Plant genomics and proteomics: applications and opportunities in Canadian crops

Faouzi Bekkaoui

Rendez-Vous Protéomiqe 2008
Montreal, 10 April 2008
Outline

• Overview of crop genomics projects
  Genome Canada projects
  Federal Genomics Initiatives
• Examples of proteomic applications
  *Brassica napus* (canola)
• Research opportunities for future projects
  Position paper and potential papers
Value of Canadian crops

• Agriculture and food is the fourth largest industry in Canada.

• Canadian agriculture will export 28.8 million tonnes of grains and oilseeds in 2007, worth $13.8 billion. Canada is one of the world’s largest exporters of agricultural products including flax, canola, lentils and durum wheat.

• Wheat crops form a $5 billion portion of agricultural production in Canada.
Importance of Canola

- Canola contributes over $11 billion to the Canadian economic activity
- More than 52,000 Canadian farmers grow canola
- Canola exports represent 75% of our annual production and bring over $2 billion back to our Canadian economy
Genome Canada Projects

• I. Functional Genomics of Abiotic Stress, Bill Crosby, U of Windsor, Genome Prairie

• II. Enhancing Canola Through Genomics, Wilf Keller, NRC-PBI, Genome Prairie

• II. Functional Genomics of Arabidopsis, John Coleman, University of Toronto, Ontario Genomics Institute

• II. The Canadian Potato Genome Project, Barry Flinn, University of New Brunswick & Sharon Regan, University of Carleton, Genome Atlantic
**Genome Canada Projects**

- III. Designing Oilseeds for Tomorrow’s Markets, Randall Weselake, U of Alberta and Gopalan Selvaraj, NRC-PBI, Genome Alberta
- III. Use of Genomic Tools for Crop Improvements in Temperate Climates, D. Brian Fowler, University of Saskatchewan, Genome Prairie
- GrapeGen - A Genomic Approach to the Identification of the Genetic and Environmental Components Underlying Berry Quality in Grapevine, Steven T. Lund, University of British Columbia
Federal Genomics Initiatives

NRC-Genomics and Health Initiative
• focus on canola (*Brassica napus*)
  – Seed Development
  – Seed Metabolism
  – Seed Quality

AAFC Crop genomics initiatives
• focus on canola, wheat, soybean and corn
  – Disease and insect resistance
  – Tolerance to stresses such as cold and drought
  – Enhanced quality attributes

Development of genomics resources
• ESTs, DNA arrays, BAC libraries, mutagenized populations etc.
**Arabidopsis thaliana vs Brassica napus**

<table>
<thead>
<tr>
<th></th>
<th>Arabidopsis thaliana</th>
<th>Brassica napus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chromosomes</td>
<td>n=5</td>
<td>n=19</td>
</tr>
<tr>
<td>Genome size</td>
<td>130 Mb</td>
<td>1200 Mb</td>
</tr>
<tr>
<td>Expressed genes</td>
<td>27,000</td>
<td>80,000-120,000</td>
</tr>
<tr>
<td>DNA Homology</td>
<td></td>
<td>85%</td>
</tr>
</tbody>
</table>
### Brassica Seed Components

<table>
<thead>
<tr>
<th>Components</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oils</td>
<td>45</td>
</tr>
<tr>
<td>Proteins</td>
<td>22</td>
</tr>
<tr>
<td>Carbohydrates</td>
<td>20</td>
</tr>
<tr>
<td>Water</td>
<td>9</td>
</tr>
<tr>
<td>others</td>
<td>4</td>
</tr>
</tbody>
</table>
Brassica Seed

Example of targets:
- Enhanced oil content
- Modified proteins
- Modified oils
- Reduced anti-nutritional factors
Seed storage proteins

Inhibition of Seed Storage Proteins
Hegedus et al

Gene expression and organisation
Sharpe et al

Agriculture and Agri-Food Canada
Agroalimentaire Canada
B. napus seed storage proteins

Napin (2S albumin) and Cruciferin (12S globulin)

Major types of seed storage proteins in B. napus

20% and 60% of total protein in mature seed
Protein Production in Seeds

Basic Requirements to Realize this Potential:

• Create space within the seed for new proteins

• Accumulate protein in a location where:
  – they are functional
  – protected from degradation
The “Empty” Seed Concept

80% of Napin, Cruciferin, Oleosin, and Dehydrin can be reduced or eliminated, allowing replacement with new proteins that offer more utility or value.
Creating Space for New Proteins
No ill-Effect of Storage Protein Depletion

Normal

CRU -

Napin -
Gene expression and organization in B. napus

- 24 *B. napus* napin genes - 4 families
- 18 *B. napus* cruciferin genes - 4 families
- 2 closely related cruciferin genes are most highly expressed
- 2 divergent napins are most highly expressed
- *B. napus* cruciferins have similar physical organisation to Arabidopsis orthologue
- *B. napus* napins have different physical organisation to Arabidopsis orthologues
Brassica Seed Development and Metabolism
Brassica Seed ESTs

