Reference Assembly of Chromosome 7A as a Platform to Study Regions of Agronomic Importance

Gabriel Keeble-Gagnere, Murdoch University
Acknowledgments

**Funding**
Grains Research Development Corporation
Bioplatforms Australia

**ACCWI group**
Rudi Appels, Hollie Webster, Shahidul Islam, Xueyan Chen, Yingjun Zhang, Johan Nystrom-Persson

**Flow-sorting DNA/BAC library construction**
Jaroslav Dolezel, Hana Simkova
Institute of Experimental Botany
Czech Republic

**Fingerprinting BAC library**
Mingcheng Luo group
UC Davis

**Physical map assembly**
Zeev Frenkel, Ambraham Korol
Haifa University

**Genetic maps**
MAGIC: Colin Cavanagh, Emma Huang, Jen Taylor (CSIRO)
MAGIC GBS: Matt Hayden (DEPI)
CSxRenan: Pierre Sourdille, Benoit Darrier (INRA)

*T. monococcum* genetic map
Population: Jorge Dubcovsky
90k chip: Matt Hayden, Kerrie Forrest

**DNA sequencing**
Matt Tinning
AGRF

**Annotation**
TriAnnot: Philippe Leroy, Aurelien Bernard (INRA)
geneID (CRG): Francisco Camara, Anna Vlasova (CRG, Spain), Juan Carlos Sanchez (ACPFG)
Storage proteins: Angela Juhasz (Hungary)
QTL mapping/Significant genome regions: Delphine Fleury (ACPFG)
Specific genes: Hui-xian Zhao (NW A&F Uni, China)

**Pseudomolecule**
Fred Choulet, Etienne Paux
INRA

**7A mate-pair sequencing of amplified DNA**
Matt Hayden, Josquin Tibbits, Sami Hakim
DEPI

**Whole-genome mate-pair data**
Andy Sharpe, David Konkin, Curtis Pozniak
NRC, Canada

**Bionano map**
Jaroslav Dolezel, Hana Simkova, Mingcheng Luo

**Supercomputing resources**
iVEC/Pawsey Supercomputing Centre
1. We have produced a high quality, genetically anchored, assembly of chromosome 7A

2. The assembly has been validated using independent genome-level information for specific regions of the chromosome

3. The assembly now forms the basis for the analysis of agronomically significant chromosome regions
Reference-level assembly of 7A

Flow-sorted DNA  Dolezel lab, Czech Republic
Reference-level assembly of 7A

Flowsorted DNA → Dolezel lab, Czech Republic

BAC library fingerprinted → Mingcheng Luo, UC Davis
Reference-level assembly of 7A

Flow-sorted DNA

Dolezel lab, Czech Republic

BAC library fingerprinted

Mingcheng Luo, UC Davis

Physical assembly with LTC

Zeev Frenkel, Korol lab, Haifa University
Reference-level assembly of 7A

Flow-sorted DNA

BAC library fingerprinted

Physical assembly with LTC

MTP defines pools of BACs to sequence

Dolezel lab, Czech Republic

Mingcheng Luo, UC Davis

Zeev Frenkel, Korol lab, Haifa University
Reference-level assembly of 7A

Flow-sorted DNA  Dolezel lab, Czech Republic
BAC library fingerprinted  Mingcheng Luo, UC Davis
Physical assembly with LTC  Zeev Frenkel, Korol lab, Haifa University
MTP defines pools of BACs to sequence
Illumina Hiseq sequencing  150bp reads, ~350bp paired-end library
Reference-level assembly of 7A

