Conservation and rearrangement: gene order in peri-centromeric regions of Triticeae genomes

Martin Mascher

IPK Gatersleben

April 13th, 2016
The polyploid wheat genome

- Genome size: \( \sim 15 - 17 \text{ Gb}, 3 \times 7 \) pairs of chromosomes
Excellent contiguity and genome representation

- IWGSC/NRGene WGS
- IWGSC CSS
- TGAC WGS
- JGI WGS

covered assembly length (Gb)

minimal sequence length

slide 3
Longer read lengths improve assembly completeness

Uniquely mappable regions as a function of k-mer/read length

fraction of uniquely mappable regions (%)

1H 2H 3H 4H 5H 6H 7H

Uniquely mappable regions as a function of k-mer/read length

500 bp 100 bp 30 bp
## Two wheat assemblies

<table>
<thead>
<tr>
<th></th>
<th>Bread wheat</th>
<th>Wild emmer</th>
</tr>
</thead>
<tbody>
<tr>
<td>ploidy</td>
<td>6x (AABBDD)</td>
<td>4x (AABB)</td>
</tr>
<tr>
<td>genome size</td>
<td>~ 15 – 17 Gb</td>
<td>~ 10 – 12 Gb</td>
</tr>
<tr>
<td>assembly size</td>
<td>14.5 Gb</td>
<td>10.5 Gb</td>
</tr>
<tr>
<td>N50</td>
<td>548 (7.4 Mb)</td>
<td>415 (7.0 Mb)</td>
</tr>
<tr>
<td>N90</td>
<td>2,276 (1.3 Mb)</td>
<td>1,828 (1.2 Mb)</td>
</tr>
<tr>
<td>genetically anchored</td>
<td>14.0 Gb (96.5 %)</td>
<td>10.1 Gb (95.6 %)</td>
</tr>
</tbody>
</table>
Alignment of the BAC-based reference of 3B

v443_0936 (4.2 Mb) vs. scaffold16560-2 (22.6 Mb)

v443_0903 (3.7 Mb) vs. scaffold123840 (10.1 Mb)
POPSEQ cannot order contigs in peri-centromeric regions

Anchored sequence along the genetic map

percentage of anchored sequence

genetic position (POPSEQ)
HiC: chromosome conformation capture sequencing

Lieberman-Aiden et al., 2009
Mapping the linear genome with 3D contact matrices

Lieberman-Aiden et al., 2009; Kaplan and Dekker, 2013
Mapping the linear genome with 3D contact matrices

Lieberman-Aiden et al., 2009; Kaplan and Dekker, 2013
<table>
<thead>
<tr>
<th></th>
<th>Bread wheat</th>
<th>Wild emmer</th>
</tr>
</thead>
<tbody>
<tr>
<td>assembly size</td>
<td>14.5 Gb</td>
<td>10.5 Gb</td>
</tr>
<tr>
<td># HiC links</td>
<td>3.7 M</td>
<td>6.5 M</td>
</tr>
<tr>
<td>ordered by HiC</td>
<td>14.0 Gb (96.3 %)</td>
<td>10.1 Gb (96.0 %)</td>
</tr>
<tr>
<td>oriented by HiC</td>
<td>–</td>
<td>9.7 Gb (92.5 %)</td>
</tr>
<tr>
<td># scaffolds per chr.</td>
<td>177</td>
<td>199</td>
</tr>
</tbody>
</table>
Rabl configuration of interphase nuclei
Rabl configuration of interphase nuclei

Simulated contact matrix

telomere
centromere
telomere
centromere
telomere
telomere
HiC contact matrices meet the expectations
Clinearity of the HiC maps of bread wheat and emmer

2B CARMA

2B HiC vs. POPSEQ

2B CS vs. emmer

Chinese Spring HiC map (bin)

2B CARMA

2B HiC vs. POPSEQ

2B CS vs. emmer

CARMA

POPSEQ (cM)

emmer HiC map (bin)
Large inversions and reduced recombination

4A CARMA

4A HiC vs. POPSEQ

4A CS vs. emmer
Centromeric rearrangements between wheat and barley

emmer 1A vs. barley 1H

emmer 1B vs. barley 1H
Outlook

- Improved HiC map of bread wheat and construction of pseudomolecules
- NRGene assembly and HiC map of durum wheat
- HiC maps for other wheat varieties to find large structural variants (inversions, deletions, insertions)
Bread wheat assembly team

- Curtis Pozniak, Andy Sharpe, University of Saskatoon
- Jesse Poland, Kansas State University
- Assaf Distelfeld, Tel Aviv University
- Gil Ronen, Omer Barad, Kobi Baruch, NRGene
- Mike Thompson, Illumina
- Fred Choulet, INRA
- Jane Rogers, Kellye Eversole, IWGSC
- Nils Stein, Axel Himmelbach, Ines Walde, IPK Gatersleben