ku_c16600_805	AG	MSV	detects poly duplicated loci	InfiniumII	Q->R	Transition	nonsynonymous	7A SvnBuild v2.0	17294483	UQ01TA7A40266	23	50
IWB38550	Ku_c16600_2646	AG	Recode AB to AA	detects poly single locus	InfiniumII	no hit	Transition	.	7A SvnBuild v2.0	17296533	UQ01TA7A40278	31	64
IWB38551	Ku_c16600_2938	AG	MSV	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7A SvnBuild v2.0	17296824	UQ01TA7A40279	34	72
IWB40615	Kukri_c11530_60	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7A SvnBuild v2.0	17532684	UQ01TA7A40820	2	46
IWB40616	Kukri_c11530_92	CT	AA Multiallelic Null	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7A SvnBuild v2.0	17532716	UQ01TA7A40821	2	37
IWB40614	Kukri_c11530_168	CT	MSV	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7A SvnBuild v2.0	17532792	UQ01TA7A40822	12	34
IWA6424	wsnp_Ku_c11530_18803034	AG	MSV	detects poly duplicated loci	InfiniumII	I->T	Transition	nonsynonymous	7A SvnBuild v2.0	17534927	UQ01TA7A40839	41	134
IWB71711	Tdurum_contig46717_2021	CT	AA Multiallelic	detects poly duplicated loci	InfiniumII	R->G	Transition	nonsynonymous	7A SvnBuild v2.0	17535426	UQ01TA7A40842	23	67
IWB10573	BS00071736_51	GT	BB Multiallelic Null	detects poly duplicated loci	InfiniumII	no hit	Transversion	.	7A SvnBuild v2.0	17644036	UQ01TA7A41144	7	142
IWB12407	BS00109393_51	AG	AA Multiallelic Null	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7A SvnBuild v2.0	17650514	UQ01TA7A41171	27	196
IWB7469	BS00023200_51	AG	BB Multiallelic Null	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7A SvnBuild v2.0	17650533	UQ01TA7A41172	30	209
IWB52831	Ra_c956_2318	CT	Codominant Null	detects poly single locus	InfiniumII	no hit	Transition	.	7A SvnBuild v2.0	17650811	UQ01TA7A41174	37	235
IWB73544	Tdurum_contig81947_779	GT	AA Multiallelic	detects poly duplicated loci	InfiniumII	no hit	Transversion	.	7A SvnBuild v2.0	17678935	UQ01TA7A41206	11	48
IWB54211	RAC875_c1606_644	AG	MSV	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7A SvnBuild v2.0	17701399	UQ01TA7A41237	14	172
IWB29740	Excalibur_rep_c101582_828	AG	Recode AB to BB	detects poly single locus	InfiniumII	C->C	Transition	synonymous	7A SvnBuild v2.0	17720970	UQ01TA7A41304	4	127
IWB41880	Kukri_c18055_1740	CT	Codominant	detects poly single locus	InfiniumII	A->A	Transition	synonymous	7A SvnBuild v2.0	17759827	UQ01TA7A41494	7	15
IWB12385	BS00108784_51	GT	Codominant	detects poly single locus	InfiniumII	no hit	Transversion	.	7A SvnBuild v2.0	1934885	UQ01TA7A06467	30	136
IWB23768	Excalibur_c22196_483	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7A SvnBuild v2.0	1934952	UQ01TA7A06468	35	142
IWB28720	Excalibur_c74442_199	CT	Mono	Mono	InfiniumII	E->E	Transition	synonymous	7A SvnBuild v2.0	1956736	UQ01TA7A06517	2	76
IWB59294	RAC875_c57541_120	AC	BB Multiallelic	detects poly duplicated loci	InfiniumII	no hit	Transversion	.	7A SvnBuild v2.0	2050043	UQ01TA7A06717	52	133
IWB59295	RAC875_c57541_193	AC	Codominant	detects poly single locus	InfiniumII	P->P	Transversion	synonymous	7A SvnBuild v2.0	2050116	UQ01TA7A06718	19	120
IWB73577	Tdurum_contig82510_556	GT	AA Multiallelic	detects poly duplicated loci	InfiniumII	no hit	Transversion	.	7A SvnBuild v2.0	2324314	UQ01TA7A07304	7	33
IWA7849	wsnp_Ra_c31237_40393880	CT	MSV	detects poly duplicated loci	InfiniumII	E->E	Transition	synonymous	7A SvnBuild v2.0	2330071	UQ01TA7A07324	21	76

IWB22256	Excalibur c13337 219	AG	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7A SvnBuild v2.0	2458344	UQ01TA7A07642	21	53
IWA1759	wsnp_Ex_c13337_21022658	AG	BB Multiallelic	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7A SvnBuild v2.0	2459081	UQ01TA7A07643	36	80
IWB20735	Ex_c4463_146	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7A SvnBuild v2.0	2502458	UQ01TA7A07755	27	85
IWB73665	Tdurum contig8441_602	CT	AA Multiallelic	detects poly duplicated loci	InfiniumII	K->K	Transition	synonymous	7A SvnBuild v2.0	2513539	UQ01TATA07796	44	148
IWB38343	Ku_c12701_1268	AG	Recode AB to AA	detects poly single locus	InfiniumII	L->L	Transition	synonymous	7A SvnBuild v2.0	2519024	UQ01TA7A07803	4	37
IWB38344	Ku_c12701_1273	AG	MSV	detects poly duplicated loci	InfiniumII	L->L	Transition	synonymous	7A SvnBuild v2.0	2519029	UQ01TA7A07804	2	36
IWA6472	wsnp_Ku_c12701_20446223	AG	Codominant	detects poly single locus	InfiniumII	D->D	Transition	synonymous	7A SvnBuild v2.0	2519846	UQ01TA7A07806	22	47
IWA6473	wsnp_Ku_c12701_20446367	AC	MSV	detects poly duplicated loci	InfiniumII	R->R	Transversion	synonymous	7A SvnBuild v2.0	2519990	UQ01TA7A07807	15	92
IWB60146	RAC875_c6805_1347	AG	Recode AB to BB	detects poly single locus	InfiniumII	R->R	Transition	synonymous	7A SvnBuild v2.0	2521636	UQ01TA7A07814	10	33
IWA3831	wsnp_Ex_c42653_49180485	AG	Codominant	detects poly single locus	InfiniumII	R->G	Transition	nonsynonymous	7A SvnBuild v2.0	2537600	UQ01TA7A07827	25	76
IWA3832	wsnp_Ex_c42653_49180603	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7A SvnBuild v2.0	2537798	UQ01TA7A07828	9	23
IWB2196	BobWhite_c2838_428	AG	Mono	Mono	InfiniumII	Y->Y	Transition	synonymous	7A SvnBuild v2.0	2624386	UQ01TA7A07992	2	25
IWA7654	wsnp_Ra_c1654_3265291	CT	MSV	detects poly duplicated loci	InfiniumII	A->A	Transition	synonymous	7A SvnBuild v2.0	2797986	UQ01TA7A08635	4	65
IWB28082	Excalibur_c6196_668	AG	Codominant	detects poly single locus	InfiniumII	F->F	Transition	synonymous	7A SvnBuild v2.0	2823326	UQ01TA7A08676	33	121
IWB21011	Ex_c6196_971	AG	Codominant	detects poly single locus	InfiniumII	F->F	Transition	synonymous	7A SvnBuild v2.0	2823682	UQ01TA7A08677	12	43
IWB58577	RAC875_c4965_891	AG	MSV	detects poly duplicated loci	InfiniumII	V->A	Transition	nonsynonymous	7A SvnBuild v2.0	302981	UQ01TA7A01219	3	113
IWA3979	wsnp_Ex_c4668_8353466	AG	MSV	detects poly duplicated loci	InfiniumII	F->F	Transition	synonymous	7A SvnBuild v2.0	312705	UQ01TA7A01255	2	18
IWA4180	wsnp_Ex_c5341_9442913	AG	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7A SvnBuild v2.0	3157452	UQ01TA7A09100	41	114
IWA3673	wsnp_Ex_c39221_46569987	AG	MSV	detects poly duplicated loci	InfiniumII	L->P	Transition	nonsynonymous	7A SvnBuild v2.0	3230701	UQ01TA7A09259	29	59
IWB68797	Tdurum contig19352_76	AG	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7A SvnBuild v2.0	3343504	UQ01TA7A09510	15	130
IWA7942	wsnp_Ra_c4418_8012732	AG	MSV	detects poly duplicated loci	InfiniumII	S->S	Transition	synonymous	7A SvnBuild v2.0	3350319	UQ01TA7A09517	24	97
IWB39741	Ku_c62485_931	AG	Codominant	detects poly single locus	InfiniumII	L->S	Transition	nonsynonymous	7A SvnBuild v2.0	3424232	UQ01TA7A09616	6	140
IWB62609	RAC875_rep_c69766_246	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7A SvnBuild v2.0	3475401	UQ01TA7A09710	6	132
IWB3129	BobWhite_c40583_146	AG	Codominant	detects poly single locus	InfiniumII	G->G	Transition	synonymous	7A SvnBuild v2.0	3476774	UQ01TA7A09711	3	92
IWB5661	BobWhite_s63403_99	AG	Recode AB to AA	detects poly single locus	InfiniumII	I->V	Transition	nonsynonymous	7A SvnBuild v2.0	3514738	UQ01TA7A09778	12	168
IWB11906	BS00094919_51	AG	MSV	detects poly duplicated loci	InfiniumII	I->V	Transition	nonsynonymous	7A SvnBuild v2.0	3514738	UQ01TA7A09778	12	168
IWB13400	CAP12_c4447_280	CT	Mono	Mono	InfiniumII	S->S	Transition	synonymous	7A SvnBuild v2.0	352619	UQ01TA7A01374	25	125
IWB62418	RAC875_rep_c117475_289	CT	Codominant	detects poly single locus	InfiniumII	I->S	Transition	nonsynonymous	7A SvnBuild v2.0	3603090	UQ01TA7A10041	7	101
IWB37440	JD_c42022_319	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7A SvnBuild v2.0	3616405	UQ01TA7A10093	51	140
IWB34627	IAAV2865	CT	Codominant	detects poly single locus	InfiniumII	A->A	Transition	synonymous	7A SvnBuild v2.0	3618540	UQ01TA7A10104	27	65
IWA1156	wsnp_CAP8_c1725_973916	GT	MSV	detects poly duplicated loci	InfiniumII	no hit	Transversion	.	7A SvnBuild v2.0	3705131	UQ01TATA10388	2	115
IWB2616	BobWhite_c33300_159	AG	Codominant Null	detects poly single locus	InfiniumII	no hit	Transition	.	7A SvnBuild v2.0	4218343	UQ01TA7A11135	42	92
IWA2945	wsnp_Ex_c25025_34285478	AG	MSV	detects poly duplicated loci	InfiniumII	N->N	Transition	synonymous	7A SvnBuild v2.0	4295051	UQ01TA7A11254	52	146
IWB47402	Kukri_c67586_509	AG	MSV	detects poly duplicated loci	InfiniumII	D->D	Transition	synonymous	7A SvnBuild v2.0	445254	UQ01TA7A01604	5	66
IWB50713	Kukri_s118416_65	AG	Codominant	detects poly single locus	InfiniumII	E->F	Transition	synonymous	7A SvnBuild v2.0	4472892	UQ01TA7A11481	19	117
IWB67909	Tdurum contig13245_443	CT	Codominant	detects poly single locus	InfiniumII	S->G	Transition	nonsynonymous	7A SvnBuild v2.0	476671	UQ01TA7A01697	4	89
IWB341	BobWhite_c1234_261	CT	Mono	Mono	InfiniumII	I->V	Transition	nonsynonymous	7A SvnBuild v2.0	490466	UQ01TA7A01836	2	71
IWB41777	Kukri_c17556_411	CT	BB Multiallelic	detects poly duplicated loci	InfiniumII	A->A	Transition	synonymous	7A SvnBuild v2.0	5209643	UQ01TA7A12447	4	173
IWB62930	RAC875_rep_c73038_974	AG	Codominant	detects poly single locus	InfiniumII	N->N	Transition	synonymous	7A SvnBuild v2.0	539411	UQ01TA7A02068	37	99
IWA5727	wsnp_Ex_rep_c71217_70021470	AC	MSV	detects poly duplicated loci	InfiniumII	K->Q	Transversion	nonsynonymous	7A SvnBuild v2.0	5525053	UQ01TA7A12834	57	148
IWA2392	wsnp_Ex_c19262_28188808	AG	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7A SvnBuild v2.0	5525695	UQ01TA7A12835	47	126
IWB46842	Kukri_c6011_3475	AG	Codominant	detects poly single locus	InfiniumII	Y->H	Transition	nonsynonymous	7A SvnBuild v2.0	554471	UQ01TA7A02136	4	81
IWB6806	BS00021936_51	CT	Codominant	detects poly single locus	InfiniumII	L->P	Transition	nonsynonymous	7A SvnBuild v2.0	5689738	UQ01TA7A13400	13	53
IWB7445	BS00023147_51	GT	MSV	detects poly duplicated loci	InfiniumII	no hit	Transversion	.	7A SvnBuild v2.0	5711639	UQ01TA7A13426	26	88
IWA834	wsnp_CAP11_rep_c4066_1921894	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7A SvnBuild v2.0	5817420	UQ01TA7A13523	47	105
IWB10386	BS00069242_51	CT	MSV	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7A SvnBuild v2.0	5820728	UQ01TA7A13539	24	74
IWB60588	RAC875_c7988_1588	CT	Codominant	detects poly single locus	InfiniumII	R->R	Transition	synonymous	7A SvnBuild v2.0	6062480	UQ01TA7A13985	2	118

