

Wheat Genomic Resources in a post-reference sequence era: a road map to establish priorities, timelines and outcomes

Prospects in Wheat Genomics --- what is needed from a research perspective

Rudi Appels, International Wheat Genome Sequencing Consortium (IWGSC)

Murdoch University (Western Australia), based at AgriBio (Victoria)

- **Background on IWGSC focused on community survey carried out in 2016**
- **Completion of the reference genome assembly for all the 21 chromosomes**
- **Assignment of genes/traits in Wheat Gene Catalogue (RA McIntosh) to reference genome sequence**
- **Enhance the current pilot EMBL-ABR project on 7A for manual annotation of the 21 chromosomes within the reference genome to include functional studies.**

Wheat Genomic Resources in a post-reference sequence era: a road map to establish priorities, timelines and outcomes

Prospects in Wheat Genomics --- what is needed from a research perspective

- **Background on IWGSC focused on community survey carried out in 2016**
- **Completion of the reference genome assembly for all the 21 chromosomes**
- **Assignment of genes/traits in Wheat Gene Catalogue (RA McIntosh) to reference genome sequence**
- **Manual annotation of the 21 chromosomes within the reference genome to include functional studies.**

IWGSC survey of wheat research community

- **High rate of responses** (14% GM, 56% CC) . General membership (GM) is ca 800 and Coordinating committee (CC) is 64.
- **7 sponsors filled out the survey**
 - 3 of them indicated that they were willing to sponsor coordination activities**
 - 5 of them indicated that they were willing to sponsor specific research activities**
- **The IWGSC should coordinate the maintenance and update of the reference sequence**
 - **The reference sequence should be maintained: high importance for 99% GM and 76% CC**
 - **The central database resource should be updated and maintained regularly (97% GM, 100% CC)**
 - **This should be done by the IWGSC (97% GM, 81% CC)**
- **Post reference genome projects that should be coordinated:**
 - **Community based manual curation and annotation (61% GM, 67% CC)**
 - **Database for the wheat genome sequence (61% GM, 67% CC)**
 - **De novo sequencing and assembly of multiple genomes (57% GM, 67% CC)**
 - **Functional annotation (67% GM, 50% CC)**
- **Members are willing to play a role**
 - **50% GM and 83% CC**
 - **As project leaders (74% GM and 75% CC)**

Wheat Genomic Resources in a post-reference sequence era: a road map to establish priorities, timelines and outcomes

Prospects in Wheat Genomics --- what is needed from a research perspective

- Background on IWGSC focused on community survey carried out in 2016
- **Completion of the reference genome assembly for all the 21 chromosomes**
- Assignment of genes/traits in Wheat Gene Catalogue (RA McIntosh) to reference genome sequence
- Manual annotation of the 21 chromosomes within the reference genome to include functional studies.

