Developing genome resources to improve wheat abiotic stress tolerance

David Edwards
University of Queensland, Australia
Dave.Edwards@uq.edu.au

Outline

- Drought in Australia
- Wheat genome sequencing
- Wheat genome diversity

THE UNIVERSITY OF QUEENSLAND
AUSTRALIA

ACPFG
AUSTRALIAN CENTRE FOR PLANT FUNCTIONAL GENOMICS PTY LTD

University of Queensland
January 2012

Drought in Australia

‘Temperatures in Australia have risen by about 0.9 °C since 1910’

‘Climate model projections for the coming decades indicate an increasing risk of below average rainfall for southern and eastern mainland Australia, higher temperatures and evaporation’

(CSIRO)

The ACPFG aims to improve wheat and barley's tolerance to environmental stresses such as drought and salinity

Challenges

- Wheat is a ‘difficult’ genome
    - Large, hexaploid, 80-90% repeats

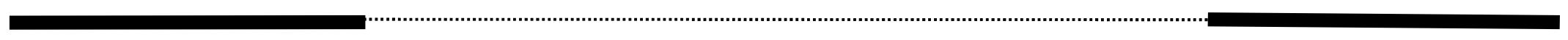
- Identify important genes/alleles
- Characterise genomic variation
- Develop tools to assist breeders improve crops

Second-generation sequencing (2GS)

- Illumina GAIIx and HiSeq2000
  - ↑↑↑ sequence
  - ↓ money
  - ↓ time

  - ↓ read-length
  - ↑ computation

Illumina paired reads



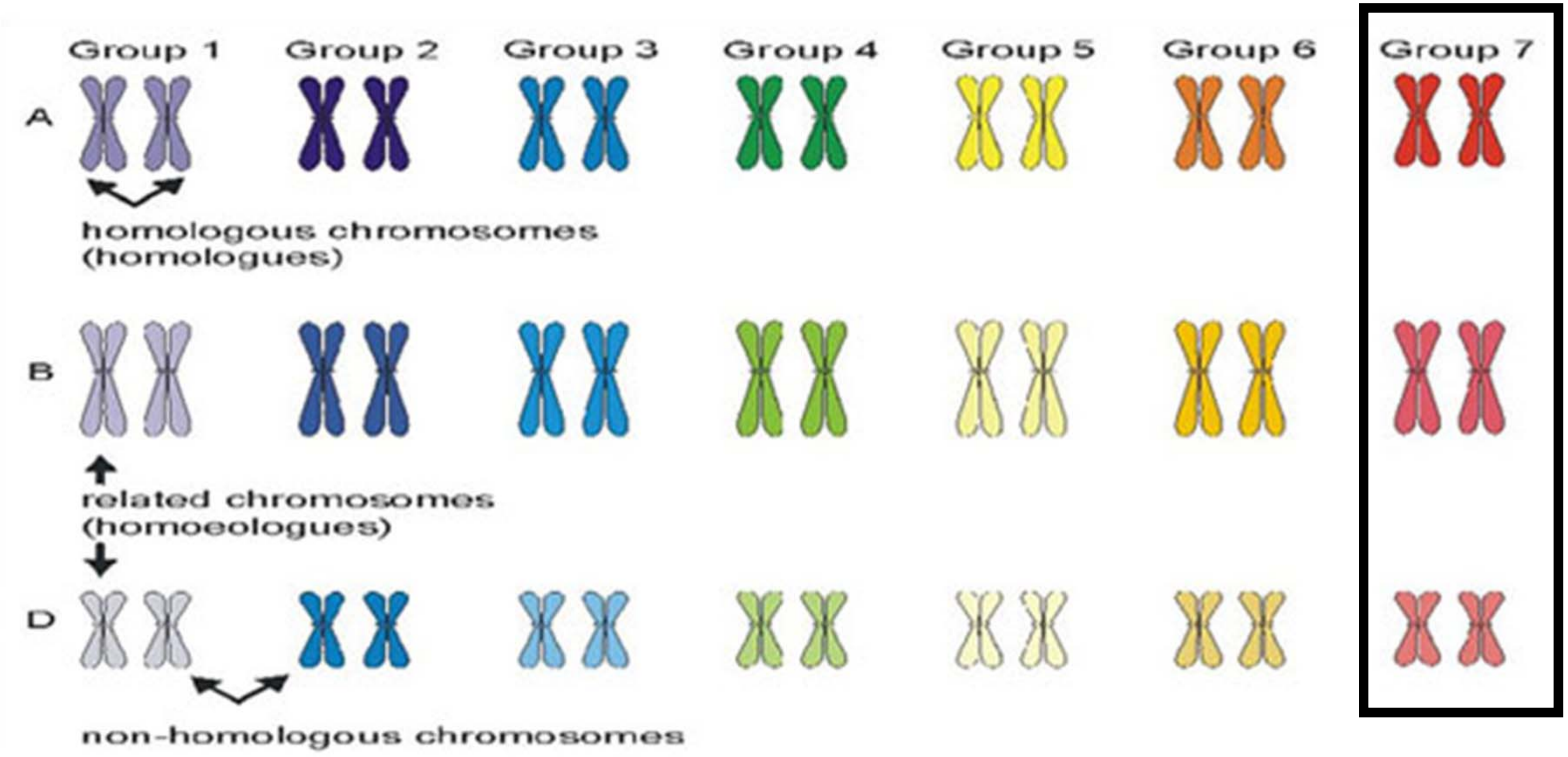
Insert size

- Illumina GAIIx/HiSeq 2000
- Read length (35 bp – 150 bp)
- Insert size up to 20 Kbp
  - ~ Normal distribution
  - Standard deviation ~ 10% mean

Problems of scale

	bp	distance	
Illumina read	100	1.88 m	me
Sanger read	900	18 m	telegraph pole
BAC	120,000	2.3 km	short walk
Canola genome	1,300,000,000	24,700 km	San Diego to Sydney (and back)
Wheat genome	17,000,000,000	323,000 km	around the world ~4 times

Hexaploid wheat genome



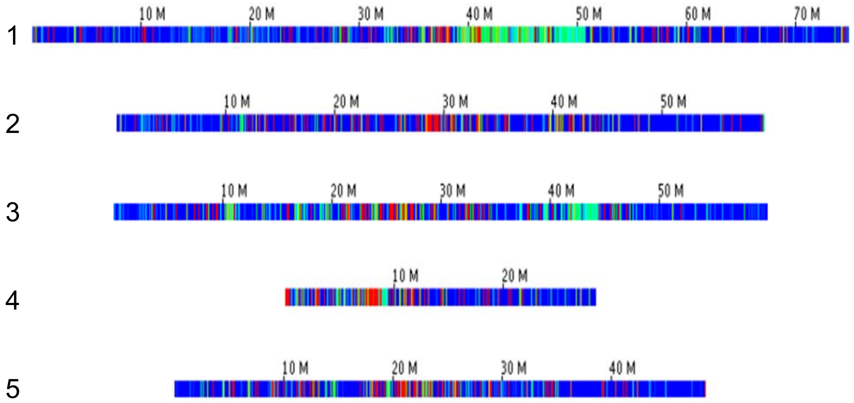
http://www.jic.ac.uk/staff/graham-moore/wheat_meiosis.htm

