A Whole Genome Assembly of Rye (*Secale cereale*)

M. Timothy Rabanus-Wallace

*Leibniz Institute of Plant Genetics and Crop Plant Research (IPK)*
Why rye?

Production share by region Average 1994 - 2014
Source: FAOSTAT (Jan 29, 2017)

Wheat
- Oceania: 3.4%
- Africa: 3.3%
- Americas: 17.5%
- Europe: 32.2%
- Asia: 43.6%

Rye
- Oceania: 0.1%
- Africa: 0.3%
- Americas: 3.1%
- Asia: 6.4%
- Europe: 90% (90% of total production)

Barley
- Oceania: 5.2%
- Africa: 3.7%
- Americas: 14%
- Asia: 15.1%
- Europe: 62.1%
Frost-damage in winter wheat

Photo: Ingrid Kristjanson (http://cropchatter.com/impact-of-frost-on-winter-wheat-fall-rye/)

QTL analysis for Survival After Winter (SAW) scores in rye:

Erath et al. 2017
Assembly challenges …

Rye

*Secale cereale*

Challenge 1) Length (7.9 Gbp)
Rye
*Secale cereale*
Challenge 1) Length (7.9 Gbp)
Challenge 2) 90+% repetitive
Rye

*Secale cereale*

Challenge 1) Length (7.9 Gbp)
Challenge 2) 90+% repetitive
Challenge 3) Obligate outcrossing
Assembly
Scaffolds
Assembly Scaffolds

Molecular Map

Genome

M T Rabanus-Wallace 2018
Major rye assembly milestones:

Martis ‘13: A Rye Proto-Genome ("Zipper")

Bauer ‘17: A Draft Genome

IRGSC ‘18: A WGS DeNovo Genome Approaching Reference Quality ...
Martis ‘13: A Rye Proto-Genome (“Zipper”)

- RNA seq (Expressed Sequence Tags; ESTs)
  - Raw sequence information
- Chromosome Survey Sequencing (CSS)
  - Chromosome-assigned sequence information
- 5K SNP-array-based genetic map
  - EST anchoring backbone
- Interspecies gene colinearity
  - Fine-scale ordering and gene identification by sequence homology
Martis ‘13: A Rye Proto-Genome (“Zipper”)

• RNA seq (Expressed Sequence Tags; ESTs)
  • *Raw sequence information*
• Chromosome Survey Sequencing (CSS)
  • *Chromosome-assigned sequence information*
• 5K SNP-array-based genetic map
  • *EST anchoring backbone*
• Interspecies gene colinearity
  • *Fine-scale ordering and gene identification by sequence homology*
Martis ‘13: A Rye Proto-Genome (“Zipper”)

- RNA seq (Expressed Sequence Tags; ESTs)
  - Raw sequence information
- Chromosome Survey Sequencing (CSS)
  - Chromosome-assigned sequence information
- 5K SNP-array-based genetic map
  - EST anchoring backbone
- Interspecies gene colinearity
  - Fine-scale ordering and gene identification by sequence homology

![Graph showing sequence in bin (bp) vs scaffold length bin (bp; bin size = 0.2 log bp)](image)
Martis ‘13: A Rye Proto-Genome (“Zipper”)  

- RNA seq (Expressed Sequence Tags; ESTs)  
  - Raw sequence information  
- Chromosome Survey Sequencing (CSS)  
  - Chromosome assignment  
- 5K SNP-array-based genetic map  
  - EST anchoring backbone  
- Interspecies gene colinearity  
  - Fine-scale EST ordering  

Bauer ‘17: A Draft Genome  

- WGS and Mate-Pair (MP) Libraries  
  - Raw sequence and hierarchical scaffolding  
- CSS  
  - Contig and mate-pair read assignment pre-scaffolding  
- High-density SNP map (iSelect Rye 600k Array)  
  - To anchor scaffolds  
- DArT seq genetic map  
  - To guide scaffolding and detect chimeras  
- Martis ’13 genome zipper (updated)
Martis ‘13: A Rye Proto-Genome (“Zipper”)

- RNA seq (Expressed Sequence Tags; ESTs)
  - Raw sequence information
- Chromosome Survey Sequencing (CSS)
  - Chromosome assignment
- 5K SNP-array-based genetic map
  - EST anchoring backbone
- Interspecies gene colinearity
  - Fine-scale EST ordering

