Sequence-based assembly of chromosome 7A and comparison to diploid progenitors

Gabriel Keeble-Gagnere[1]

J Nystrom-Persson[1], **C Cavanagh**[2], **D Fleury**[3], **H Webster**[1], **R Appels**[1]

[1] Veterinary and Life Sciences, Murdoch University, South Street, Perth, WA 6150
[2] CSIRO Plant Industry, Clunies Ross Street, Black Mountain ACT 2601
[3] Australian Centre for Plant Functional Genomics, Hartley Grove, Urrbrae, Glen Osmond SA 5064

Talk outline

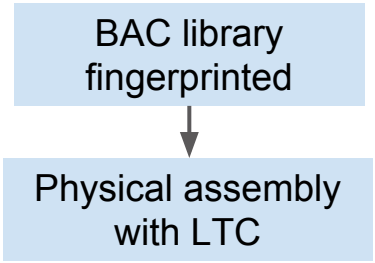
- Assembly of chromosome 7A
- Assembly results and next steps
- Post-genomics on chromosome 7A

Reference-level assembly of 7A

BAC library fingerprinted

Mingcheng Luo, UC Davis

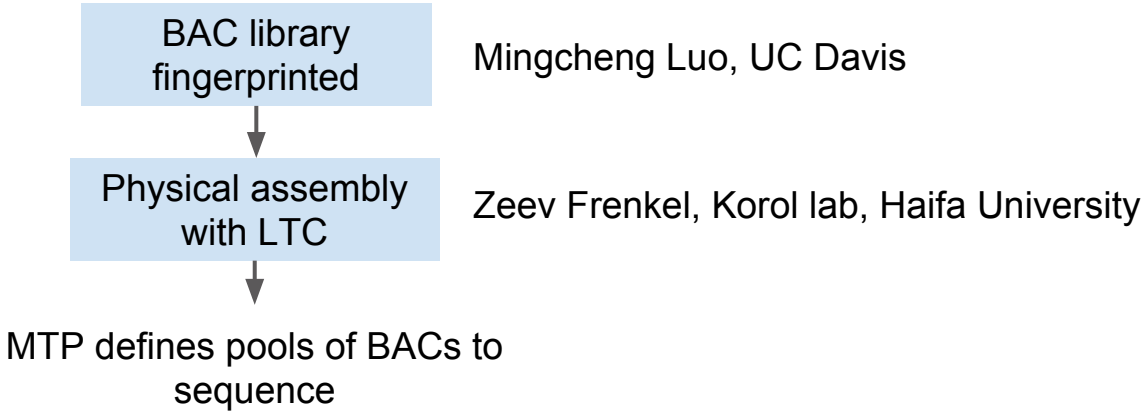
Reference-level assembly of 7A



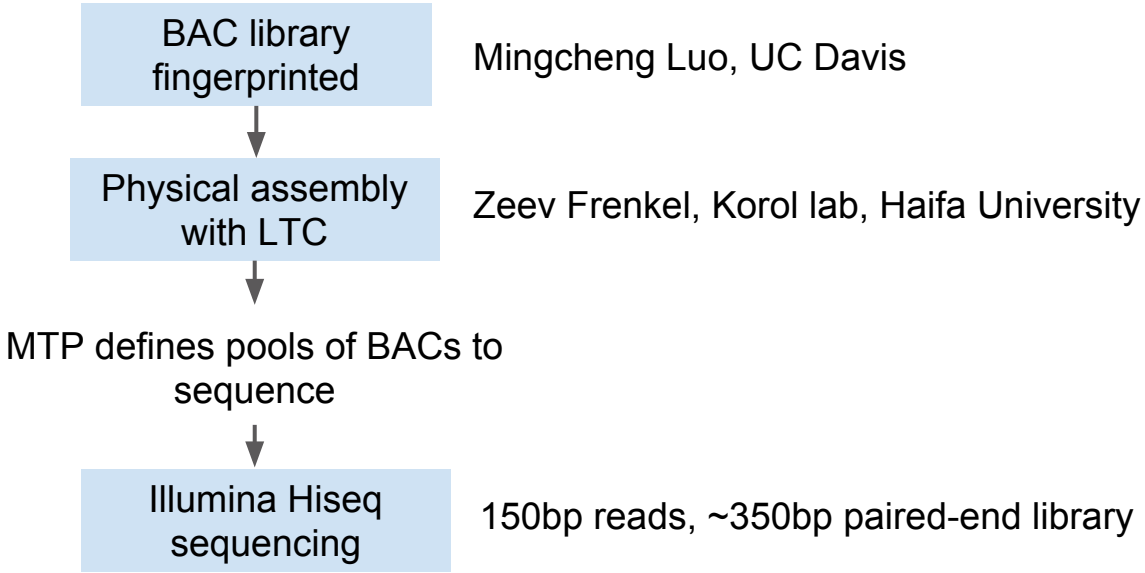
Mingcheng Luo, UC Davis

Zeev Frenkel, Korol lab, Haifa University

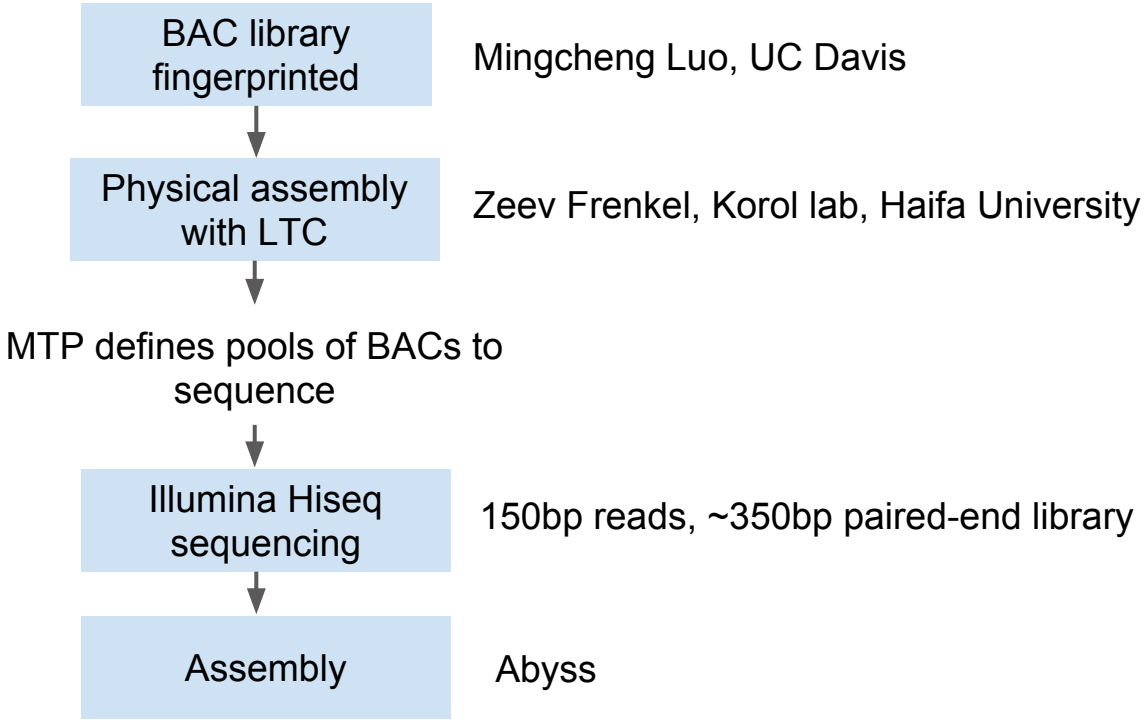
Reference-level assembly of 7A



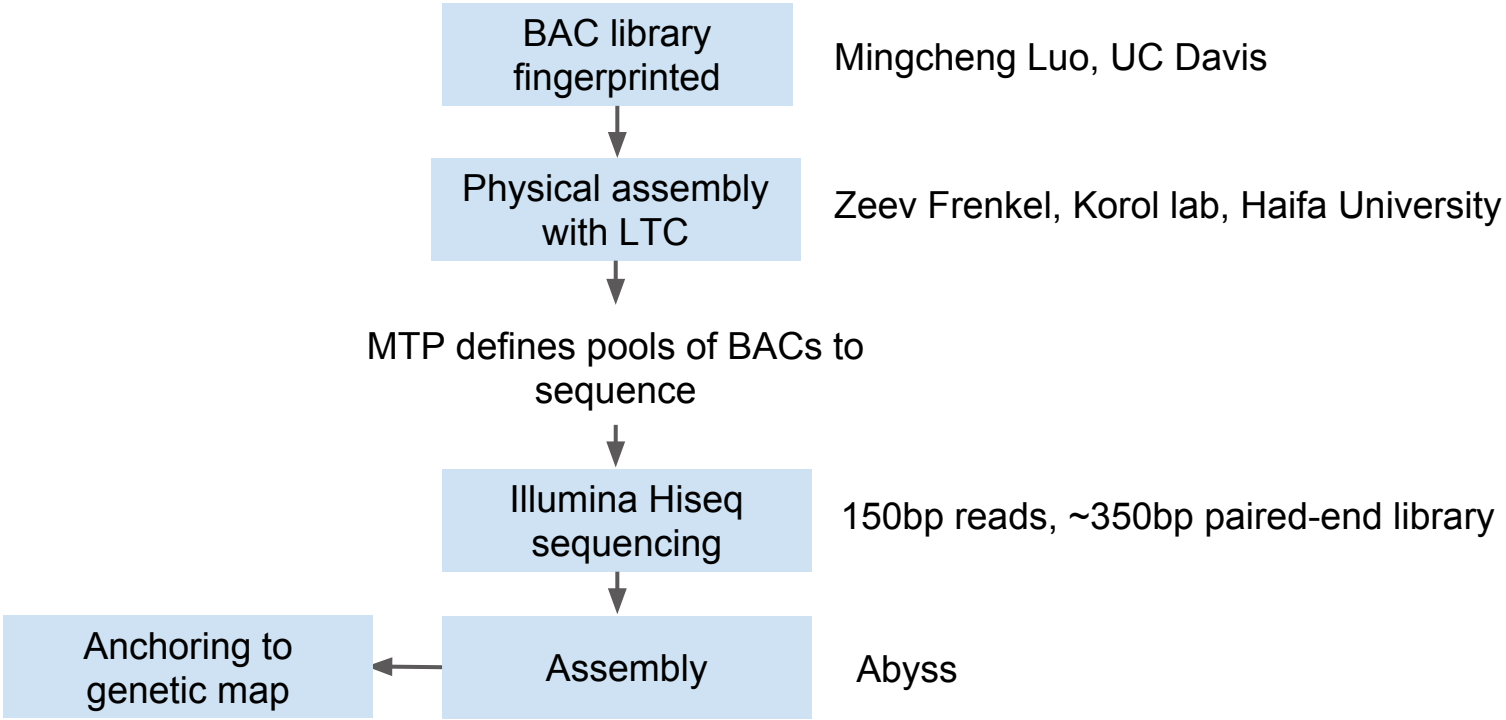
Reference-level assembly of 7A



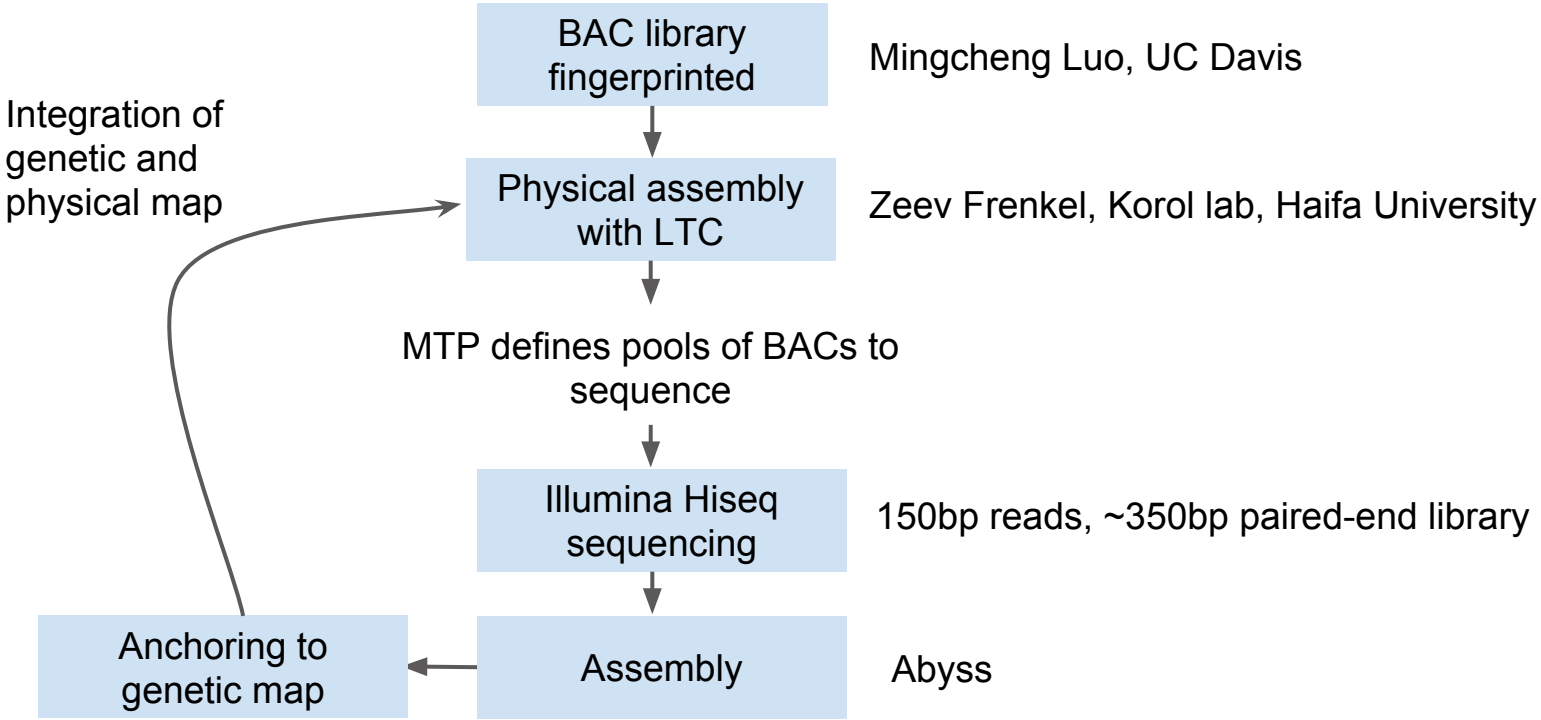
Reference-level assembly of 7A



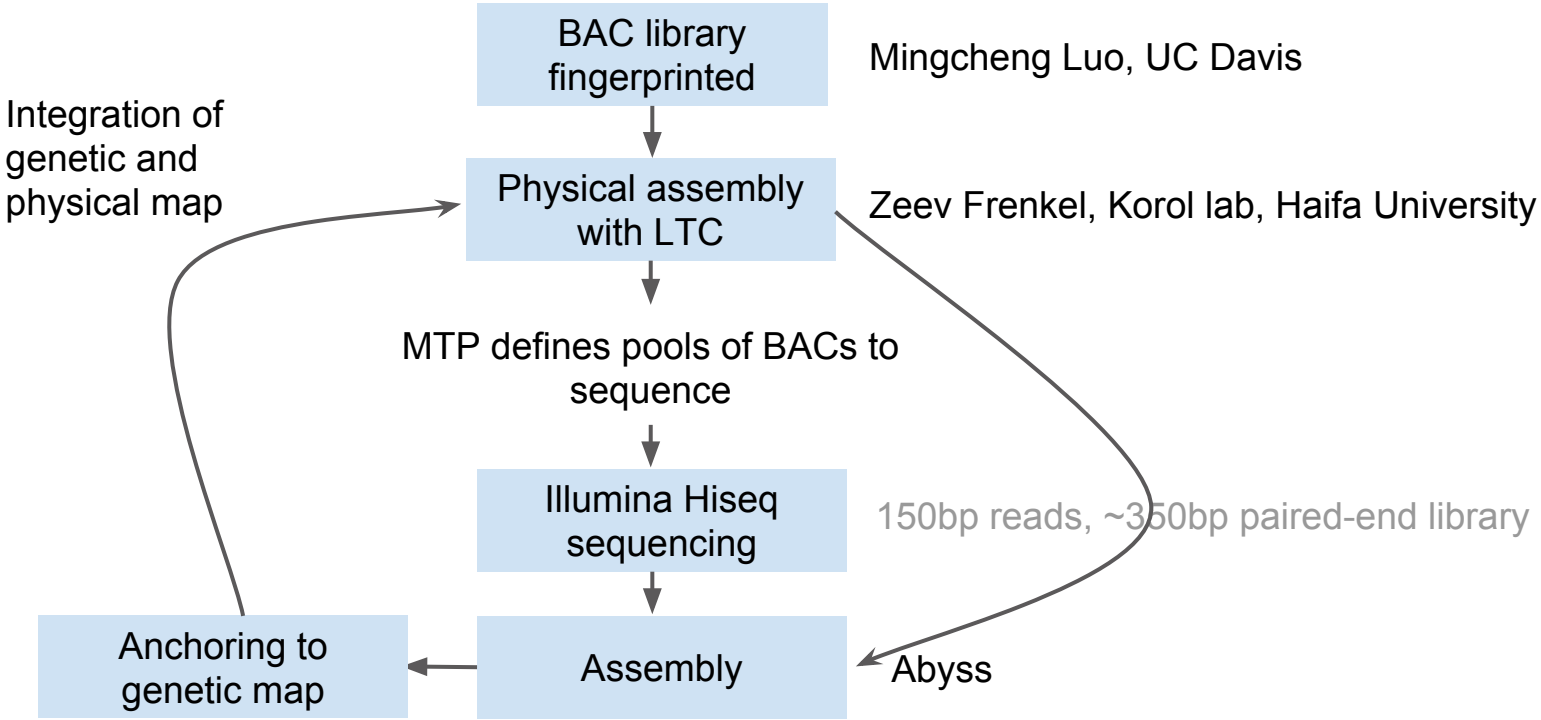
Reference-level assembly of 7A



Reference-level assembly of 7A