IWB46670	Kukri c5757 530	AG	Codominant	detects poly single locus	InfiniumII	A->A	Transition	synonymous	7A SvnBuild v2.0	6268579	UQ01TA7A14340	2	73
IWB56239	RAC875 c28842 99	AC	Codominant	detects poly single locus	InfiniumII	W->G	Transversion	nonsynonymous	7A SvnBuild v2.0	6456824	UQ01TA7A14911	24	167
IWB70641	Tdurum contig41127 265	CT	BB Multiallelic	detects poly duplicated loci	InfiniumII	I->I	Transition	synonymous	7A SvnBuild v2.0	6603599	UQ01TA7A15112	2	79
IWB5770	BS00000747 51	CT	MSV	detects poly duplicated loci	InfiniumII	F->F	Transition	synonymous	7A SvnBuild v2.0	6604257	UQ01TA7A15116	8	72
IWB3927	BobWhite c5649 344	CT	MSV	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7A SvnBuild v2.0	6604367	UQ01TA7A15117	30	116
IWA4614	wsnp_Ex c7030 12111917	AC	Codominant	detects poly single locus	InfiniumII	R->R	Transversion	synonymous	7A SvnBuild v2.0	6672142	UQ01TA7A15233	54	146
IWB7400	BS00023055 51	AG	AA Multiallelic	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7A SvnBuild v2.0	677432	UQ01TA7A02493	29	149
IWB35027	IAAV5448	CG	Codominant	detects poly single locus	InfiniumI	no hit	Transversion	.	7A SvnBuild v2.0	708116	UQ01TA7A02565	18	141
IWB33949	GENE-4419 58	CT	AA Multiallelic	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7A SvnBuild v2.0	7278337	UQ01TA7A15928	12	109
IWB43474	Kukri c27692 822	AG	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7A SvnBuild v2.0	7278817	UQ01TA7A15930	29	153
IWB34274	GENE-4953 139	AG	BB Multiallelic	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7A SvnBuild v2.0	7279963	UQ01TA7A15931	15	150
IWB21674	Excalibur c10563 523	AG	Codominant	detects poly single locus	InfiniumII	I->V	Transition	nonsynonymous	7A SvnBuild v2.0	7287786	UQ01TA7A15938	5	75
IWB56782	RAC875 c32895 211	CT	Codominant	detects poly single locus	InfiniumII	P->P	Transition	synonymous	7A SvnBuild v2.0	7316999	UQ01TA7A15994	15	65
IWB56783	RAC875 c32895 304	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7A SvnBuild v2.0	7317154	UQ01TA7A15995	19	65
IWB34201	GENE-4859 218	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7A SvnBuild v2.0	7380302	UQ01TA7A16112	11	103
IWB31316	Excalibur rep c81365 214	CT	Codominant	detects poly single locus	InfiniumII	T->T	Transition	synonymous	7A SvnBuild v2.0	7555816	UQ01TA7A16284	23	112
IWB20600	Ex c37521 670	CT	BB Multiallelic	detects poly duplicated loci	InfiniumII	N->D	Transition	nonsynonymous	7A SvnBuild v2.0	7563777	UQ01TA7A16308	43	160
IWB21030	Ex c6348 1205	AG	Recode AB to AA	detects poly single locus	InfiniumII	no hit	Transition	.	7A SvnBuild v2.0	7605144	UQ01TA7A16385	31	143
IWA1278	wsnp_Ex c10094 16590746	AG	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7A SvnBuild v2.0	7667660	UQ01TA7A16519	15	59
IWA6936	wsnp_Ku c340 706774	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7A SvnBuild v2.0	7874145	UQ01TA7A16819	31	103
IWA1842	wsnp_Ex c14009 21899923	CT	Codominant	detects poly single locus	InfiniumII	I->S	Transition	nonsynonymous	7A SvnBuild v2.0	7925633	UQ01TA7A16893	6	51
IWA3900	wsnp_Ex c44547 50572744	CT	Recode AB to AA	detects poly single locus	InfiniumII	no hit	Transition	.	7A SvnBuild v2.0	7942153	UQ01TA7A16931	13	131
IWB651	BobWhite c14471 411	CT	Codominant	detects poly single locus	InfiniumII	H->H	Transition	synonymous	7A SvnBuild v2.0	7980386	UQ01TA7A16989	17	92
IWB12917	CAP11 c4929 307	CT	Recode AB to BB	detects poly single locus	InfiniumII	no hit	Transition	.	7A SvnBuild v2.0	7999627	UQ01TA7A17020	13	64
IWB37769	JD_c7043 681	AG	MSV	detects poly duplicated loci	InfiniumII	K->E	Transition	nonsynonymous	7A SvnBuild v2.0	8038672	UQ01TA7A17048	4	133
IWA5645	wsnp_Ex rep c69838 68799256	CT	Recode AB to BB	detects poly single locus	InfiniumII	no hit	Transition	.	7A SvnBuild v2.0	8041119	UQ01TA7A17055	22	144
IWA3778	wsnp_Ex c41465 48279111	AG	MSV	detects poly duplicated loci	InfiniumII	F->L	Transition	nonsynonymous	7A SvnBuild v2.0	8063372	UQ01TA7A17100	2	54
IWA1476	wsnp_Ex c11350 18330107	AG	Codominant	detects poly single locus	InfiniumII	A->A	Transition	synonymous	7A SvnBuild v2.0	8070435	UQ01TA7A17107	25	145
IWB73946	Tdurum contig9584 455	AG	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7A SvnBuild v2.0	813016	UQ01TA7A02785	2	86
IWB73947	Tdurum contig9584 463	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7A SvnBuild v2.0	813024	UQ01TA7A02786	2	79
IWB25411	Excalibur c34091 311	AG	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7A SvnBuild v2.0	8151899	UQ01TA7A17249	61	153
IWA1871	wsnp_Ex c14173 22107343	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7A SvnBuild v2.0	8318696	UQ01TA7A17514	37	158
IWA4573	wsnp_Ex c6797 11731807	CT	Codominant	detects poly single locus	InfiniumII	G->G	Transition	synonymous	7A SvnBuild v2.0	8452107	UQ01TA7A17740	35	123
IWB6695	BS00021659 51	AC	Codominant	detects poly single locus	InfiniumII	no hit	Transversion	.	7A SvnBuild v2.0	9082799	UQ01TA7A19316	66	137
IWB996	BobWhite c17095 237	AG	Recode AB to AA	detects poly single locus	InfiniumII	no hit	Transition	.	7A SvnBuild v2.0	9117233	UQ01TA7A19452	6	158
IWB21281	Ex c7397 535	GT	Codominant	detects poly single locus	InfiniumII	K->Q	Transversion	nonsynonymous	7A SvnBuild v2.0	9119519	UQ01TA7A19462	34	111
IWB21282	Ex c7397 545	CT	Codominant	detects poly single locus	InfiniumII	A->A	Transition	synonymous	7A SvnBuild v2.0	9119529	UQ01TA7A19463	17	94
IWA4672	wsnp_Ex c7397_12674744	AG	Codominant	detects poly single locus	InfiniumII	L->S	Transition	nonsynonymous	7A SvnBuild v2.0	9119839	UQ01TA7A19464	54	111
IWA5867	wsnp_JD_c15755_15117800	AG	Codominant	detects poly single locus	InfiniumII	D->D	Transition	synonymous	7A SvnBuild v2.0	9201586	UQ01TA7A19699	27	168
IWB29354	Excalibur c916 285	CT	MSV	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7A SvnBuild v2.0	9368836	UQ01TA7A20137	7	41
IWB11349	BS00084192 51	AG	MSV	detects poly duplicated loci	InfiniumII	K->K	Transition	synonymous	7A SvnBuild v2.0	9370018	UQ01TA7A20151	30	115
IWA4887	wsnp_Ex c916 1767286	CT	AA Multiallelic	detects poly duplicated loci	InfiniumII	N->N	Transition	synonymous	7A SvnBuild v2.0	9370141	UQ01TA7A20152	35	86
IWB34840	IAAV416	CG	MSV	detects poly duplicated loci	InfiniumI	R->R	Transversion	synonymous	7A SvnBuild v2.0	9438229	UQ01TA7A20388	25	91
IWB21762	Excalibur c109258 1038	CT	Codominant	detects poly single locus	InfiniumII	S->S	Transition	synonymous	7A SvnBuild v2.0	9516757	UQ01TA7A20547	27	149
IWB72397	Tdurum contig54832 139	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7A SvnBuild v2.0	9781519	UQ01TA7A21281	44	128
IWB34237	GENE-4914 349	AG	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7A SvnBuild v2.0	9796666	UQ01TA7A21318	49	176
IWB73164	Tdurum contig70105 162	GT	Codominant	detects poly single locus	InfiniumII	no hit	Transversion	.	7A SvnBuild v2.0	9977301	UQ01TA7A21733	60	212

IWB40989	Kukri c13171 474	AC	BB Multiallelic	detects poly duplicated loci	InfiniumII	no hit	Transversion	.	7AL v2 extra contigs NODE 1075880	2525	UQ01TA7A1132960	8	45
IWB53300	RAC875 c11283 379	AG	Codominant	detects poly single locus	InfiniumII	Y->Y	Transition	synonymous	7AL v2 extra contigs NODE 1099347	1779	UQ01TA7A1140341	13	139
IWB53301	RAC875 c11283 722	AG	Recode AB to AA	detects poly single locus	InfiniumII	M->T	Transition	nonsynonymous	7AL v2 extra contigs NODE 1099347	2181	UQ01TA7A1140347	19	56
IWB13822	CAP7 c1522 255	GT	Codominant	detects poly single locus	InfiniumII	no hit	Transversion	.	7AL v2 extra contigs NODE 1143294	1755	UQ01TA7A1159652	24	103
IWA689	wsnP CAP11 c1761 958064	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7AL v2 extra contigs NODE 1158310	1170	UQ01TA7A1167253	4	16
IWB23497	Excalibur c20486 268	CT	Recode AB to BB	detects poly single locus	InfiniumII	no hit	Transition	.	7AL v2 extra contigs NODE 1172468	4305	UQ01TA7A1175417	3	118
IWB52043	Ra c41488 398	GT	BB Multiallelic	detects poly duplicated loci	InfiniumII	no hit	Transversion	.	7AL v2 extra contigs NODE 1187999	3738	UQ01TA7A1185130	56	142
IWB42688	Kukri c22645 255	CT	AA Multiallelic	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7AL v2 extra contigs NODE 1203384	920	UQ01TA7A1194564	46	93
IWB38722	Ku c19920 372	CT	Mono	Mono	InfiniumII	no hit	Transition	.	7AL v2 extra contigs NODE 1204624	1901	UQ01TA7A1195450	10	36
IWB58861	RAC875 c527 106	CT	MSV	detects poly duplicated loci	InfiniumII	D->D	Transition	synonymous	7AL v2 extra contigs NODE 1215058	5164	UQ01TA7A1201551	17	86
IWB9555	BS00065454 51	CT	Mono	Mono	InfiniumII	V->V	Transition	synonymous	7AL v2 extra contigs NODE 1215637	604	UQ01TA7A1201800	19	89
IWB21296	Ex c7569 919	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7AL v2 extra contigs NODE 1232789	1709	UQ01TA7A1211944	62	136
IWB52783	Ra c8985 557	CT	Codominant	detects poly single locus	InfiniumII	V->A	Transition	nonsynonymous	7AL v2 extra contigs NODE 1243305	6276	UQ01TA7A1218203	51	116
IWB57246	RAC875 c37085 317	AG	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7AL v2 extra contigs NODE 1287408	4688	UQ01TA7A1247341	26	123
IWB54889	RAC875 c19724 842	GT	Mono	Mono	InfiniumII	V->V	Transversion	synonymous	7AL v2 extra contigs NODE 1288704	7465	UQ01TA7A1247959	18	111
IWB58863	RAC875 c52725 218	CT	MSV	detects poly duplicated loci	InfiniumII	C->C	Transition	synonymous	7AL v2 extra contigs NODE 1313796	2015	UQ01TA7A1260582	29	116
IWB11909	BS00094965 51	AC	Codominant	detects poly single locus	InfiniumII	no hit	Transversion	.	7AL v2 extra contigs NODE 1329065	617	UQ01TA7A1266334	30	107
IWB7547	BS00024619 51	AG	AA Multiallelic	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7AL v2 extra contigs NODE 1380198	33337	UQ01TA7A1271891	8	23
IWB71670	Tdurum contig46334 832	AG	MSV	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7AL v2 extra contigs NODE 1546951	4556	UQ01TA7A1325434	59	293
IWB37021	JD c149 1700	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7AL v2 extra contigs NODE 1546951	5566	UQ01TA7A1325441	29	287
IWB3914	BobWhite c5602 291	GT	MSV	detects poly duplicated loci	InfiniumII	no hit	Transversion	.	7AL v2 extra contigs NODE 1546951	9769	UQ01TA7A1325447	13	302
IWB9574	BS00065529 51	AG	AA Multiallelic	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7AL v2 extra contigs NODE 1624776	3511	UQ01TA7A1356428	71	147
IWB9602	BS00065647 51	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7AL v2 extra contigs NODE 165174	613	UQ01TA7A676464	62	162
IWB29555	Excalibur c96483 102	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7AL v2 extra contigs NODE 1689179	10399	UQ01TA7A1363126	74	151
IWB2536	BobWhite c3232 616	CT	Mono	Mono	InfiniumII	L->S	Transition	nonsynonymous	7AL v2 extra contigs NODE 1903921	1675	UQ01TA7A1415834	3	32
IWB18064	D GB5Y7FA02G0E49 296	AG	Recode AB to AA	detects poly single locus	InfiniumII	no hit	Transition	.	7AL v2 extra contigs NODE 196645	1288	UQ01TA7A686932	18	159
IWB53267	RAC875 c1115 341	AG	AA Multiallelic	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7AL v2 extra contigs NODE 2006493	3084	UQ01TA7A1424467	20	151
IWB36931	JD c1201 631	GT	Codominant	detects poly single locus	InfiniumII	no hit	Transversion	.	7AL v2 extra contigs NODE 2006493	3434	UQ01TA7A1424468	4	212
IWB7367	BS00023003 51	AG	MSV	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7AL v2 extra contigs NODE 202156	99	UQ01TA7A688936	10	50
IWB39099	Ku c29856 132	CT	Codominant Null	detects poly single locus	InfiniumII	no hit	Transition	.	7AL v2 extra contigs NODE 2319984	888	UQ01TA7A1461601	42	115
IWB39100	Ku c29856 174	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7AL v2 extra contigs NODE 2319984	930	UQ01TA7A1461602	47	126
IWB46180	Kukri c5166 994	CT	Mono	Mono	InfiniumII	no hit	Transition	.	7AL v2 extra contigs NODE 232849	3210	UQ01TA7A699916	3	203
IWA7964	wsnP Ra c47942 53349897	CT	MSV	detects poly duplicated loci	InfiniumII	S->S	Transition	synonymous	7AL v2 extra contigs NODE 255280	5027	UQ01TA7A709209	6	29
IWB21717	Excalibur c1070 2327	CT	BB Multiallelic	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7AL v2 extra contigs NODE 255929	382	UQ01TA7A709624	3	69
IWB26786	Excalibur c46518 120	CT	MSV	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7AL v2 extra contigs NODE 303952	689	UQ01TA7A731854	37	149
IWB44173	Kukri c33036 348	AG	MSV	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7AL v2 extra contigs NODE 321071	1102	UQ01TA7A739719	6	86
IWB67094	Tdurum contig11433 636	AG	Mono	Mono	InfiniumII	C->C	Transition	synonymous	7AL v2 extra contigs NODE 330162	5862	UQ01TA7A744796	16	148
IWB27420	Excalibur c538 1152	AG	MSV	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7AL v2 extra contigs NODE 351983	2589	UQ01TA7A758077	25	144
IWB71048	Tdurum contig42260 2664	AG	MSV	detects poly duplicated loci	InfiniumII	A->A	Transition	synonymous	7AL v2 extra contigs NODE 351983	2851	UQ01TA7A758080	29	136
IWB5783	BS00001128 51	AC	MSV	detects poly duplicated loci	InfiniumII	no hit	Transversion	.	7AL v2 extra contigs NODE 359691	197	UQ01TA7A763135	39	85
IWB41395	Kukri c1552 1059	AG	Codominant	detects poly single locus	InfiniumII	P->P	Transition	synonymous	7AL v2 extra contigs NODE 36921	5170	UQ01TA7A648270	12	45
IWA6562	wsnP Ku c1552 3060297	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7AL v2 extra contigs NODE 36921	6336	UQ01TA7A648273	47	118
IWB10701	BS00073989 51	CT	AA Multiallelic Null	detects poly duplicated loci	InfiniumII	S->P	Transition	nonsynonymous	7AL v2 extra contigs NODE 369825	3013	UQ01TA7A768685	38	95
IWB66824	Tdurum contig10929 413	AG	Mono	Mono	InfiniumII	no hit	Transition	.	7AL v2 extra contigs NODE 379963	124	UQ01TA7A774436	5	19
IWB41223	Kukri c14516 224	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7AL v2 extra contigs NODE 382786	3411	UQ01TA7A776327	10	107
IWB27289	Excalibur c52115 233	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7AL v2 extra contigs NODE 419769	8158	UQ01TA7A800010	35	116
IWB36973	JD c1314 1184	GT	Codominant	detects poly single locus	InfiniumII	no hit	Transversion	.	7AL v2 extra contigs NODE 419769	9260	UQ01TA7A800017	17	37