Completion of the reference genome assembly

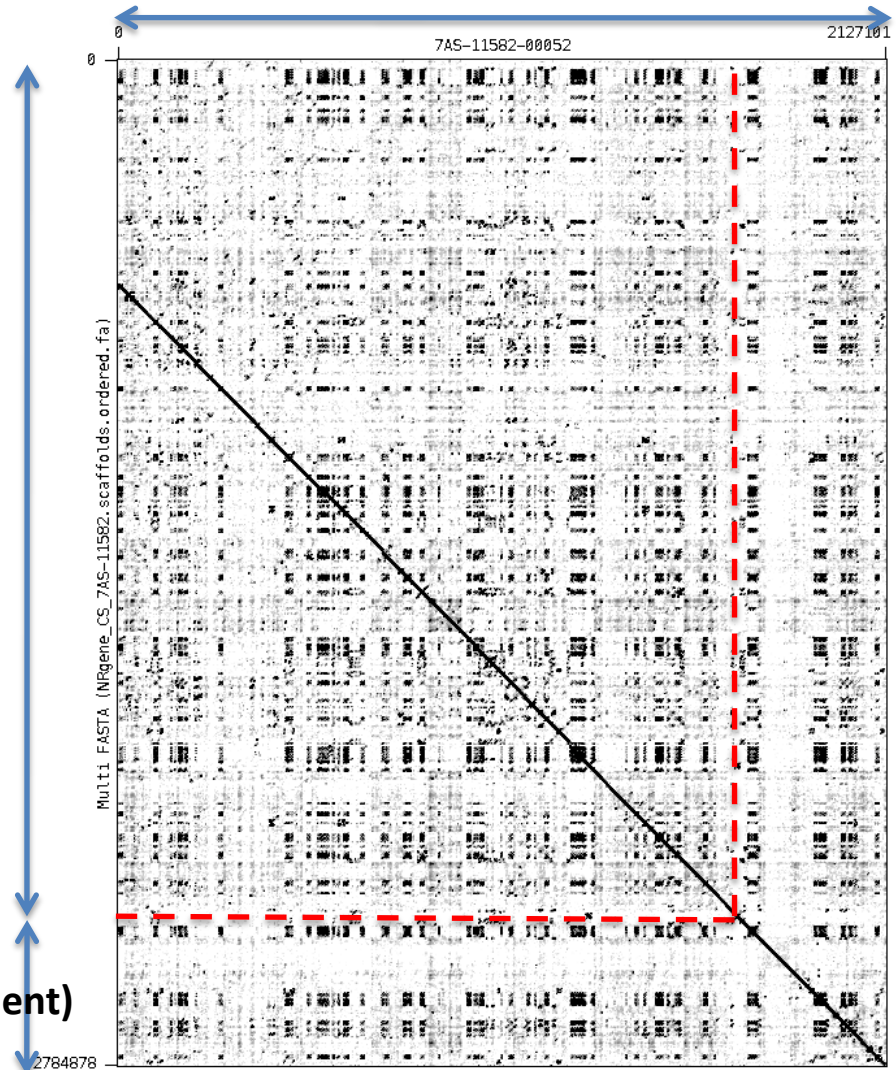
2.1Mb Gydle (g^{a}) scaffold 7AS-11582

Reference sequence for the wheat genome needs to be completed to a high quality

NRGene scaffold31394

- In this 7A example, NRGene extends the Gydle 7AS-11582 scaffold and the Gydle 7AS-11582 scaffold joins two NRGene scaffolds and corrects orientation.
- At this level a genome sequence assembly is classified as finished

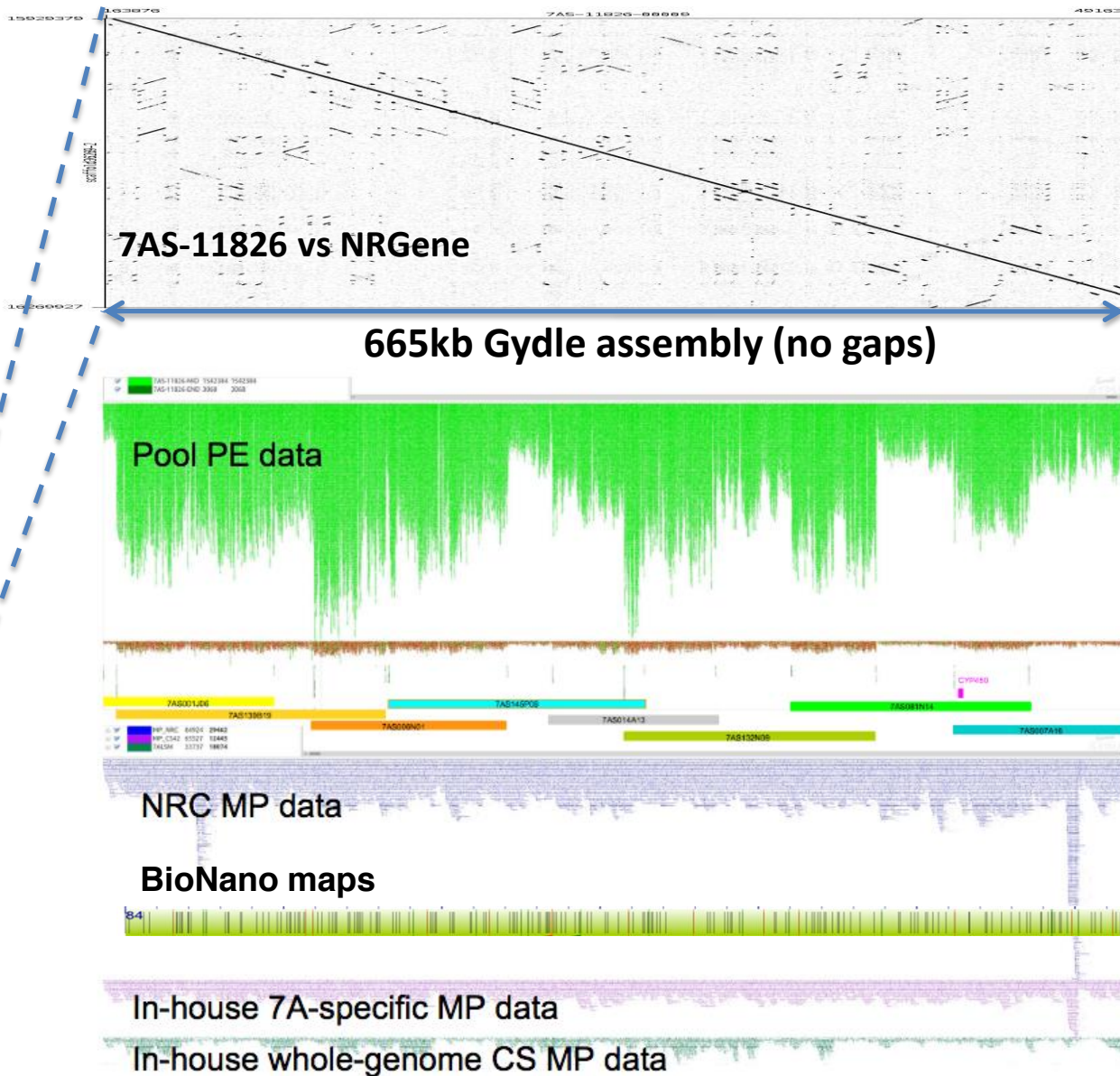
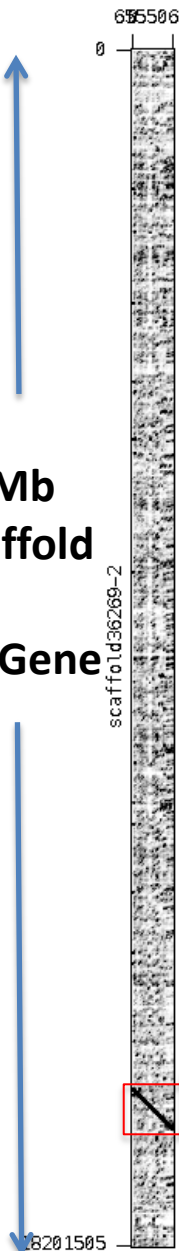
NRGene scaffold97934 (reverse complement)