Reference-level assembly of 7A



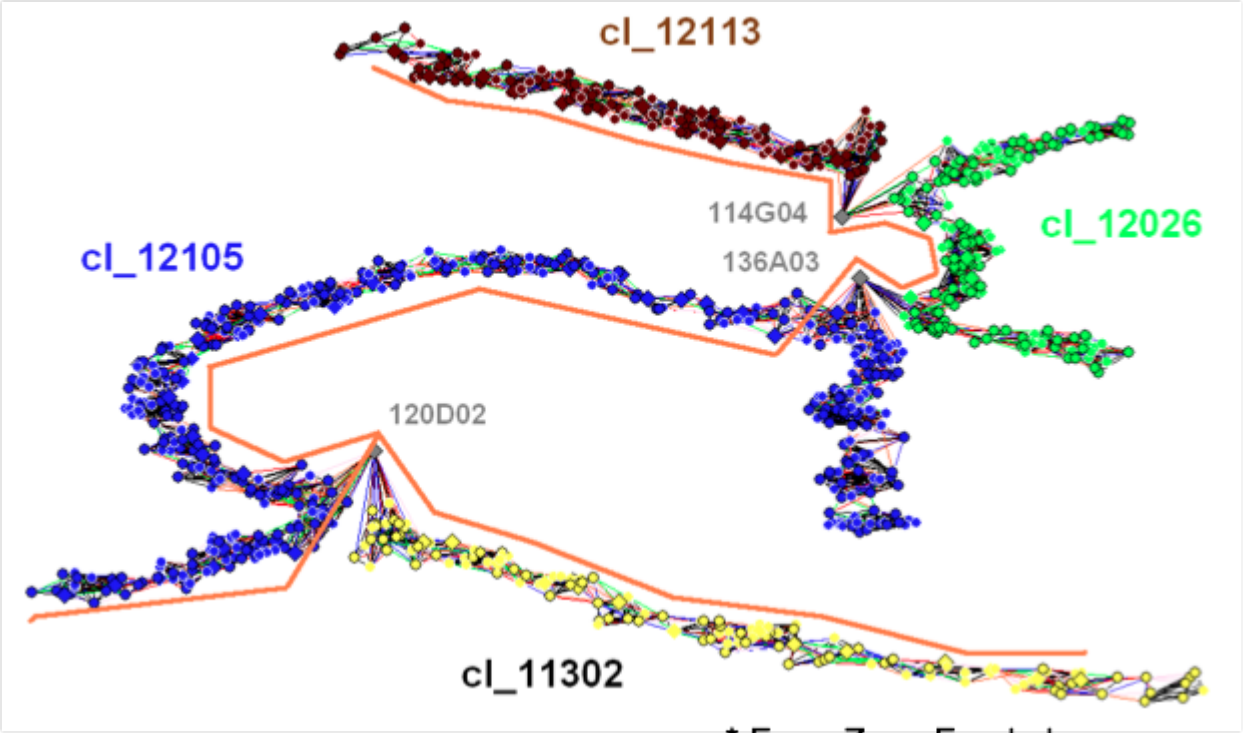
Physical map assembly

Arm	# contigs > 5 clones	Contig N50	Contig L50	Clones in MTP	Estimated total length
7AS	380	1.38Mb	81/299	5,280	353Mb
7AL	352	1.70Mb	64/288	5,832	402Mb

- 11,012 BACs in MTP
- 732 physical contigs (BAC pools) to sequence
  - Barcoded 96 pools per lane

LTC vs. FPC

A single, large physical contig from FPC assembly revealed to be 4 separate contigs joined by single clones (likely contaminated wells)



* From Zeev Frenkel

Chromosome 7A sequence assembly summary