IWB64403	RFL Contig3590_811	CT	Recode AB to AA	detects poly single locus	InfiniumII	no hit	Transition	.	7AL v2 extra contigs NODE 422658	428	UQ01TA7A802013	14	94
IWB44223	Kukri c33620_129	GT	MSV	detects poly duplicated loci	InfiniumII	no hit	Transversion	.	7AL v2 extra contigs NODE 447891	2145	UQ01TA7A819348	2	22
IWB8791	BS00059929_51	CT	AA Multiallelic	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7AL v2 extra contigs NODE 44843	1103	UQ01TA7A649544	38	114
IWB8790	BS00059928_51	AG	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7AL v2 extra contigs NODE 44843	876	UQ01TA7A649543	27	82
IWA3668	w.snp_Ex_c39119_46485649	GT	Recode AB to AA	detects poly single locus	InfiniumII	no hit	Transversion	.	7AL v2 extra contigs NODE 455819	14378	UQ01TA7A824787	59	133
IWB58339	RAC875 c4732_1521	AG	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7AL v2 extra contigs NODE 455819	9178	UQ01TA7A824773	15	34
IWB7611	BS00026122_51	CT	MSV	detects poly duplicated loci	InfiniumII	R->R	Transition	synonymous	7AL v2 extra contigs NODE 469157	1553	UQ01TA7A834178	40	90
IWB6693	BS00021657_51	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7AL v2 extra contigs NODE 477218	6458	UQ01TA7A839483	76	158
IWA8099	w.snp_Ra_c9185_15386027	AG	MSV	detects poly duplicated loci	InfiniumII	S->S	Transition	synonymous	7AL v2 extra contigs NODE 482379	913	UQ01TA7A843486	28	64
IWB10812	BS00075525_51	AG	Codominant	detects poly single locus	InfiniumII	N->N	Transition	synonymous	7AL v2 extra contigs NODE 485837	523	UQ01TA7A846168	36	109
IWB38119	Ku_c104966_604	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7AL v2 extra contigs NODE 486396	2035	UQ01TA7A846515	8	163
IWA4996	w.snp_Ex_c9982_16429661	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7AL v2 extra contigs NODE 495397	4758	UQ01TA7A852153	35	128
IWB9244	BS00064302_51	AC	Codominant	detects poly single locus	InfiniumII	no hit	Transversion	.	7AL v2 extra contigs NODE 500512	3413	UQ01TA7A855606	9	32
IWA1518	w.snp_Ex_c1159_2225557	AC	MSV	detects poly duplicated loci	InfiniumII	no hit	Transversion	.	7AL v2 extra contigs NODE 50680	641	UQ01TA7A650560	11	107
IWA4895	w.snp_Ex_c9279_15416671	CT	Mono	Mono	InfiniumII	no hit	Transition	.	7AL v2 extra contigs NODE 509251	7547	UQ01TA7A861658	64	133
IWB46043	Kukri_c50054_170	AG	MSV	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7AL v2 extra contigs NODE 509251	9387	UQ01TA7A861666	49	120
IWB7506	BS00023755_51	CT	Codominant	detects poly single locus	InfiniumII	A->A	Transition	synonymous	7AL v2 extra contigs NODE 512259	2207	UQ01TA7A863126	29	129
IWB55540	RAC875 c2359_652	AG	MSV	detects poly duplicated loci	InfiniumII	K->K	Transition	synonymous	7AL v2 extra contigs NODE 514251	2267	UQ01TA7A864414	11	70
IWB12233	BS00104760_51	CT	Recode AB to AA	detects poly single locus	InfiniumII	L->L	Transition	synonymous	7AL v2 extra contigs NODE 520499	1642	UQ01TA7A868214	4	72
IWB5923	BS00003929_51	CT	MSV	detects poly duplicated loci	InfiniumII	S->S	Transition	synonymous	7AL v2 extra contigs NODE 520499	1747	UQ01TA7A868215	5	104
IWB43008	Kukri_c24646_396	CT	Recode AB to BB	detects poly single locus	InfiniumII	no hit	Transition	.	7AL v2 extra contigs NODE 527147	4893	UQ01TA7A872068	60	130
IWB50375	Kukri_rep_c79716_729	AG	MSV	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7AL v2 extra contigs NODE 527855	2157	UQ01TA7A872502	37	127
IWA4434	w.snp_Ex_c6142_10746442	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7AL v2 extra contigs NODE 527855	4989	UQ01TA7A872536	21	83
IWB27807	Excalibur_c5839_481	CT	MSV	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7AL v2 extra contigs NODE 527855	5754	UQ01TA7A872539	4	40
IWA4364	w.snp_Ex_c5839_10246915	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7AL v2 extra contigs NODE 527855	5785	UQ01TA7A872540	12	53
IWB47548	Kukri_c7284_1859	CT	MSV	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7AL v2 extra contigs NODE 527855	5785	UQ01TA7A872540	12	53
IWB6430	BS00011532_51	CT	Recode AB to BB	detects poly single locus	InfiniumII	D->D	Transition	synonymous	7AL v2 extra contigs NODE 540966	140	UQ01TA7A880810	6	24
IWA6004	w.snp_JD_c3225_4227048	GT	Codominant	detects poly single locus	InfiniumII	no hit	Transversion	.	7AL v2 extra contigs NODE 548233	4869	UQ01TA7A884789	8	17
IWB10212	BS00068032_51	AG	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7AL v2 extra contigs NODE 548780	1973	UQ01TA7A885124	18	99
IWB10213	BS00068033_51	AC	Codominant Null	detects poly single locus	InfiniumII	no hit	Transversion	.	7AL v2 extra contigs NODE 548780	2019	UQ01TA7A885125	18	110
IWB7382	BS00023027_51	CT	Recode AB to BB	detects poly single locus	InfiniumII	no hit	Transition	.	7AL v2 extra contigs NODE 551956	875	UQ01TA7A886941	27	122
IWA7728	w.snp_Ra_c2063_4012957	AG	MSV	detects poly duplicated loci	InfiniumII	C->R	Transition	nonsynonymous	7AL v2 extra contigs NODE 676701	3206	UQ01TA7A919370	41	105
IWB4764	BobWhite_rep_c49367_405	AC	Recode AB to AA	detects poly single locus	InfiniumII	R->R	Transversion	synonymous	7AL v2 extra contigs NODE 749864	8001	UQ01TA7A953020	70	140
IWB20876	Ex_c52798_415	AC	MSV	detects poly duplicated loci	InfiniumII	K->Q	Transversion	nonsynonymous	7AL v2 extra contigs NODE 754110	1603	UQ01TA7A955797	53	141
IWA7139	w.snp_Ku_c5160_9203226	CT	MSV	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7AL v2 extra contigs NODE 769624	10953	UQ01TA7A965378	41	92
IWA7140	w.snp_Ku_c5160_9203385	GT	MSV	detects poly duplicated loci	InfiniumII	no hit	Transversion	.	7AL v2 extra contigs NODE 769624	11112	UQ01TA7A965379	12	31
IWA794	w.snp_CAP11_c639_424059	CT	MSV	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7AL v2 extra contigs NODE 778418	136	UQ01TA7A971069	9	71
IWA795	w.snp_CAP11_c639_424134	AG	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7AL v2 extra contigs NODE 778418	211	UQ01TA7A971070	19	111
IWA5800	w.snp_JD_c1219_1768165	AG	MSV	detects poly duplicated loci	InfiniumII	N->S	Transition	nonsynonymous	7AL v2 extra contigs NODE 803376	10492	UQ01TA7A987892	50	178
IWA7904	w.snp_Ra_c389_826890	AG	MSV	detects poly duplicated loci	InfiniumII	K->E	Transition	nonsynonymous	7AL v2 extra contigs NODE 803376	7843	UQ01TA7A987885	39	121
IWB6878	BS00022049_51	AC	BB Multiallelic	detects poly duplicated loci	InfiniumII	no hit	Transversion	.	7AL v2 extra contigs NODE 822055	5716	UQ01TA7A1001541	30	150
IWB14679	CAP8_c3244_208	AC	Mono	Mono	InfiniumII	no hit	Transversion	.	7AL v2 extra contigs NODE 82689	82	UQ01TA7A656088	6	37
IWB639	BobWhite_c1441_1078	AG	MSV	detects poly duplicated loci	InfiniumII	L->L	Transition	synonymous	7AL v2 extra contigs NODE 834039	8697	UQ01TA7A1010164	2	131
IWB27946	Excalibur_c60238_183	AC	BB Multiallelic	detects poly duplicated loci	InfiniumII	no hit	Transversion	.	7AL v2 extra contigs NODE 83791	402	UQ01TA7A656334	37	147
IWB27947	Excalibur_c60238_251	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7AL v2 extra contigs NODE 83791	470	UQ01TA7A656336	16	161
IWB34054	GENE-4616_496	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7AL v2 extra contigs NODE 858357	3261	UQ01TA7A1027870	53	112
IWB38869	Ku_c23689_1024	AG	MSV	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7AL v2 extra contigs NODE 870648	3572	UQ01TA7A1036810	32	112

IWB51755	Ra_c29926_625	AG	MSV	detects poly duplicated loci	InfiniumII	N->S	Transition	nonsynonymous	TAL v2 extra contigs NODE 887425	3624	UQ01TA7A1049661	18	105
IWB12152	BS000100907_51	CT	Mono	Mono	InfiniumII	L->L	Transition	synonymous	TAL v2 extra contigs NODE 890262	1798	UQ01TA7A1051618	10	78
IWB13122	CAP11 rep_c6423_103	GT	BB Multiallelic	detects poly duplicated loci	InfiniumII	no hit	Transversion	.	TAL v2 extra contigs NODE 890262	2866	UQ01TA7A1051628	36	163
IWB21365	Ex_c2861_2309	AG	MSV	detects poly duplicated loci	InfiniumII	R->R	Transition	synonymous	TAL v2 extra contigs NODE 893463	3313	UQ01TA7A1054460	21	137
IWB6983	BS00022237_51	AG	MSV	detects poly duplicated loci	InfiniumII	no hit	Transition	.	TAL v2 extra contigs NODE 895811	1603	UQ01TA7A1055909	8	147
IWA5167	wsnp_Ex rep_c108367_91621570	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	TAL v2 extra contigs NODE 919066	10411	UQ01TA7A1072690	43	113
IWB20381	Ex_c27898_414	CT	Recode AB to AA	detects poly single locus	InfiniumII	no hit	Transition	.	TAL v2 extra contigs NODE 919066	7331	UQ01TA7A1072673	34	106
IWA3128	wsnp_Ex_c27898_37058842	AG	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	TAL v2 extra contigs NODE 919066	7602	UQ01TA7A1072674	40	114
IWB70820	Tdurum contig42093_1170	AG	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	TAL v2 extra contigs NODE 928339	1466	UQ01TA7A1079166	30	143
IWA4063	wsnp_Ex_c4996_8885500	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	TAL v2 extra contigs NODE 928339	2445	UQ01TA7A1079168	3	13
IWA7600	wsnp_Ra_c1303_2598907	AG	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	TAL v2 extra contigs NODE 928339	2656	UQ01TA7A1079169	11	22
IWB48282	Kukri_c92030_109	CT	MSV	detects poly duplicated loci	InfiniumII	no hit	Transition	.	TAL v2 extra contigs NODE 941202	7382	UQ01TA7A1088962	11	102
IWB72648	Tdurum contig59467_433	AC	Codominant	detects poly single locus	InfiniumII	no hit	Transversion	.	TAL v2 extra contigs NODE 973966	2011	UQ01TA7A1109518	18	93
IWB72649	Tdurum contig59467_534	CT	MSV	detects poly duplicated loci	InfiniumII	no hit	Transition	.	TAL v2 extra contigs NODE 973966	2112	UQ01TA7A1109520	7	87
IWB43533	Kukri_c28160_220	CT	Codominant	detects poly single locus	InfiniumII	L->L	Transition	synonymous	TAL v2 extra contigs NODE 973966	2598	UQ01TA7A1109523	5	111
IWB57023	RAC875_c35010_187	CT	MSV	detects poly duplicated loci	InfiniumII	no hit	Transition	.	TAS v2 extra contigs NODE 1019805	3770	UQ01TA7A536651	9	80
IWB10939	BS00077560_51	GT	Codominant	detects poly single locus	InfiniumII	no hit	Transversion	.	TAS v2 extra contigs NODE 1024068	1667	UQ01TA7A539413	16	79
IWA8066	wsnp_Ra_c75858_73254602	AG	Codominant	detects poly single locus	InfiniumII	K->E	Transition	nonsynonymous	TAS v2 extra contigs NODE 1055976	935	UQ01TA7A555937	5	15
IWB10368	BS00068864_51	AG	BB Multiallelic	detects poly duplicated loci	InfiniumII	no hit	Transition	.	TAS v2 extra contigs NODE 1143523	457	UQ01TA7A569384	3	10
IWB34741	IAAV3589	GT	Recode AB to BB	detects poly single locus	InfiniumII	F->C	Transversion	nonsynonymous	TAS v2 extra contigs NODE 1166333	1636	UQ01TA7A574234	2	119
IWB3643	BobWhite_c48204_704	CT	Mono	Mono	InfiniumII	no hit	Transition	.	TAS v2 extra contigs NODE 1209571	4940	UQ01TA7A588176	17	101
IWB56953	RAC875_c3450_836	GT	Codominant	detects poly single locus	InfiniumII	I->R	Transversion	nonsynonymous	TAS v2 extra contigs NODE 1218878	2026	UQ01TA7A591742	54	150
IWB52818	Ra_c9427_300	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	TAS v2 extra contigs NODE 1218878	2560	UQ01TA7A591743	46	98
IWA5855	wsnp_JD_c15127_14676522	AG	Codominant	detects poly single locus	InfiniumII	K->R	Transition	nonsynonymous	TAS v2 extra contigs NODE 1239159	8734	UQ01TA7A598146	52	121
IWB21468	Ex_c9615_1202	AG	AA Multiallelic	detects poly duplicated loci	InfiniumII	no hit	Transition	.	TAS v2 extra contigs NODE 1338761	7634	UQ01TA7A610003	45	102
IWB39935	Ku_c71719_137	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	TAS v2 extra contigs NODE 1338761	7710	UQ01TA7A610004	15	122
IWA7978	wsnp_Ra_c5008_8947135	AC	Codominant	detects poly single locus	InfiniumII	C->G	Transversion	nonsynonymous	TAS v2 extra contigs NODE 1358090	14220	UQ01TA7A613845	34	103
IWA3318	wsnp_Ex_c31123_39956138	CT	Recode AB to BB	detects poly single locus	InfiniumII	no hit	Transition	.	TAS v2 extra contigs NODE 147884	11511	UQ01TA7A72164	31	148
IWB6928	BS00022146_51	AG	MSV	detects poly duplicated loci	InfiniumII	no hit	Transition	.	TAS v2 extra contigs NODE 161426	215	UQ01TA7A77093	13	86
IWB6832	BS00021973_51	CT	MSV	detects poly duplicated loci	InfiniumII	T->T	Transition	synonymous	TAS v2 extra contigs NODE 170686	4560	UQ01TA7A80914	26	83
IWB73688	Tdurum contig85377_141	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	TAS v2 extra contigs NODE 185490	1045	UQ01TA7A87103	58	160
IWB38300	Ku_c12129_1036	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	TAS v2 extra contigs NODE 22256	17720	UQ01TA7A106417	13	175
IWB23424	Excalibur_c20062_195	CT	Codominant	detects poly single locus	InfiniumII	T->T	Transition	synonymous	TAS v2 extra contigs NODE 241678	706	UQ01TA7A114872	57	138
IWB57304	RAC875_c37592_348	AG	AA Multiallelic	detects poly duplicated loci	InfiniumII	D->D	Transition	synonymous	TAS v2 extra contigs NODE 245586	77	UQ01TA7A117528	2	17
IWB23574	Excalibur_c20943_154	AG	Recode AB to BB	detects poly single locus	InfiniumII	no hit	Transition	.	TAS v2 extra contigs NODE 265451	5238	UQ01TA7A130090	4	64
IWB22378	Excalibur_c14189_156	AG	Recode AB to BB	detects poly single locus	InfiniumII	no hit	Transition	.	TAS v2 extra contigs NODE 278204	204	UQ01TA7A138716	32	93
IWB1044	BobWhite_c17467_129	AG	Recode AB to AA	detects poly single locus	InfiniumII	no hit	Transition	.	TAS v2 extra contigs NODE 289916	7472	UQ01TA7A147211	7	59
IWB12370	BS000108549_51	CT	BB Multiallelic	detects poly duplicated loci	InfiniumII	no hit	Transition	.	TAS v2 extra contigs NODE 289916	7512	UQ01TA7A147212	6	50
IWB7519	BS00023993_51	AG	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	TAS v2 extra contigs NODE 311714	3258	UQ01TA7A164314	55	122
IWB7520	BS00023994_51	CT	Recode AB to BB	detects poly single locus	InfiniumII	no hit	Transition	.	TAS v2 extra contigs NODE 311714	4045	UQ01TA7A164315	7	31
IWB64370	RFL Contig3483_512	CT	MSV	detects poly duplicated loci	InfiniumII	no hit	Transition	.	TAS v2 extra contigs NODE 319993	1094	UQ01TA7A171414	3	64
IWB64371	RFL Contig3483_628	CT	Codominant Null	detects poly single locus	InfiniumII	no hit	Transition	.	TAS v2 extra contigs NODE 319993	1210	UQ01TA7A171416	5	40
IWB64369	RFL Contig3483_411	CT	Codominant Null	detects poly single locus	InfiniumII	no hit	Transition	.	TAS v2 extra contigs NODE 319993	993	UQ01TA7A171413	10	59
IWB27983	Excalibur_c60598_158	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	TAS v2 extra contigs NODE 321967	1682	UQ01TA7A172885	8	93
IWB8251	BS00040929_51	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	TAS v2 extra contigs NODE 321967	1725	UQ01TA7A172886	6	63
IWB35451	IAAV822	AG	MSV	detects poly duplicated loci	InfiniumII	no hit	Transition	.	TAS v2 extra contigs NODE 324846	443	UQ01TA7A175244	10	33
IWB9854	BS00066651_51	AG	Recode AB to AA	detects poly single locus	InfiniumII	no hit	Transition	.	TAS v2 extra contigs NODE 330039	474	UQ01TA7A179272	6	114