- ~ 436,000 seed ESTs
  - Various stages of embryo development ~145,000
  - Early seed stages under stress conditions ~34,300
  - Microspore embryogenesis ~29,000
  - Germination ~48,900
  - Dormancy ~33,200
  - Seed Coat ~121,800
  - Endosperm ~23,500

> All ESTs were deposited in GenBank

Team:
R. Datla, F Georges, J Krochko, A Cutler, E Tsang, G Selvaraj and J Zou
Proteomic analysis of endosperm with heart stage embryo

Zou et al

Endosperm in ovules containing heart embryos exhibit all phases of endosperm development (Brown et al., 1999)
Proteome of the *B. napus* endosperm

- Total proteins identified: 809
- Chloroplast gene products: 17
- Mitochondrial gene products: 11
- Genes with no ESTs (matching BNAEN3GH): 149
- Genes with 4 or more EST members: 348
Notable entries in the proteome (Endosperm)

- Cell wall invertase
- Invertase/pectinase inhibitor
- Cytosolic beta-amylase
- Myo-inositol 1-phosphate synthase
- Various components of Golgi vesicle-mediated transport
- Heme transporter
- Lipid metabolism enzymes: fatty acyl desaturase; acyltransferase
- 13 entries correspond to embryo defective genes
B. napus Seed Coat structure and development

Selvaraj et al.
Proteomic Analysis of Black and Yellow B. napus Seed Coat

Questions to be addressed:

• What proteins constitute the total proteomes of black and yellow B. napus seed coats?

• Is there a discernable difference between the seed coat proteomes of black (N89.53) and yellow (YN01.429) B. napus seeds?
  – Do these identifiable differences contribute to the phenotypic differences?
Proteomic Analysis of *B. napus* Seed Coat Protein Extracts

Black Seed Coat

Yellow Seed Coat

Protein identification

Protein Comparison
Brassica embryo/seed development

Datla et al
B. napus Embryo Proteins

<table>
<thead>
<tr>
<th>Globular</th>
<th>Heart</th>
<th>Torpedo</th>
<th>Bent</th>
<th>Mature</th>
<th>Stages</th>
</tr>
</thead>
<tbody>
<tr>
<td>700</td>
<td>600</td>
<td>40</td>
<td>8</td>
<td>2</td>
<td>#Embryos</td>
</tr>
</tbody>
</table>

kD

750 600 40 8 2
Protein Profiles of Key Embryo Stages in *B. napus*

![Bar chart showing the number of proteins identified in different embryo stages: Globular (992), Heart (1747), Torpedo (1070), Bent (1200), Mature (428).]
Proteome of *B. napus* and *Arabidopsis* embryos (globular to mature stages)

60% overlap between identified *Arabidopsis* and *Brassica* embryo proteins
Functional Classification of Embryo Expressed Proteins
(Molecular Function)

<table>
<thead>
<tr>
<th>Absolute count</th>
<th>Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>332</td>
<td>CATALYTIC ACTIVITY</td>
</tr>
<tr>
<td>217</td>
<td>NUCLEOTIDE BINDING</td>
</tr>
<tr>
<td>212</td>
<td>MOLECULAR_FUNCTION UNKNOWN</td>
</tr>
<tr>
<td>210</td>
<td>HYDROLASE ACTIVITY</td>
</tr>
<tr>
<td>191</td>
<td>STRUCTURAL MOLECULE ACTIVITY</td>
</tr>
<tr>
<td>148</td>
<td>TRANSFERASE ACTIVITY</td>
</tr>
<tr>
<td>113</td>
<td>TRANSPORTER ACTIVITY</td>
</tr>
<tr>
<td>98</td>
<td>BINDING</td>
</tr>
<tr>
<td>95</td>
<td>DNA BINDING</td>
</tr>
<tr>
<td>72</td>
<td>RNA BINDING</td>
</tr>
<tr>
<td>57</td>
<td>CHAPERONE ACTIVITY</td>
</tr>
<tr>
<td>54</td>
<td>PROTEIN BINDING</td>
</tr>
<tr>
<td>49</td>
<td>NUCLEIC ACID BINDING</td>
</tr>
<tr>
<td>45</td>
<td>TRANSLATION FACTOR ACTIVITY, NUCLEIC ACID BINDING</td>
</tr>
<tr>
<td>39</td>
<td>KINASE ACTIVITY</td>
</tr>
<tr>
<td>28</td>
<td>MOLECULAR_FUNCTION</td>
</tr>
<tr>
<td>21</td>
<td>RECEPTOR ACTIVITY</td>
</tr>
<tr>
<td>16</td>
<td>ENZYME REGULATOR ACTIVITY</td>
</tr>
<tr>
<td>13</td>
<td>LIPID BINDING</td>
</tr>
<tr>
<td>11</td>
<td>MOTOR ACTIVITY</td>
</tr>
<tr>
<td>11</td>
<td>TRANSCRIPTION FACTOR ACTIVITY</td>
</tr>
<tr>
<td>9</td>
<td>SIGNAL TRANSDUCER ACTIVITY</td>
</tr>
<tr>
<td>7</td>
<td>TRANSCRIPTION REGULATOR ACTIVITY</td>
</tr>
<tr>
<td>4</td>
<td>NUCLEASE ACTIVITY</td>
</tr>
<tr>
<td>4</td>
<td>RECEPTOR BINDING</td>
</tr>
<tr>
<td>2</td>
<td>CHROMATIN BINDING</td>
</tr>
<tr>
<td>2</td>
<td>OXYGEN BINDING</td>
</tr>
</tbody>
</table>
Integrated *Brassica napus* and *Arabidopsis* Embryo Metabolic Network