Completion of the reference genome assembly requires the integration of multiple independent datasets

18Mb scaffold in NRGene



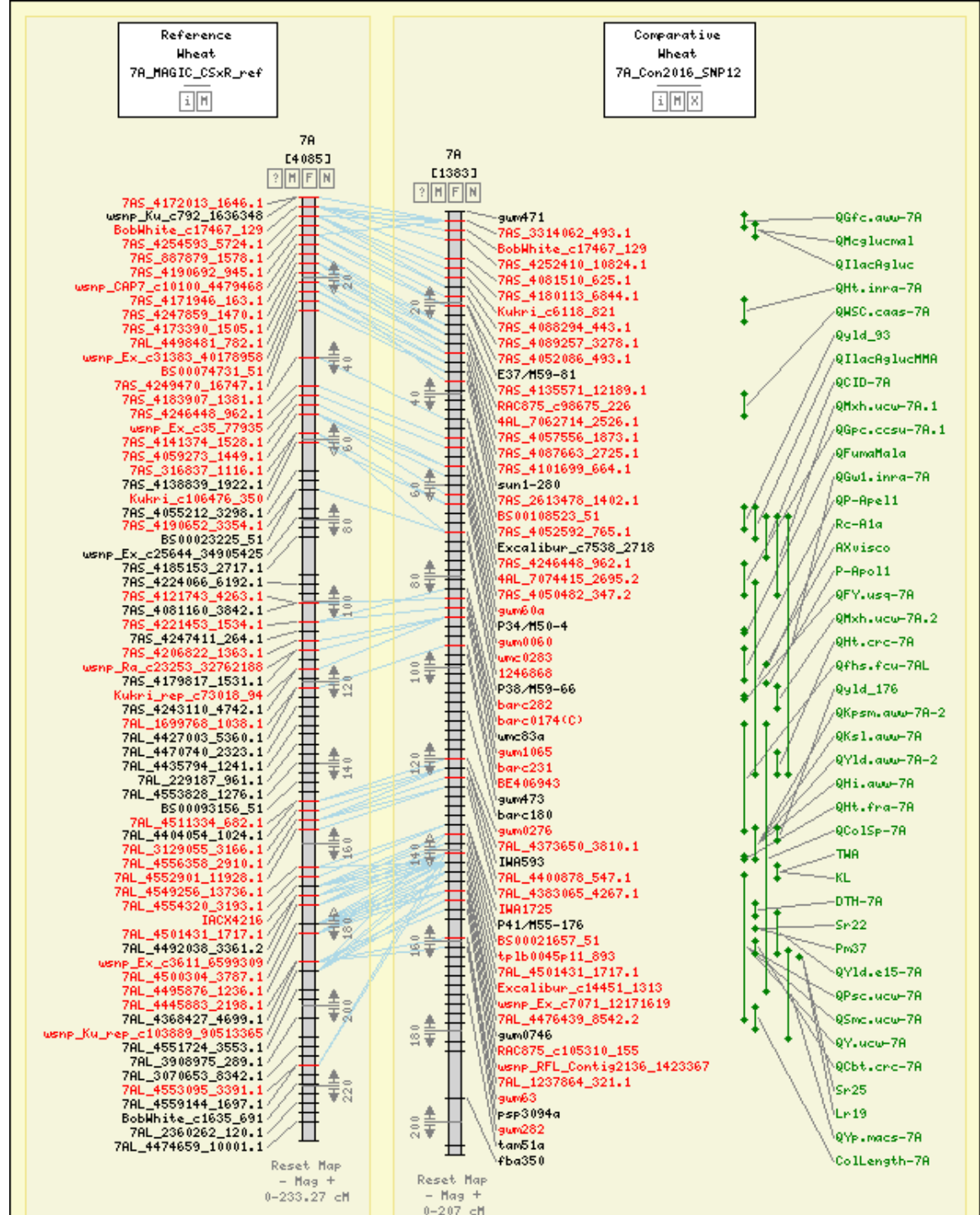
Wheat Genomic Resources in a post-reference sequence era: a road map to establish priorities, timelines and outcomes