Chromosome arm sequencing

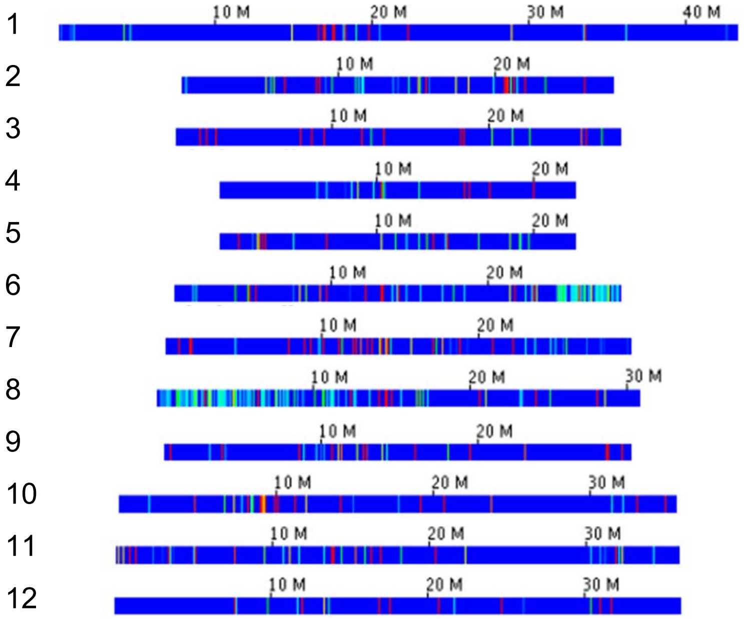
- DNA from Jaroslav Dolezel (Czech Rep.)
    - Flow sorted chromosomes

- Benefits
    - Better resolution of smaller “genome size”
    - Reduces repetitive sequence
    - Simpler assembly

Mapping 7DS reads to reference genomes

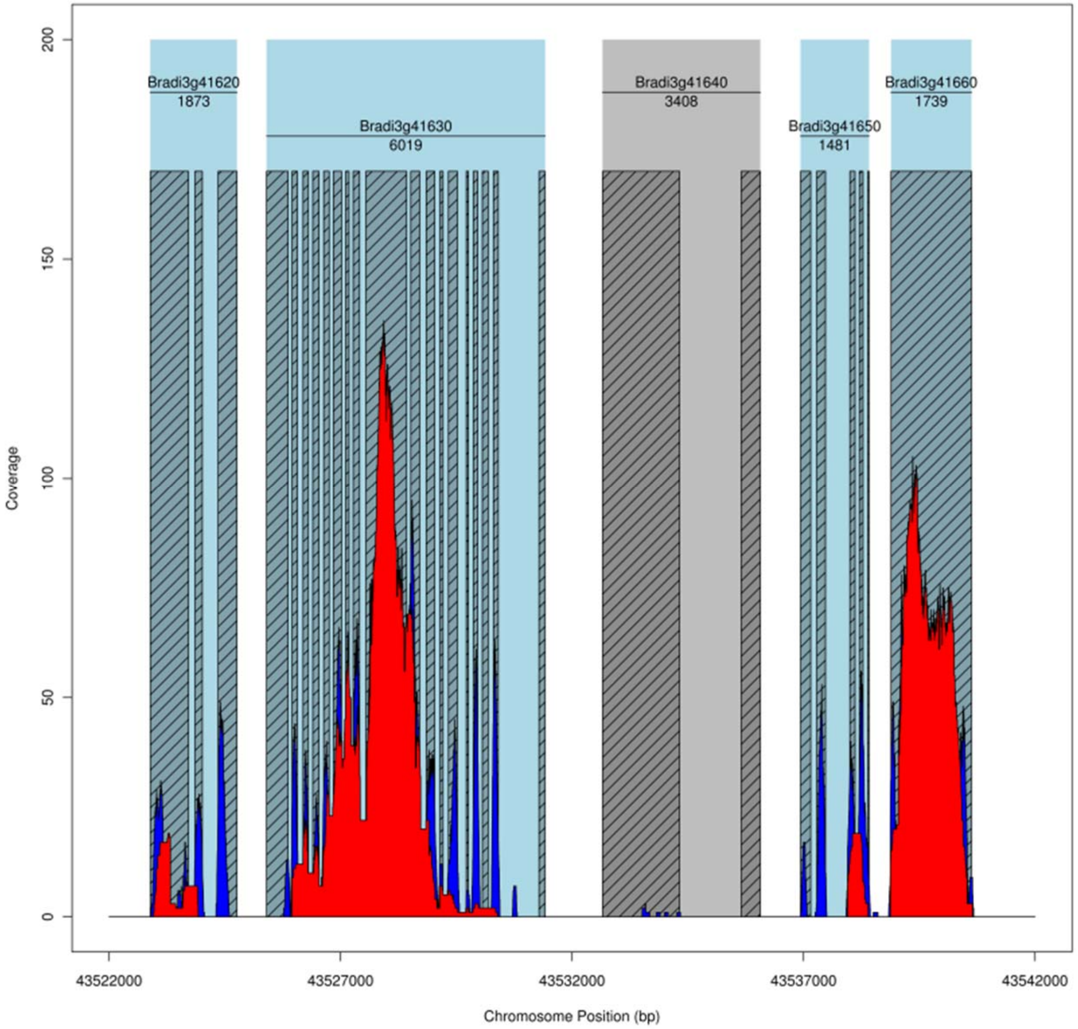


Brachypodium



rice

Mapping reads to reference genomes



7DS syntenic build

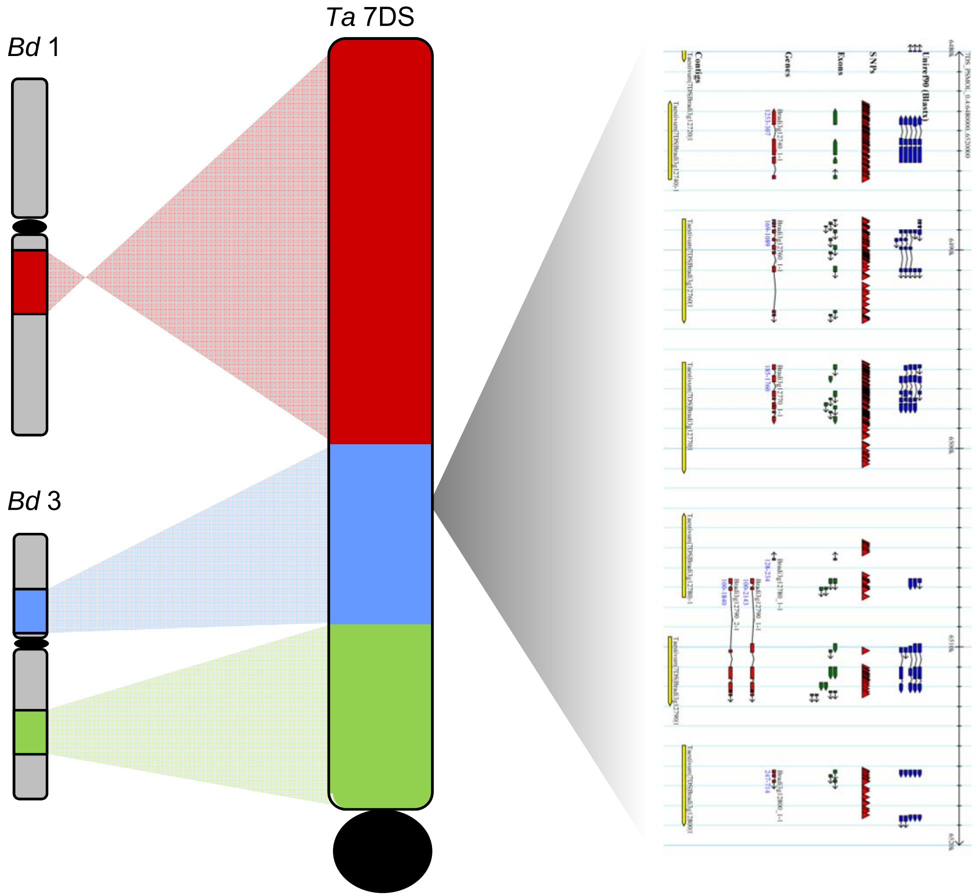
Plant Biotechnology Journal

Discover this journal online at WILEY ONLINE LIBRARY wileyonlinelibrary.com/journal/pbi

