Comparative mapping in *Salicaceae*: a tool for identifying important genes controlling adaptive traits

Véronique Jorge, Francesco Fabbrini, Patricia Faivre-Rampant, Matthias Fladung, Muriel Gaudet, Ulf Lagercrantz, Jean-Charles Leplé, Michele Morgante, Birte Pakull, Maurizio Sabatti, Véronique Storme, Gail Taylor, Nicola Vitacolonna, Jennifer De Woody, Catherine Bastien
Comparative QTL mapping

Definition of comparative mapping
Alignment of chromosomes of related species based on genetic mapping of common DNA markers

Goal
Identification of conserved/specific genomic regions.

Specific questions
How cumulated QTL data based on heterogeneous mapping experiments could help functional annotation of genomes and identification of candidate genes?

Strategies
✓ Simple graphical comparison
✓ Map construction and QTL analysis based on multiple mapping experiment (on raw data)
✓ Consensus and meta-analysis (on position data)
Methods for comparative mapping

Simple graphical comparison: matching marker names (Cmap)

Rough identification of linkage groups/chromosomes of interest

Data from http://www.evoltree.soton.ac.uk/cmap/

Forest ecosystems genomics and adaptation, 9-11 June 2010, San Lorenzo de El Escorial
Methods for comparative mapping

Map construction and QTL analysis based on multiple mapping experiment (on raw data)

Marker and phenotypic data on pedigree 1
010001101110101
110111101100110
11010011010001
11011101100110
11011101100110
25 35 21 18 20
25 26 30 21 20

Marker and phenotypic data on pedigree 2
010001101110101
110111101100110
11010011010001
11011101100110
11011101100110
25 35 21 18 20
25 26 30 21 20

- Raw data not easily available
- Connected pedigrees often needed

Blanc et al. 2006 TAG 113:206–224

Forest ecosystems genomics and adaptation, 9-11 June 2010, San Lorenzo de El Escorial
Methods for comparative mapping

Consensus map and QTL meta-analysis based on position data

- From elaborated data (maps, position and QTL confidence intervals)
- Meta-analysis: combine data from different sources in a single study
- Widely used in medical and social science

Advantages: elaborated data retrieved from published data

Steps for QTL meta-analysis

Construction of the consensus map

Projection of QTLs on the consensus map

QTL meta-analysis

For each chromosome
Simulations and analysis of real data sets showed that the **confidence intervals of the meta-QTLs are considerably reduced** compared to the observed QTLs.

**Maize example (Truntzler et al. submitted):**

- 34 QTLs for early flowering components from 15 papers.
- The best model $k = 5$ meta-QTLs.
- Colors -> quantitative assignation of QTL -> 75% of observed QTLs correspond to vgt1 and/or vgt2.
- Confidence intervals reduced: vgt1 ($\sim 4cM$) include a fine mapped QTL (Salvi et al., 2000).
Create a consensus map

In general, all genetic maps need to be connected.

Example in *Populus*: case of a LG well covered by SSR (LG_VIII)
Markers containing sequence and position information (genome):

1) SSR markers (http://www.ornl.gov/sci/ipgc/ssr_resources.htm):
   4 166 primer pairs available.

2) SNP located in genes
   - evenly distributed on the genome
   - in specific region (QTL) involved in adaptive traits (Phenology, drought resistance ...)
   - functional candidates (transcriptomics, ...).
Mapping data for *Salicaceae*

Genetic maps connected to the genome

8 different species involved in 13 intra and interspecific crosses:

(*Populus trichocarpa, P. deltoides, P. nigra, P. alba, P. angustifolia, P. fremontii, Salix vinimalis, S. schwerinii*) ...