Prospects in Wheat Genomics --- what is needed from a research perspective

- **Background on IWGSC focused on community survey carried out in 2016**
- **Completion of the reference genome assembly for all the 21 chromosomes**
- **Assignment of genes/traits in Wheat Gene Catalogue (RA McIntosh) to reference genome sequence**
- **Manual annotation of the 21 chromosomes within the reference genome to include functional studies.**

Example

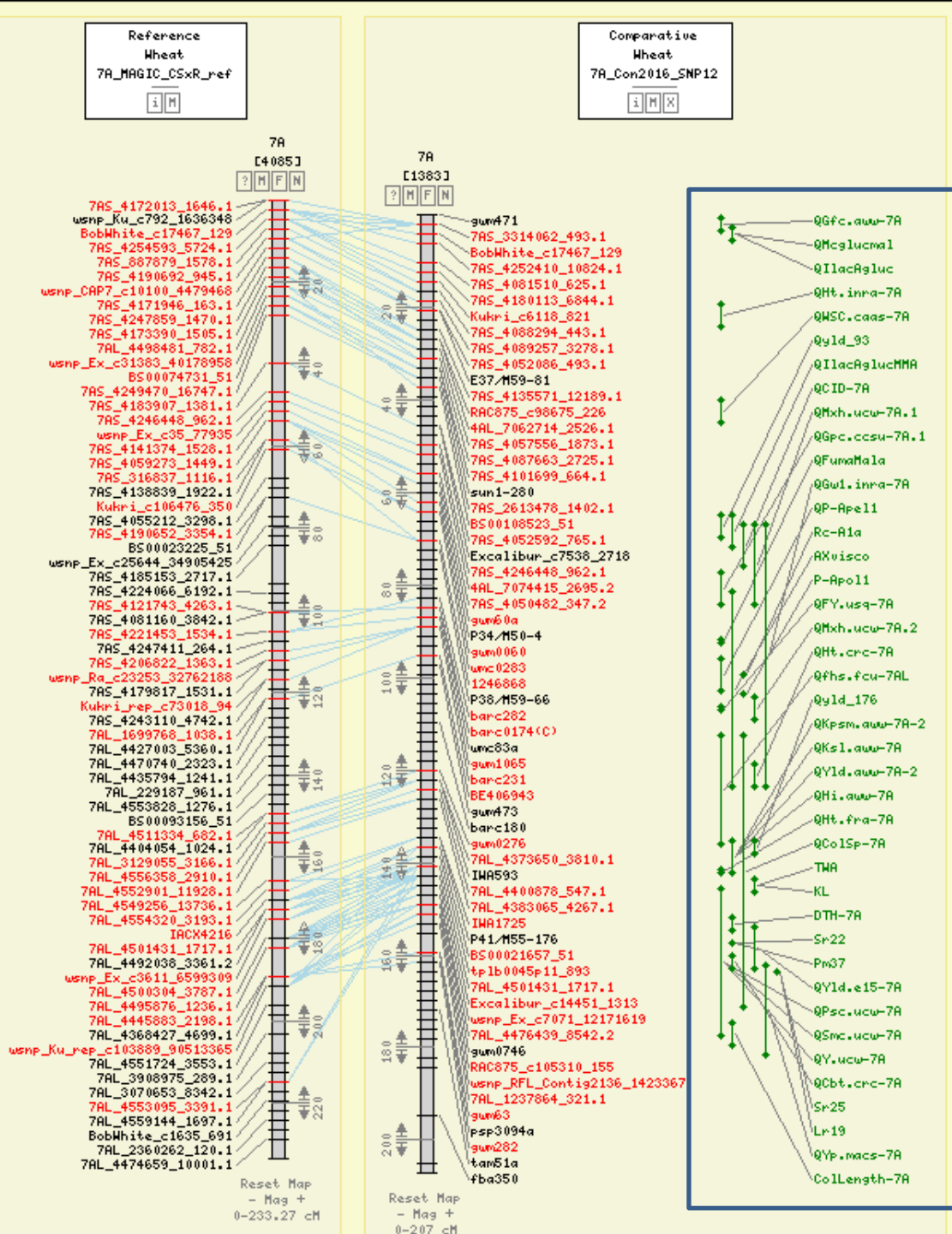
The high density-high resolution molecular genetic map for 7A that was used to anchor the genome sequence can also anchor a composite map.