Volume 9 | Issue 7
September 2011

WILEY-BLACKWELL

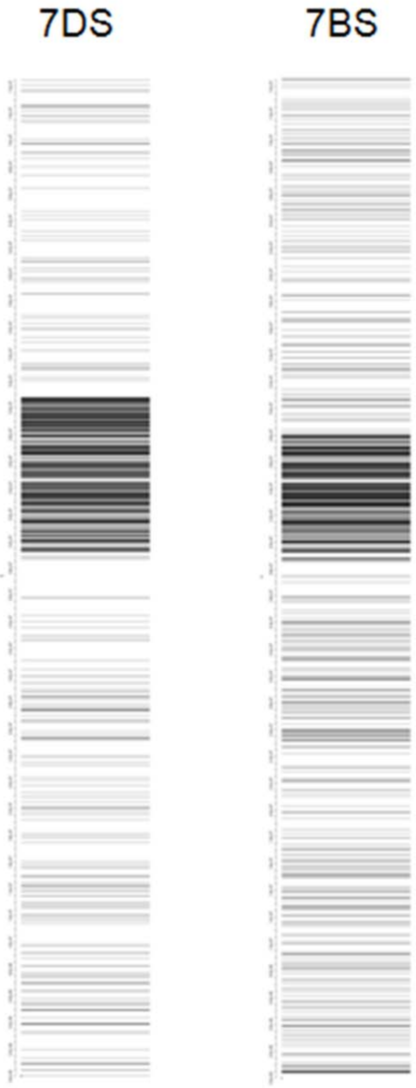
www.plantbiotechjournal.com | ISSN 1467-7644



www.wheatgenome.info

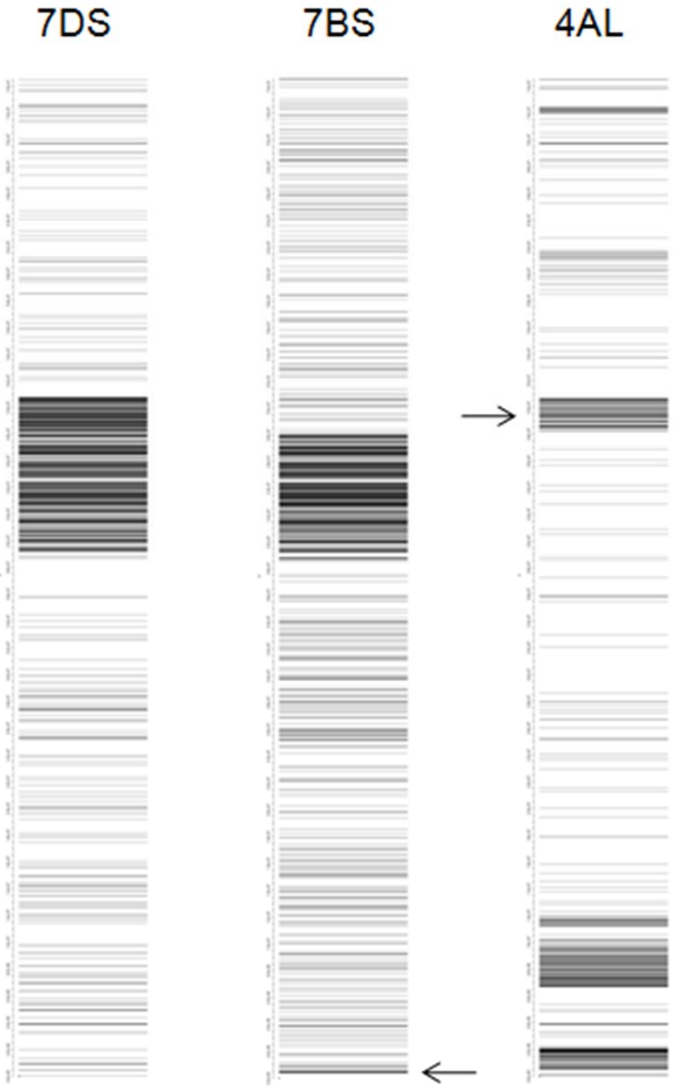
Berkman, *et al.*, Plant Biotechnology Journal (2011)

7BS/4AL translocation



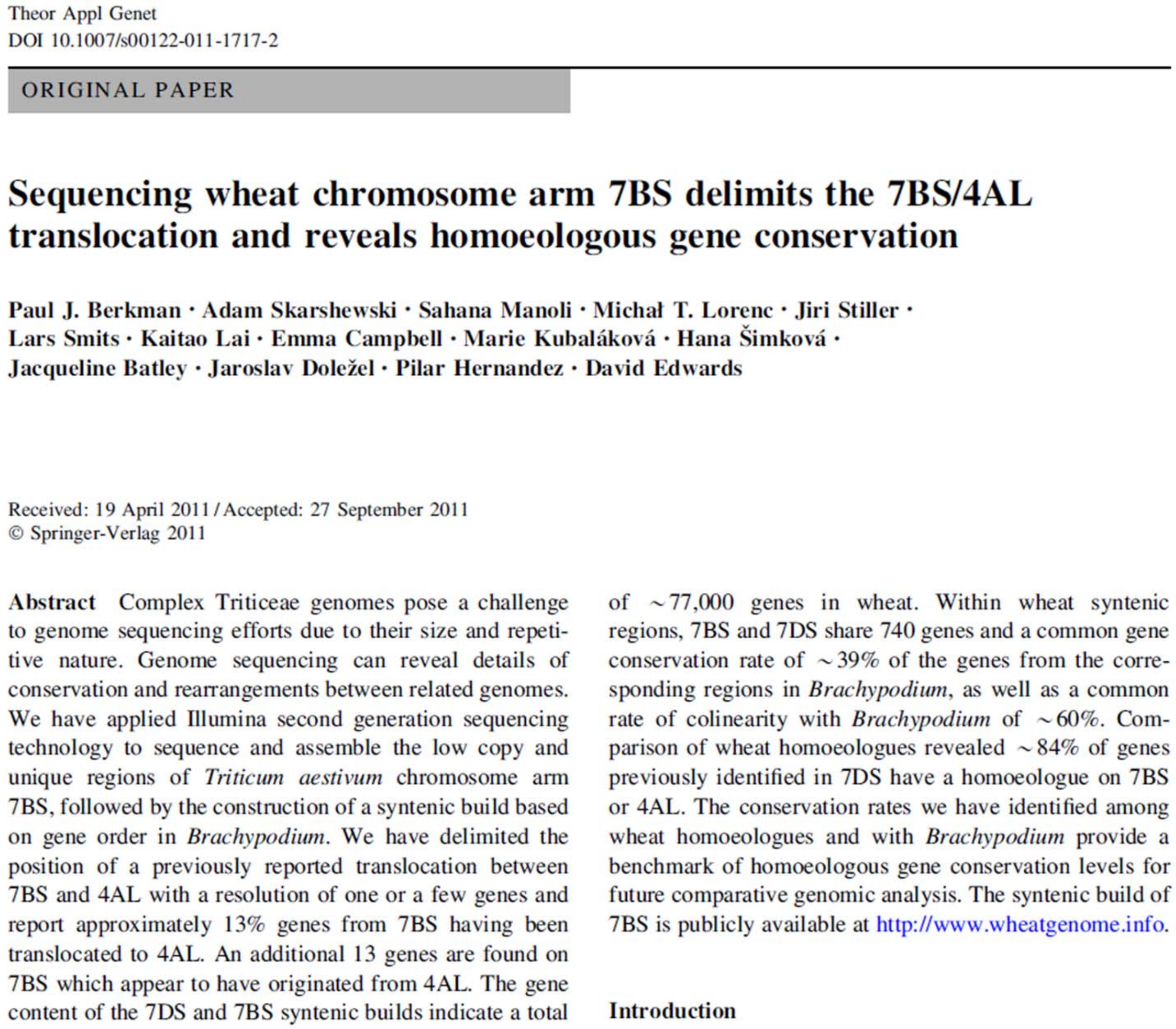
7DS and 7BL sequence similarity with Brachypodium

7BS/4AL translocation



- Translocation between Bradi1g49500 and Bradi1g49550
    - Intervening 4 genes missing from all assemblies
- ~13% genes moved from 7BS to 4AL
- 13 genes moved from 4AL to 7BS

7BS/4AL translocation



Theor Appl Genet
DOI 10.1007/s00122-011-1717-2

ORIGINAL PAPER

Sequencing wheat chromosome arm 7BS delimits the 7BS/4AL translocation and reveals homoeologous gene conservation

