

*IWGSC Workshop and Training Session
7-8 April, 2011*

Wheat Bioinformatics activities at the CCG

Matt Bellgard
Director, Centre for Comparative Genomics



CENTRE FOR
COMPARATIVE GENOMICS



Western Australia

Outline

- Contribution to the IWGSC
 - Two specific regions of chromosome 3B
 - Chromosome 7A
- Analytical environment for analysis
- Informatics issues

Outline

- Contribution to the IWGSC
 - Two specific regions of chromosome 3B
 - Chromosome 7A
- Analytical environment for analysis
- Informatics issues

Chromosome 3B

- **The ctg506 region** Selected for detailed analysis
 - Several cell wall invertase (IVR1) genes are located in this region.
 - These genes are often important in maintaining pollen viability during early development.
- **The ctg344 region** Selected for detailed analysis
 - Carries the gwm533 marker that is widely used to track disease resistance (Sr2) located on 3B
 - Potentially a region important in several disease resistances

Core Ideas for Assembly

- We know wheat is “difficult” to assemble
 - mis-assemblies are and will be common
- If we see the same sequence assembled using different assemblers and data
 - this is more likely to be correct
- Points of divergence between different assemblers need to be analysed in detail
- Information from databases such as TREP need to be considered in compiling the detailed genome sequence.
- Genome sequence information needs to relate to a high quality molecular genetic map with traits included

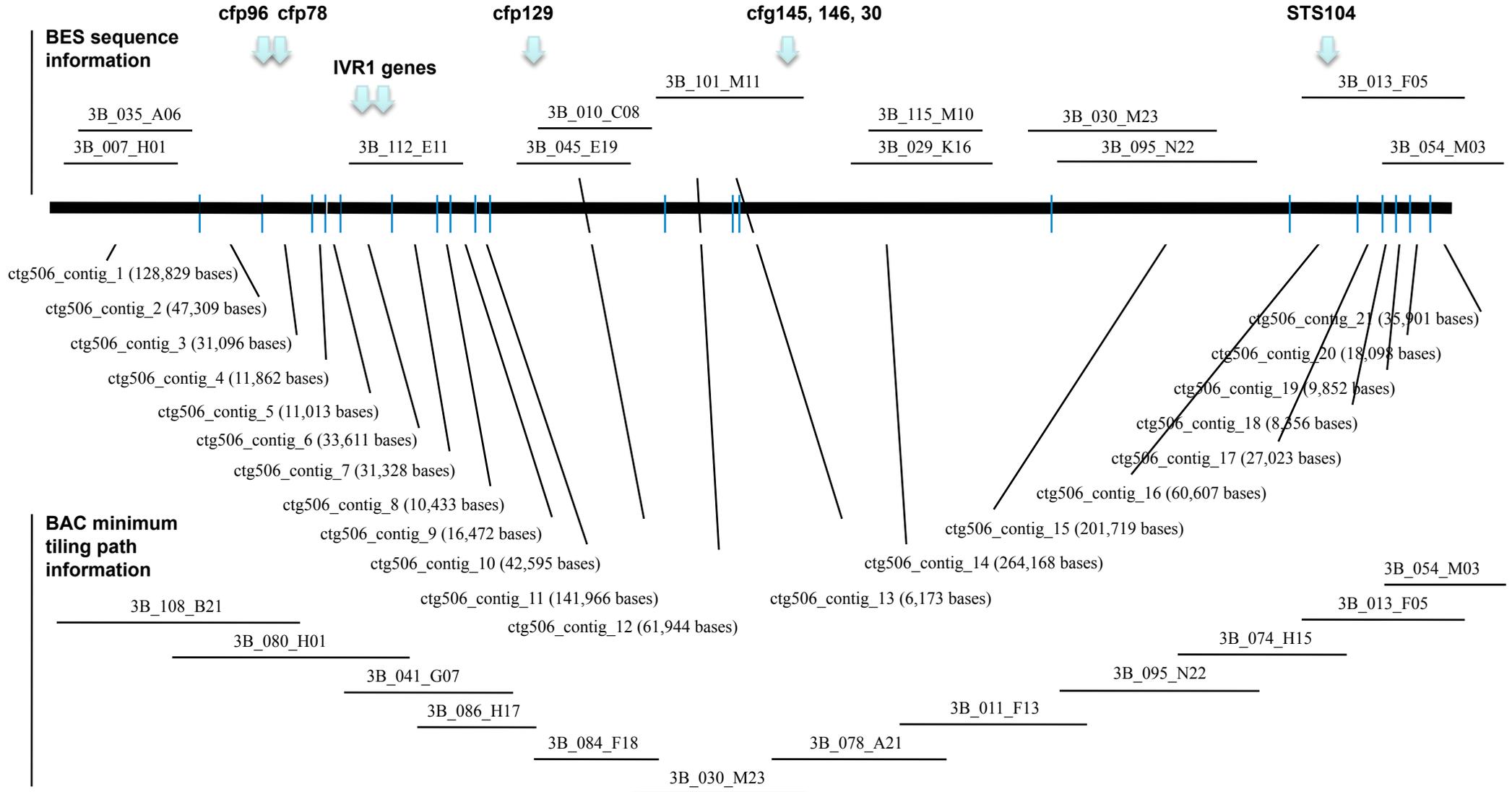
Assembly approach

- Each BAC (Illumina short reads 70bp, paired-end, Sanger) assembled separately using Velvet
 - using multiple parameter combinations
- Different assemblies of the same BAC compared (mummer, freckle)
 - Potential mis-assemblies identified (eg: assemblies that disagree at certain points)
- Using the fingerprint assembly as a guide, identifying sequence present in overlapping BACs – take contigs that agree between different BACs to be “confident”. Start with large contigs and work down to shorter contigs.
- Look at paired-end alignments (BWA + genomeview, Hawkeye) and identify potential mis-assemblies
- Use paired-end information (illumina short reads 500bp insert + Sanger 4kb insert) to try to extend contigs (in general, this is difficult)
- Use LTR information from eg: TREP to order contigs