Arm	# Scaffolds (bp)	Mean (bp)	N50 (bp)	Max scaffold (bp)	Total assembly length (Mb)
7AS	33,541	11,704	26,896	264,183	392
7AL	38,731	11,674	27,032	274,707	452
Both	72,272	11,688	26,953	274,707	844

- All 732 BAC pools (physical contigs) have now been sequenced
  - One Illumina run only pooled half the BACs for each pool - has been re-sequenced; analysis is underway

MAGIC

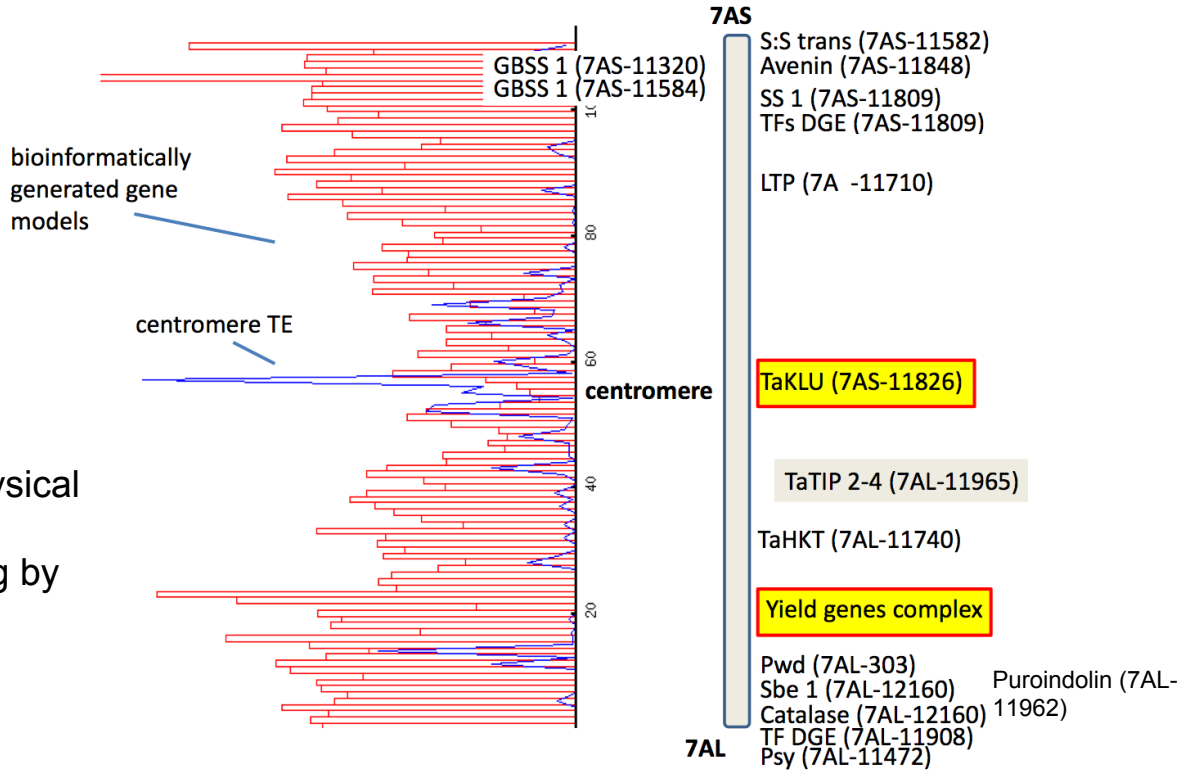
Multi-parent Advanced Generation Inter-Cross, designed to get around limitations of double-haploid populations for mapping traits.

Applied to crops for the first time by Colin Cavanagh at CSIRO.