Example

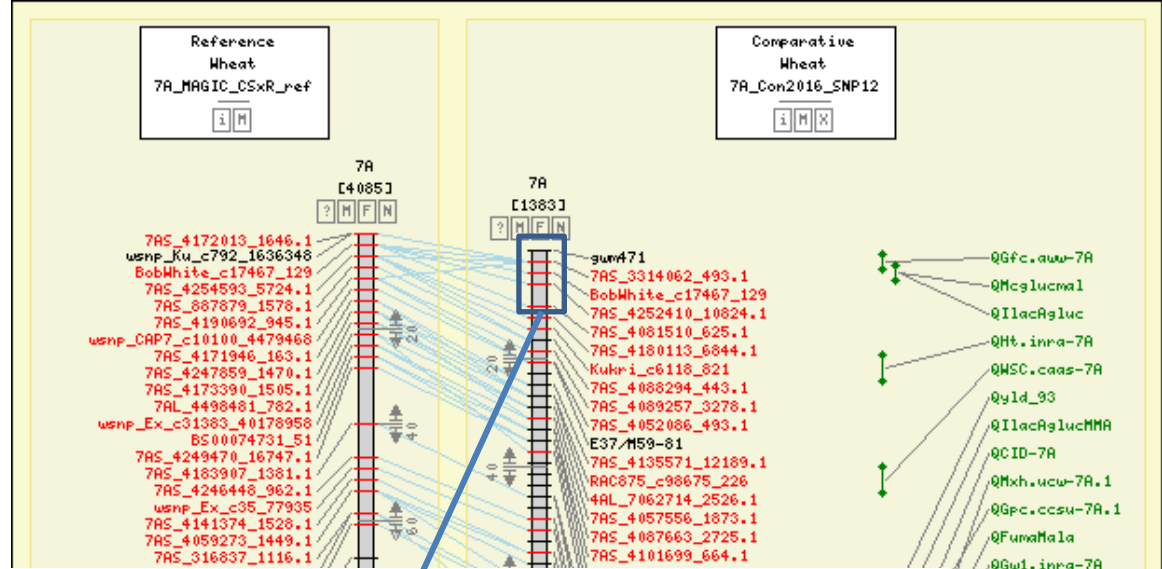
The composite map summarizes a large data set of legacy data assigning chromosome regions that control variation in phenotypic traits.

Wheat Gene Catalogue is a major source



Specify region of interest.

Link to gene models in that part of the genome



File Gene Sets View Tools Help / feedback

Home **View data**

Here you can inspect expression values for the sample groups you have defined. Click on column headers to sort data.

inra-ma:grain inra-ma:leaf inra-ma:root inra-ma:spike inra-ma:stem

Normalized count 1-25 of 324 Show More All Probes New Edit

7A 0 to 11 View all

Probe	POPSEQ dis...	inra-ma:gr...	inra-ma:l...	inra-ma:r...	inra-ma:s...
Traes_7AS_BD45B6DF6.1	7A: 0.0	301.068	100.922	100.125	700.705
Traes_7AS_C28D171B7.1	7A: 0.0	235.096			
Traes_7AS_3247D97E8.2	7A: 0.0	176.245			
Traes_7AS_EA9D6B6A6.1	7A: 0.0	95.784			
Traes_7AS_04B7C0B3F.1	7A: 0.0	80.683			
Traes_7AS_984BFFAB4.2	7A: 0.0	80.442			
Traes_7AS_BDE4D9057.2	7A: 0.0	47.185			
Traes_7AS_4E5B551E2.1	7A: 0.0	44.176			
Traes_7AS_F42F7C775.1	7A: 0.0	42.038			
Traes_7AS_EEC2FEA00.1	7A: 0.0	41.731			
Traes_7AS_BEA6C0FBA.1	7A: 0.0	37.999			
Traes_7AS_6F0ED538C.1	7A: 0.0	32.87			