Original ctg506

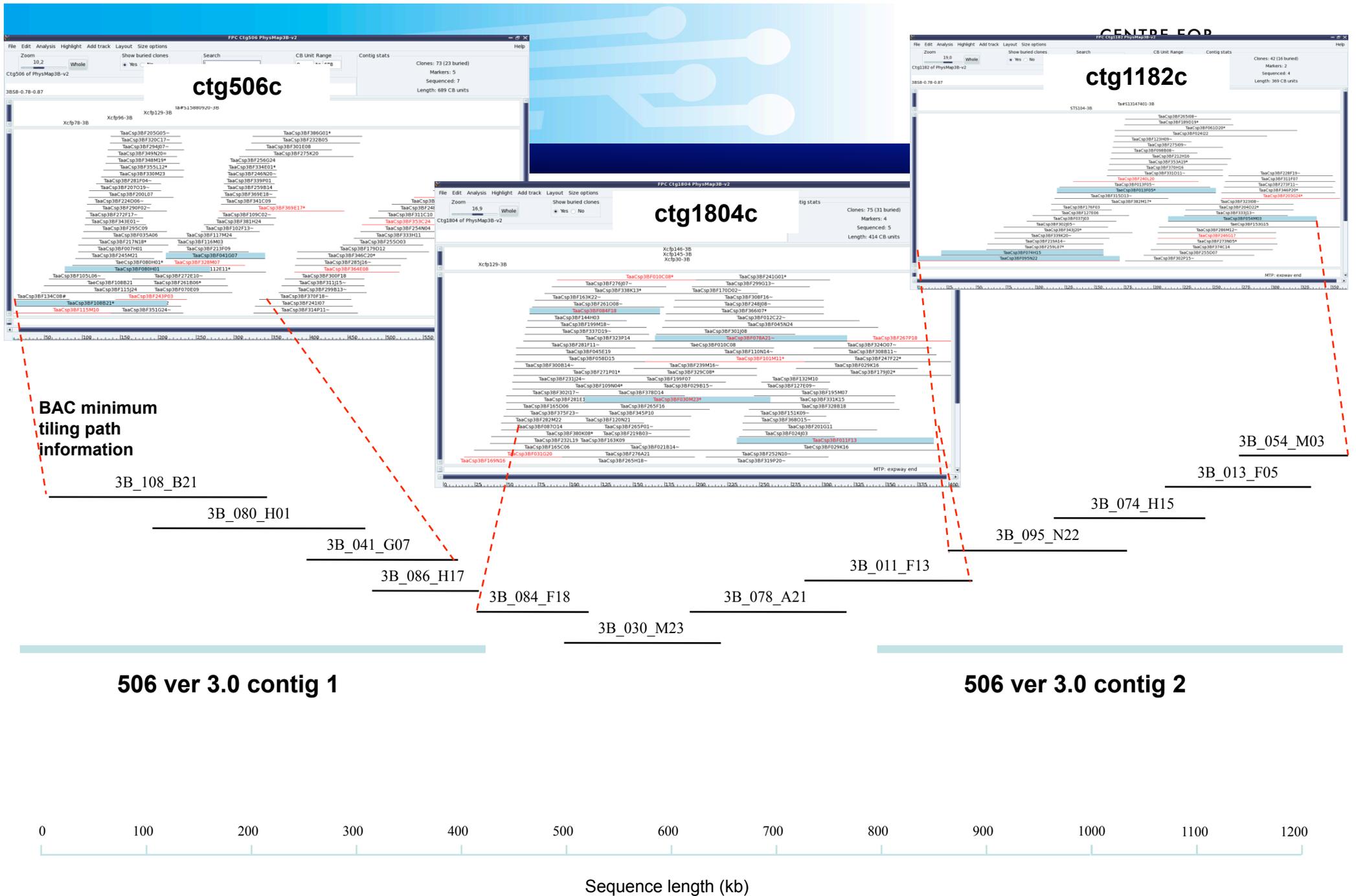


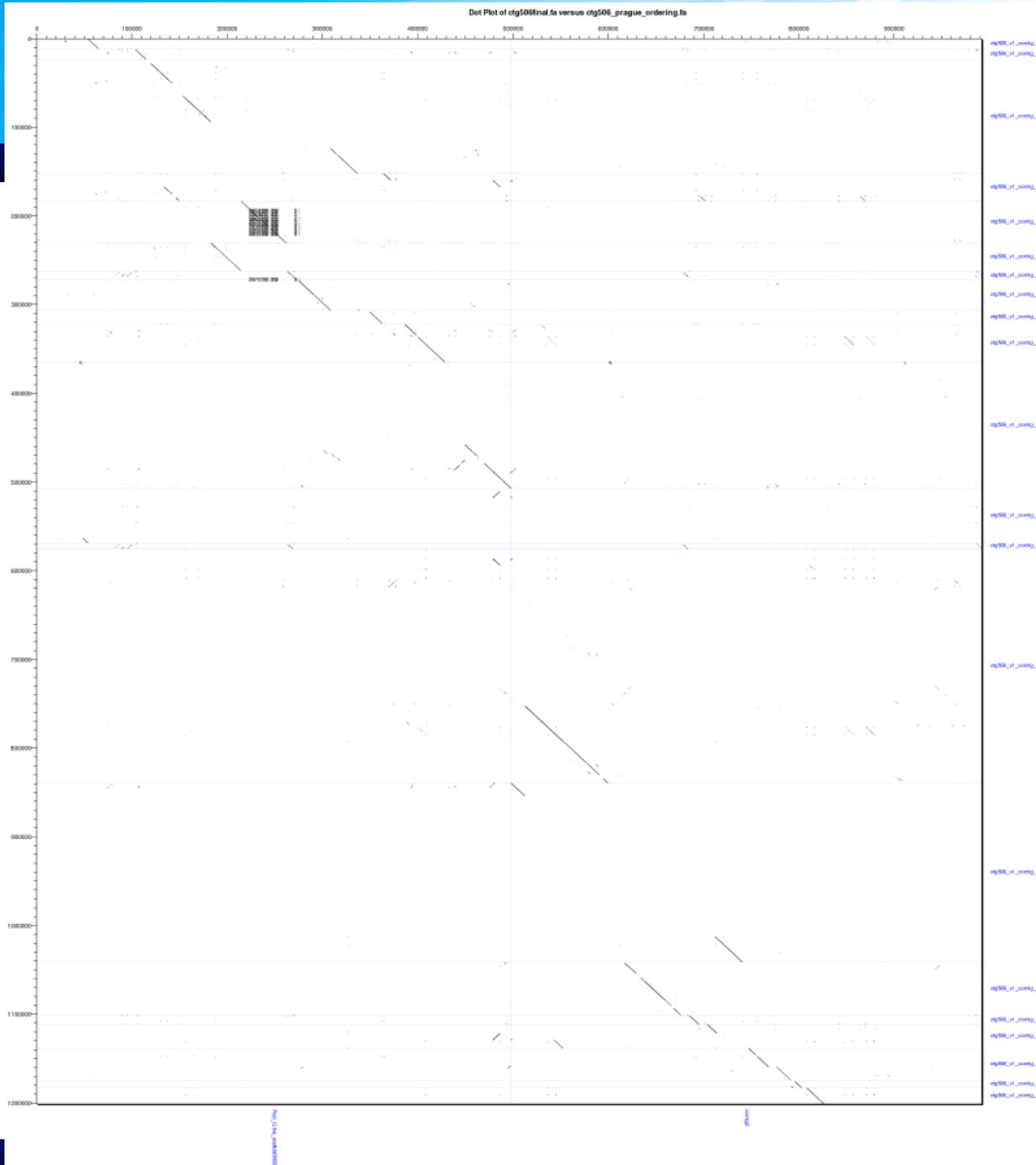
Sequence length (kb)



