Sequencing of the Aegilops tauschii genome

subspecies strangulata accession AL8/78
Genetic relationships of accession AL8/78

Hexaploid wheat

Ae. tauschii
ssp. strangulata

Ae. tauschii
ssp. tauschii
Collection site of AL8/78 and the putative geographic area of hexaploid wheat origin
Aegilops tauschii genome

- Haploid genome ~ 4 to 5 Gbp
- ~ 80 to 90% repeated sequences
Ordered-clone sequencing approach
Aegilops tauschii physical map

- 3,578 BAC contigs
- 2,263 anchored contigs (84.2% total contig length)
- MTP = 42,882 BAC clones

Luo et al., PNAS 110, 2013

1 = physical map
2 = recombination rates
3 = gene density
Division of labor among participants

- By chromosome
- By task
Tasks

- Validate and pool 8 BAC clones
- Index each pool, combine 48 pools, sequence pools
- Assemble our short pair-end reads together with BGI long pair-end reads
- Preliminarily annotate TEs and genes
- Construct nanomap and validate scaffolds
- Produce pseudomolecules
- Finally annotate TEs and genes
- Conduct community gene annotation
MiSeq MTP sequencing and assembly status

- Target
- As of January 2015

**Numbers of BAC clones**

- Chromosomes: 1D, 2D, 3D, 4D, 5D, 6D, 7D, ?, Singletons

As of January 2015:
- 1D: 4000
- 2D: 6000
- 3D: 4000
- 4D: 4000
- 5D: 4000
- 6D: 4000
- 7D: 4000
- Singletons: 10000

Target:
- 1D: 5000
- 2D: 6000
- 3D: 4000
- 4D: 4000
- 5D: 4000
- 6D: 4000
- 7D: 4000
- Singletons: 10000
Scaffold assembly v.1

Average number of scaffolds > 2Kb per BAC pool: 18

Total scaffold length: 5.7 Gb

Average Scaffold N50 length: 203 Kb

Assembly pipeline (Poster #577)
Transposable element annotation

- TE classes and families
## Gene annotation

Numbers of high-confidence genes

<table>
<thead>
<tr>
<th>Chrom.</th>
<th>Anchored BAC contigs</th>
<th>Corrected for unanchored BAC contigs</th>
<th>Wheat survey sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3,237</td>
<td>3,808</td>
<td>A 4,155</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>B 4,178</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>D 3,772</td>
</tr>
<tr>
<td>2</td>
<td>5,499</td>
<td>6,469</td>
<td>A 5,626</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>B 6,196</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>D 6,375</td>
</tr>
<tr>
<td>3 short</td>
<td>1,763</td>
<td>1,763</td>
<td>A 1,859</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>B 1,906</td>
</tr>
</tbody>
</table>
Gene annotation validation

Manual validation of annotation

- MIPS
- TriAnnot
- Maker
- Other

Annotation pipeline comparisons (Poster #576)
Pseudomolecule assembly

Challenges

- Scaffold validation, ordering, and orientating
- Ordering BAC contigs in low recombination regions
- Closing gaps
New technologies

Nanomap

Nanomaps BAC contigs
(Hastie et al., PloS ONE e55864, 2013)

Global nanomap of the Ae. tauschii genome

Nanomap construction (Poster #574)
M.C. Luo’s presentations in Tuesday’s IWGSC technical workshop
Ordering and orienting sequence scaffolds on the global nanomap

Nanomap contig

Sequence scaffolds

ctg1715 6D (76.097 cM)

ctg12344 6D (76.097 cM)

ctg6115 6D (76.097 cM)

ctg195 6D (75.686 cM)
Whole genome shotgun sequence

PacBio P6 chemistry

5 SMRT cells
Total reads 4.1 Gbp
Reads > 10 Kb 3.1 Gbp

Number of reads
Read length

10 kb
20 kb
Where can I access data?

BLAST: http://aegilops.wheat.ucdavis.edu/ATGSP/data.php

Batch download of scaffolds also available
Acknowledgements
National Science Foundation Plant Genome

Karin Deal
Pat McGuire
Ming-Cheng Luo
Naxin Huo, Yi Wang
Chad Jorgensen
Tingting Zhu, Sonny Van
Lichan Xiao, Luxia Yuan
Luis Curiel, Scott Liu
JC Rodriguez, Thanh Ngo
Armond Murray

Olin Anderson
Yong Gu

Katrien Devos
Hao Wang
Jeffrey Bennetzen

Frank You

Philippe Leroy

Shuhong Ouyang
Yong Liang
Zhenzhong Wang
Zhiyong Liu
Qixin Sun

Klaus Mayer
Matthias Pfeifer
Zhengqiang Ma

Thomas Wicker

Long Mao

Richard McCombie

Jaroslav Doležel
Cari Soderlund