- 8-way cross
  - Baxter, Yitpi and Westonia (Australia)
  - AC Barrie (Canada), Alsen (US), Pastor (Mexico), Xiaoyan 54 (China), Volcani (Israel)
- 5,000 lines
- GBS sequencing of 4,800 markers in population in 980 lines (Matt Hayden, Victoria DEPI)

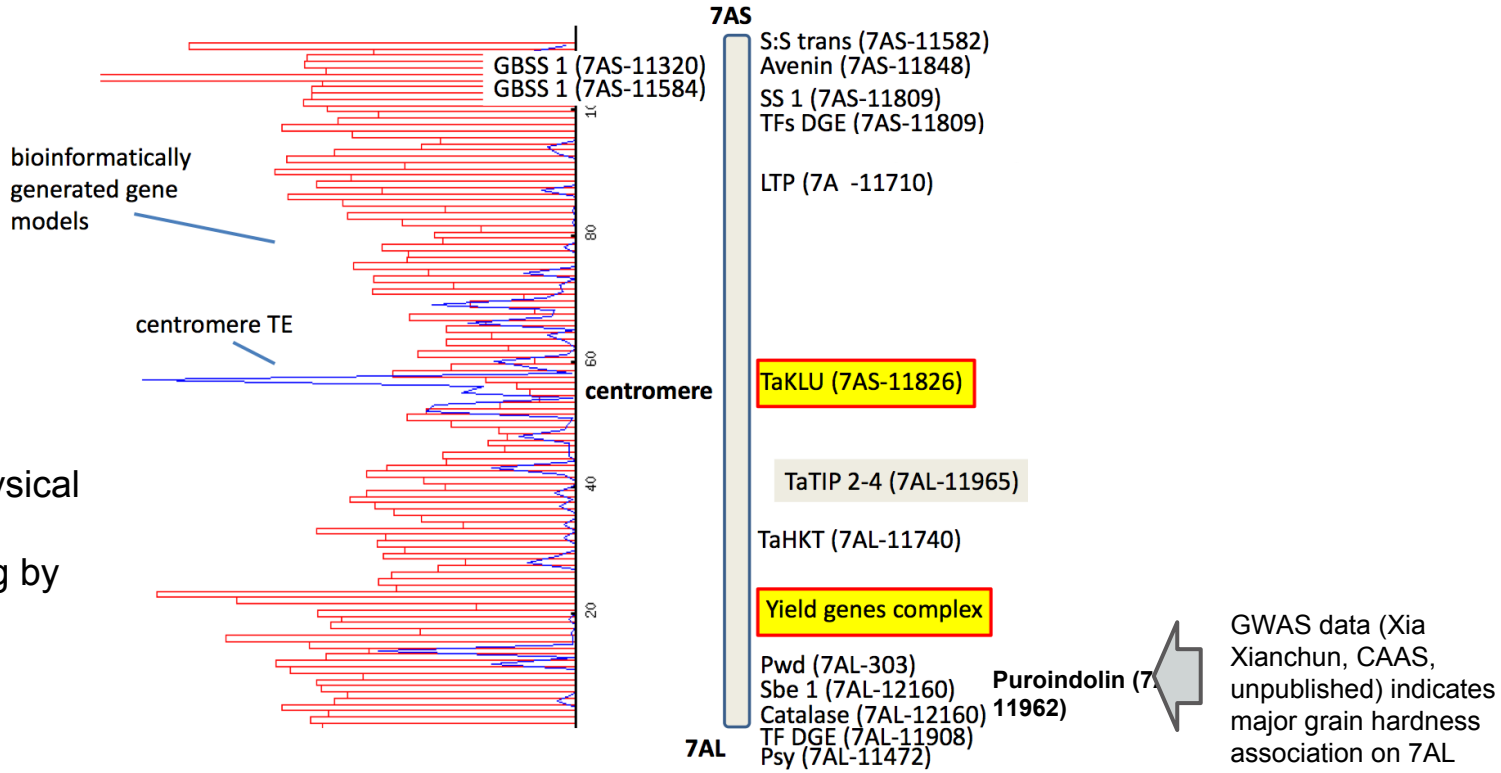
7A chromosome structure based on MAGIC

Approx. 75% of physical contigs included in preliminary ordering by MAGIC

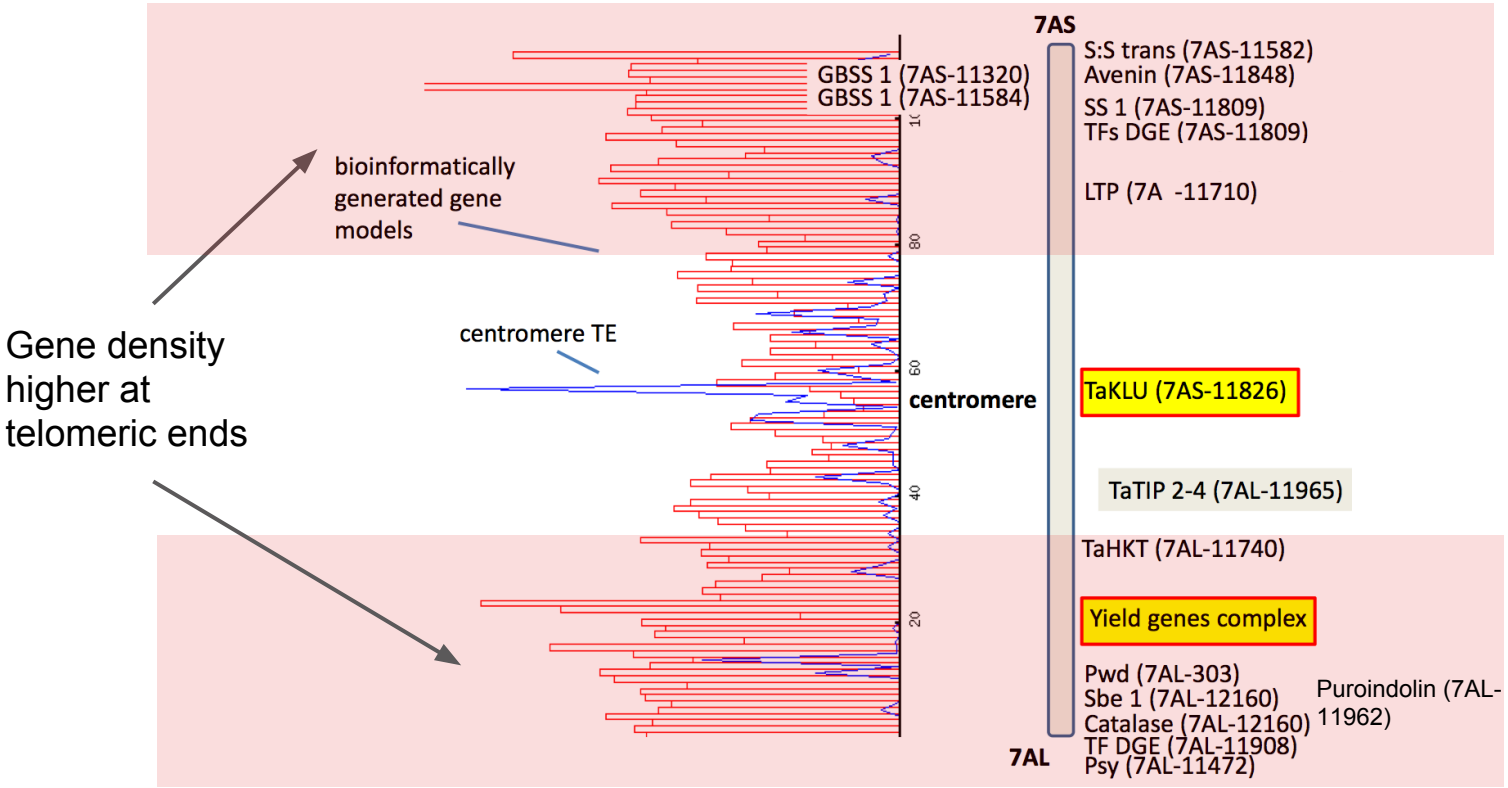


7A chromosome structure based on MAGIC

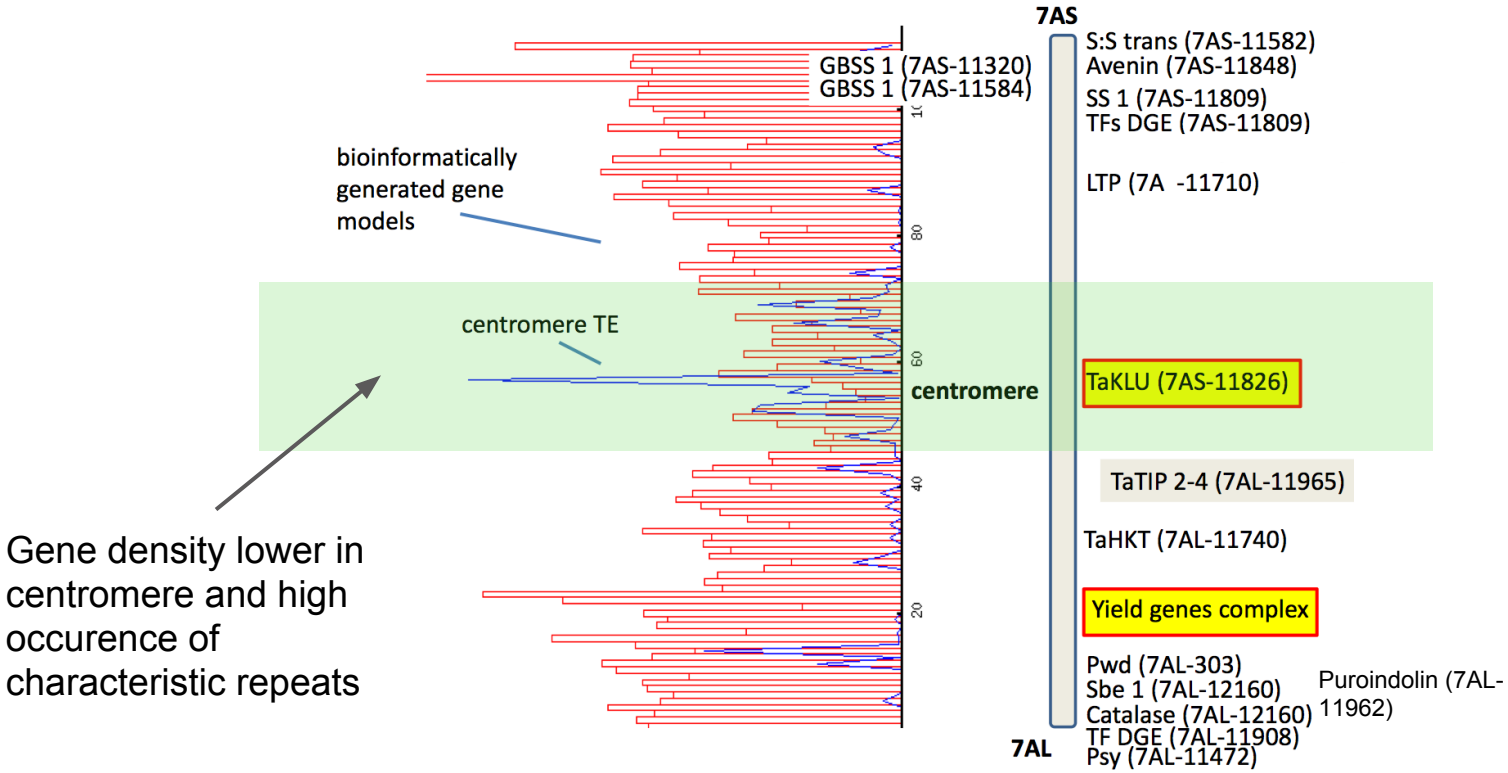
Approx. 75% of physical contigs included in preliminary ordering by MAGIC



7A chromosome structure based on MAGIC



7A chromosome structure based on MAGIC



CS x Renan population

The CS x Renan population (developed at INRA, Clermont-Ferrand) has 276 lines mapped with over 5,000 markers on chromosome 7A:

- This is being used to build a pseudo-molecule for 7A to use as a reference because the CS x Renan population was also used to build a reference pseudo-molecule for chromosome 3B (Choulet et al, Science, under submission)