<table>
<thead>
<tr>
<th>Genus</th>
<th>Family name – Pedigree type</th>
<th>Progeny size</th>
<th>Reference(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Family 13 – BC1 TD x D</td>
<td>171</td>
<td>Yin et al 2004</td>
<td></td>
</tr>
<tr>
<td>Family 331 - F2 TDxTD (POP1)</td>
<td>210</td>
<td>Tuskan et al.</td>
<td></td>
</tr>
<tr>
<td>Family 545 – F1 TxD</td>
<td>101</td>
<td>Yin et al 2008</td>
<td></td>
</tr>
<tr>
<td>Family XXX – D x DN</td>
<td>93</td>
<td>Yin et al 2008</td>
<td></td>
</tr>
<tr>
<td>Populus</td>
<td>Family BCN - A x (FxA)</td>
<td>246</td>
<td>Yin et al 2008</td>
</tr>
<tr>
<td></td>
<td>POP2 - DxD</td>
<td>330</td>
<td>Jorge et al. (2005, +in prep)*</td>
</tr>
<tr>
<td></td>
<td>POP3a - DxD</td>
<td>180</td>
<td>Cervera et al.(2001)*</td>
</tr>
<tr>
<td></td>
<td>POP3b - DxD</td>
<td>182</td>
<td>Cervera et al.(2001)*</td>
</tr>
<tr>
<td></td>
<td>POP4 - AxA</td>
<td>141</td>
<td>Paolucci et al. (unpublished)*</td>
</tr>
<tr>
<td></td>
<td>Family 52-124 - (TD)xD</td>
<td>396</td>
<td>Dorst at al 2009</td>
</tr>
<tr>
<td></td>
<td>S1 – [Vx(VS)] x V</td>
<td>463</td>
<td>Berlin et al 2010</td>
</tr>
<tr>
<td>Salix</td>
<td>S3 - VxV</td>
<td>282</td>
<td>Berlin et al 2010</td>
</tr>
<tr>
<td></td>
<td>K8 - [Vx(VS)] x [Vx(VS)]</td>
<td>471</td>
<td>Hanley et al 2006</td>
</tr>
</tbody>
</table>
Mapping data for *Salicaceae*

...but few framework genetic maps with genome anchoring markers and QTL data (growth related traits)

<table>
<thead>
<tr>
<th>Pedigree</th>
<th>Progeny size</th>
<th>Nb of anchoring markers</th>
<th># of QTLs</th>
<th>Reference(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>F2 TDxTD (POP1)</td>
<td>210</td>
<td>83</td>
<td>188, (37, 43)</td>
<td>Rae et al 2008, (2009)*</td>
</tr>
<tr>
<td>DxT (POP2)</td>
<td>330</td>
<td>74, 87</td>
<td>77</td>
<td>Jorge et al. (unpublished)</td>
</tr>
<tr>
<td>DxT, DxN (POP3)</td>
<td>180, 182</td>
<td>19, 14, 14, 16</td>
<td>21, 29, 48, 20</td>
<td>Dillen et al, 2009*</td>
</tr>
<tr>
<td>AxA (POP4)</td>
<td>141</td>
<td>55, 51</td>
<td>10, 11</td>
<td>Paolucci et al. (unpublished)*</td>
</tr>
<tr>
<td>(TxD)xD</td>
<td>396</td>
<td>178</td>
<td>63</td>
<td>Dorst at al 2009</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Novaes et al. 2009</td>
</tr>
</tbody>
</table>

*Data from http://www.evoltree.soton.ac.uk/cmap/
Lack of transferability of markers between genetic maps

<table>
<thead>
<tr>
<th></th>
<th>Populus spp.</th>
<th>Salix spp.</th>
</tr>
</thead>
<tbody>
<tr>
<td>10/05/2010</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of shared markers between genetic maps</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Berlin_S1_2010</td>
</tr>
<tr>
<td></td>
<td></td>
<td>193</td>
</tr>
<tr>
<td>Hanley_2006</td>
<td></td>
<td>48</td>
</tr>
<tr>
<td></td>
<td>Populus spp.: on average 20% fewer for P. alba pedigree (distant taxus).</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Salix spp.: 16%</td>
<td></td>
</tr>
</tbody>
</table>

**Transferability**
P.t genome -> maps

- **Populus spp.**
  - Average 20% fewer for P. alba pedigree (distant taxus).
- **Salix spp.**
  - 16%
Lack of transferability of markers between genetic maps

SSR markers
4 166 primer pairs available
(http://www.ornl.gov/sci/ipgc/ssr_resources.htm):

« Polymorphism bottleneck »