Charts

Expression vs Time, fixed Treatment: Control

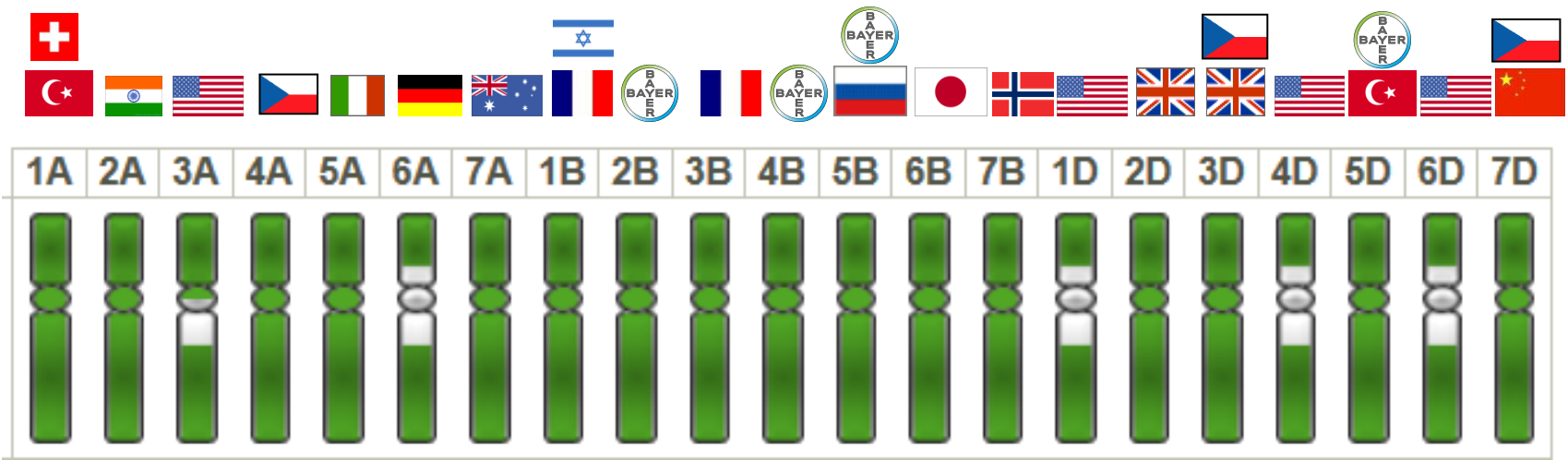
Control :ChineseSpring

Download

Tritigate

Assignment of genes/traits in Wheat Gene Catalogue (RA McIntosh) and in-house breeding programs to reference genome sequence is a major deliverable from the wheat genome sequencing effort.

This needs a coordinated effort and use of the IWGSC network



Resource such as CropStore developed for Brassica provides as a model for building the genome-breeder/research interface

Wheat Genomic Resources in a post-reference sequence era: a road map to establish priorities, timelines and outcomes

Prospects in Wheat Genomics --- what is needed from a research perspective

- **Background on IWGSC focused on community survey carried out in 2016**
- **Completion of the reference genome assembly for all the 21 chromosomes**
- **Assignment of genes/traits in Wheat Gene Catalogue (RA McIntosh) to reference genome sequence**
- **Manual annotation of the 21 chromosomes within the reference genome to include functional studies.**

Manual annotation of the 21 chromosomes within the reference genome to include functional studies is integral to a project that assigns traits to genome regions.

“Incorrect annotations poison every experiment that makes use of them.”