Paul J. Berkman · Adam Skarshewski · Sahana Manoli · Michał T. Lorenc · Jiri Stiller · Lars Smits · Kaitao Lai · Emma Campbell · Marie Kubaláková · Hana Šimková · Jacqueline Batley · Jaroslav Doležel · Pilar Hernandez · David Edwards

Received: 19 April 2011 / Accepted: 27 September 2011

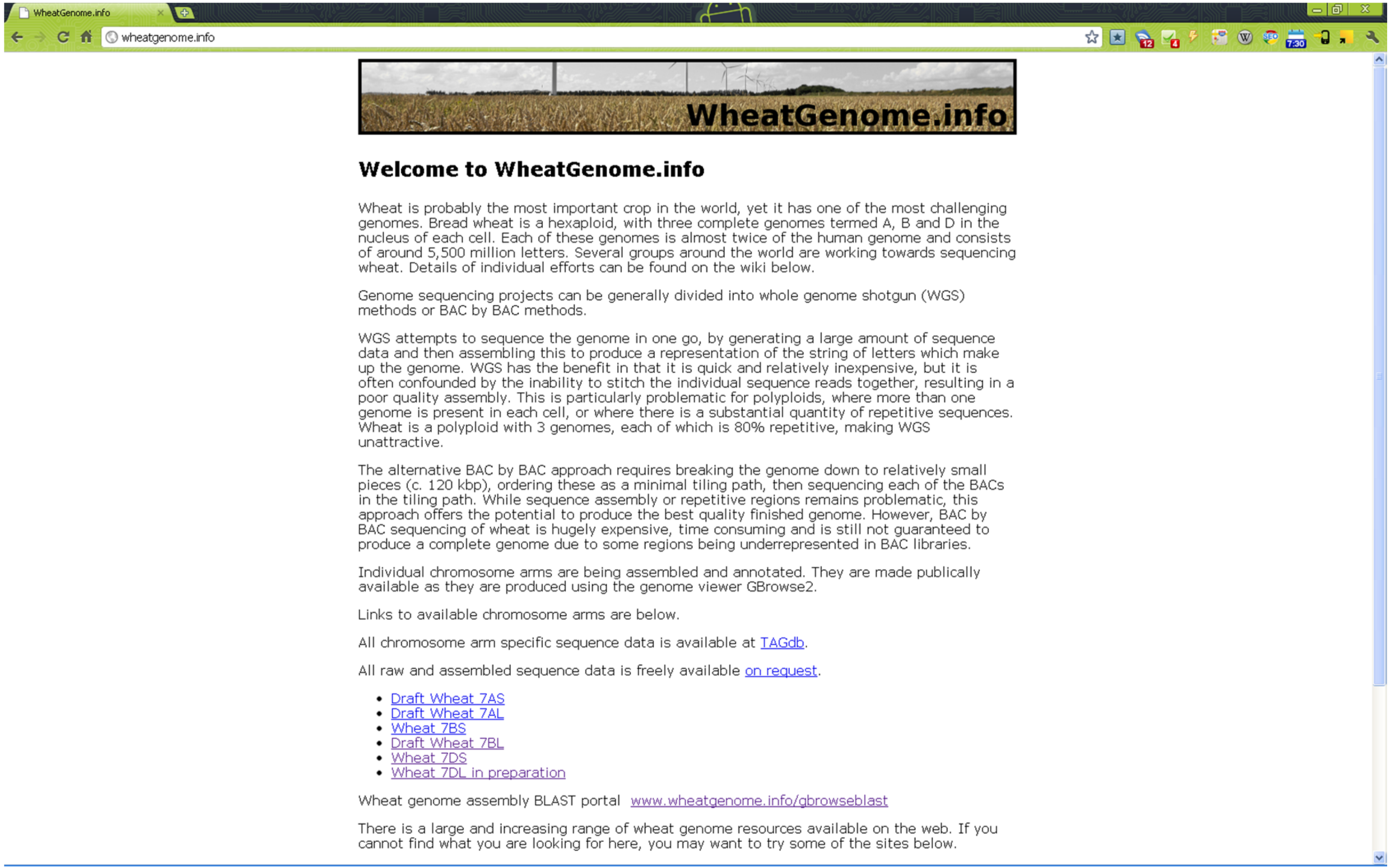
Abstract Complex Triticeae genomes pose a challenge to genome sequencing efforts due to their size and repetitive nature. Genome sequencing can reveal details of conservation and rearrangements between related genomes. We have applied Illumina second generation sequencing technology to sequence and assemble the low copy and unique regions of *Triticum aestivum* chromosome arm 7BS, followed by the construction of a syntenic build based on gene order in *Brachypodium*. We have delimited the position of a previously reported translocation between 7BS and 4AL with a resolution of one or a few genes and report approximately 13% genes from 7BS having been translocated to 4AL. An additional 13 genes are found on 7BS which appear to have originated from 4AL. The gene content of the 7DS and 7BS syntenic builds indicate a total of ~77,000 genes in wheat. Within wheat syntenic regions, 7BS and 7DS share 740 genes and a common gene conservation rate of ~39% of the genes from the corresponding regions in *Brachypodium*, as well as a common rate of colinearity with *Brachypodium* of ~60%. Comparison of wheat homoeologues revealed ~84% of genes previously identified in 7DS have a homoeologue on 7BS or 4AL. The conservation rates we have identified among wheat homoeologues and with *Brachypodium* provide a benchmark of homoeologous gene conservation levels for future comparative genomic analysis. The syntenic build of 7BS is publicly available at http://www.wheatgenome.info.

Introduction

The University Of Queensland
AUSTRALIA

ACPFG
Australian Centre for Plant Functional Genomics Pty Ltd

Group 7 chromosomes



Welcome to WheatGenome.info

Wheat is probably the most important crop in the world, yet it has one of the most challenging genomes. Bread wheat is a hexaploid, with three complete genomes termed A, B and D in the nucleus of each cell. Each of these genomes is almost twice of the human genome and consists of around 5,500 million letters. Several groups around the world are working towards sequencing wheat. Details of individual efforts can be found on the wiki below.

Genome sequencing projects can be generally divided into whole genome shotgun (WGS) methods or BAC by BAC methods.

WGS attempts to sequence the genome in one go, by generating a large amount of sequence data and then assembling this to produce a representation of the string of letters which make up the genome. WGS has the benefit in that it is quick and relatively inexpensive, but it is often confounded by the inability to stitch the individual sequence reads together, resulting in a poor quality assembly. This is particularly problematic for polyploids, where more than one genome is present in each cell, or where there is a substantial quantity of repetitive sequences. Wheat is a polyploid with 3 genomes, each of which is 80% repetitive, making WGS unattractive.

The alternative BAC by BAC approach requires breaking the genome down to relatively small pieces (c. 120 kbp), ordering these as a minimal tiling path, then sequencing each of the BACs in the tiling path. While sequence assembly or repetitive regions remains problematic, this approach offers the potential to produce the best quality finished genome. However, BAC by BAC sequencing of wheat is hugely expensive, time consuming and is still not guaranteed to produce a complete genome due to some regions being underrepresented in BAC libraries.

Individual chromosome arms are being assembled and annotated. They are made publically available as they are produced using the genome viewer GBrowse2.

Links to available chromosome arms are below.

All chromosome arm specific sequence data is available at TAGdb.