- **TWN2**
- **EMB2728**
- **TWN2**
- **EMB2024**
- **PLE2**
- **RML1**
- **EMB1144**
- **EMB1276**
- **EMB2757**
- **EMB1989**
- **EMB2284**
- **EMB1395**

- **Core reaction**
- **Non-core reaction**
- **Embryo defective**

- **Common pathway**
- **Brassica**
- **Arabidopsis**
Brassica Seed Development and Metabolism

Example of targets

- Large seed size
- Uniform seed germination
- Reduced coat thickness
Crop Genomics for a Healthy Canada
Agriculture/Plants

Genome Canada Position paper
Keller et al
Agriculture/Plants
(Crop Genomics for a Healthy Canada)

Development of Position Paper:

**Science Steering Committee:**
Faouzi Bekkaoui (NRC-PBI), Bruce Coulman (University of Saskatchewan), Adrian Cutler (NRC-PBI), Raju Datla (NRC-PBI), Wilf Keller (NRC-PBI), Isobel Parkin (AAFC-Saskatoon), Jas Singh (AAFC-Ottawa)

**Writer:**
Michael Raine

**Sponsoring Centre – Genome Prairie:**
Jerome Konesnzi (CEO), Reno Pontarollo (CSO), Faye Pagdonsolan (Admin Support)

**Workshop:**
May 31, 2007, Chateau Laurier Hotel, Ottawa
Approximately 50 participants: Academia, Government, Industry

**Facilitator:**
Pete Desai (Desai & Desai Inc.)
Position Paper: Crop Genomics for a Healthy Canada

Scope:

National: cover major Canadian agricultural crops

Key Objectives:

• Enhanced crop yield/productivity
• Enhanced health/nutritional value
• Enhanced tolerance to abiotic and biotic stress
• Environmentally friendly, sustainable production
• Diversification: support biofuel, bioproduct development
Potential Comp IV Projects – Based on Species

- *Brassica* oilseeds
- Flax
- Pulse crops (lentils, peas, chickpeas, beans)
- Wheat
- Barley
- Oats
- Corn
- Soybean
- Alfalfa (and other forages)
- Potatoes
- *Prunus* spp (cherries, etc.)
- Blueberries
- New platform development (e.g. *Camelina, Triticale*)

Questions:
- Genomics resource status?
- Overall value of crop?
- Critical mass (expertise)?
- Receptor capacity (i.e. breeding)? – Breeders must play essential role
Potential Comp IV Projects
Based on Themes

- Abiotic stress (e.g. heat/drought)
- Biotic stress (e.g. Fusarium)
- Plant productivity/yield (e.g. architecture, heterosis)
- Bioproduct development (industrial oils)
- Health product development
- Sustainability (e.g. nutrient use efficiency)
- Comparative genomics
Canola as an Example: Well Positioned for Genomics Research Projects

- $11B industry
- ~75 public sector scientists (~300 researchers)
- Active private sector (>10 companies)
- History of genetic research in Canada
- >600,000 ESTs (made in Canada)
- Microarrays (3 made in Canada)
- TILLING populations (UBC, AAFC)
- International *B. rapa* Genome Sequencing Initiative
- Arabidopsis data is relevant
Acknowledgements

For additional information www.brassicagenomics.ca

- Wilf Keller
- Raju Datla, Daoquan Xiang, Edwin Wang
- Jitao Zou
- Gopalan Selvaraj, Paula Ashe, John Kelly
- Andrew Sharpe, Dwayne Hegedus, Kevin Razwodowski, Derek Lydiate, Isobel Parkin
- DOTM and ECTG projects teams