Flow-sorted DNA

BAC library fingerprinted

Physical assembly with LTC

MTP defines pools of BACs to sequence

Illumina Hiseq sequencing

Assembly

Dolezel lab, Czech Republic

Mingcheng Luo, UC Davis

Zeev Frenkel, Korol lab, Haifa University

150bp reads, ~350bp paired-end library

Abyss
Reference-level assembly of 7A

Flow-sorted DNA

BAC library fingerprinted

Mingcheng Luo, UC Davis

Physical assembly with LTC

Zeev Frenkel, Korol lab, Haifa University

MTP defines pools of BACs to sequence

Illumina Hiseq sequencing

150bp reads, ~350bp paired-end library

Anchoring to genetic map

Assembly

Abyss

Dolezel lab, Czech Republic
Reference-level assembly of 7A

Flow-sorted DNA

BAC library fingerprinted

Physical assembly with LTC

MTP defines pools of BACs to sequence

Illumina Hiseq sequencing

150bp reads, ~350bp paired-end library

Assembly

Anchoring to genetic map

Integration of genetic and physical map

Dolezel lab, Czech Republic

Mingcheng Luo, UC Davis

Zeev Frenkel, Korol lab, Haifa University

Abyss

Integration of genetic and physical map
Reference-level assembly of 7A

Flow-sorted DNA

BAC library fingerprinted

Physical assembly with LTC

MTP defines pools of BACs to sequence

Illumina Hiseq sequencing

Assembly

Anchoring to genetic map

Dolezel lab, Czech Republic

Mingcheng Luo, UC Davis

Zeev Frenkel, Korol lab, Haifa University

150bp reads, ~350bp paired-end library

Abyss

Integration of genetic and physical map
Assembly summary

- High-density composite genetic map based on MAGIC (CSIRO) using Chinese Spring x Renan (INRA) map as anchor
  - Over 4,000 markers on 7A
Assembly summary

- High-density composite genetic map based on MAGIC using CSxRenan map as anchor
  - Over 4000 markers
- 732 physical contigs reduced to 316 scaffolds
- 676 physical contigs (92%) anchored via scaffolded physical map

* Screenshots from LTC, Zeev Frenkel
Assembly summary

- High-density composite genetic map based on MAGIC using CSxRenan map as anchor
  - Over 4000 markers
- 732 physical contigs reduced to 316 scaffolds
- 676 physical contigs (92%) anchored via scaffolded physical map

Nodes are individual BACs, edges are overlaps based on Sulston score (colour indicates confidence)

7AL-11542
Assembly summary

- High-density composite genetic map based on MAGIC using CSxRenan map as anchor
  - Over 4000 markers
- 732 physical contigs reduced to 316 scaffolds
- 676 physical contigs (92%) anchored via scaffolded physical map
Super-scaffolding

Final stats for paired-end-only (pre-mate-pair) assembly:

- 42,441 sequence scaffolds
  - Total length 940Mb
  - N50 137kb
  - Mean 22kb

A large mate-pair dataset was generated by National Research Council, Canada (Andy Sharpe) from a Chinese Spring+7EL line, including 12 insert library sizes from 1.4kb to 20kb.

The read pairs aligning perfectly (no mismatches) to our paired-end-only draft assembly were provided by David Konkin and used for super-scaffolding with SSPACE.

The minimum number of mate-pair joins required to connect two contigs (k) was explored, using k = 2 to 5.

For example, for k = 2, two scaffolds can be joined based on only two connections.
Two scaffolding approaches were explored:

1) **Chromosome-arm level scaffolding**

<table>
<thead>
<tr>
<th>k</th>
<th># Scaffolds</th>
<th>Median (bp)</th>
<th>Mean (bp)</th>
<th>N50 (bp)</th>
<th>Max scaffold (bp)</th>
<th>Total length (bp)</th>
<th>% cross-pool joins</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>23,342</td>
<td>4,732</td>
<td>38,839</td>
<td>350,507</td>
<td>2,814,297</td>
<td>906.5e6</td>
<td>3.9</td>
</tr>
<tr>
<td>3</td>
<td>27,569</td>
<td>3,941</td>
<td>32,704</td>
<td>289,304</td>
<td>2,148,657</td>
<td>904.5e6</td>
<td>1.6</td>
</tr>
<tr>
<td>4</td>
<td>30,690</td>
<td>3,631</td>
<td>29,463</td>
<td>249,246</td>
<td>2,127,911</td>
<td>904.2e6</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>33,426</td>
<td>3,449</td>
<td>27,032</td>
<td>214,649</td>
<td>2,117,720</td>
<td>903.5e6</td>
<td>0.7</td>
</tr>
</tbody>
</table>

2) **BAC pool-level scaffolding**

<table>
<thead>
<tr>
<th>k</th>
<th># Scaffolds</th>
<th>Median (bp)</th>
<th>Mean (bp)</th>
<th>N50 (bp)</th>
<th>Max scaffold (bp)</th>
<th>Total length (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>12,043</td>
<td>5,024</td>
<td>75,216</td>
<td>421,553</td>
<td>2,415,588</td>
<td>905.8e6</td>
</tr>
<tr>
<td>3</td>
<td>15,546</td>
<td>3,619</td>
<td>58,172</td>
<td>370,629</td>
<td>2,334,598</td>
<td>904.3e6</td>
</tr>
<tr>
<td>4</td>
<td>18,131</td>
<td>3,094</td>
<td>49,848</td>
<td>339,791</td>
<td>2,852,455</td>
<td>903.8e6</td>
</tr>
<tr>
<td>5</td>
<td>20,416</td>
<td>2,789</td>
<td>44,242</td>
<td>315,060</td>
<td>1,979,523</td>
<td>903.2e6</td>
</tr>
</tbody>
</table>
Two scaffolding approaches were explored:

1) Chromosome-arm level scaffolding

<table>
<thead>
<tr>
<th>k</th>
<th># Scaffolds</th>
<th>Median (bp)</th>
<th>Mean (bp)</th>
<th>N50 (bp)</th>
<th>Max scaffold (bp)</th>
<th>Total length (bp)</th>
<th>% cross-pool joins</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>23,342</td>
<td>4,732</td>
<td>38,839</td>
<td>350,507</td>
<td>2,814,297</td>
<td>906.5e6</td>
<td>3.9</td>
</tr>
<tr>
<td>3</td>
<td>27,659</td>
<td>3,941</td>
<td>32,704</td>
<td>289,304</td>
<td>2,148,657</td>
<td>904.5e6</td>
<td>1.6</td>
</tr>
<tr>
<td>4</td>
<td>30,690</td>
<td>3,631</td>
<td>29,463</td>
<td>249,246</td>
<td>2,127,911</td>
<td>904.2e6</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>33,426</td>
<td>3,449</td>
<td>27,032</td>
<td>214,649</td>
<td>2,117,720</td>
<td>903.5e6</td>
<td>0.7</td>
</tr>
</tbody>
</table>

2) BAC pool-level scaffolding

<table>
<thead>
<tr>
<th>k</th>
<th># Scaffolds</th>
<th>Median (bp)</th>
<th>Mean (bp)</th>
<th>N50 (bp)</th>
<th>Max scaffold (bp)</th>
<th>Total length (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>12,043</td>
<td>5,024</td>
<td>75,216</td>
<td>421,553</td>
<td>2,415,588</td>
<td>905.8e6</td>
</tr>
<tr>
<td>3</td>
<td>15,546</td>
<td>3,619</td>
<td>58,172</td>
<td>370,629</td>
<td>2,334,598</td>
<td>904.3e6</td>
</tr>
<tr>
<td>4</td>
<td>18,131</td>
<td>3,094</td>
<td>49,848</td>
<td>339,791</td>
<td>2,852,455</td>
<td>903.8e6</td>
</tr>
<tr>
<td>5</td>
<td>20,416</td>
<td>2,789</td>
<td>44,242</td>
<td>315,060</td>
<td>1,979,523</td>
<td>903.2e6</td>
</tr>
</tbody>
</table>

Very few scaffolds from different pools are joined.
Two scaffolding approaches were explored:

1) Chromosome-arm level scaffolding

<table>
<thead>
<tr>
<th>k</th>
<th># Scaffolds</th>
<th>Median (bp)</th>
<th>Mean (bp)</th>
<th>N50 (bp)</th>
<th>Max scaffold (bp)</th>
<th>Total length (bp)</th>
<th>% cross-pool joins</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>23,342</td>
<td>4,732</td>
<td>38,839</td>
<td>350,507</td>
<td>2,814,297</td>
<td>906.5e6</td>
<td>3.9</td>
</tr>
<tr>
<td>3</td>
<td>27,659</td>
<td>3,941</td>
<td>32,704</td>
<td>289,304</td>
<td>2,148,657</td>
<td>904.5e6</td>
<td>1.6</td>
</tr>
<tr>
<td>4</td>
<td>30,690</td>
<td>3,631</td>
<td>29,463</td>
<td>249,246</td>
<td>2,127,911</td>
<td>904.2e6</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>33,426</td>
<td>3,449</td>
<td>27,032</td>
<td>214,649</td>
<td>2,117,720</td>
<td>903.5e6</td>
<td>0.7</td>
</tr>
</tbody>
</table>

2) BAC pool-level scaffolding

<table>
<thead>
<tr>
<th>k</th>
<th># Scaffolds</th>
<th>Median (bp)</th>
<th>Mean (bp)</th>
<th>N50 (bp)</th>
<th>Max scaffold (bp)</th>
<th>Total length (bp)</th>
<th>% cross-pool joins</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>12,043</td>
<td>5,024</td>
<td>75,216</td>
<td>421,553</td>
<td>2,415,588</td>
<td>905.8e6</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>15,546</td>
<td>3,619</td>
<td>58,172</td>
<td>370,629</td>
<td>2,334,598</td>
<td>904.3e6</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>18,131</td>
<td>3,094</td>
<td>49,848</td>
<td>339,791</td>
<td>2,852,455</td>
<td>903.8e6</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>20,416</td>
<td>2,789</td>
<td>44,242</td>
<td>315,060</td>
<td>1,979,523</td>
<td>903.2e6</td>
<td></td>
</tr>
</tbody>
</table>
Two scaffolding approaches were explored:

1) **Chromosome-arm level scaffolding**

<table>
<thead>
<tr>
<th>k</th>
<th># Scaffolds</th>
<th>Median (bp)</th>
<th>Mean (bp)</th>
<th>N50 (bp)</th>
<th>Max scaffold (bp)</th>
<th>Total length (bp)</th>
<th>% cross-pool joins</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>23,342</td>
<td>4,732</td>
<td>38,839</td>
<td>350,507</td>
<td>2,814,297</td>
<td>906.5e6</td>
<td>3.9</td>
</tr>
<tr>
<td>3</td>
<td>27,659</td>
<td>3,941</td>
<td>32,704</td>
<td>289,304</td>
<td>2,148,657</td>
<td>904.5e6</td>
<td>1.6</td>
</tr>
<tr>
<td>4</td>
<td>30,690</td>
<td>3,631</td>
<td>29,463</td>
<td>249,246</td>
<td>2,127,911</td>
<td>904.2e6</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>33,426</td>
<td>3,449</td>
<td>27,032</td>
<td>214,649</td>
<td>2,117,720</td>
<td>903.5e6</td>
<td>0.7</td>
</tr>
</tbody>
</table>

2) **BAC pool-level scaffolding**

<table>
<thead>
<tr>
<th>k</th>
<th># Scaffolds</th>
<th>Median (bp)</th>
<th>Mean (bp)</th>
<th>N50 (bp)</th>
<th>Max scaffold (bp)</th>
<th>Total length (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>12,043</td>
<td>5,024</td>
<td>75,216</td>
<td>421,553</td>
<td>2,415,588</td>
<td>905.8e6</td>
</tr>
<tr>
<td>3</td>
<td>15,546</td>
<td>3,619</td>
<td>58,172</td>
<td>370,629</td>
<td>2,334,598</td>
<td>904.3e6</td>
</tr>
<tr>
<td>4</td>
<td>18,131</td>
<td>3,094</td>
<td>49,848</td>
<td>339,791</td>
<td>2,852,455</td>
<td>903.8e6</td>
</tr>
<tr>
<td>5</td>
<td>20,416</td>
<td>2,789</td>
<td>44,242</td>
<td>315,060</td>
<td>1,979,523</td>
<td>903.2e6</td>
</tr>
</tbody>
</table>

*Needs validation, eg: with Bionano maps*
From long- to short-range information

Genetic map

Assembled sequence + annotation

Physical map
Pseudomolecule

Annotation

TriAnnot (Philippe Leroy, INRA)
3897 genes predicted
(1623 “high confidence”, 2274 “low confidence”)

CRG annotation (Francisco Camara group)
24,030 predictions on an earlier draft

Many genes are unique to a particular annotation
Pseudomolecule

Gene density per 10kb (TriAnnot annotation)

TriA (Phi 7256) (3295 “high confidence”, 3961 “low confidence)

CRG annotation (Francisco Camara group)
24,030 predictions on an earlier draft

Many genes are unique to a particular annotation

Fig 1B, Choulet et al. (2014)
CDS/10Mb on chromosome 3B
Pseudomolecule

Annotation

TriAnnot
(Philippe Leroy, INRA)
7256 genes predicted
(3295 “high confidence”, 3961 “low confidence”)

CRG annotation
(Francisco Camara group)
24,030 predictions on an earlier draft

Many genes are unique to a particular annotation
Pseudomolecule

Annotation

TriAnnot
(Philippe Leroy, INRA)
7256 genes predicted
(3295 “high confidence”, 3961 “low confidence”)

CRG annotation
(Francisco Camara group)
24,030 predictions on an earlier draft

Many genes are unique to a particular annotation
Pseudomolecule genes of interest

- avenin
- granule bound starch synthase 1
- cluster of transcription factors
- starch synthase 1
- yield QTL 1
- yield QTL 2
- puroindolin
- powdery mildew resistance
- starch branching enzyme 1 x 3
Genetic map

- A composite map using the MAGIC 8-way cross population (Emma Huang, Colin Cavanagh, CSIRO and GBS by Matt Hayden, DEPI) with the Chinese Spring/Renan map (INRA) as an “anchor”. Generated with the following procedure:

1. We choose to “trust” the physical map - hence (ideally) we want all markers in a given physical contig to co-locate in the map

* Based on work done at CSIRO with Jen Taylor, Emma Huang, Penghao Wang, Stuart Stephen
Genetic map

- A composite map using the MAGIC 8-way cross population (Emma Huang, Colin Cavanagh, CSIRO and GBS by Matt Hayden, DEPI) with the Chinese Spring/Renan map (INRA) as an anchor. Generated with the following procedure:

2. For each physical contig three situations to deal with
   A) all markers are already tightly linked (which is what we want)
   B) one marker is an outlier -> remove to end up in case A
   C) multiple groups of tightly linked markers -> separate into “A” and “B” contigs to end up in case A
Genetic map

- A composite map using the MAGIC population (Emma Huang, Colin Cavanagh, CSIRO and GBS by Matt Hayden, DEPI) with the Chinese Spring/Renan map (INRA) as an anchor. Generated with the following procedure:

3. Take representative from each group, essentially collapsing contigs
4. Using this data, build clusters around framework markers in CS x Renan
5. Order markers within clusters
6. Estimate positions from full marker order
7. Expand out contigs - forces all markers within a contig to be at same position
The markers mapping to this physical contig, 7AS-12251, separate into two distinct locations in the genetic map.
Example of a split contig

The markers mapping to this physical contig, 7AS-12251, separate into two distinct locations in the genetic map. Likely caused by this repeat complex ("blob") (cf. talk by by Thomas Wicker)
Validating genetic map

7A POPSEQ v1 map (Mascher et al. 2013) shows good alignment

MAGIC/CSxR reference map shows high resolution, with increased detail around centromere
Fine Physical and Genetic Mapping of Powdery Mildew Resistance Gene MLLW172 Originating from Wild Emmer (Triticum dicoccoides)

Shuhong Ouyang1, Dong Zhang1, Jun Han1,2, Xiaojie Zhao1, Yu Cui1, Wei Song1,3, Naxin Huo4, Yong Liang1, Jingzhong Xie1, Zhenzhong Wang1, Qiuhong Wu1, Yong-Xing Chen1, Ping Lu1, De-Yun Zhang1, Lili Wang1, Hua Sun5, Tsoing Yang1, Gabriel Keeble-Gagnere6, Rudi Appels6, Jaroslav Doležel7, Hong-Qing Ling5, Mingcheng Luo8, Yongqiang Gu4, Qixin Sun1, Zhiyong Liu1x

1 State Key Laboratory for Agrobiotechnology/Beijing Key Laboratory of Crop Genetic Improvement/Key Laboratory of Crop Heterosis Research & Utilization, Ministry of Education, China Agricultural University, Beijing, China, 2 Agriculture University of Beijing, Beijing, China, 3 Maize Research Center, Beijing Academy of Agricultural and Forestry Sciences, Beijing, China, 4 USDA-ARS West Regional Research Center, Albany, California, United States of America, 5 State Key Laboratory of Plant Cell and Chromosome Engineering, Institutes of Genetics & Developmental Biology, Chinese Academy of Sciences, Beijing, China, 6 Murdoch University, Perth, Western Australia, Australia, 7 Institute of Experimental Botany, Centre of Plant Structural and Functional Genomics, Olomouc, Czech Republic, 8 Department of Plant Sciences, University of California, Davis, Davis, California, United States of America
Figure 2. Physical map of the BAC contigs and scaffolds flanking the MIIW172 locus anchored to the high-resolution genetic map. The approximate physical locations of all the newly designed markers are given on the BAC contigs or scaffolds. doi:10.1371/journal.pone.0100160.g002

Ouyang et al. 2014
Powdery mildew locus on 7AL

Adapted from Ouyang et al. 2014
Powdery mildew locus on 7AL

7AL-11771
227.4 cM

7AL-11973
227.5 cM

7AL-303
228.1 cM
This provides important validation of our map by a completely independent source.