Bauer ‘17: A Draft Genome

- WGS and Mate-Pair Libraries
  - Raw sequence and hierarchical scaffolding
- CSS
  - Contig and mate-pair assignment pre-scaffolding
- High-density SNP map (iSelect Rye 600k Array)
  - To anchor scaffolds
- DArT seq genetic map
  - To guide scaffolding and detect chimeras
- Martis ’13 genome zipper (updated)
Martis ‘13: A Rye Proto-Genome (“Zipper”)

- RNA seq (Expressed Sequence Tags; ESTs)
  - Raw sequence information
- Chromosome Survey Sequencing (CSS)
  - Chromosome assignment
- 5K SNP-array-based genetic map
  - EST anchoring backbone
- Interspecies gene colinearity
  - Fine-scale EST ordering

Bauer ‘17: A Draft Genome

- WGS and Mate-Pair Libraries
  - Raw sequence and hierarchical scaffolding
- CSS
  - Contig and mate-pair assignment pre-scaffolding
- High-density SNP map (iSelect Rye 600k Array)
  - To anchor scaffolds
- DArT seq genetic map
  - To guide scaffolding and detect chimeras
- Martis ’13 genome zipper (updated)
Martis ‘13: A Rye Proto-Genome (“Zipper”)

- RNA seq (Expressed Sequence Tags; ESTs)
  - Raw sequence information
- Chromosome Survey Sequencing (CSS)
  - Chromosome assignment
- 5K SNP-array-based genetic map
  - EST anchoring backbone
- Interspecies gene colinearity
  - Fine-scale EST ordering

Bauer ‘17: A Draft Genome

- WGS and Mate-Pair Libraries
  - Raw sequence and hierarchical scaffolding
- CSS
  - Contig and mate-pair assignment pre-scaffolding
- High-density SNP map (iSelect Rye 600k Array)
  - To anchor scaffolds
- DArT seq genetic map
  - To guide scaffolding and detect chimeras
- Martis ’13 genome zipper (updated)
2018: Approaching Reference Quality

An **NRGene DeNovoMAGIC3.0 assembly** (analogues in wheat cv. Julius and barley cv. Barke)

- WGS and mate-pair libraries
  - Raw data
- Map-anchored contigs (from Bauer ‘17)
  - Preliminary anchoring, chromosome assignment and chimera detection
- 10x Chromium molecule-linked reads
  - Long-range scaffolding information
  - Chimera breakpoint detection
- CSS
  - Chromosome assignment and chimera detection

... and upcoming ...

- PopSeq high-density genetic mapping
  - Map anchoring and chimera detection
- Chromosome-Conformation Capture Sequence (Hi-C)
  - Fine-scale ordering and orientation for pseudomolecule construction
2018: Approaching Reference Quality

An NRGene DeNovoMAGIC3.0 assembly (analogues in wheat cv. Julius and barley cv. Barke)

• WGS and mate-pair libraries
  • Raw data
• 10x Chromium molecule-linked reads
  • Scaffolding guide
  • Chimera breakpoint detection
• Map-anchored contigs (from Bauer ‘17)
  • Preliminary anchoring, chromosome assignment and chimera detection
• CSS
  • Chromosome assignment and chimera detection

...and upcoming...

• PopSeq high-density genetic mapping
  • Map anchoring and chimera detection
• Chromosome-Conformation Capture Sequence (Hi-C)
  • Fine-scale ordering and orientation for pseudomolecule construction
2018: Approaching Reference Quality

An NRGene DeNovoMAGIC 3.0 assembly (analogues in wheat cv. Julius and barley cv. Barke)

- WGS and mate-pair libraries
  - Raw data
- 10x Chromium molecule-linked reads
  - Scaffolding guide
  - Chimera breakpoint detection
- Map-anchored contigs (from Bauer ‘17)
  - Preliminary anchoring, chromosome assignment and chimera detection
- CSS
  - Chromosome assignment and chimera detection

... and upcoming ...