IWB3870	BobWhite c55017 291	GT	Codominant	detects poly single locus	InfiniumII	no hit	Transversion	.	7AS v2 extra contigs NODE 336292	121	UQ01TA7A184315	6	71
IWB3869	BobWhite c55017 267	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7AS v2 extra contigs NODE 336292	97	UQ01TA7A184314	10	64
IWB54108	RAC875 c1553 667	AG	Codominant Null	detects poly single locus	InfiniumII	I->V	Transition	nonsynonymous	7AS v2 extra contigs NODE 340823	1537	UQ01TA7A188590	46	101
IWB45320	Kukri c43234 408	AG	Mono	Mono	InfiniumII	no hit	Transition	.	7AS v2 extra contigs NODE 345288	1238	UQ01TA7A191979	7	56
IWB7949	BS00033780 51	AG	BB Multiallelic Null	detects poly duplicated loci	InfiniumII	D->G	Transition	nonsynonymous	7AS v2 extra contigs NODE 357175	4150	UQ01TA7A202478	43	174
IWB39675	Ku c5938 2248	CT	MSV	detects poly duplicated loci	InfiniumII	D->D	Transition	synonymous	7AS v2 extra contigs NODE 364263	3971	UQ01TA7A208228	40	127
IWA7200	w.snp Ku c5938 10491100	AG	MSV	detects poly duplicated loci	InfiniumII	L->L	Transition	synonymous	7AS v2 extra contigs NODE 364263	5489	UQ01TA7A208231	50	132
IWB39676	Ku c5938 4221	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7AS v2 extra contigs NODE 364263	7268	UQ01TA7A208237	43	152
IWB39677	Ku c5938 4231	CT	MSV	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7AS v2 extra contigs NODE 364263	7278	UQ01TA7A208238	44	150
IWB6757	BS00021769 51	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7AS v2 extra contigs NODE 364554	931	UQ01TA7A208507	4	113
IWB39758	Ku c6386 1034	AG	MSV	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7AS v2 extra contigs NODE 408905	2756	UQ01TA7A225780	59	127
IWB44787	Kukri c38338 485	CT	Mono	Mono	InfiniumII	no hit	Transition	.	7AS v2 extra contigs NODE 42896	2270	UQ01TA7A47292	50	173
IWB25834	Excalibur c37505 88	CT	Codominant Null	detects poly single locus	InfiniumII	no hit	Transition	.	7AS v2 extra contigs NODE 484076	1411	UQ01TA7A253424	18	63
IWB35407	IAAV7916	CT	Recode AB to BB	detects poly single locus	InfiniumII	no hit	Transition	.	7AS v2 extra contigs NODE 488064	2129	UQ01TA7A255667	53	147
IWB67628	Tdurum contig12492 680	CT	Mono	Mono	InfiniumII	H->H	Transition	synonymous	7AS v2 extra contigs NODE 513118	1577	UQ01TA7A269656	20	136
IWB72515	Tdurum contig5646 929	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7AS v2 extra contigs NODE 51570	1255	UQ01TA7A48896	20	77
IWB8106	BS00037421 51	AG	Codominant Null	detects poly single locus	InfiniumII	no hit	Transition	.	7AS v2 extra contigs NODE 51570	795	UQ01TA7A48894	22	116
IWB35136	IAAV6043	CT	BB Multiallelic	detects poly duplicated loci	InfiniumII	T->T	Transition	synonymous	7AS v2 extra contigs NODE 536431	1476	UQ01TA7A285774	17	130
IWB33152	GENE-3018 145	CT	MSV	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7AS v2 extra contigs NODE 536431	1791	UQ01TA7A285775	24	108
IWB37134	JD c19177 1284	CT	Codominant	detects poly single locus	InfiniumII	N->S	Transition	nonsynonymous	7AS v2 extra contigs NODE 536431	2254	UQ01TA7A285776	11	85
IWB38678	Ku c18978 400	CT	Mono	Mono	InfiniumII	no hit	Transition	.	7AS v2 extra contigs NODE 536431	4984	UQ01TA7A285783	3	136
IWB60049	RAC875 c66882 301	CT	MSV	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7AS v2 extra contigs NODE 542749	4979	UQ01TA7A289935	7	234
IWB9463	BS00065137 51	AG	MSV	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7AS v2 extra contigs NODE 544580	1195	UQ01TA7A291189	2	99
IWB22966	Excalibur c17447 635	GT	Codominant	detects poly single locus	InfiniumII	no hit	Transversion	.	7AS v2 extra contigs NODE 549894	5373	UQ01TA7A295050	40	113
IWB19811	Ex c13577 632	AC	MSV	detects poly duplicated loci	InfiniumII	no hit	Transversion	.	7AS v2 extra contigs NODE 561964	1690	UQ01TA7A305442	26	101
IWB19810	Ex c13577 1303	CT	MSV	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7AS v2 extra contigs NODE 561964	2422	UQ01TA7A305445	19	58
IWB2725	BobWhite c34770 148	CT	Codominant Null	detects poly single locus	InfiniumII	no hit	Transition	.	7AS v2 extra contigs NODE 566572	241	UQ01TA7A309214	23	145
IWB7484	BS00023225 51	AC	AA Multiallelic	detects poly duplicated loci	InfiniumII	no hit	Transversion	.	7AS v2 extra contigs NODE 580425	2895	UQ01TA7A321732	9	118
IWB12457	BS00110010 51	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7AS v2 extra contigs NODE 587673	2390	UQ01TA7A328204	24	71
IWB9707	BS00066054 51	AG	BB Multiallelic	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7AS v2 extra contigs NODE 588519	13128	UQ01TA7A329025	4	56
IWB73316	Tdurum contig75811 1629	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7AS v2 extra contigs NODE 588519	6318	UQ01TA7A328969	3	48
IWB59812	RAC875 c63822 185	AG	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7AS v2 extra contigs NODE 591637	3517	UQ01TA7A332309	51	121
IWB41941	Kukri c1831 269	CT	Mono	Mono	InfiniumII	I->L	Transition	synonymous	7AS v2 extra contigs NODE 596374	234	UQ01TA7A336867	4	11
IWB8446	BS00046998 51	AG	MSV	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7AS v2 extra contigs NODE 600542	201	UQ01TA7A340380	2	56
IWB75141	tblb0057f21 146	AC	AA Multiallelic	detects poly duplicated loci	InfiniumII	no hit	Transversion	.	7AS v2 extra contigs NODE 600603	3829	UQ01TA7A340457	29	95
IWB18514	D GBQ4KXB01CK7BP 333	CT	Mono	Mono	InfiniumII	I->I	Transition	synonymous	7AS v2 extra contigs NODE 602971	2991	UQ01TA7A342610	19	69
IWB52741	Ra c8425 1331	AG	AA Multiallelic	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7AS v2 extra contigs NODE 606884	10070	UQ01TA7A346381	36	123
IWB28435	Excalibur c662 1047	AG	MSV	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7AS v2 extra contigs NODE 614091	8316	UQ01TA7A353703	36	73
IWB48426	Kukri c96054 123	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7AS v2 extra contigs NODE 616680	7422	UQ01TA7A356370	73	147
IWB11228	BS00082055 51	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7AS v2 extra contigs NODE 640300	406	UQ01TA7A376388	2	31
IWB52245	Ra c54443 444	GT	AA Multiallelic Null	detects poly duplicated loci	InfiniumII	no hit	Transversion	.	7AS v2 extra contigs NODE 640548	4555	UQ01TA7A376679	27	59
IWB71070	Tdurum contig42297 5306	AG	Mono	Mono	InfiniumII	Y->Y	Transition	synonymous	7AS v2 extra contigs NODE 678911	18930	UQ01TA7A386476	2	117
IWB10483	BS00070642 51	CT	Codominant Null	detects poly single locus	InfiniumII	no hit	Transition	.	7AS v2 extra contigs NODE 680597	1415	UQ01TA7A386818	7	24
IWB10484	BS00070643 51	GT	Codominant	detects poly single locus	InfiniumII	no hit	Transversion	.	7AS v2 extra contigs NODE 680597	1417	UQ01TA7A386819	7	24
IWB48383	Kukri c94705 239	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7AS v2 extra contigs NODE 701928	1859	UQ01TA7A392487	4	40
IWA3267	w.snp Ex c30239 39179460	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7AS v2 extra contigs NODE 741816	4914	UQ01TA7A406156	73	155
IWB26590	Excalibur c44794 122	CT	MSV	detects poly duplicated loci	InfiniumII	L->L	Transition	synonymous	7AS v2 extra contigs NODE 746460	7090	UQ01TA7A408150	2	7

IWB58109	RAC875 c45031 574	AG	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7AS v2 extra contigs NODE 749404	11574	UQ01TA7A409336	5	17
IWB22185	Excalibur c12989 3283	CT	Mono	Mono	InfiniumII	no hit	Transition	.	7AS v2 extra contigs NODE 751372	63	UQ01TA7A410235	4	54
IWB50393	Kukri rep c80771 409	AC	MSV	detects poly duplicated loci	InfiniumII	A->A	Transversion	synonymous	7AS v2 extra contigs NODE 774889	6693	UQ01TA7A422600	42	91
IWB36315	IACX8614	AT	MSV	detects poly duplicated loci	InfiniumII	no hit	Transversion	.	7AS v2 extra contigs NODE 793763	2108	UQ01TA7A435239	5	92
IWB10977	BS00077952 51	CT	Codominant Null	detects poly single locus	InfiniumII	N->N	Transition	synonymous	7AS v2 extra contigs NODE 804548	1297	UQ01TA7A442803	6	22
IWB53270	RAC875 c11156 137	AG	MSV	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7AS v2 extra contigs NODE 804548	2131	UQ01TA7A442807	5	39
IWB53271	RAC875 c11156 164	AG	MSV	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7AS v2 extra contigs NODE 804548	2158	UQ01TA7A442808	6	42
IWB8826	BS00061123 51	AG	Codominant Null	detects poly single locus	InfiniumII	no hit	Transition	.	7AS v2 extra contigs NODE 804548	2740	UQ01TA7A442813	10	73
IWB60377	RAC875 c75528 355	AG	BB Multiallelic	detects poly duplicated loci	InfiniumII	I->V	Transition	nonsynonymous	7AS v2 extra contigs NODE 811718	6066	UQ01TA7A448322	29	76
IWA797	wsnp CAP11 c651 429263	AG	BB Multiallelic	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7AS v2 extra contigs NODE 813965	218	UQ01TA7A449822	9	58
IWA796	wsnp CAP11 c651 429138	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7AS v2 extra contigs NODE 813965	93	UQ01TA7A449821	11	83
IWB44597	Kukri c36710 1007	CT	Codominant	detects poly single locus	InfiniumII	T->T	Transition	synonymous	7AS v2 extra contigs NODE 815400	80	UQ01TA7A450823	8	47
IWB9696	BS00066015 51	AG	MSV	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7AS v2 extra contigs NODE 828054	2475	UQ01TA7A460215	75	162
IWB33243	GENE-3129 436	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7AS v2 extra contigs NODE 837533	871	UQ01TA7A467551	8	167
IWB44901	Kukri c39230 362	CT	BB Multiallelic	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7AS v2 extra contigs NODE 859542	54	UQ01TA7A483657	7	33
IWB56005	RAC875 c26917 144	CT	Mono	Mono	InfiniumII	Y->Y	Transition	synonymous	7AS v2 extra contigs NODE 860307	3272	UQ01TA7A484096	2	97
IWB12011	BS00097659 51	AG	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7AS v2 extra contigs NODE 924331	400	UQ01TA7A500368	20	103
IWB11685	BS00091003 51	CT	Codominant Null	detects poly single locus	InfiniumII	no hit	Transition	.	7AS v2 extra contigs NODE 959921	3026	UQ01TA7A510860	6	31
IWB33938	GENE-4375 382	AG	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7AS v2 extra contigs NODE 960071	1203	UQ01TA7A510937	28	121
IWB5993	BS00004971 51	CT	Recode AB to AA	detects poly single locus	InfiniumII	no hit	Transition	.	7AS v2 extra contigs NODE 96325	3610	UQ01TA7A57798	8	16
IWB9203	BS00064143 51	AG	Recode AB to AA	detects poly single locus	InfiniumII	no hit	Transition	.	7AS v2 extra contigs NODE 969799	7223	UQ01TA7A514273	23	77
IWB10141	BS00067759 51	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7AS v2 extra contigs NODE 970257	11571	UQ01TA7A514452	36	115
IWA7784	wsnp Ra c250 526345	AG	MSV	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7AS v2 extra contigs NODE 970628	1279	UQ01TA7A514623	27	105
IWB72199	Tdurum contig51645 137	CT	Recode AB to AA	detects poly single locus	InfiniumII	no hit	Transition	.	7AS v2 extra contigs NODE 985136	155	UQ01TA7A520402	7	25
IWB58718	RAC875 c511 216	AG	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7B SvnBuild v2.0	10142144	UQ01TA7B23841	2	115
IWB13850	CAP7 c1748 201	GT	MSV	detects poly duplicated loci	InfiniumII	no hit	Transversion	.	7B SvnBuild v2.0	10275107	UQ01TA7B24058	5	49
IWB2567	BobWhite c3269 141	AG	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7B SvnBuild v2.0	10743034	UQ01TA7B24789	18	59
IWB2568	BobWhite c3269 191	AG	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7B SvnBuild v2.0	10743084	UQ01TA7B24790	36	88
IWB37345	JD c3269 342	AG	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7B SvnBuild v2.0	10743235	UQ01TA7B24792	51	118
IWB56714	RAC875 c3229 165	CT	Codominant	detects poly single locus	InfiniumII	A->A	Transition	synonymous	7B SvnBuild v2.0	10743572	UQ01TA7B24794	73	146
IWB343	BobWhite c12355 1548	AG	Recode AB to BB	detects poly single locus	InfiniumII	N->N	Transition	synonymous	7B SvnBuild v2.0	10758907	UQ01TA7B24829	47	105
IWB344	BobWhite c12355 1590	CT	MSV	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7B SvnBuild v2.0	10758949	UQ01TA7B24830	29	74
IWB71981	Tdurum contig49575 1207	CT	MSV	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7B SvnBuild v2.0	10758949	UQ01TA7B24830	29	74
IWB45661	Kukri c46447 1738	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7B SvnBuild v2.0	10759573	UQ01TA7B24834	19	56
IWB71916	Tdurum contig48695 527	CT	Codominant	detects poly single locus	InfiniumII	N->D	Transition	nonsynonymous	7B SvnBuild v2.0	10881837	UQ01TA7B25078	41	142
IWB71464	Tdurum contig43945 501	CT	Codominant	detects poly single locus	InfiniumII	T->A	Transition	nonsynonymous	7B SvnBuild v2.0	11147107	UQ01TA7B25649	17	176
IWB71499	Tdurum contig44171 1744	AC	Codominant	detects poly single locus	InfiniumII	no hit	Transversion	.	7B SvnBuild v2.0	11170376	UQ01TA7B25710	8	85
IWA1524	wsnp_Ex c11636 18742884	CT	Codominant	detects poly single locus	InfiniumII	N->D	Transition	nonsynonymous	7B SvnBuild v2.0	11210811	UQ01TA7B25747	10	127
IWB71444	Tdurum contig43589 825	CT	Codominant	detects poly single locus	InfiniumII	N->S	Transition	nonsynonymous	7B SvnBuild v2.0	11289575	UQ01TA7B25945	5	75
IWB72398	Tdurum contig54833 200	AC	Codominant	detects poly single locus	InfiniumII	no hit	Transversion	.	7B SvnBuild v2.0	11297497	UQ01TA7B25956	10	179
IWB22540	Excalibur c15029 350	AG	Codominant	detects poly single locus	InfiniumII	M->T	Transition	nonsynonymous	7B SvnBuild v2.0	11396499	UQ01TA7B26064	2	16
IWB45513	Kukri c45242 1220	AG	Codominant Null	detects poly single locus	InfiniumII	S->S	Transition	synonymous	7B SvnBuild v2.0	11416644	UQ01TA7B26153	9	181
IWA6717	wsnp_Ku c21752 31528824	AG	Codominant	detects poly single locus	InfiniumII	V->A	Transition	nonsynonymous	7B SvnBuild v2.0	11642240	UQ01TA7B26611	74	175
IWB4816	BobWhite rep c49910 432	AG	Codominant	detects poly single locus	InfiniumII	R->R	Transition	synonymous	7B SvnBuild v2.0	11650707	UQ01TA7B26622	2	104
IWB71924	Tdurum contig48847 202	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7B SvnBuild v2.0	11809133	UQ01TA7B26963	14	108
IWB69052	Tdurum contig25380 69	AG	BB Multiallelic	detects poly duplicated loci	InfiniumII	Y->Y	Transition	synonymous	7B SvnBuild v2.0	11901157	UQ01TA7B27170	4	146
IWB73494	Tdurum contig81318 116	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7B SvnBuild v2.0	12017275	UQ01TA7B27432	28	116