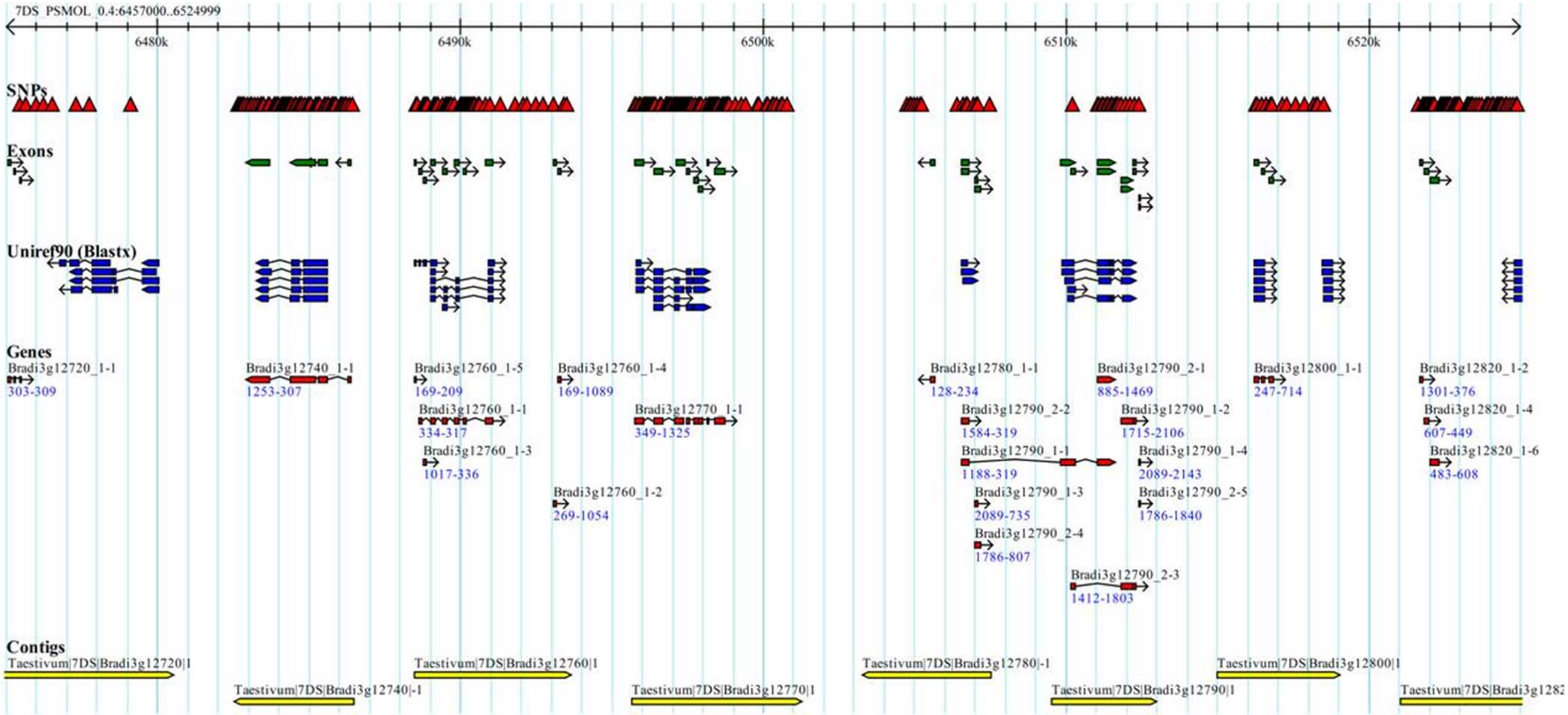
All raw and assembled sequence data is freely available on request.

- Draft Wheat 7AS
- Draft Wheat 7AL
- Wheat 7BS
- Draft Wheat 7BL
- Wheat 7DS
- Wheat 7DL in preparation

Wheat genome assembly BLAST portal www.wheatgenome.info/gbrowseblast

There is a large and increasing range of wheat genome resources available on the web. If you cannot find what you are looking for here, you may want to try some of the sites below.

GBrowse
http://wheatgenome.info/



Wheat summary

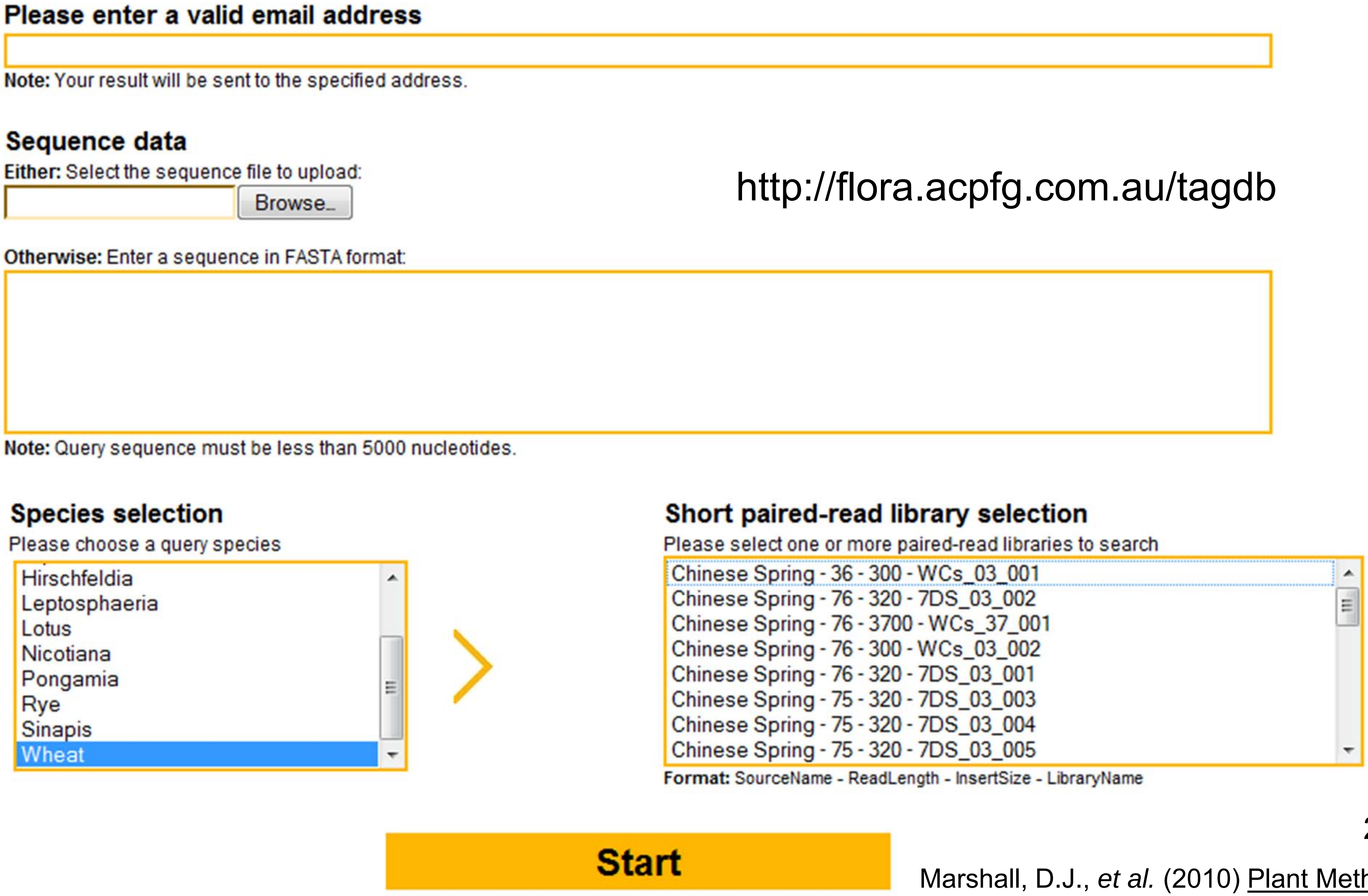
- Shotgun assembly of 7A, 7B and 7D
- Contains all or nearly all genes
- Model for identification of all wheat genes
- ~13% of 7BS genes translocated to 4AL
- Gene movement is consistent between arms
- We estimate ~77,000 genes in wheat

Australian wheat diversity

Researchers in Australia established a consortium to develop genomic resources in wheat, funded by Bioplatforms Australia.

- Generated data for wheat group 7 chromosome arms
- Generated 10x coverage whole genome sequence data for 16 wheat varieties, representing diversity in Australian bread wheat

TAGdb



Please enter a valid email address

Note: Your result will be sent to the specified address.

Sequence data

Either: Select the sequence file to upload:

Otherwise: Enter a sequence in FASTA format:

Note: Query sequence must be less than 5000 nucleotides.