INRA 454/8kb of Ctg506

- CCG-illumina/Sanger sequencing compared to INRA 454/8kb mate pairs sequencing:
- The INRA scaffolds for ctg506
 - Closed 6 small gaps of Ns in CCG assembly
 - Improved the ordering of contigs in the CCG sequencing
 - Highlights miss-assemblies (at least 6)
- The CCG sequencing
 - Orient large INRA scaffolds
 - Closed 30/83 gaps in INRA sequencing assembly
 - Most of these gaps were in the introns of regions annotated as coding sequences.

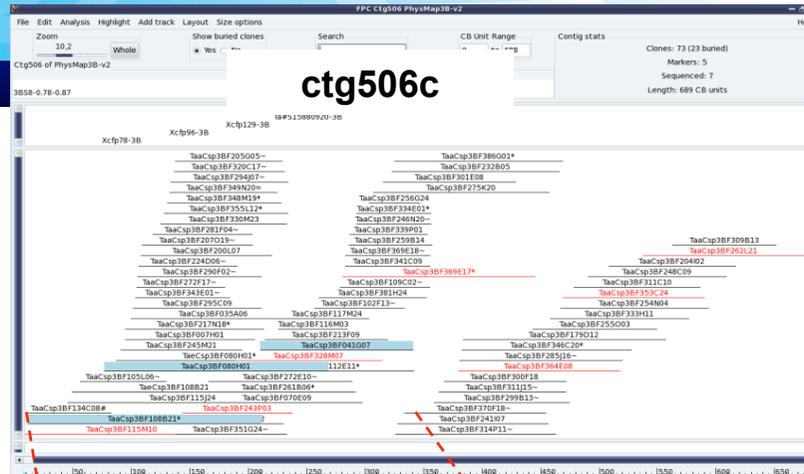




Alignment of CCG short read +
Sanger assembly (y-axis) against
version 3 (combined CCG + INRA)
(x-axis)

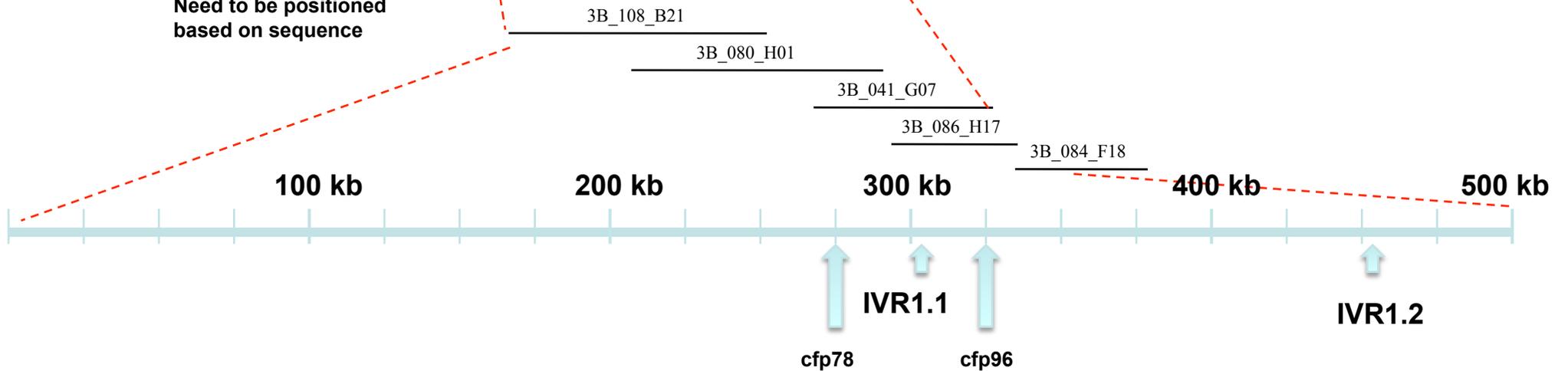
The CCG contigs have been re-
arranged in an attempt to fit the order
revealed by INRA scaffolds