IWB5325	BobWhite rep c63008 468	AC	Codominant	detects poly single locus	InfiniumII	S->S	Transversion	synonymous	7B SvnBuild v2.0	12092655	UQ01TA7B27699	7	151
IWB63652	RFL Contig1472 821	CT	MSV	detects poly duplicated loci	InfiniumII	D->D	Transition	synonymous	7B SvnBuild v2.0	12096549	UQ01TA7B27703	2	51
IWB71582	Tdurum contig45195 117	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7B SvnBuild v2.0	12143353	UQ01TA7B27806	5	163
IWB26679	Excalibur c4556 776	AG	Codominant	detects poly single locus	InfiniumII	S->S	Transition	synonymous	7B SvnBuild v2.0	12264248	UQ01TA7B28206	14	97
IWB73208	Tdurum contig71786 231	AC	MSV	detects poly duplicated loci	InfiniumII	L->L	Transversion	synonymous	7B SvnBuild v2.0	12264985	UQ01TA7B28209	13	155
IWB62078	RAC875 rep c111726 114	CT	MSV	detects poly duplicated loci	InfiniumII	Q->Q	Transition	synonymous	7B SvnBuild v2.0	12326339	UQ01TA7B28366	5	169
IWB56242	RAC875 c2887 52	AG	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7B SvnBuild v2.0	12337234	UQ01TA7B28402	24	54
IWB40901	Kukri c12822 132	AG	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7B SvnBuild v2.0	12436579	UQ01TA7B28614	64	152
IWB73035	Tdurum contig65979 289	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7B SvnBuild v2.0	12442336	UQ01TA7B28622	31	136
IWB72254	Tdurum contig52239 120	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7B SvnBuild v2.0	12452664	UQ01TA7B28641	37	151
IWA3986	wsnp_Ex c47153 52447514	AG	Codominant	detects poly single locus	InfiniumII	S->S	Transition	synonymous	7B SvnBuild v2.0	12459998	UQ01TA7B28668	19	72
IWB56558	RAC875 c31078 188	AG	Recode AB to AA	detects poly single locus	InfiniumII	M->T	Transition	nonsynonymous	7B SvnBuild v2.0	12485064	UQ01TA7B28744	14	126
IWB72923	Tdurum contig62981 468	AG	MSV	detects poly duplicated loci	InfiniumII	V->V	Transition	synonymous	7B SvnBuild v2.0	12551978	UQ01TA7B28904	36	142
IWB72924	Tdurum contig62981 470	CT	Codominant	detects poly single locus	InfiniumII	I->V	Transition	nonsynonymous	7B SvnBuild v2.0	12551980	UQ01TA7B28905	36	145
IWB68767	Tdurum contig19022 1524	GT	BB Multiallelic	detects poly duplicated loci	InfiniumII	H->Q	Transversion	nonsynonymous	7B SvnBuild v2.0	12624532	UQ01TA7B29038	16	131
IWB68768	Tdurum contig19022 1555	AG	Codominant	detects poly single locus	InfiniumII	T->A	Transition	nonsynonymous	7B SvnBuild v2.0	12624563	UQ01TA7B29039	14	147
IWB33622	GENE-3849 389	AC	Recode AB to AA	detects poly single locus	InfiniumII	no hit	Transversion	.	7B SvnBuild v2.0	12661849	UQ01TA7B29116	14	100
IWB30564	Excalibur rep c111831 114	AG	AA Multiallelic	detects poly duplicated loci	InfiniumII	S->S	Transition	synonymous	7B SvnBuild v2.0	12661957	UQ01TA7B29117	10	65
IWB61623	RAC875 rep c106651 490	CT	Codominant	detects poly single locus	InfiniumII	A->A	Transition	synonymous	7B SvnBuild v2.0	12662014	UQ01TA7B29118	12	61
IWB39070	Ku c28853 1518	AG	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7B SvnBuild v2.0	12823312	UQ01TA7B29573	40	91
IWB6855	BS00022009_51	CT	Recode AB to AA	detects poly single locus	InfiniumII	O->O	Transition	synonymous	7B SvnBuild v2.0	12918759	UQ01TA7B29839	39	88
IWA4250	wsnp_Ex c54863 57588264	AG	Codominant	detects poly single locus	InfiniumII	S->S	Transition	synonymous	7B SvnBuild v2.0	12956925	UQ01TA7B29895	15	41
IWA3437	wsnp_Ex c33461 41945399	AG	Codominant	detects poly single locus	InfiniumII	L->P	Transition	nonsynonymous	7B SvnBuild v2.0	12971890	UQ01TA7B29920	54	158
IWB52533	Ra c71101 755	CT	Recode AB to AA	detects poly single locus	InfiniumII	I->I	Transition	synonymous	7B SvnBuild v2.0	1304661	UQ01TA7B02265	8	168
IWA3965	wsnp_Ex c46274 51831129	CT	AA Multiallelic	detects poly duplicated loci	InfiniumII	A->A	Transition	synonymous	7B SvnBuild v2.0	1304718	UQ01TA7B02266	64	164
IWB53763	RAC875 c13696 1384	AG	Recode AB to AA	detects poly single locus	InfiniumII	D->D	Transition	synonymous	7B SvnBuild v2.0	13048228	UQ01TA7B30105	15	42
IWB33941	GENE-4376 518	CT	Recode AB to AA	detects poly single locus	InfiniumII	no hit	Transition	.	7B SvnBuild v2.0	1305519	UQ01TA7B02268	12	125
IWA8177	wsnp_Ra rep c71101 69119989	AC	Codominant	detects poly single locus	InfiniumII	V->V	Transversion	synonymous	7B SvnBuild v2.0	1305871	UQ01TA7B02271	48	107
IWA3852	wsnp_Ex c43096 49510017	CT	Recode AB to AA	detects poly single locus	InfiniumII	R->R	Transition	synonymous	7B SvnBuild v2.0	13066563	UQ01TA7B30161	14	59
IWA3853	wsnp_Ex c43096 49510056	AG	MSV	detects poly duplicated loci	InfiniumII	H->H	Transition	synonymous	7B SvnBuild v2.0	13066602	UQ01TA7B30162	4	13
IWA3854	wsnp_Ex c43096 49510164	AG	MSV	detects poly duplicated loci	InfiniumII	Y->Y	Transition	synonymous	7B SvnBuild v2.0	13066774	UQ01TA7B30163	12	30
IWB59498	RAC875_c60161_281	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7B SvnBuild v2.0	13205354	UQ01TA7B30525	45	115
IWB59499	RAC875_c60161_448	AG	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7B SvnBuild v2.0	13205521	UQ01TA7B30528	40	112
IWA8021	wsnp_Ra c60161 61164295	AG	Codominant	detects poly single locus	InfiniumII	S->S	Transition	synonymous	7B SvnBuild v2.0	13205707	UQ01TA7B30529	9	97
IWA8022	wsnp_Ra c60161 61164325	CT	Codominant	detects poly single locus	InfiniumII	S->S	Transition	synonymous	7B SvnBuild v2.0	13205737	UQ01TA7B30530	7	99
IWB43515	Kukri c2796 1436	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7B SvnBuild v2.0	13206271	UQ01TA7B30535	18	80
IWB28152	Excalibur c62837 164	AG	BB Multiallelic	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7B SvnBuild v2.0	13206391	UQ01TA7B30536	6	76
IWB40249	Kukri c10108_115	AG	Codominant_Null	detects poly single locus	InfiniumII	no hit	Transition	.	7B SvnBuild v2.0	13351001	UQ01TA7B31065	14	105
IWB4278	BobWhite c7082_346	AG	Codominant	detects poly single locus	InfiniumII	N->N	Transition	synonymous	7B SvnBuild v2.0	13351203	UQ01TA7B31066	6	81
IWB62881	RAC875 rep c72524_90	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7B SvnBuild v2.0	13359637	UQ01TA7B31115	3	73
IWB42156	Kukri c1957_581	AG	Codominant	detects poly single locus	InfiniumII	F->F	Transition	synonymous	7B SvnBuild v2.0	13359832	UQ01TA7B31120	9	72
IWB73363	Tdurum contig76289_1530	AG	MSV	detects poly duplicated loci	InfiniumII	D->D	Transition	synonymous	7B SvnBuild v2.0	13360271	UQ01TA7B31122	15	73
IWA640	wsnp_CAP11_c103_134545	CT	MSV	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7B SvnBuild v2.0	13389221	UQ01TA7B31252	13	125
IWB5785	BS00001144_51	AG	MSV	detects poly duplicated loci	InfiniumII	G->G	Transition	synonymous	7B SvnBuild v2.0	13417299	UQ01TA7B31378	43	145
IWB34141	GENE-4746_978	CT	MSV	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7B SvnBuild v2.0	13524481	UQ01TA7B31771	8	72
IWB48361	Kukri c9405_379	AG	BB Multiallelic	detects poly duplicated loci	InfiniumII	Y->Y	Transition	synonymous	7B SvnBuild v2.0	13541570	UQ01TA7B31825	6	70
IWB11629	BS00089938_51	CT	MSV	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7B SvnBuild v2.0	13775480	UQ01TA7B32541	15	32

IWB6876	BS00022045_51	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7B_SvnBuild_v2.0	13788909	UQ01TA7B32588	27	96
IWB39590	Ku_c5351_1820	AG	Codominant	detects poly single locus	InfiniumII	S->S	Transition	synonymous	7B_SvnBuild_v2.0	13825355	UQ01TA7B32662	11	64
IWB61497	RAC875_reo_c105766_652	AG	Codominant	detects poly single locus	InfiniumII	C->R	Transition	nonsynonymous	7B_SvnBuild_v2.0	14075862	UQ01TA7B33371	4	32
IWB40231	Kukri_c100592_82	CT	Codominant	detects poly single locus	InfiniumII	T->T	Transition	synonymous	7B_SvnBuild_v2.0	14140780	UQ01TA7B33586	32	98
IWA1297	wsnp_Ex_c10193_16730126	CT	AA Multiallelic	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7B_SvnBuild_v2.0	14154847	UQ01TA7B33643	55	119
IWB35241	IAAV6740	AG	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7B_SvnBuild_v2.0	14154958	UQ01TA7B33644	15	123
IWB71577	Tdurum contig44993_359	AG	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7B_SvnBuild_v2.0	14323983	UQ01TA7B34248	6	168
IWB68305	Tdurum contig14821_751	CT	AA Multiallelic	detects poly duplicated loci	InfiniumII	T->T	Transition	synonymous	7B_SvnBuild_v2.0	14333489	UQ01TA7B34270	22	124
IWB71728	Tdurum contig46877_173	AG	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7B_SvnBuild_v2.0	1433428	UQ01TA7B02613	7	63
IWB73273	Tdurum contig75479_559	CT	Mono	Mono	InfiniumII	no hit	Transition	.	7B_SvnBuild_v2.0	14398472	UQ01TA7B34398	6	128
IWB34981	IAAV5167	AG	MSV	detects poly duplicated loci	InfiniumII	W->R	Transition	nonsynonymous	7B_SvnBuild_v2.0	14677401	UQ01TA7B35224	2	34
IWB71465	Tdurum contig43954_1287	AC	Codominant	detects poly single locus	InfiniumII	V->V	Transversion	synonymous	7B_SvnBuild_v2.0	14752224	UQ01TA7B35645	2	94
IWB71466	Tdurum contig43954_2291	AG	Codominant	detects poly single locus	InfiniumII	L->L	Transition	synonymous	7B_SvnBuild_v2.0	14753664	UQ01TA7B35652	9	41
IWA7907	wsnp_Ra_c39394_47110214	AC	Codominant	detects poly single locus	InfiniumII	L->L	Transversion	synonymous	7B_SvnBuild_v2.0	14775362	UQ01TA7B35717	34	77
IWB69655	Tdurum contig28884_302	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7B_SvnBuild_v2.0	14796764	UQ01TA7B35794	2	72
IWB69657	Tdurum contig28884_460	AC	Codominant	detects poly single locus	InfiniumII	S->S	Transversion	synonymous	7B_SvnBuild_v2.0	14797145	UQ01TA7B35797	33	78
IWA2824	wsnp_Ex_c2365_4431185	CT	AA Multiallelic	detects poly duplicated loci	InfiniumII	V->V	Transition	synonymous	7B_SvnBuild_v2.0	14862376	UQ01TA7B35976	44	139
IWB22759	Excalibur_c16245_801	AG	MSV	detects poly duplicated loci	InfiniumII	D->D	Transition	synonymous	7B_SvnBuild_v2.0	14873015	UQ01TA7B36004	10	109
IWB22760	Excalibur_c16245_840	AG	Recode AB to BB	detects poly single locus	InfiniumII	H->H	Transition	synonymous	7B_SvnBuild_v2.0	14873054	UQ01TA7B36005	39	91
IWB73519	Tdurum contig81683_217	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7B_SvnBuild_v2.0	1491560	UQ01TA7B02757	2	26
IWB71507	Tdurum contig44206_433	GT	BB Multiallelic	detects poly duplicated loci	InfiniumII	E->A	Transversion	nonsynonymous	7B_SvnBuild_v2.0	15057324	UQ01TA7B36547	19	38
IWB38489	Ku_c15539_433	AC	Codominant	detects poly single locus	InfiniumII	S->A	Transversion	nonsynonymous	7B_SvnBuild_v2.0	15069948	UQ01TA7B36607	22	105
IWB14974	CAP8_c8935_183	CT	MSV	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7B_SvnBuild_v2.0	15132654	UQ01TA7B36860	3	54
IWB14321	CAP7_c9291_206	CT	Recode AB to BB	detects poly single locus	InfiniumII	no hit	Transition	.	7B_SvnBuild_v2.0	15132672	UQ01TA7B36862	19	53
IWB58601	RAC875_c49954_785	AG	Codominant	detects poly single locus	InfiniumII	N->N	Transition	synonymous	7B_SvnBuild_v2.0	15134094	UQ01TA7B36880	12	182
IWB71593	Tdurum contig45280_451	CT	MSV	detects poly duplicated loci	InfiniumII	D->G	Transition	nonsynonymous	7B_SvnBuild_v2.0	15134810	UQ01TA7B36889	7	163
IWB58602	RAC875_c49954_992	AG	MSV	detects poly duplicated loci	InfiniumII	A->A	Transition	synonymous	7B_SvnBuild_v2.0	15134839	UQ01TA7B36890	65	153
IWB58600	RAC875_c49954_1172	GT	Codominant	detects poly single locus	InfiniumII	R->S	Transversion	nonsynonymous	7B_SvnBuild_v2.0	15135117	UQ01TA7B36895	13	146
IWB1478	BobWhite_c21129_830	CT	Recode AB to BB	detects poly single locus	InfiniumII	S->S	Transition	synonymous	7B_SvnBuild_v2.0	15139205	UQ01TA7B36917	37	87
IWB7586	BS00025724_51	AG	AA Multiallelic Null	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7B_SvnBuild_v2.0	15175494	UQ01TA7B37012	23	77
IWB57014	RAC875_c34939_86	AC	Codominant	detects poly single locus	InfiniumII	L->L	Transversion	synonymous	7B_SvnBuild_v2.0	15274623	UQ01TA7B37417	25	108
IWB42818	Kukri_c2348_2340	AG	Codominant	detects poly single locus	InfiniumII	T->A	Transition	nonsynonymous	7B_SvnBuild_v2.0	15276730	UQ01TA7B37430	2	73
IWA3958	wsnp_Ex_c46061_51675763	AG	Codominant	detects poly single locus	InfiniumII	A->A	Transition	synonymous	7B_SvnBuild_v2.0	153750	UQ01TA7B00380	8	131
IWB71980	Tdurum contig49572_643	CT	Codominant	detects poly single locus	InfiniumII	A->A	Transition	synonymous	7B_SvnBuild_v2.0	1550066	UQ01TA7B02923	91	186
IWB70127	Tdurum contig30909_76	GT	Codominant	detects poly single locus	InfiniumII	no hit	Transversion	.	7B_SvnBuild_v2.0	15508470	UQ01TA7B38081	51	143
IWB51062	Ra_c1352_219	AC	Mono	Mono	InfiniumII	no hit	Transversion	.	7B_SvnBuild_v2.0	1576443	UQ01TA7B02974	79	173
IWB3402	BobWhite_c4481_96	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7B_SvnBuild_v2.0	1623038	UQ01TA7B03107	59	150
IWB3251	BobWhite_c42613_62	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7B_SvnBuild_v2.0	1672343	UQ01TA7B03185	2	129
IWB10705	BS00074083_51	AG	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7B_SvnBuild_v2.0	1803455	UQ01TA7B03733	16	156
IWB6255	BS00010616_51	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7B_SvnBuild_v2.0	1837113	UQ01TA7B03844	5	123
IWB7527	BS00024215_51	CT	MSV	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7B_SvnBuild_v2.0	2189013	UQ01TA7B04591	18	158
IWB7285	BS00022841_51	AG	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7B_SvnBuild_v2.0	2199415	UQ01TA7B04626	17	136
IWB34706	IAAV3391	CG	Codominant	detects poly single locus	InfiniumII	no hit	Transversion	.	7B_SvnBuild_v2.0	2204630	UQ01TA7B04634	13	151
IWA4129	wsnp_Ex_c52259_55922750	AG	MSV	detects poly duplicated loci	InfiniumII	Y->Y	Transition	synonymous	7B_SvnBuild_v2.0	238484	UQ01TA7B00580	49	166
IWB27984	Excalibur_c60612_236	CT	BB Multiallelic	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7B_SvnBuild_v2.0	245405	UQ01TA7B00605	2	179
IWB67979	Tdurum contig13431_127	CT	Codominant	detects poly single locus	InfiniumII	K->R	Transition	nonsynonymous	7B_SvnBuild_v2.0	2513038	UQ01TA7B05088	7	145
IWB72316	Tdurum contig5352_556	CT	AA Multiallelic	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7B_SvnBuild_v2.0	254132	UQ01TA7B00646	15	65