Species selection

Please choose a query species

Hirschfeldia
Leptosphaeria
Lotus
Nicotiana
Pongamia
Rye
Sinapis
Wheat

Short paired-read library selection

Please select one or more paired-read libraries to search

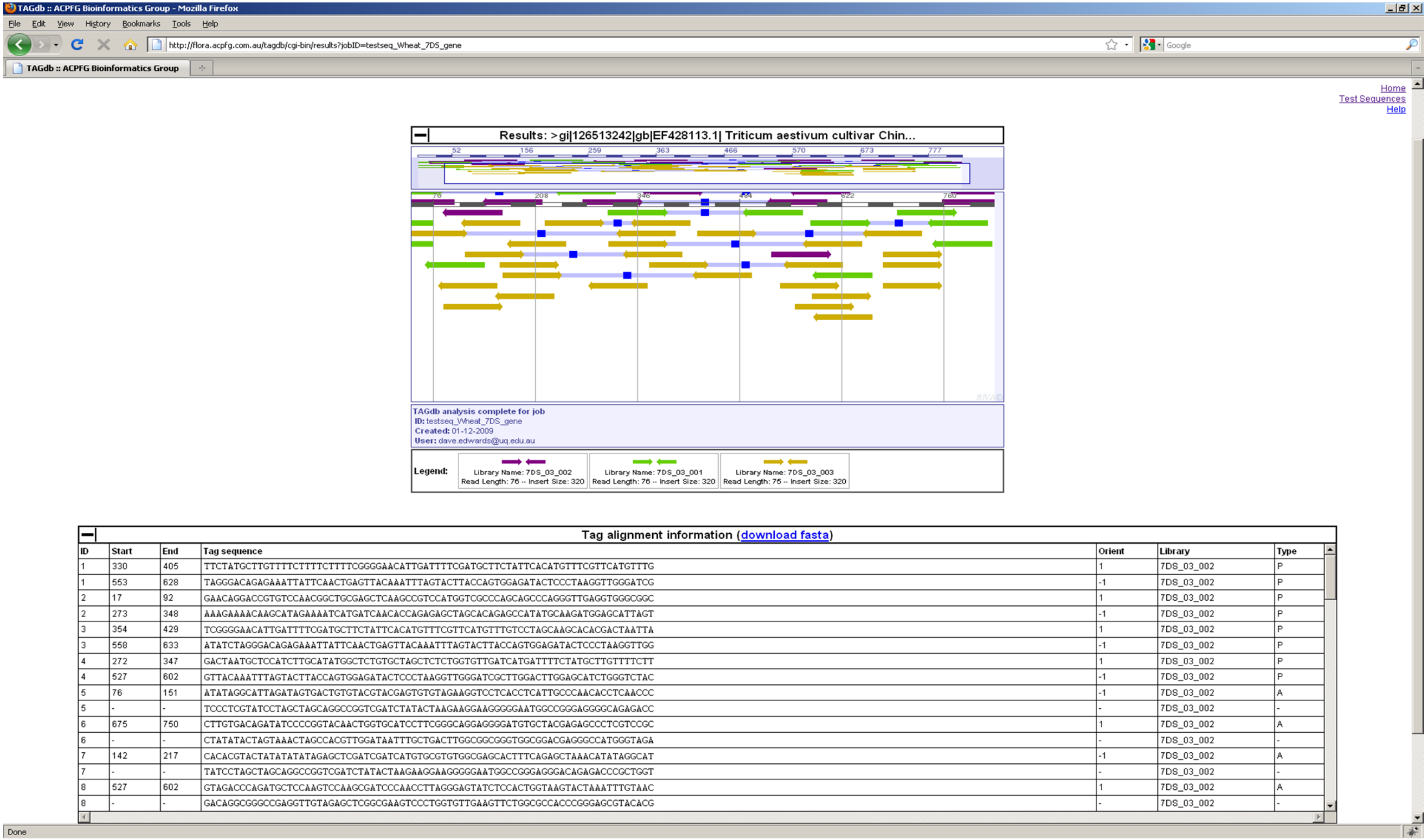
Chinese Spring - 36 - 300 - WCs_03_001
Chinese Spring - 76 - 320 - 7DS_03_002
Chinese Spring - 76 - 3700 - WCs_37_001
Chinese Spring - 76 - 300 - WCs_03_002
Chinese Spring - 76 - 320 - 7DS_03_001
Chinese Spring - 75 - 320 - 7DS_03_003
Chinese Spring - 75 - 320 - 7DS_03_004
Chinese Spring - 75 - 320 - 7DS_03_005

Format: SourceName - ReadLength - InsertSize - LibraryName

http://flora.acpfg.com.au/tagdb

Marshall, D.J., *et al.* (2010) Plant Methods. 6:19

TAGdb output



Results: >gi|126513242|gb|EF428113.1| Triticum aestivum cultivar Chin...

TAGdb analysis complete for job
ID: testseq_Wheat_7DS_gene
Created: 01-12-2009
User: dave.edwards@uq.edu.au

Legend:
Library Name: 7DS_03_002 Read Length: 76 -- Insert Size: 320
Library Name: 7DS_03_001 Read Length: 76 -- Insert Size: 320
Library Name: 7DS_03_003 Read Length: 75 -- Insert Size: 320

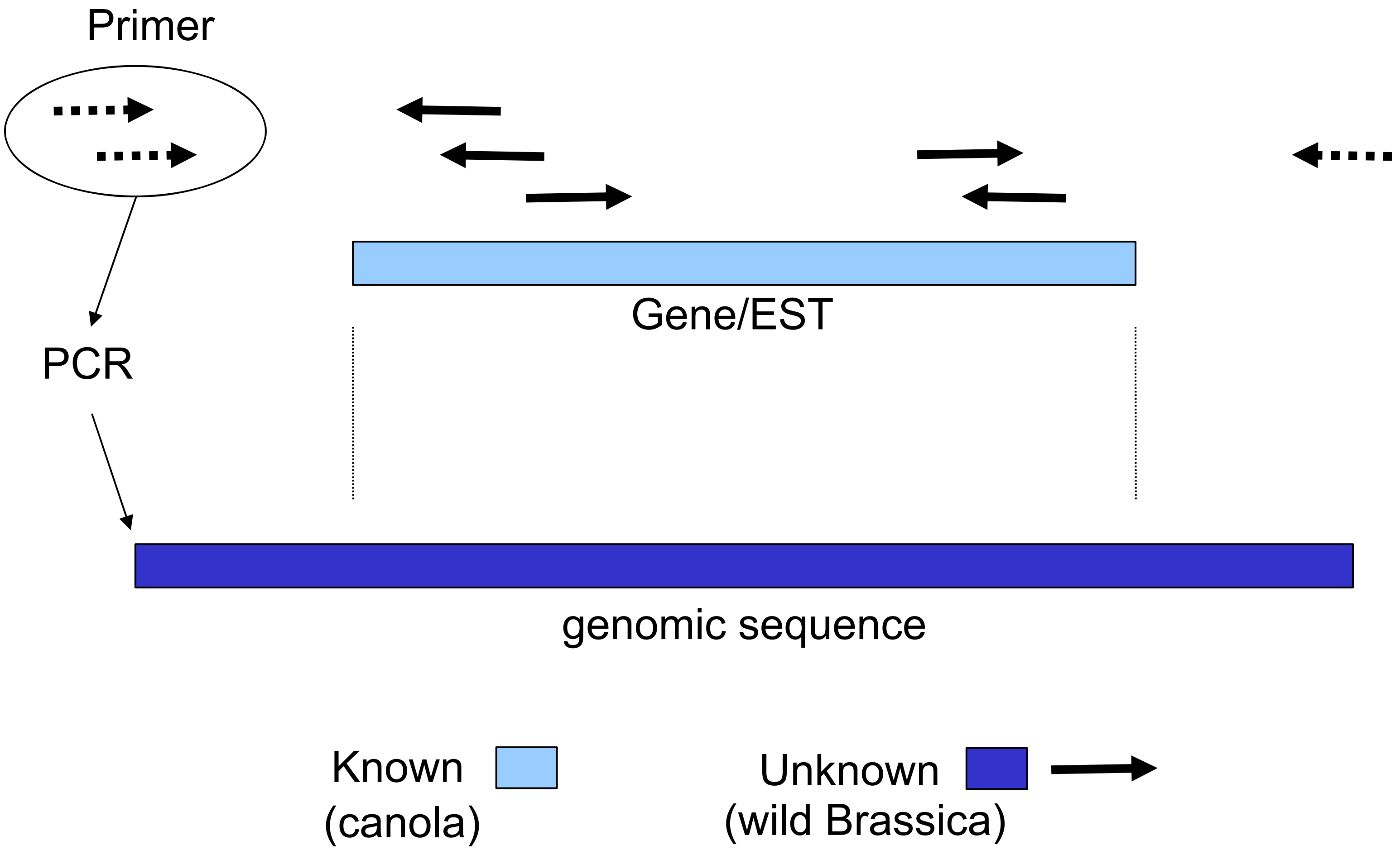
Tag alignment information (download fasta)

