Cereal endosperm; lessons learnt from molecular biology, cell biology, genomics and transcriptomics analysis.

Odd-Arne Olsen
Department of plant sciences
Norwegian university of life sciences
The cereal endosperm from different species have the same structure.

The dicot endosperm is similar, but evolved independently.

A. thaliana

Brown et al., 1999
The endosperm cell types are very different and are run by separate developmental programs.

Developmental stages:
- **1-3 DAP**: Syncytial
- **3-5 DAP**: Cellularization
- **6-10 DAP**: Differentiation
- **15-30 DAP**: Maturation

Part I. A review of endosperm developmental programming gained over the last decades.

1. The endosperm development program is locked in either before or right after fertilization.
2. Sucrose is a trigger for endosperm development.

*In vitro grown* maize endosperm grown in high sucrose runs the *in planta* developmental program.

*In planta*  
*In vitro culture (12% sucrose)*  

Conventional maize endosperm culture; callus-like growth with some aleurone cells.
2. Sucrose is a trigger for endosperm development, cont’d.

A tripiple florence marker line to help identify endosperm cell types.

*In vitro* grown maize endosperm autonomously organize with a surface layer of aleurone cells and interior starchy endosperm cells.

*Whole endosperm*  
*Mini-endosperm*
3. Starchy endosperm cell fate is default; lack of signaling for aleurone and transfer cells gives an endosperm with only starchy endosperm cells.
4. Transfer cell specification needs maternal signal input.

Marker transcript (*End1*) for transfer cells present at the syncytial stage depend on maternal signal input.

*In planta* Marker transcript (*End1*) for transfer cells present at the syncytial stage depend on maternal signal input.

5. Ectopic expression of *MRP-1* (myb-related protein) and MEG1 (Cystein-rich peptide) is sufficient to induce transfer cell differentiation in maize endosperm.

*ProAl9-MRP-1*

Gómez et al, Plant Cell 2009

Maternal signal

6. Endosperm cellularization occurs by an ancient mechanism.
7. Aleurone cell fate is specified by surface cell position and is DEK1 dependent. DEK1 is a membrane-anchored calpain protease.

WT maize endosperm in vitro with aleurone

dek1 maize endosperm in vitro lacking aleurone cells

1. The endosperm developmental program is locked early in development
2. Sucrose is a trigger for endosperm development
3. Starchy endosperm is the default cell fate
4. Transfer cell fate specification needs maternal signal input
5. Ectopic expression of MRP-1 and MEG1 induce transfer cell differentiation
6. Endosperm cellularization occurs by an ancient mechanism
7. Aleurone cell fate is specified by surface cell position and is DEK1 dependent
Part II. Wheat genomics and transcriptomics
The International Wheat Genome Sequencing Consortium

2016

Board of Directors
Role: Overall strategy and organization
Up to 10 directors including at least 1 Outside Director

Leadership Team
Role: Daily management
Members: Appels, Eversole, Feuillet, Keller, & Rogers

Coordinating Committee
Role: Scientific strategy
Members: Sponsors & Leaders of IWGSC Projects & Initiatives

Members
Role: Participation & Input
Members: Open to Anyone

COUNTRIES
59

MEMBERS
1400

SPONSORS
19

INSTITUTES/COMPANIES
434

www.wheatgenome.org
A chromosome-based approach

Jaroslav Doležel et al.

Dissection of the genome to single chromosomes (arms) representing individual (sub)genomes

Norway: chromosome 7B

Doležel et al., Chromosome Res. 15: 51, 2007

- Chromosomes: 605 - 995 Mbp (3.6 – 5.9% of the genome)
- Chromosome arms: 225 - 585 Mbp (1.3 – 3.4% of the genome)

Amplified DNA for chromosome survey (Nov 2011)
Illumina sequencing of DNA from individual chromosomes

IWGSC CSS v2 (2014)

- First identification of A, B and D homeologues in hexaploid bread wheat.

- 124,201 gene loci annotated across all 21 chromosomes.
Using the WGSC CSS v2 sequence to re-evaluate the evolutionary history of bread wheat.

- The A and the B genomes diverged from a common ancestor 7 million year ago.

- The D genome arose by homoploid hybridization speciation between the A and the B genomes 1.2 million years later.
Using the WGSC CSS v2 sequence to characterize the bread wheat endosperm transcriptome

Sampling and dissection;

- Chinese Spring, 15°C night/20°C day
- 30 samples, 4 biological replicates per cell type and time point (two replicates per room) and two additional technical replicates.
- RNAseq pair end libraries: 200bp, Illumina Hi-seq 2000
Are the A, B and D genomes equally expressed in endosperm?

Investigating the expression of 6,576 *triplets* of homeologous genes (A,B,D).