- The 330 CS x Renan bins from this 7A pseudo-molecule are currently the targets for refinement using the MAGIC 7A mapping data
- The BAC pools (= MTP contigs) in the CS x Renan bins are also being re-examined in light of the sequence data using LTC (Zeev Frenkel, Korol lab, Haifa University)

Gene annotation of genome assembly to date

A collaborative effort to annotate the 7A genome assembly has been established:

- Philippe Leroy and INRA group (TriAnnot) at Clermont-Ferrand (France)
- Francisco Camara group (geneID), CRG (Spain)
- Angela Juhasz (Martonvásár, Hungary)
- Colleagues in Adelaide (ACPFG, Delphine Fleury, Diane Mather, Ute Baumann), Canberra (CSIRO, Jen Taylor) and Perth (Murdoch University, Michael Francki, Shahid Islam)

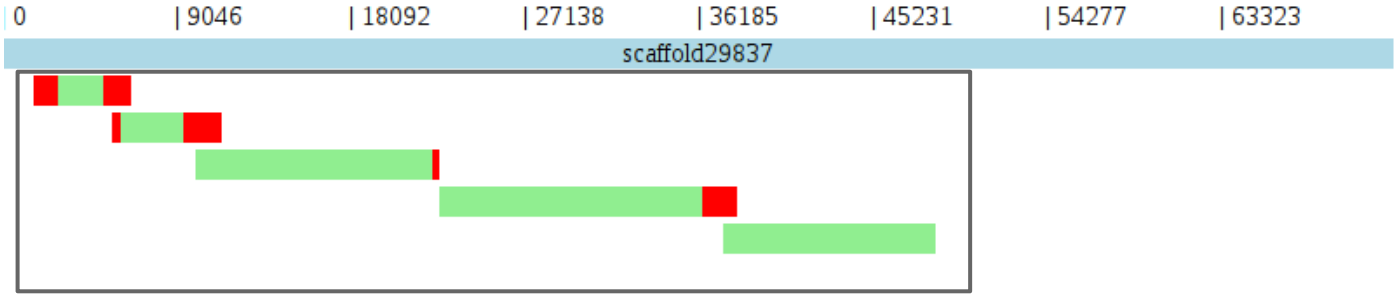
Using *T. urartu* sequence to guide scaffolding

Whole-genome shotgun assembly of *T. urartu* used 8 library sizes (200bp-20kb) for scaffolding (Ling et al, 2013).