ID	Start	End	Tag sequence	Orient	Library	Type
1	330	405	TTCTATGCTTGTTTTCTTTTCTTTTCGGGGAACATTGATTTTCGATGCTTCTATTCACATGTTTCGTTCATGTTTG	1	7DS_03_002	P
1	553	628	TAGGGACAGAGAAATTATTCAACTGAGTTACAAATTTAGTACTTACCAGTGGAGATACTCCCTAAGGTTGGGATCG	-1	7DS_03_002	P
2	17	92	GAACAGGACCGTGTCCAACGGCTGCGAGCTCAAGCCGTCCATGGTCGCCCAGCAGCCCAGGGTTGAGGTGGGCGGC	1	7DS_03_002	P
2	273	348	AAAGAAAACAAGCATAGAAAATCATGATCAACACCAGAGAGCTAGCACAGAGCCATATGCAAGATGGAGCATTAGT	-1	7DS_03_002	P
3	354	429	TCGGGGAACATTGATTTTCGATGCTTCTATTCACATGTTTCGTTCATGTTTGTCCTAGCAAGCACACGACTAATTA	1	7DS_03_002	P
3	558	633	ATATCTAGGGACAGAGAAATTATTCAACTGAGTTACAAATTTAGTACTTACCAGTGGAGATACTCCCTAAGGTTGG	-1	7DS_03_002	P
4	272	347	GACTAATGCTCCATCTTGCATATGGCTCTGTGCTAGCTCTCTGGTGTTGATCATGATTTTCTATGCTTGTTTTCTT	1	7DS_03_002	P
4	527	602	GTTACAAATTTAGTACTTACCAGTGGAGATACTCCCTAAGGTTGGGATCGCTTGGACTTGGAGCATCTGGGTCTAC	-1	7DS_03_002	P
5	76	151	ATATAGGCATTAGATAGTGACTGTGTACGTACGAGTGTGTAGAAGGTCCTCACCTCATTGCCCAACACCTCAACCC	-1	7DS_03_002	A
5	-	-	TCCCTCGTATCCTAGCTAGCAGGCCGGTCGATCTATACTAAGAAGGAAGGGGGAATGGCCGGGAGGGGCAGAGACC	-	7DS_03_002	-
6	675	750	CTTGTGACAGATATCCCCGGTACAACTGGTGCATCCTTCGGGCAGGAGGGGATGTGCTACGAGAGCCCTCGTCCGC	1	7DS_03_002	A
6	-	-	CTATATACTAGTAAACTAGCCACGTTGGATAATTTGCTGACTTGGCGGCGGGTGGCGGACGAGGGCCATGGGTAGA	-	7DS_03_002	-
7	142	217	CACACGTACTATATATATAGAGCTCGATCGATCATGTGCGTGTGGCGAGCACTTTCAGAGCTAAACATATAGGCAT	-1	7DS_03_002	A
7	-	-	TATCCTAGCTAGCAGGCCGGTCGATCTATACTAAGAAGGAAGGGGGAATGGCCGGGAGGGACAGAGACCCGCTGGT	-	7DS_03_002	-
8	527	602	GTAGACCCAGATGCTCCAAGTCCAAGCGATCCCAACCTTAGGGAGTATCTCCACTGGTAAGTACTAAATTTGTAAC	1	7DS_03_002	A
8	-	-	GACAGGCGGGCCGAGGTTGTAGAGCTCGGCGAAGTCCCTGGTGTTGAAGTTCTGGCGCCACCCGGGAGCGTACACG	-	7DS_03_002	-

THE UNIVERSITY OF QUEENSLAND
AUSTRALIA

ACPFG
AUSTRALIAN CENTRE FOR PLANT FUNCTIONAL GENOMICS PTY LTD

TAGdb – Gene discovery



Primer

PCR

Gene/EST

genomic sequence

Known (canola)

Unknown (wild Brassica)

Marshall, D.J., *et al.* (2010) Plant Methods. 6:19

THE UNIVERSITY OF QUEENSLAND
AUSTRALIA

ACPFG
AUSTRALIAN CENTRE FOR PLANT FUNCTIONAL GENOMICS PTY LTD

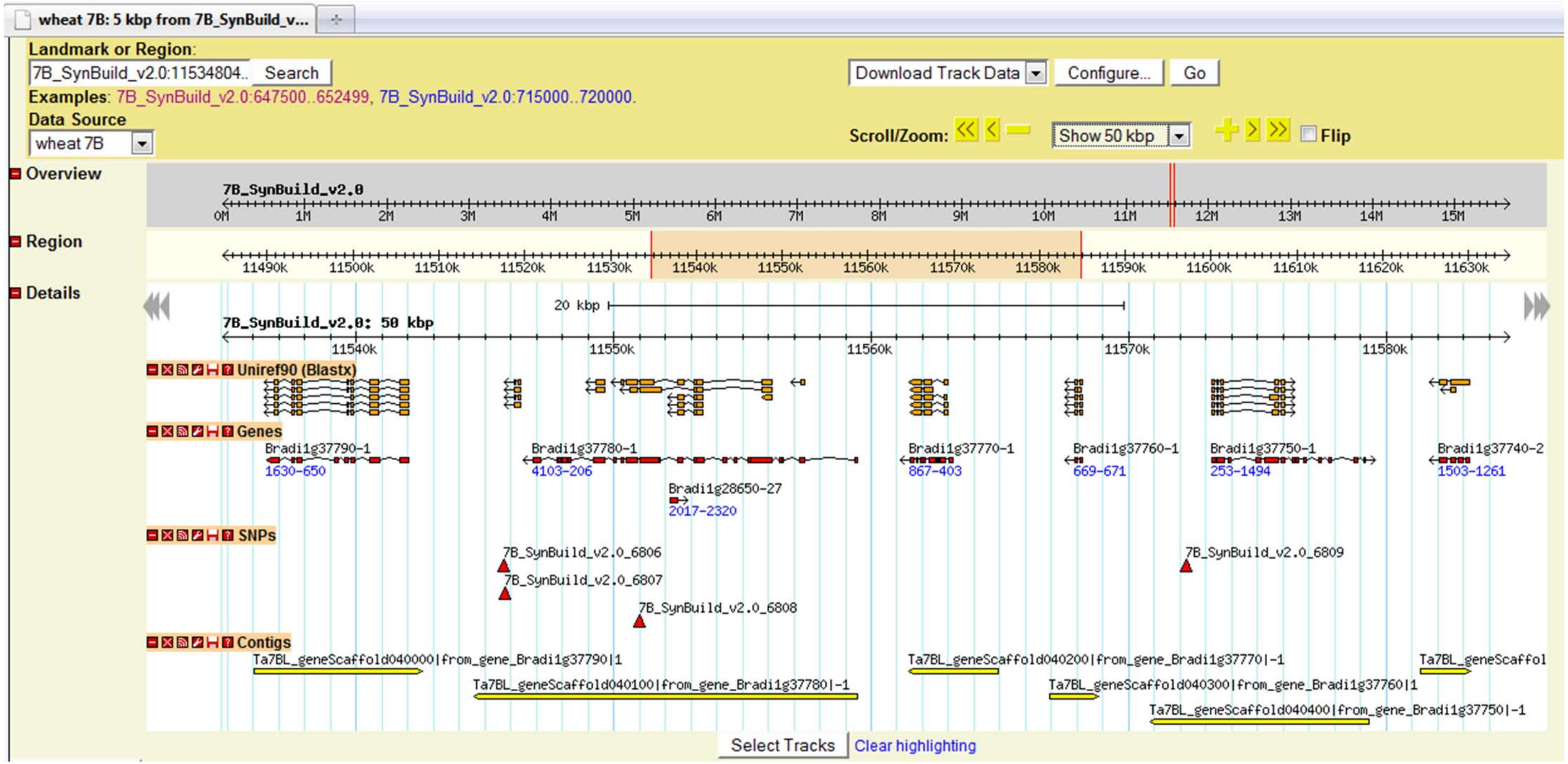
Wheat genomic SNP discovery

- Need all homoeologs to identify varietal SNPs
- Initial four Australian varieties
- >900,000 confident SNPs across group 7
- SNP density varies between genomes A>B>>D