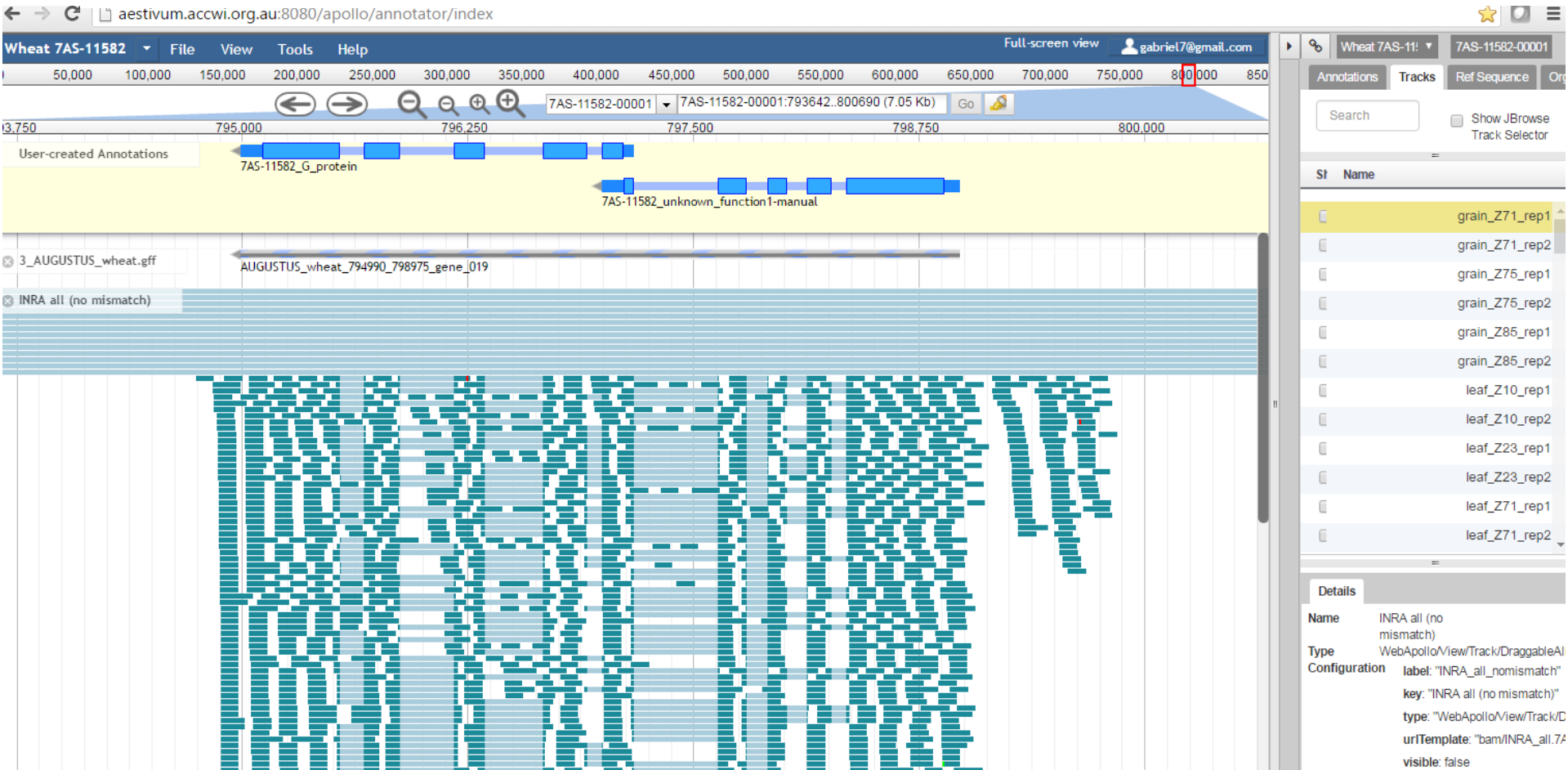
[Yandell and Ence \(2012\).](#)

i5K project: <https://i5k.nal.usda.gov/manual-curation-overview>

Preferred option (supported by i5K project):