Detailed analysis of IVR gene region



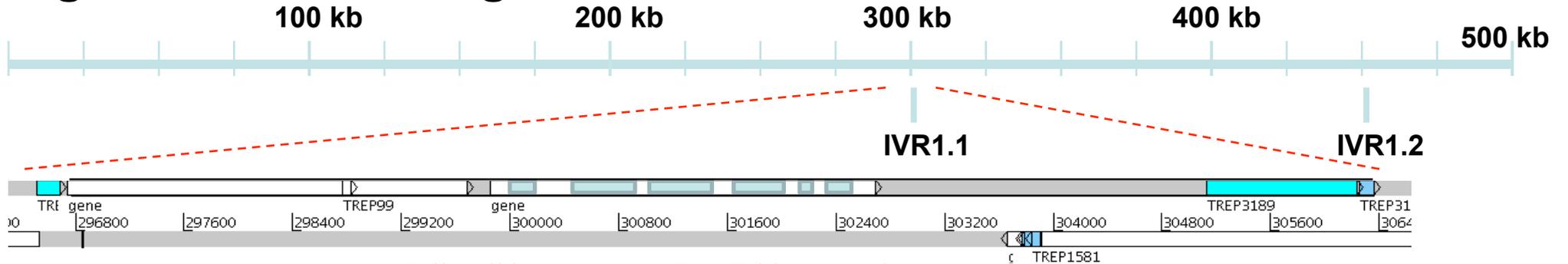
Need to be positioned
based on sequence

BAC minimum
tiling path
information

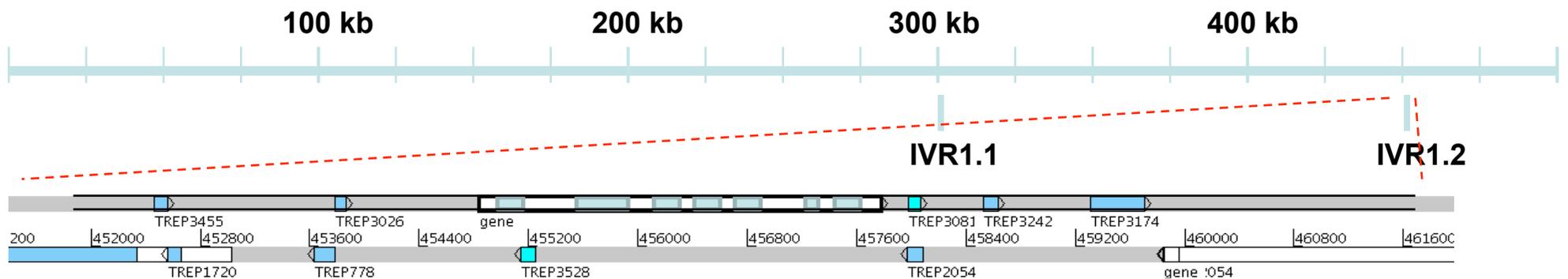


Detailed analysis of IVR gene region

ctg506c ver3.0 contig1



Cell wall invertase n=1 Tax=Triticum aestivum
RepID=O81118_WHEAT"; Evalue "0.0"; Identity "81%
(476/586)";



Cell wall invertase n=1 Tax=Triticum aestivum
RepID=O81118_WHEAT"; Evalue "1e-114"; Identity "54% (235/432)";

Summary – Chromosome 3B

- Have a good appreciation of sequence assembly issues for wheat genome
- Comparative genomics of a repetitive protein kinase locus in ctg344 and biological studies of cell wall invertase genes on ctg506
- Waiting for ctg1804c and ctg344 scaffolds from INRA

Outline

- Contribution to the IWGSC
 - Two specific regions of chromosome 3B
 - Chromosome 7A
- Analytical environment for analysis
- Informatics issues

Chromosome 7A

Overall aims

- Detailed analysis of QTL regions that are important to Australian agriculture
- Trait information linked directly to genome sequence
- ISBP identification of new markers for wheat breeding and selection for traits of interest

Tasks

- BAC-based physical map assembly of 7AS and 7AL
 - BAC end sequencing
 - Survey sequencing of 7A
 - In-depth sequencing of QTL regions
 - Anchor to genetic map
- 



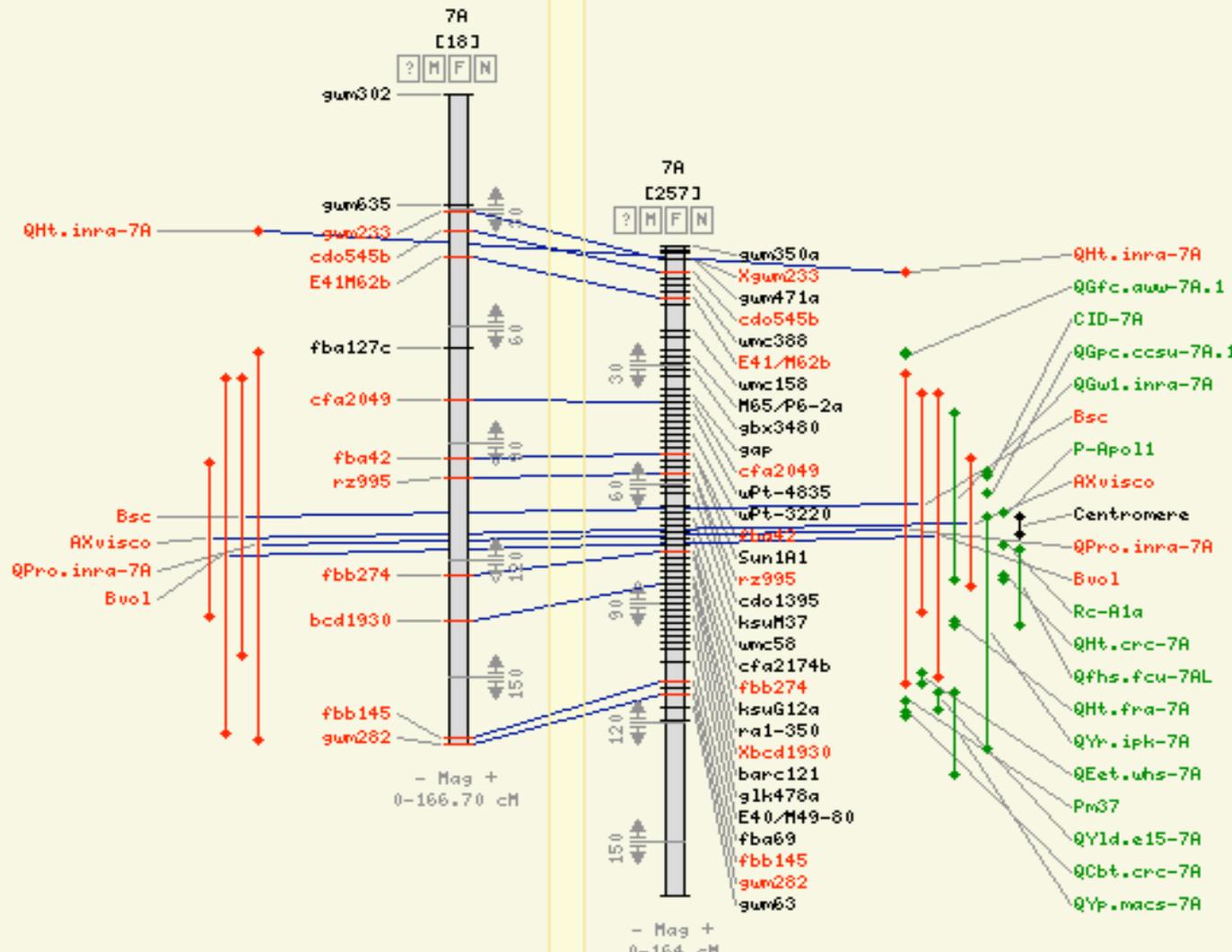
7A genome
sequence

Comparative
Wheat
Renan*Recital Groos 07

[i M X]

Reference
Wheat
7A Consensus May 09

[i M]



- Cmap (at CCG) compiles published QTL/trait data onto composite map.

