Large-scale RNA-Seq Transcriptomics Studies in *P. tremula*: Projects and Resource Development

Nathaniel Street
The UmAsp Sex Project

Agilent arrays: 10 male, 10 female (2008)

RNA-Seq: 8 male, 8 female + 4 poplars (2008+2010)

Phenotypic data: leaf traits + arthropods (2010)
Signal or Noise?

gw1.III.2363.1
• V1.0 – present
• V1.1 – absent but region in assembly
• V2.x – region absent in assembly

Multiple analysis methods
• Univariate
• Machine learning

RNA-Seq
Read mapping

- Mapping to *P. trichocarpa* transcriptome
  - Cufflinks alignment to genome results in unstable expression profiles
  - RSEM much more reliable

- Raw data:
  - ~25% map to *P. trichocarpa*
  - ~50% map to *P. tremula*

- With QC trimming:
  - Over 80% map to *P. tremula*
  - ~50% map to *P. trichocarpa*
RNA-Seq Sample Network

- Sex ✗
- Year ✗
- Species ✔
Sexual Equality

Leaf area

Arthropod abundance
The AspWood Project

• 14 group leaders involved:
  • Catherine Bellini, Rishi Bhalerao, Stefan Jansson, Jan Karlsson, Ewa Mellerowicz, Thomas Moritz, Totte Niittyla, Ove Nilsson, Johan Trygg, Hannele Tuominen; Gunnar Wingsle, Nathaniel Street, Björn Sundberg, Torgeir Hvidsten

• Financial support:
  • The SSF centre for Developmental Biology of Plants
  • UPSC Berzelii Centre for Forest Biotechnology
Sampling:
- 30 sections from phloem to cell death
- five trees

Reverse engineering

RNA-Seq
GC- and LC-MS

Phloem
Camellia
Xylem

Division
Expansion
Maturation
Cell death

Transverse Section
Sample section
30µm
17mm
2-2.5mm
The trees

- July 7th, 2010 at 10.00-12.00. Vindeln/Mullkälen.
- 10 trees, 15 m high, about 47 years old.
- Leaves were also sampled for genotyping: genetically identical
- Propagated from roots for leaf developmental series (2011)
Samples

- Tangential cryo-sections of 20 μm sampled from phloem to cell death
- Sections for mRNA sequencing are pooled into 30 samples
- Sections for small RNA sequencing are pooled into 5 samples
- Min 20 M mRNA RNA-Seq read pairs
- Current analysis based on alignment to *P. trichocarpa* transcriptome