- PopSeq high-density genetic mapping
  - Map anchoring and chimera detection
- Chromosome-Conformation Capture Sequence (Hi-C)
  - Fine-scale ordering and orientation for pseudomolecule construction
<table>
<thead>
<tr>
<th>NRGene DeNovoMAGIC 3.0 Assemblies</th>
<th>Rye</th>
<th>Wheat (Julius)</th>
<th>Barley (Barke)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total length (Gbp) (Genome Size)</td>
<td>6.67 (7.9)</td>
<td>14.38 (16)</td>
<td>4.18 (5.1)</td>
</tr>
<tr>
<td>Map-Anchored</td>
<td>6.16</td>
<td>14.20</td>
<td>4.03</td>
</tr>
<tr>
<td>N50 length (Mbp)</td>
<td>22.49</td>
<td>38.03</td>
<td>38.37</td>
</tr>
<tr>
<td>Map-Anchored</td>
<td>24.15</td>
<td>39.36</td>
<td>40.48</td>
</tr>
<tr>
<td>Number of scaffolds</td>
<td>107580</td>
<td>99465</td>
<td>17669</td>
</tr>
<tr>
<td>Map-Anchored</td>
<td>1099</td>
<td>3211</td>
<td>522</td>
</tr>
<tr>
<td>Proportion sequence anchored</td>
<td>.92</td>
<td>.98</td>
<td>.96</td>
</tr>
<tr>
<td>Proportion complete BUSCOs</td>
<td>.98</td>
<td>.98</td>
<td>.98</td>
</tr>
</tbody>
</table>
Quality Validation:

*Leveraging molecule-linked reads and CSS to identify chimeric scaffolds …*

Identification by CSS:

Identification by 10x molecule linked reads:

Break point!

Chromosomal and scaffold identification:

Chromosome A

A chimeric scaffold

Chromosome B

Mapped CSS Reads/Contigs

Chromosomal origin

Scaffold

Chromosomal scaffold
In reality:
*Scaffold951*

**Identification by CSS:**

**Identification by 10x molecule linked reads:**

Position in scaffold (Mbp)
<table>
<thead>
<tr>
<th>NRGene DeNovoMAGIC 3.0 Assemblies</th>
<th>Rye</th>
<th>Wheat (Julius)</th>
<th>Barley (Barke)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Total length (Gbp) (Genome Size)</strong></td>
<td>6.67 (7.9)</td>
<td>14.38 (16)</td>
<td>4.18 (5.1)</td>
</tr>
<tr>
<td><em>Map-Anchored</em></td>
<td>6.16</td>
<td>14.20</td>
<td>4.03</td>
</tr>
<tr>
<td><strong>N50 length (Mbp)</strong></td>
<td>22.49</td>
<td>38.03</td>
<td>38.37</td>
</tr>
<tr>
<td><em>Map-Anchored</em></td>
<td>24.15</td>
<td>39.36</td>
<td>40.48</td>
</tr>
<tr>
<td><strong>Number of scaffolds</strong></td>
<td>107580</td>
<td>99465</td>
<td>17669</td>
</tr>
<tr>
<td><em>Map-Anchored</em></td>
<td>1099</td>
<td>3211</td>
<td>522</td>
</tr>
<tr>
<td><strong>Proportion sequence anchored</strong></td>
<td>.92</td>
<td>.98</td>
<td>.96</td>
</tr>
<tr>
<td><strong>Bad CSS flagged scaffolds per ten thousand (Number)</strong></td>
<td>4.74 (51)</td>
<td>6.24 (62)</td>
<td>10.75 (19)</td>
</tr>
<tr>
<td><strong>Auto-IDd breaks (10x) per Mbp (Number)</strong></td>
<td>0.181 (1206)</td>
<td>-</td>
<td>.0103 (43)</td>
</tr>
</tbody>
</table>
Quality Validation:

Assessment of gene colinearity ...
Assessment of gene colinearity ...
Quality Validation:

Assessment of gene colinearity …
Quality Validation:

Assessment of gene colinearity …
Quality Validation:

Assessment of gene colinearity ...
Confirmation by 10x and CSS ...

Inferred coverage (10X molecules)

Rye Scaffold Position

Illumina CSS reads

H. vulgare gene models

Chromosome of origin

Rye Scaffold468 Position

30 million bp
2018: Approaching Reference Quality

An NRGene DeNovoMAGIC3.0 assembly (analogues in wheat cv. Julius and barley cv. Barke)

- WGS and mate-pair libraries
  - Raw data
- 10x Chromium molecule-linked reads
  - Scaffolding guide
  - Chimera breakpoint detection
- Map-anchored contigs (from Bauer ‘17)
  - Preliminary anchoring, chromosome assignment and chimera detection
- CSS
  - Chromosome assignment and chimera detection

...and upcoming...