<table>
<thead>
<tr>
<th>Pedigrees</th>
<th>POP2 (DxT)</th>
<th>POP4 (AxA)</th>
<th>POP5 (NxN)</th>
</tr>
</thead>
<tbody>
<tr>
<td># SSR tested</td>
<td>317</td>
<td>273</td>
<td>386</td>
</tr>
<tr>
<td># SSR not amplifying/monomorphic</td>
<td>144</td>
<td>170</td>
<td>259</td>
</tr>
<tr>
<td># SSR polymorphic/mapped</td>
<td>173</td>
<td>103</td>
<td>127</td>
</tr>
</tbody>
</table>

Forest ecosystems genomics and adaptation, 9-11 June 2010, San Lorenzo de El Escorial
Lack of transferability of markers between genetic maps

Substitution

P. nigra SNP

P. deltoides SNP

SNP are highly specific
Good anchoring on the genome sequence

Status of the alignment of Salicaceae genetic maps on the *Populus trichocarpa* genome

*Populus* spp.

Salix spp.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>178</td>
<td>83</td>
<td>19</td>
<td>14</td>
<td>14</td>
<td>16</td>
<td>74</td>
<td>87</td>
<td>55</td>
<td>51</td>
<td>259</td>
<td>177</td>
<td>139</td>
</tr>
</tbody>
</table>

Number of anchoring markers between genetic maps and the *P. trichocarpa* genome 10/05/2010

Forest ecosystems genomics and adaptation, 9-11 June 2010, San Lorenzo de El Escorial
Strategy for comparative QTL mapping in *Salicaceae*

Anchoring to the genome sequence and bp/cM ratio calculation

Relationship between pair wise genetic and physical distances between genome anchored markers

\[ y = R \times x \]

Number of bp / cM
Strategy for comparative QTL mapping in *Salicaceae*

Calculation of start and stop of the QTL projected on the genome using the nearest markers and R (projection geometry)

\[
\text{Start(pb)} = 11238978 - \left[|26.4 - 7.9| \times R\right] = 7 \, 142 \, 811 \, \text{pb}
\]

\[
\text{Stop(pb)} = 15003259 - \left[|43.4 - 36.6| \times R\right] = 13 \, 497 \, 405 \, \text{pb}
\]
Strategy for comparative QTL mapping in *Salicaceae*

QTL projection from each map to the genome sequence thanks to genome anchoring markers

Meta-analysis

Future development of Biomercator3, Joets et al. In development

San Lorenzo de El Escorial
Example of physical QTL intervals calculation and corresponding number of genes

Leaf nitrogen content:
8 QTL, 9.7% - 34.7% of the total PVE

---

Functional analysis of large gene lists
Functional analysis of large gene lists

Annotation use a classification based on Gene Ontology (GO)

- Molecular function (ex: oxidoreductase activity, ...)
- Biological process (ex: photosynthesis light harvesting, ...)
- Cellular component (ex: membrane, chloroplast ...)

Forest ecosystems genomics and adaptation, 9-11 June 2010, San Lorenzo de El Escorial
Functional analysis of large gene lists

Methods:
Bioinformatic Enrichment Tools

- gene list 1
  - Annotation: Gene ontology
  - Comparison: Statistical tests
  - Significant GO terms

- gene list 2
  - Annotation: Gene ontology

Statistical tests
Significant GO terms

## Functional analysis of large gene lists

5 QTLs explaining 2.8 to 9% of the variation for nb of sylleptic branches

2906 genes included in the 5 QTL confidence intervals

45 555 genes in the genome sequence v1.1 (~60% have GO terms)