THE UNIVERSITY OF QUEENSLAND
AUSTRALIA

ACPFG
AUSTRALIAN CENTRE FOR PLANT FUNCTIONAL GENOMICS PTY LTD

http://wheatgenome.info/



wheat 7B: 5 kbp from 7B_SynBuild_v...

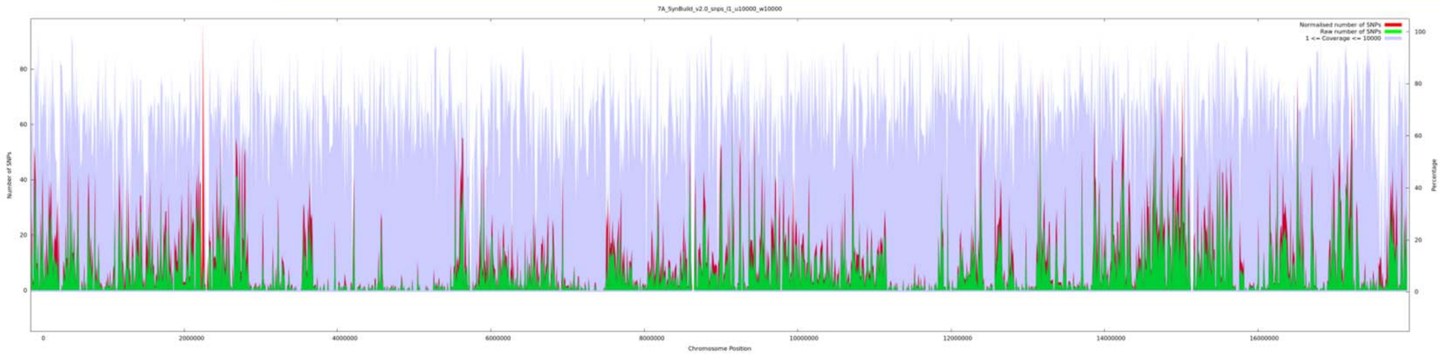
Landmark or Region: 7B_SynBuild_v2.0:11534804..

Examples: 7B_SynBuild_v2.0:647500..652499, 7B_SynBuild_v2.0:715000..720000.

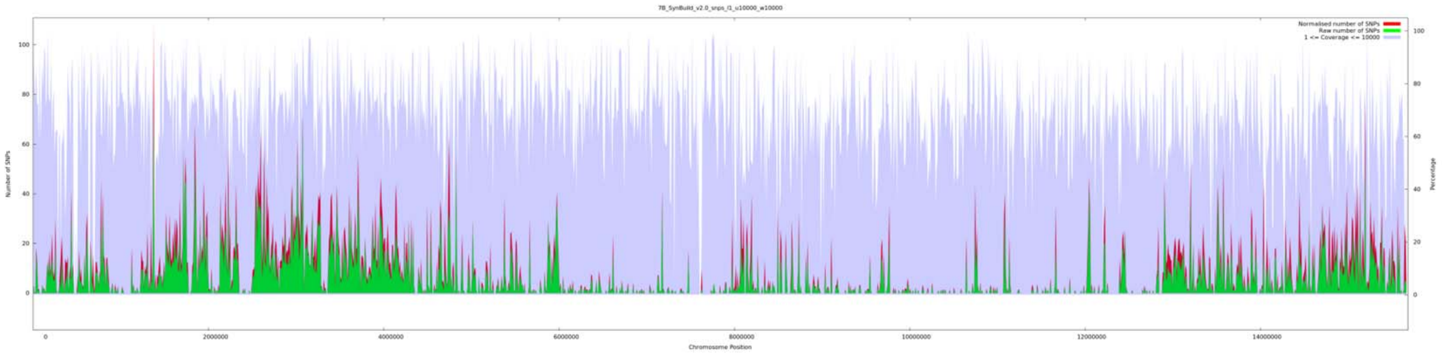
Data Source: wheat 7B

Scroll/Zoom: Show 50 kbp ☐ Flip

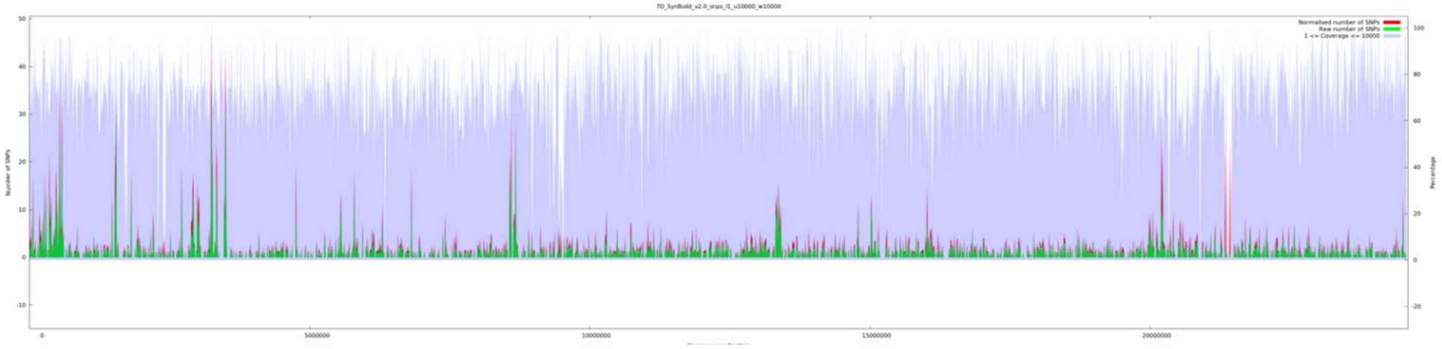
Wheat genomic SNP discovery



A



B



D

How can this accelerate breeding?

- Provide high density SNPs for mapping/association
- Identify candidate genes in assemblies near mapped traits
- Have all favourable alleles been fixed?
- Have unfavourable alleles been dragged along for the ride?
- Use this genotype information for selecting parents/progeny
- Improve drought tolerance in Australian wheat

The University of Queensland
Australia

ACPFG
Australian Centre for Plant Functional Genomics Pty Ltd

Acknowledgements

Australian Government
Australian Research Council

The University of Queensland
Australia

GRDC
Grains Research & Development Corporation

ACPFG
Australian Centre for Plant Functional Genomics Pty Ltd

agrf
Australian Genome Research Facility Ltd

illumina

qcif

Bioplatforms Australia

Paul Berkman
Kenneth Chan
Michael Imelfort
Kaitao Lai
Hong Lee
Michal Lorenc
Sahana Manoli
Jiri Stiller
Paul Visendi

Chris Duran
Dominic Eales
Adam Skarshewski
Lars Smits

Jacqueline Batley
Alice Hayward
Emma Campbell

Hana Šimková
Marie Kubaláková
Jaroslav Doležel
Pilar Hernandez

Contact:
Dave.Edwards@uq.edu.au

BIOMATTERS

geneious

BAYER
Bayer CropScience