- The QTL/trait data includes the information from the Wheat Gene Catalogue (McIntosh et al).

- The composite map is built from sections of published maps that share common markers to allow their integration into a master map. Trait data is incorporated based on flanking molecular markers. The molecular markers allow projection on to the genomic DNA sequence

Genome sequence

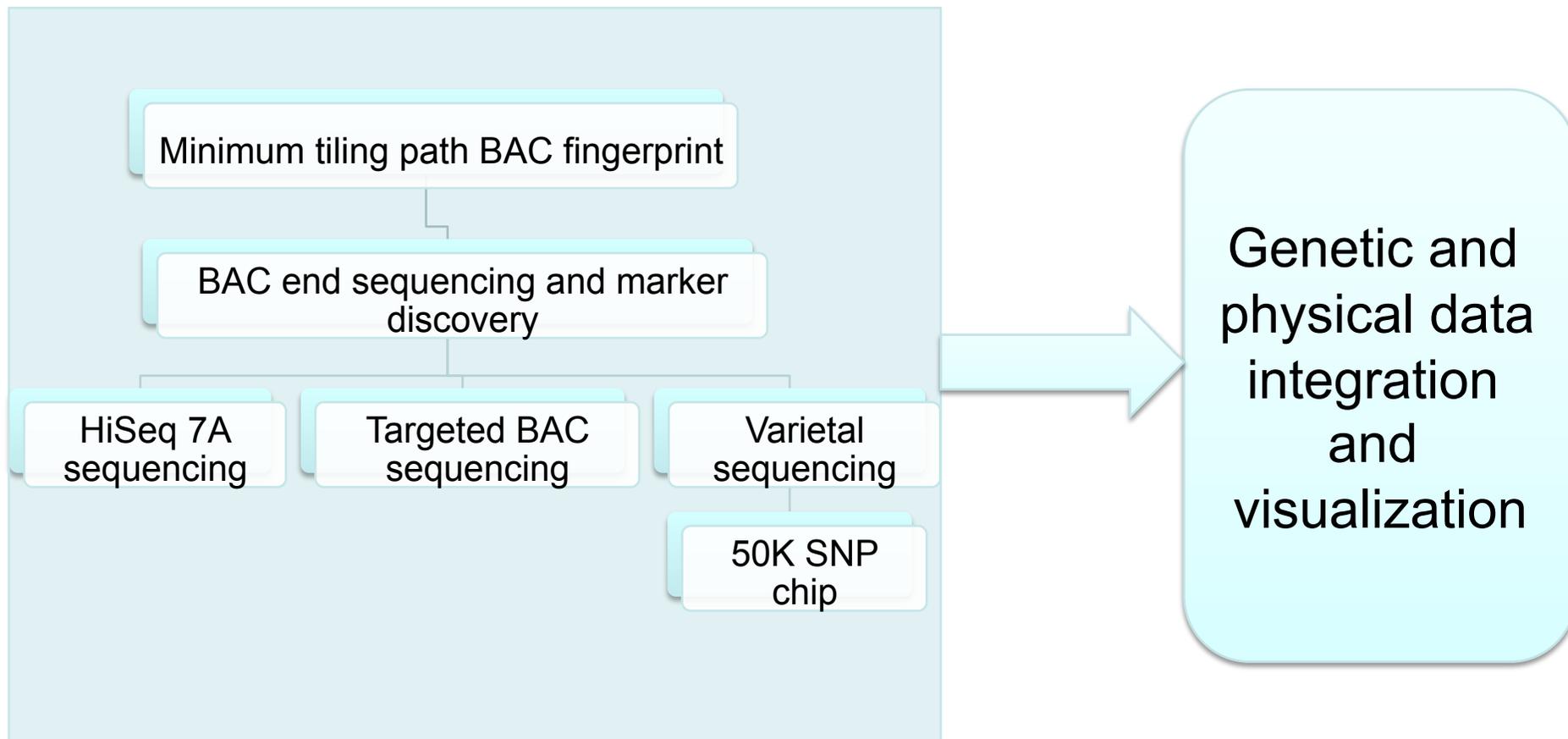
Molecular genetic maps

Traits

7A timeline

- **2010** Funding finalised (GRDC and BioPlatforms Australia)
 - 7A physical map, support to the IWGSC, survey sequencing, QTL region sequencing
- **March 2011** 7AS BAC library (58,000 BAC clones) produced by Dolezel lab
 - UC Davis for DNA fingerprinting (Mingcheng Luo)
- **May 2011** LTC/FPC 7AS fingerprint assembly
- **June 2011** compile BAC contigs and define minimum tiling paths (MTP) for physical map
- **May/June 2011** 7AL library from Dolezel lab (underway) shipped to UC Davis for DNA fingerprinting
- **June** commence BAC end sequencing (BACs from MTPs)
- **2011** Targeted BAC sequencing
- **2012** anchoring of BAC contigs to genetic maps

CCG: Wheat Bioinformatics 7A (Data integration)



Outline

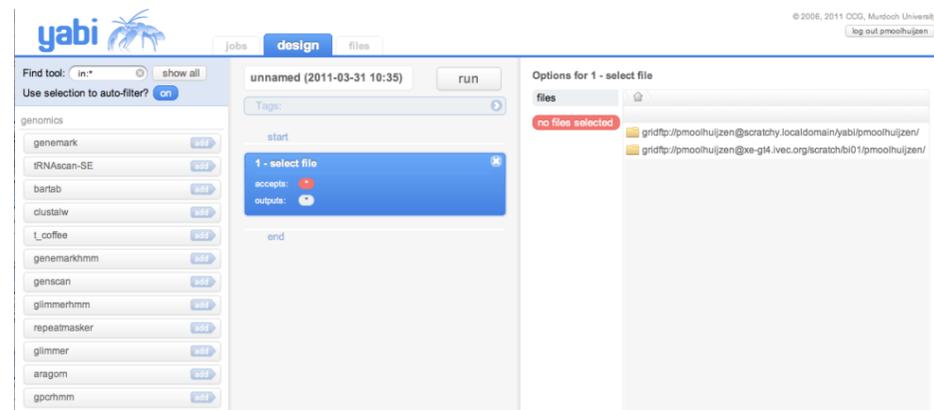
- Contribution to the IWGSC
 - Two specific regions of chromosome 3B
 - Chromosome 7A
- Analytical environment for analysis
- Informatics issues