<table>
<thead>
<tr>
<th>Sections #</th>
<th>Sample #</th>
<th>Sample Name</th>
<th>sRNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>1--3</td>
<td>3</td>
<td>K9-00</td>
<td>Old Phloem</td>
</tr>
<tr>
<td>4--8</td>
<td>5</td>
<td>K9-01</td>
<td>Phloem</td>
</tr>
<tr>
<td>9--9</td>
<td>1</td>
<td>K9-02</td>
<td>Cambium1</td>
</tr>
<tr>
<td>10--10</td>
<td>1</td>
<td>K9-03</td>
<td>Cambium2</td>
</tr>
<tr>
<td>11--11</td>
<td>1</td>
<td>K9-04</td>
<td>Cambium3</td>
</tr>
<tr>
<td>12--12</td>
<td>1</td>
<td>K9-05</td>
<td>Cambium4</td>
</tr>
<tr>
<td>13--13</td>
<td>1</td>
<td>K9-06</td>
<td>Cambium5</td>
</tr>
<tr>
<td>14--14</td>
<td>1</td>
<td>K9-07</td>
<td>Cambium6</td>
</tr>
<tr>
<td>15--15</td>
<td>1</td>
<td>K9-08</td>
<td>Cambium7</td>
</tr>
<tr>
<td>16--16</td>
<td>1</td>
<td>K9-09</td>
<td>Cambium8</td>
</tr>
<tr>
<td>17--19</td>
<td>3</td>
<td>K9-10</td>
<td>Exp/DevXyl1</td>
</tr>
<tr>
<td>20--22</td>
<td>3</td>
<td>K9-11</td>
<td>Exp/DevXyl2</td>
</tr>
<tr>
<td>23--25</td>
<td>3</td>
<td>K9-12</td>
<td>Exp/DevXyl3</td>
</tr>
<tr>
<td>26--28</td>
<td>3</td>
<td>K9-13</td>
<td>Exp/DevXyl4</td>
</tr>
<tr>
<td>29--31</td>
<td>3</td>
<td>K9-14</td>
<td>Exp/DevXyl5</td>
</tr>
<tr>
<td>32--34</td>
<td>3</td>
<td>K9-15</td>
<td>Exp/DevXyl6</td>
</tr>
<tr>
<td>35--37</td>
<td>3</td>
<td>K9-16</td>
<td>Exp/DevXyl7</td>
</tr>
<tr>
<td>38--40</td>
<td>3</td>
<td>K9-17</td>
<td>Exp/DevXyl8</td>
</tr>
<tr>
<td>41--43</td>
<td>3</td>
<td>K9-18</td>
<td>Exp/DevXyl9</td>
</tr>
<tr>
<td>44--46</td>
<td>3</td>
<td>K9-19</td>
<td>Exp/DevXyl10</td>
</tr>
<tr>
<td>47--55</td>
<td>9</td>
<td>K9-20</td>
<td>MatXyl1</td>
</tr>
<tr>
<td>56--64</td>
<td>9</td>
<td>K9-21</td>
<td>MatXyl2</td>
</tr>
<tr>
<td>65--73</td>
<td>9</td>
<td>K9-22</td>
<td>MatXyl3</td>
</tr>
<tr>
<td>74--82</td>
<td>9</td>
<td>K9-23</td>
<td>MatXyl4</td>
</tr>
<tr>
<td>83--91</td>
<td>9</td>
<td>K9-24</td>
<td>MatXyl5</td>
</tr>
<tr>
<td>92--100</td>
<td>9</td>
<td>K9-25</td>
<td>MatXyl6</td>
</tr>
<tr>
<td>101--109</td>
<td>9</td>
<td>K9-26</td>
<td>MatXyl7</td>
</tr>
<tr>
<td>110--118</td>
<td>9</td>
<td>K9-27</td>
<td>MatXyl8</td>
</tr>
<tr>
<td>119--127</td>
<td>9</td>
<td>K9-28</td>
<td>MatXyl9</td>
</tr>
<tr>
<td>128--135</td>
<td>8</td>
<td>K9-29</td>
<td>MatXyl10</td>
</tr>
</tbody>
</table>
Expression consistency

Correlations for gene POPTR_0004s23690

- K2 K3 0.97
- K1 K9 0.96
- K1 K2 0.96
- K3 K9 0.96
- K2 K9 0.95
- K1 K3 0.95
- K1 K5 0.88
- K3 K5 0.88
- K2 K5 0.85
- K5 K9 0.82

With at least one mapped read:

- Genes: 29148
- Transcription factors: 1467

With consistent expression and std. dev > 0.75:

- Genes: 8747
- Transcription factors: 521
Example:
POPTR_0004s23690, Median = 0.95
mRNA: Genes

Hertzberg et al, (2001), PNAS, 98,14732–14737
32 lignin genes

Phenylalanine ammonia-lyase

MYB

Peroxidase

http://statgen.ncsu.edu/lignin/
mRNA: Transcription factors

521 transcription factors
Network Analysis
Network Patterns

Street et al. 2008
Comparative regulomics

Identify conserved regulatory control
- Public data (arabidopsis, rice, poplar)
- Spruce Genome Project
- UPSC aspen projects
Spruce Genome Project

- Phase two projects
  - Wood development series (90)
  - Compression wood formation (90)
  - Diurnal expression (72)
  - Seedling development and megagametophyte (30)
  - Somatic culture $\rightarrow$ differentiation series (50)
  - Tracheary element differentiation (54)
PopGenIE – Case study

• All UPSC RNA-Seq data will be made available and integrated into PopGenIE (and spruce into ConGenIE)

• Public Galaxy server
• Community based functional annotation
• Literature incorporation

• Example use:
  • POPTR_0019s13500: secretory carrier-associated membrane protein
  • Part of the BioImprove collection
Acknowledgements

• Torgeir Hvidsten
  • Sergiu Netotea
  • Jenny Önskog

• Björn Sundberg

• Stefan Jansson
• Benedicte Albrectsen
  • Kathryn Robinson

• Pär Ingvarsson
  • Agneta Andersson

• SciLifeLab, Uppmax
• Spruce Genome Assembly Team