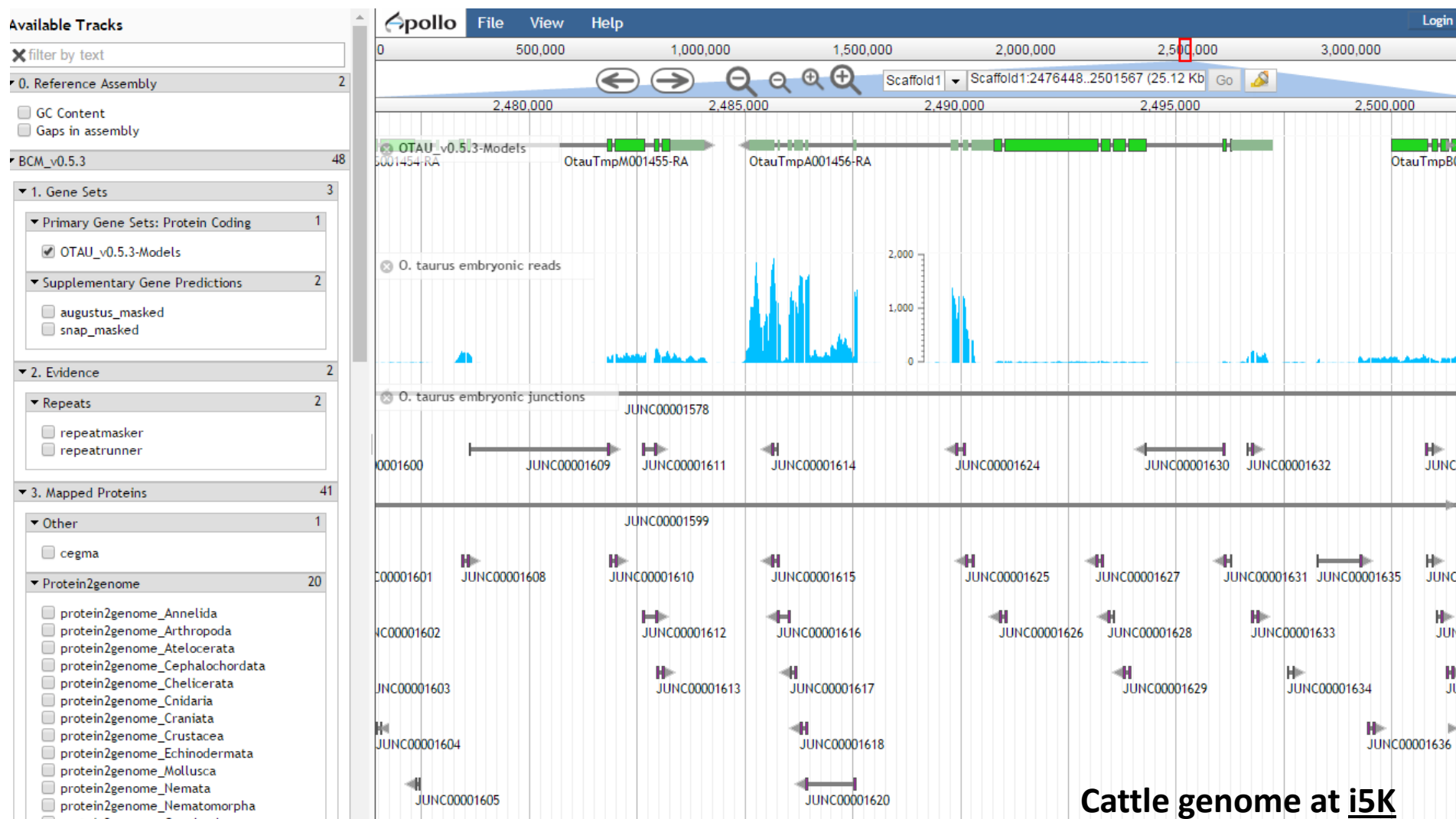
The genome community generate an OGS (Official Gene Set) from the manually curated genes and a single, ‘primary’ set of computationally predicted genes.

Manual annotation of the 21 chromosomes within the reference genome to include functional studies is integral to a project that assigns traits to genome regions.



The screenshot shows the Apollo genome annotation interface. The main window displays a genomic track for Wheat 7AS-11582, with a scale from 50,000 to 850,000. A specific region is highlighted, showing coordinates 795,000 to 800,000. The track includes user-created annotations for '7AS-11582_G_protein' and '7AS-11582_unknown_function1-manual'. Below this, the AUGUSTUS gene prediction track shows 'AUGUSTUS_wheat_794990_798975_gene_019'. The INRA all (no mismatch) track is also visible. The right sidebar shows a list of tracks, including 'grain_Z71_rep1' through 'leaf_Z71_rep2'. The details panel for the selected track shows the name 'INRA all (no mismatch)', type 'WebApolloViewTrack/DruggableAI', and configuration options like 'label: "INRA_all_nomismatch"', 'key: "INRA all (no mismatch)"', 'type: "WebApolloViewTrack/D...', 'urlTemplate: "bam/INRA_all.7...', and 'visible: false'.

Manual annotation of the 21 chromosomes within the reference genome to include functional studies is integral to a project that assigns traits to genome regions.



Cattle genome at 5K

Prospects in Wheat Genomics --- what is needed from a research perspective

- Background on IWGSC focused on community survey carried out in 2016

The IWGSC has a well established network of collaborators covering most of the wheat wheat community globally

- Completion of the reference genome assembly for all the 21 chromosomes

It is feasible to produced a finished genome sequence using multiple datasets, as a base for re-sequencing other wheat varieties

- Assignment of genes/traits in Wheat Gene Catalogue (RA McIntosh) to reference genome sequence

It is feasible through a coordinated, chromosome-based effort to establish a breeder/research/industry interface for gaining maximum value from the reference genome

- Manual annotation of the 21 chromosomes within the reference genome to include functional studies.

Manual annotation raises issues that require functional studies and the experience from other genome project is that it needs a well-organized coordinated effort