- PopSeq high-density genetic mapping
  - Map anchoring and chimera detection
- Chromosome-Conformation Capture Sequence (Hi-C)
  - Fine-scale ordering and orientation for pseudomolecule construction
**PopSeq**

*High-density genetic mapping on the cheap…*

- Low-coverage WGS data used to call SNPs in assembly scaffolds in a mapping population

---

**Chromosome Conformation Capture Sequencing (Hi-C)**

*High-density distance information for mapping/scaffolding*

---

**Population**

**Assembly Scaffolds**

Genotype Calls

- Parent A
- Parent B
- Missing

---

Mascher et al. 2017
<table>
<thead>
<tr>
<th>Country</th>
<th>Institution</th>
<th>Scientist</th>
</tr>
</thead>
<tbody>
<tr>
<td>Germany</td>
<td>IPK Gatersleben</td>
<td>Uwe Scholz, Martin Mascher, Andreas Houben, Andreas Börner, Andreas Graner, Nils Stein</td>
</tr>
<tr>
<td></td>
<td>JKI Groß-Lüsewitz</td>
<td>Bernd Hackauf</td>
</tr>
<tr>
<td></td>
<td>JKI Quedlinburg</td>
<td>Frank Ordon</td>
</tr>
<tr>
<td></td>
<td>HMGU</td>
<td>Klaus Mayer</td>
</tr>
<tr>
<td></td>
<td>KWS LOCHOW GMBH</td>
<td>Viktor Korzun</td>
</tr>
<tr>
<td></td>
<td>Hybro Saatzucht</td>
<td>Joachim Fromme</td>
</tr>
<tr>
<td></td>
<td>TUM</td>
<td>Bauer</td>
</tr>
<tr>
<td>Canada</td>
<td>AAC/AAFC</td>
<td>André Laroche</td>
</tr>
<tr>
<td></td>
<td>USASK/GIFS/NRC</td>
<td>Curtis Pozniac</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Sharpe</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Konkin</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Bekkaoui</td>
</tr>
<tr>
<td>Poland</td>
<td>West Pomeranian University of Technology Szczecin</td>
<td>Stefan Stojalowski</td>
</tr>
<tr>
<td></td>
<td>Warsaw University of Life Sciences</td>
<td>Hanna Bolibok-Bragoszewska</td>
</tr>
<tr>
<td></td>
<td>Warsaw University of Life Sciences</td>
<td>Monika Rakoczy-Trojanowska</td>
</tr>
<tr>
<td>Czech Republic</td>
<td>Institute of Experimental Botany</td>
<td>Jaroslav Doležel</td>
</tr>
<tr>
<td>Finland</td>
<td>University of Helsinki</td>
<td>Alan Schulman</td>
</tr>
<tr>
<td>USA</td>
<td>The Samuel Roberts Noble Foundation</td>
<td>Xuefeng Ma</td>
</tr>
<tr>
<td></td>
<td>KSU</td>
<td>Jesse Poland</td>
</tr>
<tr>
<td></td>
<td>MSU</td>
<td>Hikmet Budak</td>
</tr>
<tr>
<td></td>
<td>UMD</td>
<td>Vijay K Tiwari</td>
</tr>
<tr>
<td>UK</td>
<td>2Blades</td>
<td>Lynne Reuber</td>
</tr>
<tr>
<td></td>
<td>EI</td>
<td>Hall</td>
</tr>
<tr>
<td>China</td>
<td>CAAS</td>
<td>Jizeng Jia</td>
</tr>
<tr>
<td>Switzerland</td>
<td>Zürich University</td>
<td>Beat Keller</td>
</tr>
<tr>
<td>Turkey</td>
<td>Cukurova University</td>
<td>Hakan Özkan</td>
</tr>
<tr>
<td>Israel</td>
<td>NRGene</td>
<td>Gil Ronen</td>
</tr>
<tr>
<td></td>
<td>NRGene</td>
<td>Kobi Baruch</td>
</tr>
</tbody>
</table>
... Thanks!

M. Timothy Rabanus-Wallace
Leibniz Institute of Plant Genetics and Crop Plant Research (IPK)

M T Rabanus-Wallace 2018