55% of genes expressed in endosperm as a whole; A, B and D approx. equally represented.

Approx. equal number of A,B,D homeologs expressed in the endosperm samples

Pfeifer et al., Science (2014)
Examples of expression asymmetry between sub-genomes: baking quality genes.

“A” sub-genome transcripts are strongly underrepresented for LMW-Glu, HMV-Glu as well as for Puroindolines.

Pfeifer et al., Science (2014)
Endosperm development is driven by distinct expression modules with many genes.

GO-terms for cluster I:
- 214 establishment of localization
- 38 organic acid biosynthetic process

Detailed studies of hub genes and their promoters in these clusters may eventually identify the temporal and functional network driving endosperm development.
Using “Weighted correlation network (WGCNA) analysis to investigate the transcriptional dynamics of A,B and D homeologs.

25 co-expression modules identified using network clustering. Each dot: a triplet of A,B and D homeologue expression
Modules function in developmental stages and cell types as shown by color-codes. GO-term labeling indicate dominating functions for selected modules.

Pfeifer et al., Science (2014)
Coloring for sub-genome identifies sub-genome dominance

Pfeifer et al., Science (2014)
The asymmetric contribution of the three sub-genomes A, B and D to the endosperm transcriptome may explain the unique characteristics of bread wheat grains.

Semantic representation of GO-categories colored according to the significant overrepresentation in co-expression modules (green, A; purple, B; orange, D).

Pfeifer et al., Science (2014)
Roadmap to the Wheat Genome Sequence, next step

Illumina sequencing of individual chromosomes

- IWGSC CSS v2 (2014)
  - Whole genome mate pairs
    - IWGSC CSS v3 (2016)

Physical maps of individual chromosomes

- MTP sequencing
  - Pseudomolecule assembly
    - 100%
    - Chromosome 3B (2014)
    - 20 chromosomes (2016)

NRGene-Illumina WGS

- IWGSC Whole Genome Assembly v0.4 (2016)
  - 100%

Radiation Hybrid, Hi-C, genetic, LD maps

BioNanoGenomics optical maps

MTP sequence tags…..

Reference Genome Sequence (2017)
RefSeq v1.0 statistics

<table>
<thead>
<tr>
<th>Description</th>
<th>RefSeqv1.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number / coverage of scaffolds/contigs</td>
<td>138,665 / 14.5 Gb</td>
</tr>
<tr>
<td>Number / coverage of scaffolds/contigs &gt;=100kb</td>
<td>4,443 / 14.2 Gb</td>
</tr>
<tr>
<td>N50 scaffolds / superscaffolds</td>
<td>22.8 Mb</td>
</tr>
<tr>
<td>L50 (no. sequences (\rightarrow) N50)</td>
<td>186</td>
</tr>
<tr>
<td>N90 scaffolds / superscaffolds</td>
<td>4.1 Mb</td>
</tr>
<tr>
<td>L90 (no. sequences (\rightarrow) N50)</td>
<td>718</td>
</tr>
<tr>
<td>Gaps filled with BAC sequences</td>
<td>183 (1.7 Mb)</td>
</tr>
<tr>
<td>Average size of inserted BAC sequences</td>
<td>9.5 kb</td>
</tr>
<tr>
<td>Sequence assigned to chromosomes</td>
<td>14.1 Gb (96.8%)</td>
</tr>
<tr>
<td>Sequence assigned to chromosomes (&gt;=100kb)</td>
<td>14.1 Gb (99.1%)</td>
</tr>
<tr>
<td>No. scaffolds / superscaffolds on chromosomes</td>
<td>1,601</td>
</tr>
<tr>
<td>No. oriented scaffolds / superscaffolds</td>
<td>1,243</td>
</tr>
<tr>
<td>Oriented sequence</td>
<td>13.8 Gb (95%)</td>
</tr>
<tr>
<td>Oriented sequence &gt;=100kb</td>
<td>13.8 Gb (97.3%)</td>
</tr>
</tbody>
</table>

RefSeq v1.0 contains ~ 75 scaffolds per chromosome
Using IWGSC RefSeq v1.0

Assessing the homeologue allele combination of two production bread wheat cultivars; a first glance.

UNPUBLISHED

Belova et al., unpublished
Using IWGSC RefSeq v1.0, cont´d

Genome distribution of coeliac disease associated prolamin gene families and associated non-prolamin gene families.

UNPUBLISHED

A. Juhász et al., unpublished
Pre-publication data access:

IWGSC RefSeq v1.0: January 14, 2017: Assembly made available for pre-publication access under Toronto agreement.

Gene models completed: March 2017

Final analyses completed: April/May 2017

Manuscript submission: Summer 2017

https://wheat-urgi.versailles.inra.fr/Seq-Repository/Assemblies
Thank you!