Whole Genome Profiling (WGP) for physical mapping of wheat chromosome 6A

Naser Poursarebani

Poursarebani@ipk-gatersleben.de

PBP Group, IPK-Gatersleben, Germany

IWGSC Workshop - 2012
Outline:

- **Introduction**
  - Importance of wheat for German agriculture,
  - Why wheat chromosome 6A?
  - And the important genes localized on this wheat chromosome

- **Methodologies of making a BAC based physical map**
  - High Information Content Fingerprinting (HICF) based physical map
  - Whole Genome Profiling (WGP™) based physical map (KeyGene Company)

- **Preliminary results _ WGP based 6A assembly**

- **Anchoring resources**

- **Summary and Outlook**
Wheat in Germany!

Wheat is the number 1 crops regarding harvested acreage cultivated are 3.2 mio ha, 7.2 t/h

Why Chromosome 6A

- last-remained chromosome

IWGSC = International Wheat Genome Sequencing Consortium,
http://www.wheatgenome.org
Selected Traits associated with Chromosome 6A

Yield
- QTL for Yield and TKW (ca. 10-15%; Snape et al. 2007; Röder et al. unpubli., centromere)

Development
- Seedling vigor (Spielmeyer et al. 2007)

Disease Resistances
- Powdery mildew, Pm21 (Muranty et al. 2009)
- Stem rust, Sr8, Sr13 (Ug99), Sr26, (McIntosh et al. 1995/ Belayneh Admassu, 2010)
- Aphid resistance (Castro et al. 2005; centromere)
- Septoria tritici, Stb15 (Arraiano et al. 2007)
- Fusarium head blight, major QTL (Holzapfel et al. 2008)

List possibly not complete
BAC based Physical map

*Basic principles and the new technology*
Get the chromosome sequenced; From BAC library to Physical mapping and Sequencing

Chromosome Sorting by Flow Cytometry

- Fragmented Chromosomal DNA
- A FP contig (FingerPrinted contig) with two Clones
- Cloned into *Escherichia coli* F-factor (Bacterial Artificial Chromosome; BAC)
- Desired Chromosome
- Other Chromosomes

Sequencing

Anchoring to the chromosome

Genetic map

Modified from Nils Stein; 2007
Fingerprinting strategies

High Information content fingerprinting (HICF)

I: Multiple RE digestion
- XbaI, BamH1, XhoI, EcoR1 and HaeIII

II: Enzymatic digested fragments: BAC fingerprints

III: Fragment labeling using fluorescent SNaPshot chemistry: required for size calling

IV: BAC fragment based profile: Between 40 to 250 bans per BAC

IV: BAC overlap identification and contig assembly
- tolerance of 0.4 bp

Whole Genome Profiling (WGP™)

I: Single RE digestion
- HindIII

II: Sequencing adaptor ligation

III: End sequencing of enzymatic digested fragments

IV: BAC sequence based profile: on average 50 sequence tags of ~100bp per BAC (126kb, 1 HindIII site/2.5Kb)

IV: BAC overlap identification and contig assembly
- tolerance of 0.0 bp

Modified from Catherine Feuillet; PAG, 2012
A pooling scheme was utilized for WGP of the 6A BACs

BAC Clones are cleverly pooled

Pooled BAC DNA are digested and sequenced with Illumina GAII

Sequence tags are de-convoluted and assigned to individual BAC (2-6 kb apart)

Jan von Overen et al. Genome.Res 2011; Arabidopsis

http://www.keygene.com/services/technologies_whole_genome_profiling.php
Comparison of HICF and WGP using wheat 3B BAC clones
(Philippe et al., 2012)

- 11,238 BAC clones (230 Mb) including 1380 fully sequenced BACs (12 Mb) were used for both HICF and WGP approaches

**Physical map assembly using HICF**

- Initial Assembly
- Automated Assembly (Single to end & end to end Merging, DQing)

**Physical map assembly using WGP**

- Assembly step
Comparison of HICF and WGP

<table>
<thead>
<tr>
<th></th>
<th>HICF (1e-25)</th>
<th>WGP\textsuperscript{TM} (1e-11)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimated coverage in length</td>
<td>236 Mb +- 65</td>
<td>199 Mb +- 42</td>
</tr>
<tr>
<td>Number of contigs</td>
<td>631</td>
<td>434</td>
</tr>
<tr>
<td>Average contig size (kb)</td>
<td>374</td>
<td>469</td>
</tr>
<tr>
<td>N50 (kb)</td>
<td>455</td>
<td>567</td>
</tr>
</tbody>
</table>

Comparison to 12 sequenced contigs

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Coverage percentage in length</td>
<td>95.8%</td>
<td>94.9%</td>
</tr>
<tr>
<td>Number of chimerical contigs for 10 Mb</td>
<td>0.6%</td>
<td>0.6%</td>
</tr>
<tr>
<td>Percentage of mis-assembled BACs</td>
<td>9.5%</td>
<td>2.7%</td>
</tr>
</tbody>
</table>