7AL-11771
227.4 cM

7AL-11973
227.5 cM

7AL-303
228.1 cM
Two genes stand out as candidate genes for powdery mildew resistance:

*Disease resistance protein RPP8*

*Putative disease resistance protein RGA4*

Adapted from Ouyang et al. 2014
Two genes stand out as candidate genes for powdery mildew resistance:

- Disease resistance protein RPP8
- Putative disease resistance protein RGA4

* Evidence for Pwd gene in 7AL-11973 also supported by data from Kuldeep Singh

Adapted from Ouyang et al. 2014
Next steps

- Bionano optical mapping data is being generated (Hana Simkova/Jaroslav Dolezel, Mingcheng Luo) from flow-sorted DNA (Dolezel lab)
- Annotation - manual effort
- Diversity analysis and comparison to *T. urartu/T. monococcum* assembly

7A map vs. *T. monococcum* 90k SNP map (DNA from Jorge Dubcovsky, SNP map by Kerrie Forrest and Matt Hayden)
Next steps

- Bionano optical mapping data is being generated (Hana Simkova/Jaroslav Dolezel, Mingcheng Luo) from flow-sorted DNA (Dolezel lab)
- Annotation - manual effort
- Diversity analysis and comparison to *T. urartu/T. monococcum* assembly

7A map vs. *T. monococcum* 90k SNP map (DNA from Jorge Dubcovsky, SNP map by Kerrie Forrest and Matt Hayden)

Large inversion?
Summary of achievements

1. We have produced a high quality, genetically anchored, assembly of chromosome 7A.

2. The assembly has been validated using independent genome-level information for specific regions of the chromosome.

3. The assembly now forms the basis for the analysis of agronomically significant chromosome regions.
Acknowledgments

**Funding**
Grains Research Development Corporation
Bioplatforms Australia

**ACCWI group**
Rudi Appels, Hollie Webster, Shahidul Islam, Xueyan Chen, Yingjun Zhang, Johan Nystrom-Persson

**Flow-sorting DNA/BAC library construction**
Jaroslav Dolezel, Hana Simkova
Institute of Experimental Botany
Czech Republic

**Fingerprinting BAC library**
Mingcheng Luo group
UC Davis

**Physical map assembly**
Zeev Frenkel, Amraham Korol
Haifa University

**Genetic maps**
MAGIC: Colin Cavanagh, Emma Huang, Jen Taylor (CSIRO)
MAGIC GBS: Matt Hayden (DEPI)
CSxRenan: Pierre Sourdille, Benoit Darrier (INRA)

T. monococcum genetic map
Population: Jorge Dubcovsky
90k chip: Matt Hayden, Kerrie Forrest

**DNA sequencing**
Matt Tinning
AGRF

**Annotation**
TriAnnot: Philippe Leroy, Aurelien Bernard (INRA)
geneID (CRG): Francisco Camara, Anna Vlasova (CRG, Spain), Juan Carlos Sanchez (ACPFG)
Storage proteins: Angela Juhasz (Hungary)
QTL mapping/Significant genome regions: Delphine Fleury (ACPFG)
Specific genes: Hui-xian Zhao (NW A&F Uni, China)

**Pseudomolecule**
Fred Choulet, Etienne Paux
INRA

**7A mate-pair sequencing of amplified DNA**
Matt Hayden, Josquin Tibbits, Sami Hakim
DEPI

**Whole-genome mate-pair data**
Andy Sharpe, David Konkin, Curtis Pozniak
NRC, Canada

**Bionano map**
Jaroslav Dolezel, Hana Simkova, Mingcheng Luo

**Supercomputing resources**
iVEC/Pawsey Supercomputing Centre
Assembly summary

● 42,441 sequence scaffolds
  ○ Total length 940Mb
    ○ N50 137kb
    ○ Mean 22kb

● High-density composite genetic map based on MAGIC using CSxRenan map as anchor
  ○ Over 4000 markers

● 732 physical contigs
  ○ 316 physical contig scaffolds

● 535 anchored to genetic map
  ○ 676 anchored in scaffolded physical map
Genetic map

- A composite map using the MAGIC population (Emma Huang, Colin Cavanagh, CSIRO and GBS by Matt Hayden, DEPI) with the Chinese Spring/Renan map (INRA) as an anchor. Generated with the following procedure:

Example of B) case

One problem marker
Genetic map

- A composite map using the MAGIC population (Emma Huang, Colin Cavanagh, CSIRO and GBS by Matt Hayden, DEPI) with the Chinese Spring/Renan map (INRA) as an anchor. Generated with the following procedure:

Example of C) case
The Starch Branching Enzyme 1 (SBE1) locus
Starch branching enzyme (SBE1)

The Starch Branching Enzyme 1 (SBE1) locus

Starch branching enzyme (SBE1)

* TriAnnot annotation viewed in GBrowse
Starch branching enzyme (SBE1)

Structures of the 7D and 7A SBE1 loci are similar