IWB72335	Tdurum contig53986_316	AC	Codominant	detects poly single locus	InfiniumII	no hit	Transversion	.	7B_SvnBuild_v2.0	2918965	UQ01TA7B06256	9	123
IWA1404	wsnp_Ex_c10903_17717179	AG	MSV	detects poly duplicated loci	InfiniumII	T->T	Transition	synonymous	7B_SvnBuild_v2.0	2929742	UQ01TA7B06280	14	95
IWB54563	RAC875_c17861_221	GT	MSV	detects poly duplicated loci	InfiniumII	S->A	Transversion	nonsynonymous	7B_SvnBuild_v2.0	296803	UQ01TA7B00702	3	37
IWB21814	Excalibur_c11146_1322	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7B_SvnBuild_v2.0	3154074	UQ01TA7B07175	34	126
IWB54221	RAC875_c1610_485	AG	MSV	detects poly duplicated loci	InfiniumII	G->G	Transition	synonymous	7B_SvnBuild_v2.0	317013	UQ01TA7B00749	3	13
IWB56153	RAC875_c28057_144	CT	MSV	detects poly duplicated loci	InfiniumII	L->L	Transition	synonymous	7B_SvnBuild_v2.0	317543	UQ01TA7B00759	4	10
IWB34209	GENE-4867_470	AG	Mono	Mono	InfiniumII	no hit	Transition	.	7B_SvnBuild_v2.0	3189555	UQ01TA7B07326	12	149
IWA3121	wsnp_Ex_c27666_36847022	CT	Codominant	detects poly single locus	InfiniumII	D->G	Transition	nonsynonymous	7B_SvnBuild_v2.0	3259181	UQ01TA7B07587	47	106
IWB2872	BobWhite_c36864_159	AG	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7B_SvnBuild_v2.0	3294487	UQ01TA7B07783	9	136
IWB46358	Kukri_c53648_585	GT	Codominant	detects poly single locus	InfiniumII	R->R	Transversion	synonymous	7B_SvnBuild_v2.0	3400158	UQ01TA7B08073	31	62
IWB59735	RAC875_c6303_529	CT	MSV	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7B_SvnBuild_v2.0	3473394	UQ01TA7B08329	47	147
IWB61065	RAC875_c9267_2130	AC	Mono	Mono	InfiniumII	N->K	Transversion	nonsynonymous	7B_SvnBuild_v2.0	3497540	UQ01TA7B08406	4	85
IWB73957	Tdurum contig9619_68	AG	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7B_SvnBuild_v2.0	3700635	UQ01TA7B08932	5	83
IWB73956	Tdurum contig9619_123	GT	Codominant	detects poly single locus	InfiniumII	no hit	Transversion	.	7B_SvnBuild_v2.0	3700690	UQ01TA7B08933	43	131
IWB55522	RAC875_c23521_589	AG	BB Multiallelic	detects poly duplicated loci	InfiniumII	I->V	Transition	nonsynonymous	7B_SvnBuild_v2.0	380631	UQ01TA7B00871	28	118
IWB59446	RAC875_c5965_317	AG	Codominant	detects poly single locus	InfiniumII	V->V	Transition	synonymous	7B_SvnBuild_v2.0	4171099	UQ01TA7B10488	3	59
IWB4969	BobWhite rep_c51665_281	CT	MSV	detects poly duplicated loci	InfiniumII	C->C	Transition	synonymous	7B_SvnBuild_v2.0	4530996	UQ01TA7B11540	5	44
IWB6883	BS00022056_51	AG	AA Multiallelic	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7B_SvnBuild_v2.0	4531530	UQ01TA7B11541	19	122
IWB54370	RAC875_c16839_188	CT	Codominant	detects poly single locus	InfiniumII	S->S	Transition	synonymous	7B_SvnBuild_v2.0	4968972	UQ01TA7B12700	14	130
IWB48960	Kukri rep_c104213_234	AG	Codominant	detects poly single locus	InfiniumII	P->P	Transition	synonymous	7B_SvnBuild_v2.0	4969274	UQ01TA7B12701	5	107
IWB62205	RAC875.rep_c113331_55	GT	Codominant	detects poly single locus	InfiniumII	no hit	Transversion	.	7B_SvnBuild_v2.0	5000283	UQ01TA7B12791	12	112
IWB48576	Kukri rep_c100925_359	AC	Codominant	detects poly single locus	InfiniumII	no hit	Transversion	.	7B_SvnBuild_v2.0	5110216	UQ01TA7B13102	4	114
IWB42211	Kukri_c19823_491	AG	Codominant	detects poly single locus	InfiniumII	A->A	Transition	synonymous	7B_SvnBuild_v2.0	5167528	UQ01TA7B13187	49	155
IWB52	BobWhite_c10364_57	AG	Codominant	detects poly single locus	InfiniumII	T->A	Transition	nonsynonymous	7B_SvnBuild_v2.0	5188482	UQ01TA7B13229	9	144
IWB54284	RAC875_c1638_165	AG	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7B_SvnBuild_v2.0	5213697	UQ01TA7B13308	6	151
IWB34169	GENE-4790_279	AC	AA Multiallelic	detects poly duplicated loci	InfiniumII	no hit	Transversion	.	7B_SvnBuild_v2.0	5820953	UQ01TA7B14701	4	134
IWA3886	wsnp_Ex_c4408_7939986	CT	Codominant	detects poly single locus	InfiniumII	V->A	Transition	nonsynonymous	7B_SvnBuild_v2.0	5921609	UQ01TA7B14960	8	116
IWA8418	wsnp_RFL_Contig3054_2955094	CT	Recode AB to AA	detects poly single locus	InfiniumII	V->A	Transition	nonsynonymous	7B_SvnBuild_v2.0	5950234	UQ01TA7B15016	27	167
IWA2272	wsnp_Ex_c1790_3378771	AC	BB Multiallelic	detects poly duplicated loci	InfiniumII	no hit	Transversion	.	7B_SvnBuild_v2.0	5951717	UQ01TA7B15020	32	99
IWB73331	Tdurum contig75931_1967	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7B_SvnBuild_v2.0	6144424	UQ01TA7B15500	15	102
IWA2079	wsnp_Ex_c15972_24385702	GT	Codominant	detects poly single locus	InfiniumII	S->A	Transversion	nonsynonymous	7B_SvnBuild_v2.0	6185860	UQ01TA7B15592	4	123
IWB34256	GENE-4929_245	CT	Mono	Mono	InfiniumII	no hit	Transition	.	7B_SvnBuild_v2.0	6297859	UQ01TA7B15812	12	98
IWB34623	IAAV285	CG	Codominant	detects poly single locus	InfiniumII	T->T	Transversion	synonymous	7B_SvnBuild_v2.0	6318433	UQ01TA7B15840	6	21
IWB34131	GENE-4720_644	AG	Codominant Null	detects poly single locus	InfiniumII	no hit	Transition	.	7B_SvnBuild_v2.0	6325267	UQ01TA7B15847	10	135
IWB58816	RAC875_c52266_76	CT	Codominant	detects poly single locus	InfiniumII	I->I	Transition	synonymous	7B_SvnBuild_v2.0	6339169	UQ01TA7B15882	16	121
IWB57783	RAC875_c4186_1198	AG	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7B_SvnBuild_v2.0	6358713	UQ01TA7B15922	21	131
IWA2964	wsnp_Ex_c2539_4733110	AG	Codominant	detects poly single locus	InfiniumII	A->A	Transition	synonymous	7B_SvnBuild_v2.0	6464330	UQ01TA7B16187	54	174
IWB74635	tblb0037k05_1475	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7B_SvnBuild_v2.0	6500880	UQ01TA7B16286	12	125
IWB50273	Kukri_rep_c73612_444	AG	Recode AB to BB	detects poly single locus	InfiniumII	no hit	Transition	.	7B_SvnBuild_v2.0	6643938	UQ01TA7B16573	28	172
IWB34162	GENE-4775_115	GT	MSV	detects poly duplicated loci	InfiniumII	no hit	Transversion	.	7B_SvnBuild_v2.0	6900949	UQ01TA7B16958	22	126
IWB54910	RAC875_c19880_1414	AG	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7B_SvnBuild_v2.0	6926370	UQ01TA7B17007	4	196
IWB61628	RAC875.rep_c106698_86	CT	Recode AB to AA	detects poly single locus	InfiniumII	S->P	Transition	nonsynonymous	7B_SvnBuild_v2.0	6977383	UQ01TA7B17153	8	86
IWB54688	RAC875_c18513_376	AG	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7B_SvnBuild_v2.0	7055473	UQ01TA7B17367	36	119
IWB49605	Kukri_rep_c113231_244	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7B_SvnBuild_v2.0	7383627	UQ01TA7B18050	19	128
IWB424	BobWhite_c12889_390	CT	Recode AB to AA	detects poly single locus	InfiniumII	C->R	Transition	nonsynonymous	7B_SvnBuild_v2.0	7439884	UQ01TA7B18179	22	145
IWB7155	BS00022562_51	AG	Codominant Null	detects poly single locus	InfiniumII	N->D	Transition	nonsynonymous	7B_SvnBuild_v2.0	74634	UQ01TA7B00205	19	100
IWB51255	Ra_c17151_576	AG	Recode AB to AA	detects poly single locus	InfiniumII	no hit	Transition	.	7B_SvnBuild_v2.0	7800437	UQ01TA7B18964	14	114

IWB74799	tplb004419_1180	AG	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7B_SynBuild_v2.0	7889650	UQ01TA7B19172	18	128
IWB13248	CAP12_c1816_325	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7B_SynBuild_v2.0	852298	UQ01TA7B01659	14	36
IWB73668	Tdurum contig08448_700	AG	MSV	detects poly duplicated loci	InfiniumII	L->P	Transition	nonsynonymous	7B_SynBuild_v2.0	8599111	UQ01TA7B20815	11	123
IWB40480	Kukri_c109962_396	AG	MSV	detects poly duplicated loci	InfiniumII	M->T	Transition	nonsynonymous	7B_SynBuild_v2.0	878542	UQ01TA7B01700	16	65
IWB73909	Tdurum contig94390_406	AG	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7B_SynBuild_v2.0	8792232	UQ01TA7B21369	9	192
IWB19554	Ex_c101666_634	GT	Codominant	detects poly single locus	InfiniumII	K->T	Transversion	nonsynonymous	7B_SynBuild_v2.0	879416	UQ01TA7B01702	33	124
IWB40562	Kukri_c11274_506	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7B_SynBuild_v2.0	8856493	UQ01TA7B21519	16	468
IWB12413	BS00109533_51	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7B_SynBuild_v2.0	9064746	UQ01TA7B21860	31	175
IWB6907	BS00022106_51	AG	BB Multiallelic	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7B_SynBuild_v2.0	9064811	UQ01TA7B21861	23	154
IWB43119	Kukri_c2539_371	AG	Recode AB to AA	detects poly single locus	InfiniumII	G->G	Transition	synonymous	7B_SynBuild_v2.0	9843377	UQ01TA7B23314	5	142
IWB24749	Excalibur_c28715_447	CT	Codominant	detects poly single locus	InfiniumII	V->A	Transition	nonsynonymous	7BL_v2_extra contigs NODE_1021167	2700	UQ01TA7B1539279	35	132
IWB74566	tplb0035h03_1251	CT	BB Multiallelic	detects poly duplicated loci	InfiniumII	D->D	Transition	synonymous	7BL_v2_extra contigs NODE_1024853	3326	UQ01TA7B1541765	14	117
IWB12159	BS00101087_51	AG	MSV	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7BL_v2_extra contigs NODE_1033245	3213	UQ01TA7B1547353	16	169
IWB71473	Tdurum contig43995_370	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7BL_v2_extra contigs NODE_1045763	4595	UQ01TA7B1556081	8	49
IWB71474	Tdurum contig43995_611	AG	Recode AB to AA	detects poly single locus	InfiniumII	L->L	Transition	synonymous	7BL_v2_extra contigs NODE_1045763	4890	UQ01TA7B1556082	11	29
IWB8002	BS00035234_51	AG	MSV	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7BL_v2_extra contigs NODE_1059531	7433	UQ01TA7B1566105	2	11
IWB9120	BS00063744_51	AG	MSV	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7BL_v2_extra contigs NODE_1092768	1525	UQ01TA7B1593620	11	44
IWB25486	Excalibur_c3468_324	AG	MSV	detects poly duplicated loci	InfiniumII	F->F	Transition	synonymous	7BL_v2_extra contigs NODE_1095557	2082	UQ01TA7B1595952	15	97
IWB71571	Tdurum contig44948_1132	AG	BB Multiallelic	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7BL_v2_extra contigs NODE_1100461	4681	UQ01TA7B1599558	8	54
IWB31273	Excalibur rep_c77206_397	AG	BB Multiallelic	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7BL_v2_extra contigs NODE_1100461	4808	UQ01TA7B1599559	14	66
IWB35376	IAAV7708	AT	Codominant	detects poly single locus	InfiniumI	no hit	Transversion	.	7BL_v2_extra contigs NODE_1100461	4893	UQ01TA7B1599560	14	58
IWB59453	RAC875_c59682_105	AG	MSV	detects poly duplicated loci	InfiniumII	S->S	Transition	synonymous	7BL_v2_extra contigs NODE_1100461	5729	UQ01TA7B1599567	3	55
IWB59454	RAC875_c59682_144	CT	Recode AB to AA	detects poly single locus	InfiniumII	P->P	Transition	synonymous	7BL_v2_extra contigs NODE_1100461	5768	UQ01TA7B1599568	3	47
IWB31241	Excalibur rep_c75066_126	AG	BB Multiallelic	detects poly duplicated loci	InfiniumII	L->L	Transition	synonymous	7BL_v2_extra contigs NODE_1100461	8927	UQ01TA7B1599577	3	64
IWB38949	Ku_c25443_1454	AG	MSV	detects poly duplicated loci	InfiniumII	L->L	Transition	synonymous	7BL_v2_extra contigs NODE_1100637	1594	UQ01TA7B1599810	4	109
IWA4191	wsnp_Ex_c53725_56865973	AC	Codominant	detects poly single locus	InfiniumII	no hit	Transversion	.	7BL_v2_extra contigs NODE_1108948	1170	UQ01TA7B1606997	30	61
IWA8387	wsnp_RFL_Contig2766_2515703	GT	Codominant	detects poly single locus	InfiniumII	M->R	Transversion	nonsynonymous	7BL_v2_extra contigs NODE_1108948	439	UQ01TA7B1606995	7	17
IWB39616	Ku_c556_1056	GT	MSV	detects poly duplicated loci	InfiniumII	no hit	Transversion	.	7BL_v2_extra contigs NODE_1114550	6874	UQ01TA7B1611508	48	131
IWA1354	wsnp_Ex_c10571_17258682	CT	MSV	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7BL_v2_extra contigs NODE_1114550	7600	UQ01TA7B1611511	46	122
IWB44463	Kukri_c35601_253	CT	MSV	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7BL_v2_extra contigs NODE_1116866	42	UQ01TA7B1613388	2	13
IWB5727	BobWhite_s67603_103	AG	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7BL_v2_extra contigs NODE_1132070	6919	UQ01TA7B1626275	30	159
IWB11033	BS00078785_51	GT	AA Multiallelic	detects poly duplicated loci	InfiniumII	no hit	Transversion	.	7BL_v2_extra contigs NODE_1147830	1131	UQ01TA7B1637777	14	30
IWB7643	BS00027058_51	AG	AA Multiallelic	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7BL_v2_extra contigs NODE_1147830	2917	UQ01TA7B1637786	22	67
IWB7642	BS00027054_51	AG	AA Multiallelic	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7BL_v2_extra contigs NODE_1147830	2981	UQ01TA7B1637787	48	97
IWB31227	Excalibur rep_c74234_183	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7BL_v2_extra contigs NODE_1147830	3438	UQ01TA7B1637792	16	68
IWB7218	BS00022700_51	CT	BB Multiallelic	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7BL_v2_extra contigs NODE_1147830	3878	UQ01TA7B1637798	7	38
IWB34634	IAAV2921	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7BL_v2_extra contigs NODE_1149295	2130	UQ01TA7B1639002	50	160
IWB59198	RAC875_c5646_774	AG	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7BL_v2_extra contigs NODE_1150440	32	UQ01TA7B1639761	4	15
IWB9405	BS00064933_51	GT	AA Multiallelic Null	detects poly duplicated loci	InfiniumII	G->G	Transversion	synonymous	7BL_v2_extra contigs NODE_127603	1077	UQ01TA7B941963	7	63
IWB27414	Excalibur_c5374_1252	AG	MSV	detects poly duplicated loci	InfiniumII	L->L	Transition	synonymous	7BL_v2_extra contigs NODE_127603	396	UQ01TA7B941954	7	103
IWB4352	BobWhite_c7544_545	AG	Recode AB to BB	detects poly single locus	InfiniumII	no hit	Transition	.	7BL_v2_extra contigs NODE_132478	1333	UQ01TA7B943920	16	151
IWB22870	Excalibur_c16856_432	CT	Mono	Mono	InfiniumII	Y->Y	Transition	synonymous	7BL_v2_extra contigs NODE_1342799	1514	UQ01TA7B1705991	4	91
IWA7190	wsnp_Ku_c5745_10169129	AG	Codominant	detects poly single locus	InfiniumII	F->S	Transition	nonsynonymous	7BL_v2_extra contigs NODE_1343286	12809	UQ01TA7B1706374	39	113
IWB50374	Kukri_rep_c79716_389	AG	Codominant	detects poly single locus	InfiniumII	T->T	Transition	synonymous	7BL_v2_extra contigs NODE_1346683	1113	UQ01TA7B1708177	4	72
IWB47549	Kukri_c7284_674	GT	Codominant	detects poly single locus	InfiniumII	I->S	Transversion	nonsynonymous	7BL_v2_extra contigs NODE_1346683	1205	UQ01TA7B1708178	6	27
IWB50375	Kukri_rep_c79716_729	AG	MSV	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7BL_v2_extra contigs NODE_1346683	1690	UQ01TA7B1708181	7	146
IWA1736	wsnp_Ex_c1318_2519998	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7BL_v2_extra contigs NODE_1365268	525	UQ01TA7B1719549	6	20