- Equivalent coverage in length
- Equivalent number of chimerical contigs
- Less miss-assembled BACs in WGP
- Physical map fully enriched by sequence info in WGP
- Less time and effort (in case of keygene collaboration) in WGP
- Same cost
Wheat 6A physical map was decided to be performed in collaboration with KeyGene using Whole Genome Profiling (WGP) approach
### 6A BAC library resource utilized

<table>
<thead>
<tr>
<th>No of BAC clone</th>
<th>Long Arm (369 Mb)</th>
<th>Short Arm (336 Mb)</th>
<th>Chromosome coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complete BAC library</td>
<td>55,296 (123 kb)</td>
<td>46.080 (130 kb)</td>
<td>~16x</td>
</tr>
<tr>
<td>BACs for WGP™</td>
<td>23,040</td>
<td>24,576</td>
<td>~8 - 9x</td>
</tr>
<tr>
<td>WGP™ Output (FPC ready _keygene)</td>
<td>18,660</td>
<td>19,183</td>
<td>6.2 - 7.4x</td>
</tr>
</tbody>
</table>

Library RE: HindIII

WGP RE: HindIII/MseI
Detail results of the WGP & Preliminary results on the **automated assembly** (6AL and 6AS)
Overview of general WGP output parameters and sequence data processing (Illumina HiSeq2000 sequencing)

<table>
<thead>
<tr>
<th>WGP parameter</th>
<th>6AS</th>
<th>6AL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimated size of chromosome</td>
<td>336 Mbp</td>
<td>369 Mbp</td>
</tr>
<tr>
<td># BACs tested</td>
<td>23,040</td>
<td>24,576</td>
</tr>
<tr>
<td>tagged BACs (FPC ready)</td>
<td>19,183 (83.3%)</td>
<td>18,660 (75.9%)</td>
</tr>
<tr>
<td>Enzyme combination</td>
<td>HindIII/MseI</td>
<td>HindIII/MseI</td>
</tr>
<tr>
<td>% deconvolutable reads</td>
<td>51.6%</td>
<td>47.30%</td>
</tr>
<tr>
<td># unique WGP tags (FPC ready)</td>
<td>85,013</td>
<td>108,811</td>
</tr>
<tr>
<td>average # WGP tags/ BAC</td>
<td>27.1</td>
<td>27</td>
</tr>
<tr>
<td>average # reads/ tag</td>
<td>122.4</td>
<td>111.1</td>
</tr>
</tbody>
</table>
Automated physical map assembly using WGP fingerprints (performed by FPC tool)

- Initial Assembly
- Automated Assembly (Single to end & end to end Merging, Dqing, Rebuilding)
6AL Stepwise assembly _ reduction of Contigs and Singletons
(17309 BACs with 6 to 68 tags, 5.7 x)
6AS _stepwise assembly_ Reduction of Contigs and Singletons
(17853 BACs with 6 to 68 tags, 6.9x)
Anchoring strategies

Recombination based genetic maps

- Genotyping By Sequencing based (GBS) markers (Jesse Poland, PLoS One 2012);
  - 1400 genetic markers from 6A with known sequences,
  - 215 Double haploid individuals (‘Opata / Syn W9784 )

Whole chromosome 6A shotgun sequence contigs

- SNP markers, Röder et al. unpubli.,

- Wheat KASPar SNP database
  - currently 93 SNPs mapped on chr 6A
  - ca. 4000 SNPs mapped by the end of this year (in total)
  - monthly update of the SNP database
Anchoring strategies
Recombination independent genetic maps
(Radiation Hybrid)

- Physical mapping based on radiation induced chromosome breakage and reconstruction of marker order based on their co-retention pattern

Wheat A genome RH map is being produced!

(Dr. Shahryar Kianian, North Dakota U, Fargo, USA)

Summary

- The WGP approach is shown to be potentially applicable to make a robust physical map.
- The HindIII enzyme utilized in the WGP panel increased the tag density.
- The assembly length of the 6AS is corresponding well to the estimated size of the arm while this is not the case for 6AL.

and the outlook

- Perform the genetic anchoring using marker info mentioned.
- Start the manual editing and merging using the marker info from earlier step of the assembly (e.g. cutoff e-30).
Acknowledgment

IPK, Germany
Thorsten Schnurbusch
Nils Stein
Uwe Scholz / Group
Jenny Knibbiche
KeyGene, Wageningen
Hanneke Witsenboer
Jan van Oeveren

IEB, Czech Republic
Jaroslav Dolezel / Group
INRA, France
Helene Berges / Group
MIPS, Germany
Klaus Mayer / Group

Funding:
BMBF in the frame of TRITEX project