Identification and characterisation of metabolic QTL in *Arabidopsis thaliana*

—

Metabolome analyses of heterosis in maize roots

Bioen Workshop on Metabolomics of Sugarcane
Jan Lisec, 07.12.2009
Metabolomics allows a new perspective on plant systems

- Metabolite Quantitative Trait Loci (mQTL) enable to connect phenotype and genotype in *Arabidopsis thaliana*
  - Recombinant Inbred Line and Introgression Line populations
  - Heterosis and quantitative genetics
  - Identification and characterization of (heterotic) mQTL
  - Relation of metabolic profiles and plant biomass

- Metabolome analyses suggest a different view on heterosis in maize roots
  - Metabolic correlation networks
  - Heterosis on the molecular and whole plant level

lisec@mpimp-golm.mpg.de

São Paulo, 07.12.2009
Plants are the primary producers of biomass

- Food
- Clothing
- Health
- Cosmetics
- Decoration
- Renewable Resources
- Carbon Cycle
From seed to seed – monitoring plant development on a molecular level to better understand growth
Arabidopsis and maize are excellent model plants for quantitative genetics and heterosis, respectively

- short life cycle
- modest growth requirements
- small and fully sequenced genome
- knockout lines, transgenic lines
- plethora of available data
- of low economic importance

- more than 100% best-parent heterosis for yield
- heterosis effect agronomical used since more than 100 years
- of high economic importance
- difficult to cultivate in higher number in a controlled environment

Arabidopsis thaliana

Zea mais
GC-ToF-MS was used to measure metabolic profiles

- Metabolome: The entirety of small molecules within a cell.
- GC-MS: a combination of a separation method (Gas Chromatography) with an detection system (Mass Spectrometry)
- Metabolome mirrors the biological endpoint or actual state of a plant closer than its genes.

lisec@mpimp-golm.mpg.de

São Paulo, 07.12.2009
Data Processing for large scale metabolomics was established

- R-Script 1
  - ChromaTOF
  - TargetSearch
  - TagFinder
  - XCMS
  - SpectConnect
  - BinBase

- R-Script 2
  - ChromaTOF
  - TargetSearch
  - TagFinder
  - XCMS
  - SpectConnect
  - BinBase

- Allowed the evaluation of >2000 samples for 181 metabolic traces (85 known)
Recombinant Inbred Lines and Introgression Lines were used to identify Quantitative Trait Loci (QTL)

- Introgression Lines (Near Isogenic Lines)
- Recombinant Inbred Lines
- Test Crosses (TC)

\[ V_P = V_G + V_E + V_G \times V_E \]
Recombinant Inbred Lines and Introgression Lines were used to identify Quantitative Trait Loci (QTL)

• QTL indicates a region of the genome bearing loci which influence the phenotype of an observed trait (e.g. metabolite concentration or biomass)

• quantitative traits – usually polygenic inherited

IL population:
- 97 ILs (6 repl.)
- 45 IL-TCs (6 repl.)
- \(P_{1,2}, F_{1-a,1-b}\) (~50 repl.)

RIL population:
- 369 RILs (1 repl.)
- 735 RIL-TCs (1 repl.)
- \(P_{1,2}, F_{1-a,1-b}\) (~10 repl.)
A case study: *myo*-Inositol has 4 QTL

Lisec et al. (2008, Plant J)
São Paulo, 07.12.2009
A birds view on RIL QTL for known metabolites

- 157 mQTL (for 84 Metab.)
- max. 6 per Metab.
- PVE: 7.1% (1.7% - 52.1%)
- Hotspots at 4/4 and 5/76
- average QTL width: 6cM

---

**Lisec et al. (2008, Plant J)**

lise@mpimp-golm.mpg.de

Max Planck Institut für Molekulare Pflanzenphysiologie

São Paulo, 07.12.2009
Candidate genes for metabolic QTL were identified using AraCyc 3.5 database

**Inositol**

- # mQTL: 6
- # path: 6
- # dAGI: 12
- # pAGI: 42

**LOD**

Chr1 Chr2 Chr3 Chr4 Chr5

15 45 75 15 45 75 15 45 75 cM

- mQTL
- direct AGI
- pathway AGI

**IL confirmation level**

** approximate LOD
Metabolic QTL are enriched for enzyme candidate genes

24% direct candidate genes
67% pathway candidate genes

lisec@mpimp-golm.mpg.de

São Paulo, 07.12.2009
Introgression Line analysis confirms RIL QTL and reveals additional effects

Expected Differences:
- statistical power and resolution
- linked QTL (in coupling/ in repulsion)
- small effect QTL
- epistatic effects