IWB60181	RAC875 c68398 75	AG	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7BL v2 extra contigs NODE 1374711	656	UQ01TA7B1724424	42	137
IWB1060	BobWhite c17544 776	GT	Codominant	detects poly single locus	InfiniumII	V->V	Transversion	synonymous	7BL v2 extra contigs NODE 1382551	948	UQ01TA7B1728499	14	166
IWB59287	RAC875 c5744 115	CT	MSV	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7BL v2 extra contigs NODE 1387700	6531	UQ01TA7B1731078	51	123
IWB62656	RAC875 rep c70147 132	CT	MSV	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7BL v2 extra contigs NODE 1400152	18992	UQ01TA7B1737459	15	154
IWB68549	Tdurum contig16482_252	AG	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7BL v2 extra contigs NODE 1416075	1570	UQ01TA7B1743745	2	37
IWB44831	Kukri c38676 251	CT	Recode AB to BB	detects poly single locus	InfiniumII	no hit	Transition	.	7BL v2 extra contigs NODE 1416075	2273	UQ01TA7B1743759	5	13
IWB44832	Kukri c38676 278	CT	MSV	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7BL v2 extra contigs NODE 1416075	2300	UQ01TA7B1743760	2	6
IWB40602	Kukri c11467 1107	CT	MSV	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7BL v2 extra contigs NODE 145880	22728	UQ01TA7B950103	3	28
IWA660	wsnp CAP11 c1196 692246	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7BL v2 extra contigs NODE 1497034	5842	UQ01TA7B1753152	35	114
IWB59569	RAC875 c61016 110	GT	MSV	detects poly duplicated loci	InfiniumII	K->Q	Transversion	nonsynonymous	7BL v2 extra contigs NODE 1605467	2833	UQ01TA7B1784521	13	106
IWB69685	Tdurum contig29016_234	AG	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7BL v2 extra contigs NODE 1633763	139	UQ01TA7B1793845	8	133
IWB42703	Kukri c22826 137	CT	MSV	detects poly duplicated loci	InfiniumII	F->L	Transition	nonsynonymous	7BL v2 extra contigs NODE 173725	816	UQ01TA7B963302	22	90
IWB71622	Tdurum contig45585_432	CT	MSV	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7BL v2 extra contigs NODE 174775	1549	UQ01TA7B963943	10	26
IWB73659	Tdurum contig8402_638	AC	Codominant	detects poly single locus	InfiniumII	no hit	Transversion	.	7BL v2 extra contigs NODE 1777966	418	UQ01TA7B1811035	5	20
IWB25295	Excalibur c3309_1180	AG	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7BL v2 extra contigs NODE 1813654	4271	UQ01TA7B1819081	27	139
IWA3423	wsnp_Ex c3309_6096114	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7BL v2 extra contigs NODE 1813654	4319	UQ01TA7B1819082	35	148
IWA4857	wsnp_Ex c8963_14948293	AG	MSV	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7BL v2 extra contigs NODE 184363	80	UQ01TA7B968843	30	73
IWB53550	RAC875 c12549_892	CT	MSV	detects poly duplicated loci	InfiniumII	D->G	Transition	nonsynonymous	7BL v2 extra contigs NODE 1844683	5687	UQ01TA7B1826059	15	38
IWB51787	Ra c3092_741	CT	AA Multiallelic Null	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7BL v2 extra contigs NODE 190958	64	UQ01TA7B972056	9	33
IWB12828	CAP11 c31_213	CT	MSV	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7BL v2 extra contigs NODE 200047	74	UQ01TA7B977133	4	17
IWB58429	RAC875_c4834_694	CT	MSV	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7BL v2 extra contigs NODE 2006345	419	UQ01TA7B1841061	10	78
IWB38566	Ku c16895_793	CT	MSV	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7BL v2 extra contigs NODE 205045	1909	UQ01TA7B979762	27	104
IWB38567	Ku c16895_803	AG	MSV	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7BL v2 extra contigs NODE 205045	1919	UQ01TA7B979763	26	105
IWB10676	BS00073560_51	CT	Codominant Null	detects poly single locus	InfiniumII	no hit	Transition	.	7BL v2 extra contigs NODE 208994	941	UQ01TA7B981989	14	151
IWB12357	BS00108264_51	AG	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7BL v2 extra contigs NODE 213988	594	UQ01TA7B985068	14	172
IWB7456	BS00023166_51	AC	Codominant Null	detects poly single locus	InfiniumII	no hit	Transversion	.	7BL v2 extra contigs NODE 213988	650	UQ01TA7B985069	37	150
IWA5001	wsnp_Ex rep c101269_86664147	CT	MSV	detects poly duplicated loci	InfiniumII	F->F	Transition	nonsynonymous	7BL v2 extra contigs NODE 245681	156	UQ01TA7B1003824	38	92
IWB52394	Ra c66200_955	GT	Mono	Mono	InfiniumII	D->E	Transversion	nonsynonymous	7BL v2 extra contigs NODE 245681	195	UQ01TA7B1003825	2	105
IWB71690	Tdurum contig4658_106	CT	MSV	detects poly duplicated loci	InfiniumII	G->G	Transition	synonymous	7BL v2 extra contigs NODE 269967	323	UQ01TA7B1019732	23	91
IWB18673	D GBUVHFX02F4VT5_101	CT	Codominant Null	detects poly single locus	InfiniumII	C->R	Transition	nonsynonymous	7BL v2 extra contigs NODE 283464	377	UQ01TA7B1029183	5	136
IWA4803	wsnp_Ex c8400_14157318	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7BL v2 extra contigs NODE 288716	536	UQ01TA7B1033218	36	80
IWB48538	Kukri c99107_143	CT	Codominant	detects poly single locus	InfiniumII	M->V	Transition	nonsynonymous	7BL v2 extra contigs NODE 289666	1930	UQ01TA7B1033848	32	90
IWB45705	Kukri c4682_1095	CT	Mono	Mono	InfiniumII	no hit	Transition	.	7BL v2 extra contigs NODE 290999	160	UQ01TA7B1034989	2	45
IWB54467	RAC875 c1742_2710	GT	AA Multiallelic	detects poly duplicated loci	InfiniumII	L->R	Transversion	nonsynonymous	7BL v2 extra contigs NODE 295291	694	UQ01TA7B1037989	50	124
IWB41511	Kukri c16034_113	AG	Recode AB to AA	detects poly single locus	InfiniumII	I->V	Transition	nonsynonymous	7BL v2 extra contigs NODE 297728	159	UQ01TA7B1039860	12	82
IWB38052	Ku c10179_1837	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7BL v2 extra contigs NODE 316615	2201	UQ01TA7B1054889	12	75
IWB62671	RAC875 rep c70325_264	CT	MSV	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7BL v2 extra contigs NODE 323585	234	UQ01TA7B1060667	4	132
IWB29358	Excalibur c91733_95	AG	Codominant	detects poly single locus	InfiniumII	O->R	Transition	nonsynonymous	7BL v2 extra contigs NODE 329611	119	UQ01TA7B1065761	20	40
IWB10797	BS00075332_51	GT	Codominant	detects poly single locus	InfiniumII	no hit	Transversion	.	7BL v2 extra contigs NODE 333006	892	UQ01TA7B1068969	14	54
IWB11215	BS00081841_51	AG	Recode AB to AA	detects poly single locus	InfiniumII	V->V	Transition	synonymous	7BL v2 extra contigs NODE 345214	358	UQ01TA7B1080302	54	122
IWB14408	CAP7 rep c5216_143	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7BL v2 extra contigs NODE 347233	3444	UQ01TA7B1082102	15	71
IWB29313	Excalibur c90713_226	CT	BB Multiallelic	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7BL v2 extra contigs NODE 365016	437	UQ01TA7B1098543	15	132
IWB43532	Kukri c28160_2017	AG	Recode AB to AA	detects poly single locus	InfiniumII	R->R	Transition	synonymous	7BL v2 extra contigs NODE 378528	171	UQ01TA7B1111809	8	40
IWB57972	RAC875 c4377_461	AG	Mono	Mono	InfiniumII	T->T	Transition	synonymous	7BL v2 extra contigs NODE 420692	2142	UQ01TA7B1153734	10	23
IWB29419	Excalibur c93217_61	AG	MSV	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7BL v2 extra contigs NODE 421157	617	UQ01TA7B1154317	8	107
IWB28473	Excalibur c6738_2072	AG	Codominant	detects poly single locus	InfiniumII	D->D	Transition	synonymous	7BL v2 extra contigs NODE 424880	537	UQ01TA7B1157679	32	96
IWB7576	BS00025305_51	AC	MSV	detects poly duplicated loci	InfiniumII	no hit	Transversion	.	7BL v2 extra contigs NODE 427503	4554	UQ01TA7B1160639	19	67

IWB39778	Ku c6550 1698	CT	MSV	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7BL v2 extra contigs NODE 431988	626	UQ01TA7B1165309	8	123
IWA8570	wsnp RFL Contig4236 4881643	AG	MSV	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7BL v2 extra contigs NODE 438174	2633	UQ01TA7B1171238	163	393
IWB4739	BobWhite rep c49196 319	AG	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7BL v2 extra contigs NODE 444281	1216	UQ01TA7B1177198	17	104
IWB9914	BS00066920 51	AG	MSV	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7BL v2 extra contigs NODE 446490	116	UQ01TA7B1179165	3	25
IWB8624	BS00053286 51	AC	Recode AB to AA	detects poly single locus	InfiniumII	no hit	Transversion	.	7BL v2 extra contigs NODE 447528	1254	UQ01TA7B1180148	22	126
IWB8625	BS00053287 51	AG	AA Multiallelic	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7BL v2 extra contigs NODE 447528	1260	UQ01TA7B1180149	20	127
IWB51650	Ra c27077 468	CT	Mono	Mono	InfiniumII	no hit	Transition	.	7BL v2 extra contigs NODE 451828	3242	UQ01TA7B1184008	10	79
IWB2166	BobWhite c28058 232	AG	Recode AB to BB	detects poly single locus	InfiniumII	no hit	Transition	.	7BL v2 extra contigs NODE 456113	426	UQ01TA7B1188012	32	186
IWB72960	Tdurum contig63792 549	CT	MSV	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7BL v2 extra contigs NODE 469358	1758	UQ01TA7B1200074	2	69
IWB7378	BS00023023 51	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7BL v2 extra contigs NODE 470299	120	UQ01TA7B1200985	24	85
IWB12171	BS00101364 51	CT	BB Multiallelic	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7BL v2 extra contigs NODE 470299	89	UQ01TA7B1200984	36	74
IWB56854	RAC875 c3361 180	AG	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7BL v2 extra contigs NODE 519449	631	UQ01TA7B1219613	19	124
IWB56856	RAC875 c3361 403	AG	Codominant	detects poly single locus	InfiniumII	D->D	Transition	synonymous	7BL v2 extra contigs NODE 519449	854	UQ01TA7B1219616	12	79
IWA2767	wsnp Ex c22955 32173776	GT	Codominant	detects poly single locus	InfiniumII	no hit	Transversion	.	7BL v2 extra contigs NODE 525759	1331	UQ01TA7B1221192	87	186
IWB13946	CAP7 c2649 283	GT	MSV	detects poly duplicated loci	InfiniumII	no hit	Transversion	.	7BL v2 extra contigs NODE 52792	778	UQ01TA7B916802	44	94
IWB59247	RAC875 c570 302	CT	Codominant	detects poly single locus	InfiniumII	I->T	Transition	nonsynonymous	7BL v2 extra contigs NODE 645131	929	UQ01TA7B1281960	17	83
IWB2670	BobWhite c34068 854	CT	MSV	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7BL v2 extra contigs NODE 660590	3643	UQ01TA7B1293402	3	83
IWB73339	Tdurum contig76013 352	AC	Codominant	detects poly single locus	InfiniumII	no hit	Transversion	.	7BL v2 extra contigs NODE 674357	1534	UQ01TA7B1304213	33	176
IWB73340	Tdurum contig76013 605	GT	Codominant	detects poly single locus	InfiniumII	no hit	Transversion	.	7BL v2 extra contigs NODE 674357	2564	UQ01TA7B1304220	2	42
IWB71733	Tdurum contig46922 814	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7BL v2 extra contigs NODE 705211	172	UQ01TA7B1330707	9	64
IWB74097	tolb0021f14 1700	AG	MSV	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7BL v2 extra contigs NODE 733961	1490	UQ01TA7B1356842	9	46
IWB33401	GENE-3452 1116	AG	AA Multiallelic	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7BL v2 extra contigs NODE 744329	3796	UQ01TA7B1366828	35	135
IWB55234	RAC875 c21760 200	AG	MSV	detects poly duplicated loci	InfiniumII	N->N	Transition	synonymous	7BL v2 extra contigs NODE 755119	469	UQ01TA7B1377748	2	9
IWB10769	BS00074919 51	AG	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7BL v2 extra contigs NODE 759072	2018	UQ01TA7B1381700	30	77
IWB69818	Tdurum contig29522 232	AG	Mono	Mono	InfiniumII	no hit	Transition	.	7BL v2 extra contigs NODE 759264	1605	UQ01TA7B1381863	16	69
IWB4770	BobWhite rep c49390 455	CT	AA Multiallelic	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7BL v2 extra contigs NODE 761140	1295	UQ01TA7B1383542	15	37
IWB9596	BS00065624 51	AG	BB Multiallelic	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7BL v2 extra contigs NODE 762063	8605	UQ01TA7B1384391	10	50
IWB67554	Tdurum contig12326 232	AG	AA Multiallelic	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7BL v2 extra contigs NODE 764828	6361	UQ01TA7B1386824	47	126
IWB28513	Excalibur c6871 217	CT	MSV	detects poly duplicated loci	InfiniumII	N->N	Transition	synonymous	7BL v2 extra contigs NODE 773587	4401	UQ01TA7B1396062	3	25
IWB40924	Kukri c12901 706	AG	MSV	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7BL v2 extra contigs NODE 773587	6755	UQ01TA7B1396071	8	50
IWA1737	wsnp Fx c1318 2520706	AC	Codominant	detects poly single locus	InfiniumII	no hit	Transversion	.	7BL v2 extra contigs NODE 774393	225	UQ01TA7B1396963	2	14
IWB10527	BS00071100 51	AG	AA Multiallelic	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7BL v2 extra contigs NODE 777322	5795	UQ01TA7B1400235	7	162
IWB27012	Excalibur c48976 396	AG	Recode AB to BB	detects poly single locus	InfiniumII	no hit	Transition	.	7BL v2 extra contigs NODE 779550	801	UQ01TA7B1402226	38	147
IWB21716	Excalibur c1070 1978	AG	MSV	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7BL v2 extra contigs NODE 788570	8855	UQ01TA7B1410727	31	71
IWB21717	Excalibur c1070 2327	CT	BB Multiallelic	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7BL v2 extra contigs NODE 788570	9262	UQ01TA7B1410731	8	37
IWB46640	Kukri c57138 106	AG	MSV	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7BL v2 extra contigs NODE 796122	4266	UQ01TA7B1419205	21	61
IWB13220	CAP12 c1587 70	AC	MSV	detects poly duplicated loci	InfiniumII	no hit	Transversion	.	7BL v2 extra contigs NODE 806448	548	UQ01TA7B1429479	2	43
IWB13219	CAP12 c1587 142	AC	MSV	detects poly duplicated loci	InfiniumII	no hit	Transversion	.	7BL v2 extra contigs NODE 806448	620	UQ01TA7B1429480	2	51
IWB70977	Tdurum contig42194 1242	AG	Mono	Mono	InfiniumII	no hit	Transition	.	7BL v2 extra contigs NODE 817385	1510	UQ01TA7B1440316	11	48
IWB36376	IACX9353	CG	AA Multiallelic	detects poly duplicated loci	InfiniumII	S->W	Transversion	nonsynonymous	7BL v2 extra contigs NODE 830803	146	UQ01TA7B1452726	3	7
IWB72241	Tdurum contig52079 1174	AG	Codominant	detects poly single locus	InfiniumII	R->R	Transition	synonymous	7BL v2 extra contigs NODE 841629	3090	UQ01TA7B1463567	21	121
IWA832	wsnp CAP11 rep c4027 1902057	AG	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7BL v2 extra contigs NODE 841731	350	UQ01TA7B1463622	63	148
IWA4249	wsnp Ex c54863 57586730	GT	Recode AB to AA	detects poly single locus	InfiniumII	S->S	Transversion	synonymous	7BL v2 extra contigs NODE 846330	861	UQ01TA7B1468240	3	6
IWB56649	RAC875 c31791 559	AG	MSV	detects poly duplicated loci	InfiniumII	I->I	Transition	synonymous	7BL v2 extra contigs NODE 849596	4183	UQ01TA7B1470681	16	115
IWB56835	RAC875 c33407 350	CT	Mono	Mono	InfiniumII	no hit	Transition	.	7BL v2 extra contigs NODE 852676	424	UQ01TA7B1473631	8	60
IWB11413	BS00085556 51	GT	MSV	detects poly duplicated loci	InfiniumII	no hit	Transversion	.	7BL v2 extra contigs NODE 860357	2427	UQ01TA7B1478933	9	78
IWB31237	Excalibur rep c74778 252	CT	MSV	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7BL v2 extra contigs NODE 860357	2451	UQ01TA7B1478934	8	83