CCG: Wheat Bioinformatics activities 3B

HTP Assembly (Ctg344 and Ctg506)

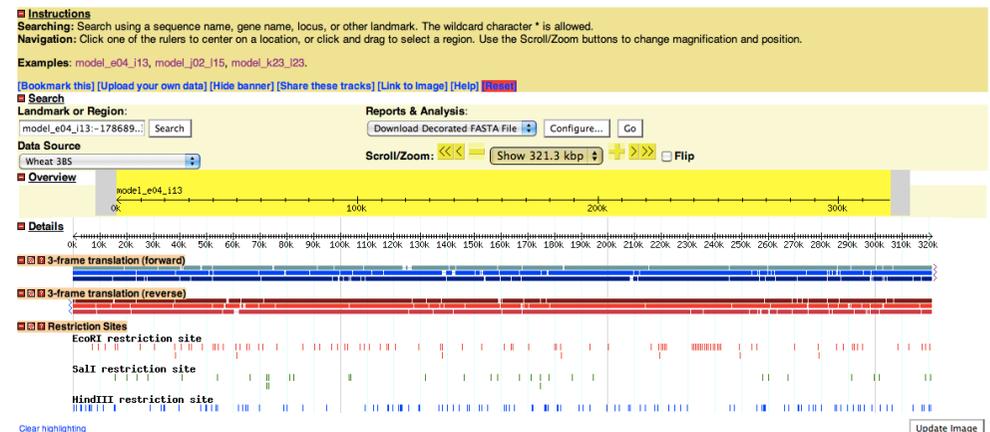
Annotation (HTP pipeline)

Visualization (Gbrowse)

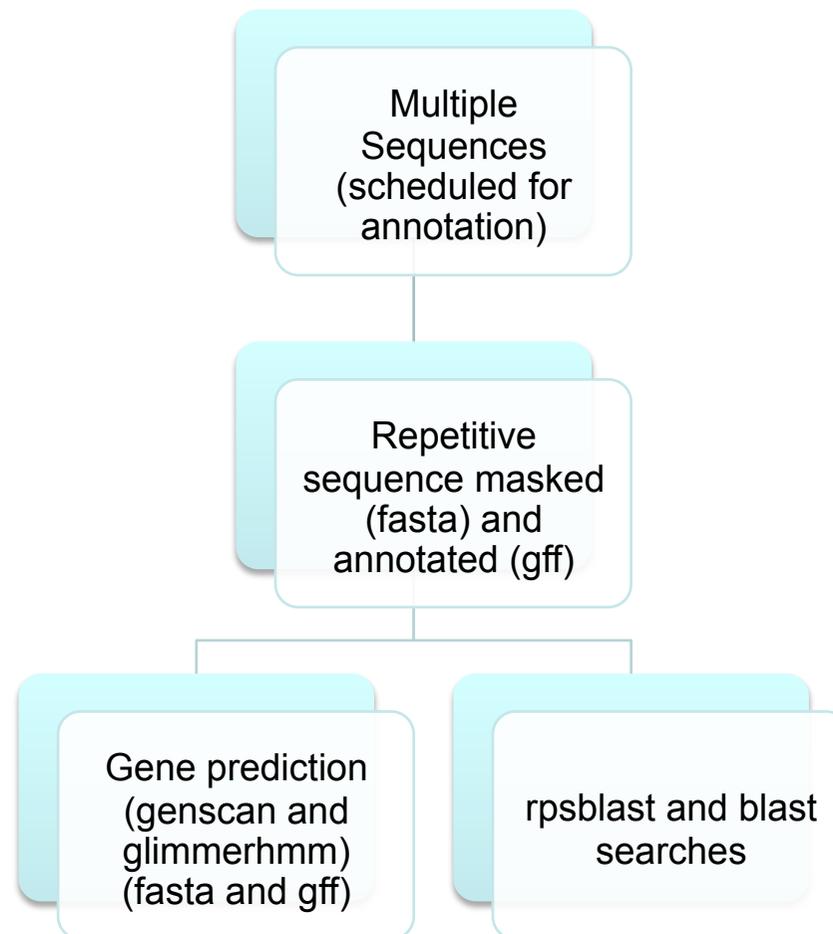


Wheat 3BS

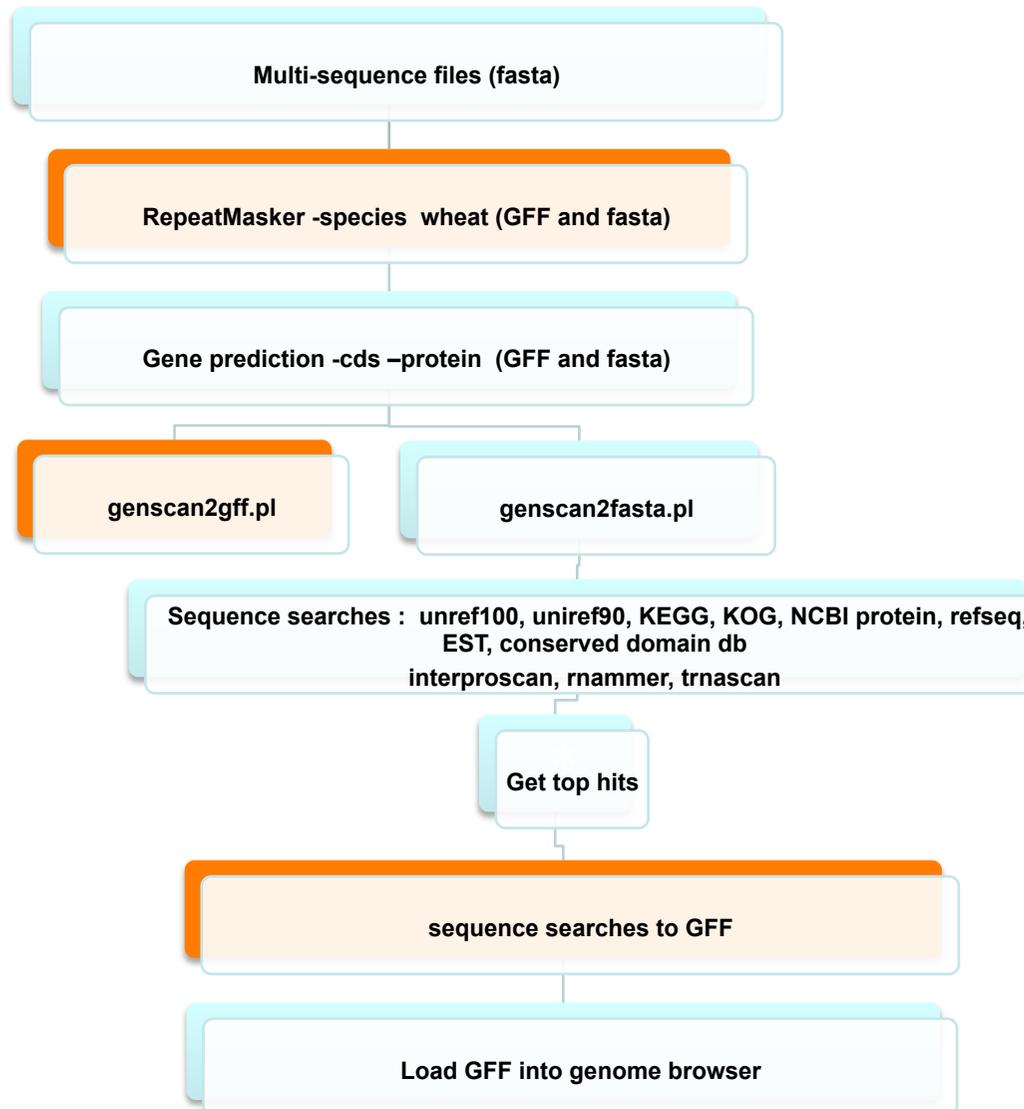
Showing 321.3 kbp from model_e04_i13, positions 1 to 321,310