<table>
<thead>
<tr>
<th>ID</th>
<th>Name</th>
<th>p-Value</th>
<th>p-Value (Adj)</th>
<th>Study Count</th>
<th>Population Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>GO:0010422</td>
<td>regulation of brassinosteroid biosynthetic process</td>
<td>0.0000</td>
<td>0.0000</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>GO:0010023</td>
<td>proanthocyanidin biosynthetic process</td>
<td>0.0000</td>
<td>0.0000</td>
<td>6</td>
<td>10</td>
</tr>
<tr>
<td>GO:0009641</td>
<td>shade avoidance</td>
<td>0.0000</td>
<td>0.0000</td>
<td>6</td>
<td>11</td>
</tr>
<tr>
<td>GO:0016595</td>
<td>glutamate binding</td>
<td>0.0001</td>
<td>0.0001</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>GO:0009741</td>
<td>response to brassinosteroid stimulus</td>
<td>0.0003</td>
<td>0.0003</td>
<td>10</td>
<td>38</td>
</tr>
<tr>
<td>GO:0009788</td>
<td>negative regulation of abscisic acid mediated signaling pathway</td>
<td>0.0006</td>
<td>0.0006</td>
<td>5</td>
<td>11</td>
</tr>
<tr>
<td>GO:0006122</td>
<td>mitochondrial electron transport, ubiquinol to cytochrome c</td>
<td>0.0006</td>
<td>0.0006</td>
<td>5</td>
<td>11</td>
</tr>
<tr>
<td>GO:0004406</td>
<td>H3/H4 histone acetyltransferase activity</td>
<td>0.0008</td>
<td>0.0008</td>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td>GO:0048830</td>
<td>adventitious root development</td>
<td>0.0008</td>
<td>0.0008</td>
<td>4</td>
<td>7</td>
</tr>
</tbody>
</table>
Comparison between linkage maps and QTLs on Linkage Group X

Functional analysis of large gene lists
Functional analysis of large gene lists

**Gene list 1**
QTL set
2337 genes included in LG_X

**Gene list 2**
Genome set
45 555 genes in the genome sequence v1.1 (~60% have GO terms)

<table>
<thead>
<tr>
<th>ID</th>
<th>Name</th>
<th>P value</th>
<th>P value (Adj.)</th>
<th>Study count</th>
<th>Pop. count</th>
</tr>
</thead>
<tbody>
<tr>
<td>GO:0008825</td>
<td>cyclopropane-fatty-acyl-phospholipid synthase activity</td>
<td>5.88E-05</td>
<td>5.88E-05</td>
<td>5</td>
<td>9</td>
</tr>
<tr>
<td>GO:0005575</td>
<td>cellular_component</td>
<td>9.97E-05</td>
<td>9.97E-05</td>
<td>1258</td>
<td>21116</td>
</tr>
<tr>
<td>GO:000287</td>
<td>magnesium ion binding</td>
<td>3.06E-04</td>
<td>3.06E-04</td>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td>GO:0008150</td>
<td>biological_process</td>
<td>7.95E-04</td>
<td>7.95E-04</td>
<td>1433</td>
<td>24560</td>
</tr>
<tr>
<td>GO:0009536</td>
<td>plastid</td>
<td>0.001270407</td>
<td>0.001270407</td>
<td>232</td>
<td>3413</td>
</tr>
<tr>
<td>GO:0009507</td>
<td>chloroplast</td>
<td>0.001524634</td>
<td>0.001524634</td>
<td>226</td>
<td>3327</td>
</tr>
<tr>
<td>GO:0044444</td>
<td>cytoplasmic part</td>
<td>0.001912735</td>
<td>0.001912735</td>
<td>400</td>
<td>6246</td>
</tr>
<tr>
<td>GO:0044275</td>
<td><strong>cellular carbohydrate catabolic process</strong></td>
<td>0.002304709</td>
<td>0.002304709</td>
<td>19</td>
<td>164</td>
</tr>
<tr>
<td>GO:0005982</td>
<td><strong>starch metabolic process</strong></td>
<td>0.002975107</td>
<td>0.002975107</td>
<td>7</td>
<td>35</td>
</tr>
<tr>
<td>GO:0044265</td>
<td>cellular macromolecule catabolic process</td>
<td>0.003129754</td>
<td>0.003129754</td>
<td>38</td>
<td>422</td>
</tr>
</tbody>
</table>

*Forest ecosystems genomics and adaptation, 9-11 June 2010, San Lorenzo de El Escorial*
Future developments
Combining meta QTL analysis and gene enrichment approaches

• Biomercator3: projection of QTLs on the genome and QTL meta-analysis

• Gene list analysis: tools crossing information from GO + bibliography + biochemical pathways + various omics ...
Thank you!