IWB56260	RAC875 c29004 652	AG	AA Multiallelic	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7BS v2 extra contigs NODE 863447	3209	UQ01TA7B1481263	42	113
IWB42040	Kukri c18871 1338	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7BS v2 extra contigs NODE 897013	1504	UQ01TA7B1487268	21	144
IWB26145	Excalibur c40516 119	GT	MSV	detects poly duplicated loci	InfiniumII	no hit	Transversion	.	7BS v2 extra contigs NODE 948809	4203	UQ01TA7B1503004	7	63
IWB36761	Jagger c7930 97	CT	Mono	Mono	InfiniumII	N->N	Transition	synonymous	7BS v2 extra contigs NODE 1011288	543	UQ01TA7B818214	6	14
IWB442	BobWhite c13013 81	CT	Mono	Mono	InfiniumII	N->N	Transition	synonymous	7BS v2 extra contigs NODE 1011288	543	UQ01TA7B818214	6	14
IWB4714	BobWhite rep c48912 53	CT	Mono	Mono	InfiniumII	N->N	Transition	synonymous	7BS v2 extra contigs NODE 1011288	543	UQ01TA7B818214	6	14
IWA6742	w.snp Ku c23549 33473349	AC	MSV	detects poly duplicated loci	InfiniumII	no hit	Transversion	.	7BS v2 extra contigs NODE 126437	3160	UQ01TA7B88763	27	150
IWB71523	Tdurum contig44382 932	GT	Codominant	detects poly single locus	InfiniumII	no hit	Transversion	.	7BS v2 extra contigs NODE 126437	3808	UQ01TA7B88765	22	168
IWB70551	Tdurum contig35652 348	AG	MSV	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7BS v2 extra contigs NODE 134655	929	UQ01TA7B93066	85	187
IWA1642	w.snp Ex c12535 19963035	AG	BB Multiallelic	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7BS v2 extra contigs NODE 135709	5845	UQ01TA7B93755	14	83
IWB21280	Ex c7356 639	AG	Mono	Mono	InfiniumII	no hit	Transition	.	7BS v2 extra contigs NODE 176686	2302	UQ01TA7B123055	35	108
IWA7033	w.snp Ku c4067 7419106	CT	Codominant	detects poly single locus	InfiniumII	K->R	Transition	nonsynonymous	7BS v2 extra contigs NODE 19119	3527	UQ01TA7B42657	20	129
IWB71851	Tdurum contig47854 142	CT	Codominant Null	detects poly single locus	InfiniumII	no hit	Transition	.	7BS v2 extra contigs NODE 198321	1327	UQ01TA7B140641	8	104
IWB13169	CAP12 c1039 114	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7BS v2 extra contigs NODE 200009	1091	UQ01TA7B142209	21	115
IWB13775	CAP7 c122 102	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7BS v2 extra contigs NODE 216371	4329	UQ01TA7B157652	37	109
IWB69576	Tdurum contig28644 281	AG	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7BS v2 extra contigs NODE 216371	4432	UQ01TA7B157654	2	41
IWB60430	RAC875 c7671 148	AG	AA Multiallelic	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7BS v2 extra contigs NODE 228338	2625	UQ01TA7B170091	10	82
IWB47762	Kukri c78330 327	AG	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7BS v2 extra contigs NODE 228338	3258	UQ01TA7B170093	16	115
IWB8021	BS00035640 51	CT	AA Multiallelic	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7BS v2 extra contigs NODE 228338	5053	UQ01TA7B170094	23	131
IWB63248	RAC875 rep c86713 495	AG	Recode AB to AA	detects poly single locus	InfiniumII	no hit	Transition	.	7BS v2 extra contigs NODE 237531	171	UQ01TA7B180457	5	78
IWB13264	CAP12 c1975 313	CT	Mono	Mono	InfiniumII	no hit	Transition	.	7BS v2 extra contigs NODE 254696	6474	UQ01TA7B200943	23	124
IWB65103	RFL Contig5734 2005	GT	Codominant	detects poly single locus	InfiniumII	no hit	Transversion	.	7BS v2 extra contigs NODE 269425	2211	UQ01TA7B219930	64	190
IWB13418	CAP12 c4844 191	AG	Mono	Mono	InfiniumII	no hit	Transition	.	7BS v2 extra contigs NODE 282974	238	UQ01TA7B239696	4	59
IWB10628	BS00072941 51	AG	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7BS v2 extra contigs NODE 296233	14333	UQ01TA7B258258	13	97
IWB34962	IAAV503	CT	AA Multiallelic	detects poly duplicated loci	InfiniumII	K->E	Transition	nonsynonymous	7BS v2 extra contigs NODE 304442	12606	UQ01TA7B270610	51	131
IWA814	w.snp CAP11 c847 522893	AG	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7BS v2 extra contigs NODE 307323	2116	UQ01TA7B274910	16	105
IWB70094	Tdurum contig30714 123	AG	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7BS v2 extra contigs NODE 31379	1346	UQ01TA7B46302	9	122
IWB26632	Excalibur c45099 1048	AG	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7BS v2 extra contigs NODE 323866	3359	UQ01TA7B298802	37	120
IWB71430	Tdurum contig43523 359	CT	MSV	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7BS v2 extra contigs NODE 410118	3422	UQ01TA7B356761	10	105
IWB48553	Kukri c99538 516	AC	Recode AB to AA	detects poly single locus	InfiniumII	no hit	Transversion	.	7BS v2 extra contigs NODE 503773	2945	UQ01TA7B428299	38	120
IWB8597	BS00050993 51	AC	AA Multiallelic	detects poly duplicated loci	InfiniumII	M->L	Transversion	nonsynonymous	7BS v2 extra contigs NODE 509122	8156	UQ01TA7B433856	14	68
IWA3915	w.snp Ex c44814 50770533	AG	Codominant	detects poly single locus	InfiniumII	N->D	Transition	nonsynonymous	7BS v2 extra contigs NODE 524285	3387	UQ01TA7B451711	10	87
IWB41364	Kukri c15333 672	CT	MSV	detects poly duplicated loci	InfiniumII	Y->Y	Transition	synonymous	7BS v2 extra contigs NODE 524285	3449	UQ01TA7B451712	55	114
IWB474	BobWhite c13269 917	AG	Mono	Mono	InfiniumII	W->R	Transition	nonsynonymous	7BS v2 extra contigs NODE 527502	20597	UQ01TA7B456043	2	138
IWA4873	w.snp Ex c908 1754208	CT	Codominant	detects poly single locus	InfiniumII	S->S	Transition	synonymous	7BS v2 extra contigs NODE 570869	2310	UQ01TA7B514073	31	113
IWB9254	BS00064343 51	GT	MSV	detects poly duplicated loci	InfiniumII	C->G	Transversion	nonsynonymous	7BS v2 extra contigs NODE 589289	4187	UQ01TA7B540985	16	150
IWB73292	Tdurum contig75644 871	AG	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7BS v2 extra contigs NODE 605216	2186	UQ01TA7B565460	9	148
IWB8672	BS00055761 51	CT	Codominant	detects poly single locus	InfiniumII	D->D	Transition	synonymous	7BS v2 extra contigs NODE 712884	1329	UQ01TA7B623325	5	47
IWA2534	w.snp Ex c204 400545	CT	AA Multiallelic	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7BS v2 extra contigs NODE 718878	2318	UQ01TA7B627377	35	100
IWB20228	Ex c24068 652	GT	Codominant	detects poly single locus	InfiniumII	->E	Transversion	nonsynonymous	7BS v2 extra contigs NODE 726732	583	UQ01TA7B633015	19	58
IWB6919	BS00022127 51	AG	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7BS v2 extra contigs NODE 726732	781	UQ01TA7B633018	39	119
IWB72566	Tdurum contig57220 173	GT	MSV	detects poly duplicated loci	InfiniumII	no hit	Transversion	.	7BS v2 extra contigs NODE 744885	4630	UQ01TA7B647251	22	125
IWB65212	RFL Contig6075 618	CT	MSV	detects poly duplicated loci	InfiniumII	R->R	Transition	synonymous	7BS v2 extra contigs NODE 754814	1758	UQ01TA7B656434	14	111
IWB3050	BobWhite c39364 231	AG	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7BS v2 extra contigs NODE 781212	3853	UQ01TA7B682352	7	19
IWB27459	Excalibur c54242 704	AG	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7BS v2 extra contigs NODE 781212	4557	UQ01TA7B682355	38	92
IWB4698	BobWhite rep c48793 750	CT	BB Multiallelic	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7BS v2 extra contigs NODE 783694	119	UQ01TA7B684796	44	162
IWA1543	w.snp Ex c11860 19030807	CT	Codominant	detects poly single locus	InfiniumII	L->L	Transition	synonymous	7BS v2 extra contigs NODE 809462	4836	UQ01TA7B715484	37	131

IWA2353	w.snp_Ex_c18800_27681277	AC	Codominant	detects poly single locus	InfiniumII	no hit	Transversion	.	7BS v2 extra contigs NODE 810284	1228	UQ01TA7B716128	13	137
IWB71357	Tdurum contig42813_285	AG	Recode AB to BB	detects poly single locus	InfiniumII	no hit	Transition	.	7BS v2 extra contigs NODE 84777	3775	UQ01TA7B67337	13	144
IWB71358	Tdurum contig42813_397	AC	MSV	detects poly duplicated loci	InfiniumII	no hit	Transversion	.	7BS v2 extra contigs NODE 84777	3887	UQ01TA7B67338	27	177
IWB25263	Excalibur_c32859_652	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7BS v2 extra contigs NODE 886285	6320	UQ01TA7B752203	6	25
IWB13589	CAP12 rep_c4079_97	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7BS v2 extra contigs NODE 905888	4101	UQ01TA7B757075	38	140
IWA8525	w.snp_RFL Contig3854_4205716	AG	BB Multiallelic	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7BS v2 extra contigs NODE 957498	17982	UQ01TA7B781948	32	136
IWB52695	Ra_c7974_1192	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7BS v2 extra contigs NODE 957498	19247	UQ01TA7B781953	21	56
IWA7881	w.snp_Ra_c3450_6434387	GT	MSV	detects poly duplicated loci	InfiniumII	no hit	Transversion	.	7BS v2 extra contigs NODE 985712	1940	UQ01TA7B800855	53	140
IWA208	w.snp_Ex_c17346_26030825	AG	Codominant	detects poly single locus	InfiniumII	V->A	Transition	nonsynonymous	7D SvnBuild v2.0	12070678	UQ01TA7D09742	16	52
IWB12476	BS000101024_51	AG	MSV	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7D SvnBuild v2.0	1526916	UQ01TA7D01663	6	95
IWB22539	Excalibur_c15029_144	AG	BB Multiallelic	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7D SvnBuild v2.0	15897042	UQ01TA7D12861	2	153
IWA1537	w.snp_Ex_c11813_18968198	GT	Codominant	detects poly single locus	InfiniumII	D->A	Transversion	nonsynonymous	7D SvnBuild v2.0	17019413	UQ01TA7D13736	38	96
IWB60753	RAC875_c83928_222	AG	Recode AB to BB	detects poly single locus	InfiniumII	S->S	Transition	synonymous	7D SvnBuild v2.0	18997842	UQ01TA7D15577	3	123
IWB27037	Excalibur_c49272_174	AG	MSV	detects poly duplicated loci	InfiniumII	T->T	Transition	synonymous	7D SvnBuild v2.0	22240411	UQ01TA7D18422	4	43
IWB39179	Ku_c32426_324	GT	MSV	detects poly duplicated loci	InfiniumII	no hit	Transversion	.	7D SvnBuild v2.0	22564748	UQ01TA7D18675	68	147
IWB6836	BS00021979_51	CT	MSV	detects poly duplicated loci	InfiniumII	P->P	Transition	synonymous	7D SvnBuild v2.0	23619840	UQ01TA7D19655	3	61
IWA2522	w.snp_Ex_c20320_29383710	AC	Codominant	detects poly single locus	InfiniumII	no hit	Transversion	.	7D SvnBuild v2.0	3341506	UQ01TA7D03276	12	39
IWA2523	w.snp_Ex_c20320_29383733	AG	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7D SvnBuild v2.0	3341529	UQ01TA7D03277	15	57
IWA2524	w.snp_Ex_c20320_29384395	CT	Recode AB to AA	detects poly single locus	InfiniumII	S->G	Transition	nonsynonymous	7D SvnBuild v2.0	3342569	UQ01TA7D03279	15	117
IWA7610	w.snp_Ra_c13568_21427471	AG	Codominant	detects poly single locus	InfiniumII	G->G	Transition	synonymous	7D SvnBuild v2.0	3343089	UQ01TA7D03282	16	41
IWB3933	BobWhite_c5654_231	AG	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7D SvnBuild v2.0	352532	UQ01TA7D00383	55	169
IWA4132	w.snp_Ex_c5231_9256869	AC	MSV	detects poly duplicated loci	InfiniumII	A->A	Transversion	synonymous	7D SvnBuild v2.0	379113	UQ01TA7D00459	2	8
IWB20865	Ex_c5231_1655	CT	MSV	detects poly duplicated loci	InfiniumII	S->S	Transition	synonymous	7D SvnBuild v2.0	380141	UQ01TA7D00462	43	104
IWB38316	Ku_c12427_1367	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7D SvnBuild v2.0	4055429	UQ01TA7D03786	5	113
IWB47761	Kukri_c78306_115	CT	Mono	Mono	InfiniumII	P->P	Transition	synonymous	7D SvnBuild v2.0	4126922	UQ01TA7D03842	2	25
IWB14320	CAP7_c9278_185	AG	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7D SvnBuild v2.0	4207769	UQ01TA7D03917	6	134
IWB17630	D_F5XZDLF02H192C_184	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7D SvnBuild v2.0	474809	UQ01TA7D00580	12	81
IWB63607	RFL Contig1323_544	AG	Codominant	detects poly single locus	InfiniumII	T->A	Transition	nonsynonymous	7D SvnBuild v2.0	5789447	UQ01TA7D05012	84	169
IWB52752	Ra_c8680_450	AG	AA Multiallelic	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7D SvnBuild v2.0	6574735	UQ01TA7D05624	30	123
IWB35409	IAAV7925	GT	Codominant	detects poly single locus	InfiniumII	no hit	Transversion	.	7D SvnBuild v2.0	8542249	UQ01TA7D06873	10	125
IWB58272	RAC875_c46691_110	AG	Mono	Mono	InfiniumII	H->R	Transition	nonsynonymous	7D SvnBuild v2.0	8658247	UQ01TA7D07036	6	35
IWB54329	RAC875_c16641_411	CT	MSV	detects poly duplicated loci	InfiniumII	V->A	Transition	nonsynonymous	7D SvnBuild v2.0	8986275	UQ01TA7D07351	4	43
IWA2772	w.snp_Ex_c23001_32223579	AG	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7D SvnBuild v2.0	9095497	UQ01TA7D07422	6	81
IWB6288	BS00010794_51	AG	MSV	detects poly duplicated loci	InfiniumII	T->T	Transition	synonymous	7DL v2 extra contigs NODE 1085479	8564	UQ01TA7D575055	20	79
IWB12410	BS00109445_51	AG	AA Multiallelic	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7DL v2 extra contigs NODE 1115857	2350	UQ01TA7D581793	5	139
IWB21665	Excalibur_c10528_347	CT	Recode AB to BB	detects poly single locus	InfiniumII	no hit	Transition	.	7DL v2 extra contigs NODE 1355457	2587	UQ01TA7D622673	3	165
IWB3654	BobWhite_c48395_143	AC	Codominant Null	detects poly single locus	InfiniumII	no hit	Transversion	.	7DL v2 extra contigs NODE 1466591	9726	UQ01TA7D634593	22	62
IWB38312	Ku_c1234_733	AG	Codominant	detects poly single locus	InfiniumII	M->V	Transition	nonsynonymous	7DL v2 extra contigs NODE 151886	9393	UQ01TA7D381201	20	121
IWB39518	Ku_c48337_1131	AG	Codominant	detects poly single locus	InfiniumII	I->I	Transition	synonymous	7DL v2 extra contigs NODE 679425	4428	UQ01TA7D499752	22	137
IWB16155	D contig23893_348	CT	Mono	Mono	InfiniumII	->Q	Transition	nonsynonymous	7DL v2 extra contigs NODE 699588	4557	UQ01TA7D505887	2	169
IWB7447	BS00023150_51	CT	MSV	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7DL v2 extra contigs NODE 866122	3184	UQ01TA7D531821	8	130
IWB44562	Kukri_c36345_281	CT	Mono	Mono	InfiniumII	no hit	Transition	.	7DL v2 extra contigs NODE 928346	177	UQ01TA7D551155	5	12
IWB23164	Excalibur_c18558_376	CT	MSV	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7DS v2 extra contigs NODE 1265492	36	UQ01TA7D295832	4	11
IWB45523	Kukri_c45368_300	AG	MSV	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7DS v2 extra contigs NODE 156088	3360	UQ01TA7D53072	17	185
IWB24296	Excalibur_c25315_702	AC	MSV	detects poly duplicated loci	InfiniumII	no hit	Transversion	.	7DS v2 extra contigs NODE 156088	3456	UQ01TA7D53073	6	151
IWB15569	D contig10382_335	CT	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7DS v2 extra contigs NODE 1713077	148	UQ01TA7D340429	6	66
IWB19067	D_GDEEGVY02G3Q15_336	CT	MSV	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7DS v2 extra contigs NODE 2166035	35	UQ01TA7D357769	2	9

IWB15080	CAP8 rep c9420 186	AG	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7DS v2 extra contigs NODE 286040	1572	UQ01TA7D99178	5	26
IWB63737	RFL Contig1820 712	AG	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7DS v2 extra contigs NODE 345970	886	UQ01TA7D122768	3	18
IWB45999	Kukri c49522 313	CT	Recode AB to AA	detects poly single locus	InfiniumII	M->V	Transition	nonsynonymous	7DS v2 extra contigs NODE 387807	103	UQ01TA7D139144	6	48
IWA5972	wsnp JD c2734 3667052	AC	MSV	detects poly duplicated loci	InfiniumII	L->W	Transversion	nonsynonymous	7DS v2 extra contigs NODE 413505	1633	UQ01TA7D147858	59	155
IWB8938	BS00062860 51	AG	Mono	Mono	InfiniumII	no hit	Transition	.	7DS v2 extra contigs NODE 55643	42	UQ01TA7D28565	2	22
IWA6139	wsnp JD c6436 7600132	AG	Codominant	detects poly single locus	InfiniumII	no hit	Transition	.	7DS v2 extra contigs NODE 764765	52	UQ01TA7D223985	3	19
IWB61909	RAC875 rep c109720 265	GT	Mono	Mono	InfiniumII	I->I	Transversion	synonymous	7DS v2 extra contigs NODE 770388	5195	UQ01TA7D225655	5	99
IWB71933	Tdurum contig4885 588	AG	MSV	detects poly duplicated loci	InfiniumII	no hit	Transition	.	7DS v2 extra contigs NODE 973563	7191	UQ01TA7D254587	3	137