Bioinformatics annotation workflows URGI (triannot) and YABI



Bioinformatics workflows - Annotate sequences (YABI and Triannot)



YABI - Front-end

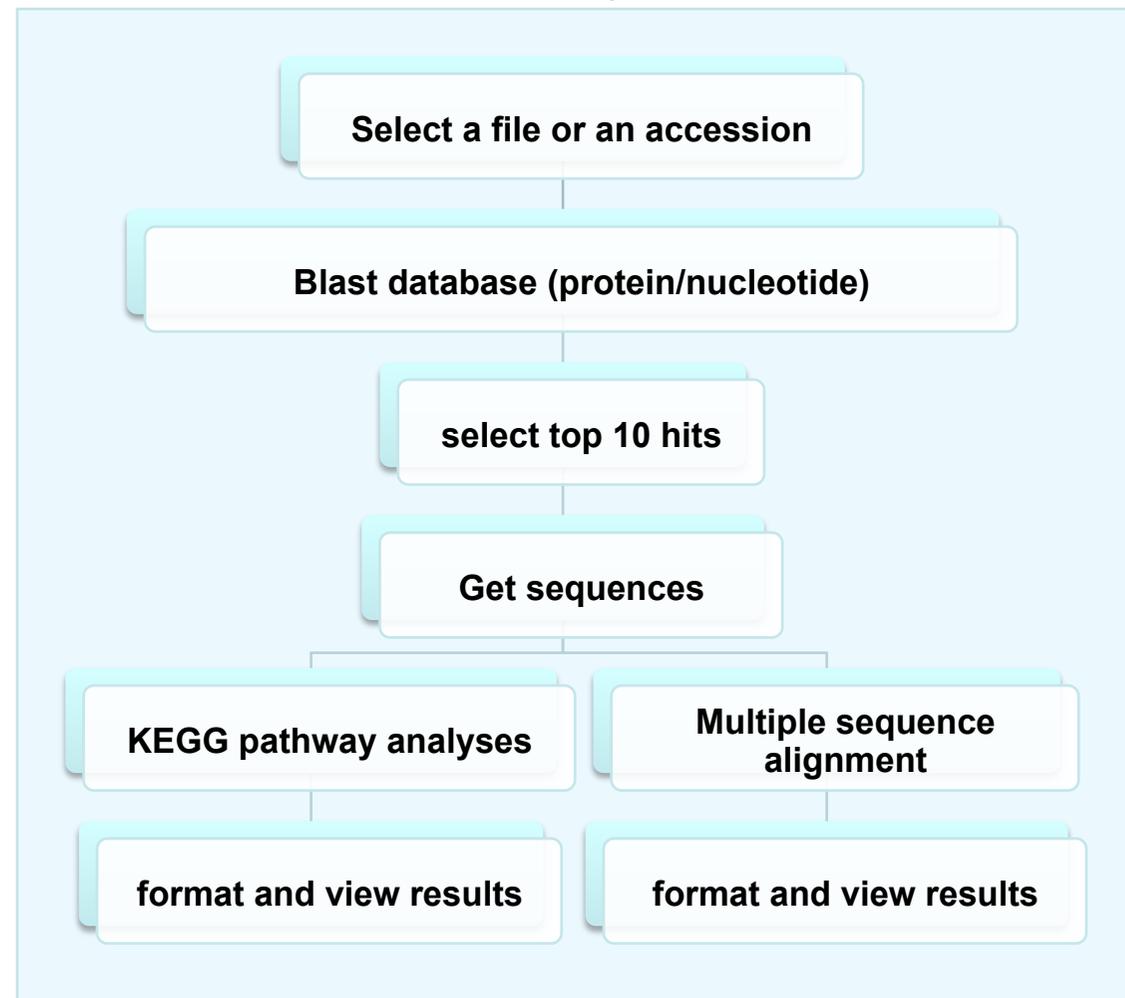
- Designed for a non-technical users
 - jobs tab, design tab, files tab
- Simple and easy to use as possible
- Reuse workflows
- User access
 - When you log in, user sees what they have access to
 - Allow scientists to work together
- Drag and drop tools

YABI - Front-end (2)

- Usability
 - Users warned if tools dragged out of place
 - Anticipate file extensions require for a given tool
 - Tools filtered
 - Tags – capture meta data
 - Errors trapped by system
- File manager
 - Drag and drop files
 - File copying via streaming
 - not via front-end
- Command-line
 - Power users

YABI demo

Simple Blast search and further analyses



CCG: Hardware



- Stage 1A Pawsey Centre
- Ranked 87 in the world
- 9600 cores



CCG: Software development

CENTRE FOR
COMPARATIVE GENOMICS



Western Australia



PB Web

Australian National Duchenne Muscular Dystrophy Registry

Background

Recent scientific and medical advances have improved the lives of many neuromuscular disease patients, and for some of these treatments, more than one country are already using them.

The Australian National Duchenne Registry was developed to collate a patient register for Duchenne Muscular Dystrophy. The Registry provides a platform for researchers and clinicians to quickly identify patients and develop strategies that target specific genetic and clinical subgroups about new trials and studies.