We found we can use the scaffolds from *T. urartu* to order our scaffolds from 7A:

- First, align our scaffolds to *T. urartu* scaffolds
- Use alignment to determine potential ordering of our own scaffolds
- Check our own mate-pair data for evidence supporting the join

Using *T. urartu* sequence to guide scaffolding

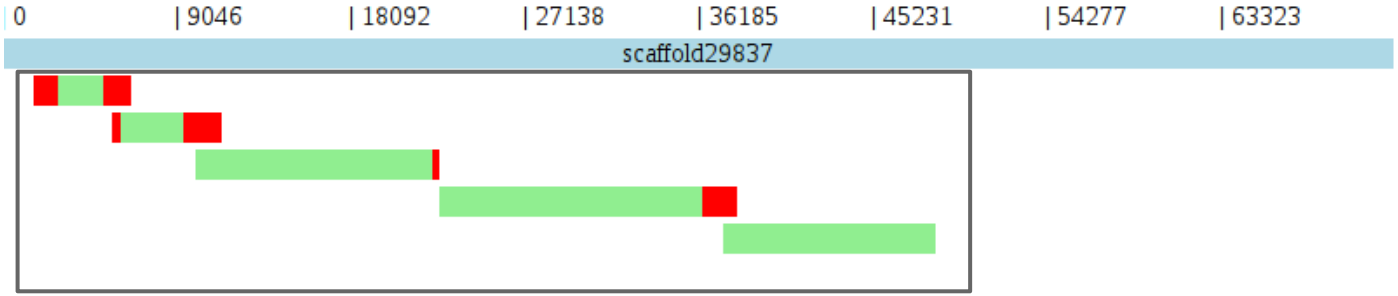


T. urartu scaffold

Wheat 7A scaffolds

Query	QS	QE	SS	SE	Comm
7AS-11582.draft080314.scaffold0227	1282	3591	3052	5360	MPs: 2
7AS-11582.draft080314.scaffold0062	430	3361	6270	9575	MPs: 2
7AS-11582.draft080314.scaffold0019	1	12186	10152	22464	MPs: 3
7AS-11582.draft080314.scaffold0012	1	15479	22854	36495	MPs: 6
7AS-11582.draft080314.scaffold0248	1	10364	37602	48668	MPs: 0

Using *T. urartu* sequence to guide scaffolding



T. urartu scaffold

Wheat 7A scaffolds

Query	QS	QE	SS	SE	Comm
7AS-11582.draft080314.scaffold0227	1282	3591	3052	5360	MPs: 2
7AS-11582.draft080314.scaffold0062	430	3361	6270	9575	MPs: 2
7AS-11582.draft080314.scaffold0019	1	12186	10152	22464	MPs: 3
7AS-11582.draft080314.scaffold0012	1	15479	22854	36495	MPs: 6
7AS-11582.draft080314.scaffold0248	1	10364	37602	48668	MPs: 0

Mate-pairs from wheat required confirmed join

Post-genomics on chromosome 7A

From genome to metabolome:

- Transcriptome (RNA-seq)
  - Hollie Webster PhD. thesis, ACCWI
- Proteome (8-plex iTRAQ)
  - Shahid Islam, ACCWI
- Metabolomics
  - Camilla Hill, Ute Roessner (Melbourne University)

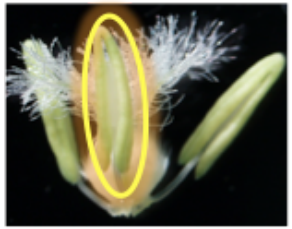
RNA-seq study (Hollie Webster)

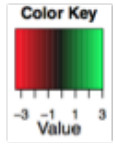
Investigating the effects of drought on the developing head, in two double haploid lines, D02-105 (drought intolerant) and D08-299 (drought tolerant).

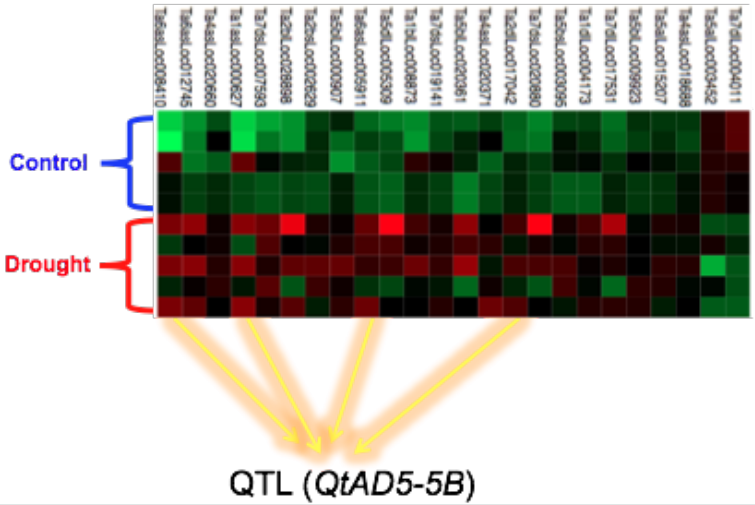
- Two years: 2011, 2012
- Two varieties: D02-105, D08-299 (plus Westonia, Kauz in 2011)
- Four timepoints: AR05, AR10, AR15, FHE
- Two experimental factors: Control, Drought
- Between 3 and 6 biological replicates per sample, and up to 3 technical replicates per biological replicate
    - After filtering, around 200 technical replicates in total (for 2012 experiment)

RNA-seq study (Hollie Webster)

- A total of 230 anther-specific rice genes identified by Deveshwar et al (2011) could be identified in our RNA-Seq data from the developing spike of wheat
- 24 of these genes were differentially expressed in response to water deficit early in spike development
- 4 of these 24 genes were located in a small region on 5BL that also defines a QTL for the timing for the start of head development.
- These genes are currently under further investigation



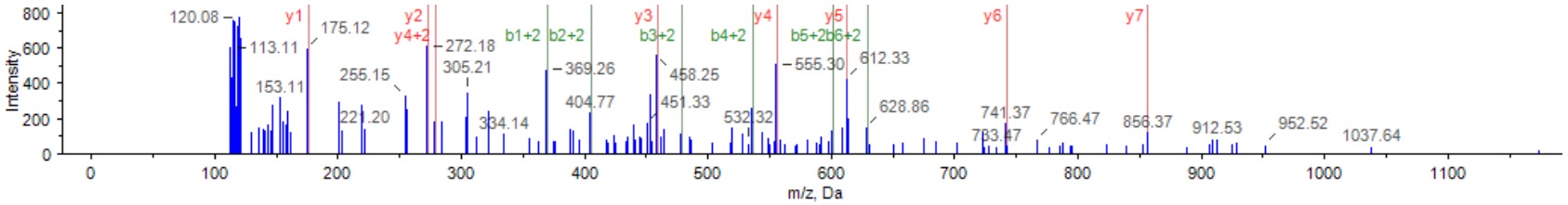




iTRAQ proteomics (Shahid Islam)

18, 8-plex, iTRAQ experiments which replicates the experimental design of Hollie Webster's RNA-seq study, plus two extra time points (7- and 30-days post-anthesis).

- Largest shotgun proteomics study in wheat
- Approximately 1,500 peptides per experiment



Metabolomics

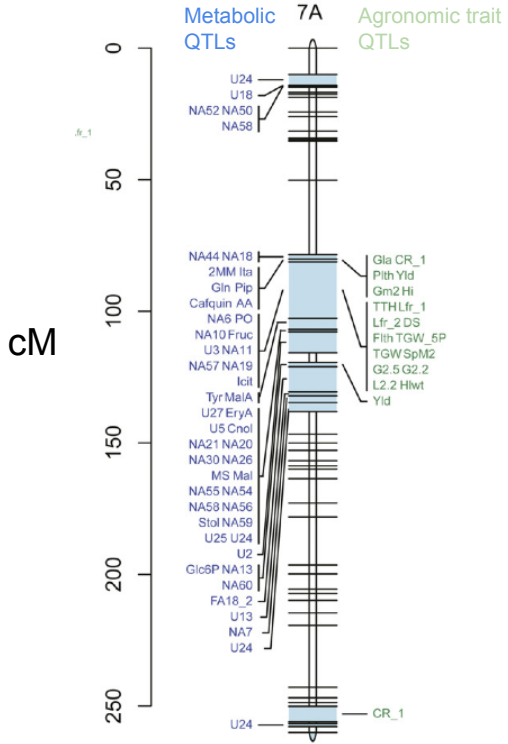
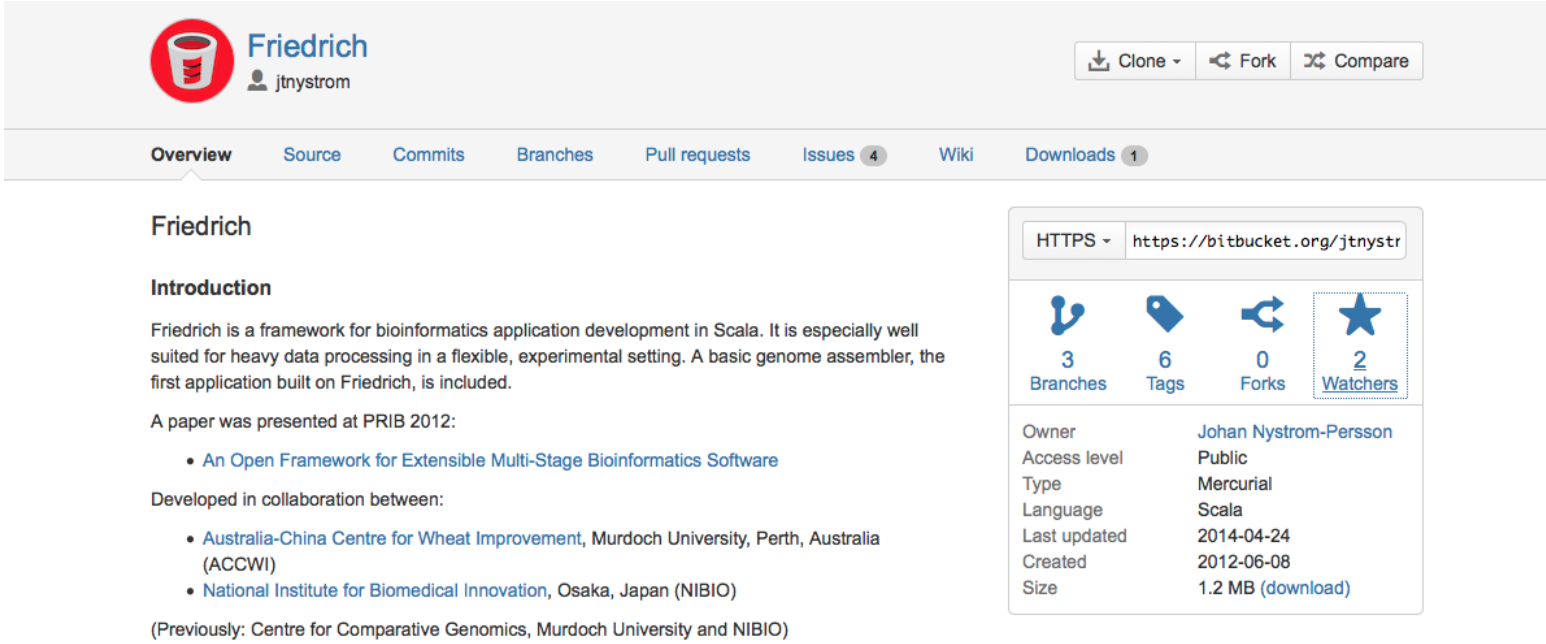


Figure 4 from Hill et al. 2013, *Plant Physiology*

- Metabolomics work carried out by Camilla Hill in Ute Roessner lab, Melbourne University
- GC-MS study of drought stress in Excalibur/Kukri double haploid population published in 2013 showed highest number of metabolic QTLs (mQTLs) mapped to chromosome 7A
- LC-MS study manuscript in preparation
  - High number of mQTLs on chromosome 7A
  - First mapping of mQTL to genomic sequence in wheat

Friedrich



https://bitbucket.org/jtnystrom/friedrich/
Open source, under intensive development...

Thanks

Funding:

- Bioplatforms Australia (BPA)
- Grain Research Development Corporation (GRDC)

Collaborators:

- Australian Genome Research Facility (AGRF)
- Department of Environment and Primary Industries, Melbourne (DEPI)
  - Matt Hayden, Josquin Tibbits
- CSIRO
- INRA

ACCWI software developer: Johan Nystrom-Persson