Purpose

To provide Australian Duchenne patients with improved outcomes of individualized care.

Metabolomics Australia User and Quote Manager

ID	U...	Email	First Name	Last Name	Phone
199		jennheitz@yahoo.com	Jennifer	Heitz	04065842
198		c.schmitz-peffer@parvan.org.au	Carsten	Schmitz-Peffer	29295821
197		tpower@ccg.murdoch.edu.au	Brad	Power	12345556
195		mpornice@unimelb.edu.au	Maria Pia	Cornio	36297645
193		corneali@pms.edu.au	Carla	Costana	98-2-3544
192		elwin.verheij@zo.nl	Elwin	Verheij	+31 88 88
188		Donna.Glassop@csiro.au	Donna	Glassop	
176		kelle@spooowaycentre.org	Kelle	L...	
198		Deepak.karshen@cmr.edu.au	Deepak	Di...	
195		tpower@ccg.murdoch.edu.au	Brad	P...	
193		tpower@ccg.murdoch.edu.au	Brad	P...	
162		phaj@cbicgenetics.com	Philip	S...	
156		arjandi@yahoo.com	Mohammad	Aj...	
153		n.turner@parvan.org.au	Ngel	T...	
148		luthern.zuk@univie.ac.at	Kathlene	Z...	
145		tpower@ccg.murdoch.edu.au	Brad	P...	
143		andrew.brown@ngc.com.au	Andrew	B...	
140		sdayalan@gaenna.cs.rmit.edu.au	Saravanan	D...	
124		tpower@ccg.murdoch.edu.au	Brad	P...	
123		tpower@ccg.murdoch.edu.au	Brad	P...	
115		tpower@ccg.murdoch.edu.au	Brad	P...	
113		nicketal@mac.com	test	te...	
111		saravanan.dayalan@gmail.com	Saravanan	D...	
110		tpower@ccg.murdoch.edu.au	Brad	P...	
108		saravanan.dayalan@gmail.com	Saravanan	D...	
107		saravanan.dayalan@gmail.com	Saravanan	D...	
101		sdayalan@gaenna.cs.rmit.edu.au	Saravanan	D...	
100		sdayalan@gaenna.cs.rmit.edu.au	Saravanan	D...	

yabi

Username:

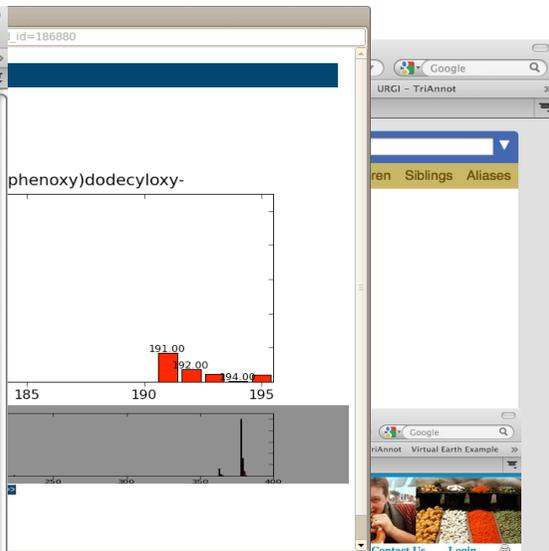
Password:

Help/Support: Please direct all enquiries to yabi@ccg.murdoch.edu.au

Demonstration Account: A minimal demo account may be accessed with username "yabidemo" and password "yabidemo". All files and workflows will be periodically removed from the demo account.

Screencasts: Click a link below to watch the screencast via YouTube.

- Marine Science workflow
- Blast and ClustalW
- GetORF and gpcchmm
- Chipsseq
- Chipsseq 2
- Trans Proteomic Pipeline (Peptide Prophet)



LOVD Gene Home

General information

Gene name:

Gene symbol:

Database location:

Curator:

PubMed references:

Date of creation:

Last update:

Version:

Add sequence variant:

First time submitters:

Reference sequence:

GenBank reference:

Total number of unique DNA variants reported: 61

Total number of individuals with variant(s): 9401

Total number of variants reported: 9401

Subscribe to updates of this gene:

NOTE:

Graphical displays and utilities

- Summary tables
- UCSC Genome Browser
- Ensembl Genome Browser
- NCBI Sequence Viewer

YABI is best viewed in standards compliant browsers:

Authorised by: Prof Matthew Bellgard; Director, Centre for Comparative Genomics
Disclaimer & Copyright Notice © 2008, 2011 Murdoch University
Contact: Site administrator:
Logo font 'fontless' by Top Wanders

Done	1586	6	None	None
	1585	5	None	None
	1584	4	None	None
	1583	3	None	None

provided by Grain Foods CRC Limited, Australia.

Information for cereal grains and pulses with applications to find the database easy to use with useful tools such as comparative genomics.

u/pms at a

Google Calendar

Username:

Password:

Forgot your password? Changed your details?

Contact Us
Arch Centres Program
Genomics, Murdoch University.



Username:

Next generation sequencing analysis projects

- De novo genome/transcriptome assembly and annotation
 - Wheat, barley, rat mutant, Wine yeast, cane toad, Rhizopertha, Campylobacter, Euphorbia, Cattle tick, dog tick, rhizobium, spirochetes
- Transcriptomics/epigenomics/metagenomics
 - microRNAs (human, Arabidopsis, cattle tick)
 - Epigenomics in Arabidopsis
 - Metagenomics of environmental samples (ancient DNA)
- Disease association
 - LPK rat mutant, human disorders, diagnosis assays

Outline

- Contribution to the IWGSC
 - Two specific regions of chromosome 3B
 - Chromosome 7A
- Analytical environment for analysis
- Informatics issues

Informatics considerations

- Process/timing to integrate 3B CCG results back to IWGSC
- Protocols for BAC physical assembly
- Process/timing to integrate 7A results back to IWGSC
 - Linkages to CCG resources
- CCG could assist in scoping LTC software porting requirements

Acknowledgements

CCG Team

- Rudi Appels
- Gabriel Keeble
- Adam Hunter
- Andrew McGregor
- Paula Moolhuijzen

Collaborators

- Catherine Feuillet, Etienne Paux and Frederic Choulet (INRA, France)
- Jaroslav Dolezel (Czech Republic)
- Mingcheng Luo (UC Davis)

Funding

- GRDC
- BioPlatforms Australia