<table>
<thead>
<tr>
<th>Significance level</th>
<th>Number of significant changes</th>
<th>FDR (%)</th>
<th>Number of confirmed RIL QTL</th>
<th>Confirmed RIL QTL (%)</th>
<th>Average $R^2$ of confirmed RIL QTL (%)</th>
<th>Average $R^2$ of non confirmed RIL QTL (%)</th>
<th>Confirmed allelic effect</th>
<th>Confirmed allelic effect (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.001</td>
<td>177</td>
<td>9.61</td>
<td>17</td>
<td>11.33</td>
<td>11.62</td>
<td>6.67</td>
<td>16</td>
<td>94</td>
</tr>
<tr>
<td>0.01</td>
<td>773</td>
<td>22.01</td>
<td>41</td>
<td>27.33</td>
<td>10.17</td>
<td>6.12</td>
<td>38</td>
<td>93</td>
</tr>
<tr>
<td>0.05</td>
<td>2511</td>
<td>33.88</td>
<td>83</td>
<td>55.33</td>
<td>7.79</td>
<td>6.54</td>
<td>68</td>
<td>82</td>
</tr>
<tr>
<td>0.1</td>
<td>3941</td>
<td>43.17</td>
<td>99</td>
<td>66.00</td>
<td>7.45</td>
<td>6.79</td>
<td>80</td>
<td>81</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Direct candidate gene</th>
<th>IL confirmation rate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>&gt;0.05</td>
</tr>
<tr>
<td>no</td>
<td>40</td>
</tr>
<tr>
<td>yes</td>
<td>10</td>
</tr>
</tbody>
</table>

Lisec et al. (2008, Plant J)

São Paulo, 07.12.2009

lisec@mpimp-golm.mpg.de
Resequencing of 8 mQTL candidate genes reveals amino acid substitutions

Candidate Genes to test selected according to:

- Confirmation level in IL-population
- Number of known genes in AraCyc 3.5
- Catalyzed reaction

<table>
<thead>
<tr>
<th>AG</th>
<th>LOD value</th>
<th>Substitution</th>
<th>Synonymus</th>
<th>Intron</th>
<th>bp</th>
<th>Metabolite</th>
<th>Gene function</th>
<th>LODRIL</th>
<th>PIL</th>
</tr>
</thead>
<tbody>
<tr>
<td>AT1G49520</td>
<td>5</td>
<td></td>
<td></td>
<td>1</td>
<td>189</td>
<td>Inositol</td>
<td>MIOX1 (myo-inositol oxygenase)</td>
<td>6.5</td>
<td>0.047</td>
</tr>
<tr>
<td>AT5G39570</td>
<td>1</td>
<td></td>
<td></td>
<td>0</td>
<td>2240</td>
<td>Tyrosine</td>
<td>tyrosine aminotransferase</td>
<td>9.6</td>
<td>0.054</td>
</tr>
<tr>
<td>AT1G43710</td>
<td>0</td>
<td></td>
<td>1</td>
<td>1</td>
<td>2064</td>
<td>Ethanolamine</td>
<td>glutamate decarboxylase</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT3G44740</td>
<td>0</td>
<td></td>
<td></td>
<td>0</td>
<td>1446</td>
<td>Glycine</td>
<td>glycyl-tRNA synthetase</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT4G15210</td>
<td>0</td>
<td></td>
<td>?</td>
<td>4</td>
<td>3132</td>
<td>Maltose</td>
<td>beta-amylase activity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT2G38400</td>
<td>0</td>
<td></td>
<td>?</td>
<td>2</td>
<td>2660</td>
<td>4-Aminobutyric acid</td>
<td>glyoxylate aminotransfera</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT4G05632</td>
<td>0</td>
<td></td>
<td>?</td>
<td>4</td>
<td>747</td>
<td>Glucose 1-phosphate</td>
<td>unknown (G3P DH)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT5G15600</td>
<td>1</td>
<td></td>
<td></td>
<td>0</td>
<td>891</td>
<td>Nicotinic acid</td>
<td>unknown (Nitrilase)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Metabolite QTL Candidate genes can be confirmed using knock-out lines

Fumaric acid

Tyrosine

lisec@mpimp-golm.mpg.de

David Riewe, IPK Gatersleben

São Paulo, 07.12.2009
In total 157 metabolic QTL (for 84 out of 181 metabolites) were found using Recombinant Inbred Lines.

Introgression Lines allowed the independent confirmation for 11-55% of all mQTL and revealed 160 additional effects at p ≤ 0.001.

Candidate genes (using AraCyc 3.5) were identified for 24-67% of all mQTL.

Amino acid substitutions were determined to be present in four of eight re-sequenced candidate genes.

Experiments with knock-out lines showed significant changes in the respective metabolite levels for several candidate genes.
Heterosis – the hybrid vigour

• Heterosis: increased fitness of a hybrid cross compared to its homozygous parents
• ‘fitness’: biomass, size, yield, fertility, speed of development and stress resistance

• Best-Parent-Heterosis: the hybrid increase over the better performing parent
• Mid-Parent-Heterosis: the hybrid deviation from the parental mean
Two classic heterosis theories

- the **dominance** theory:
  a complementation of deleterious alleles

<table>
<thead>
<tr>
<th>Inbred Parent 1</th>
<th>F1 Hybrid</th>
<th>Inbred Parent 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genotype</td>
<td>Phenotype</td>
<td>Genotype</td>
</tr>
<tr>
<td>a</td>
<td>a</td>
<td>a</td>
</tr>
<tr>
<td>B</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td>C</td>
<td>C</td>
<td>C</td>
</tr>
</tbody>
</table>

- the **overdominance** theory:
  an interaction of two alleles

<table>
<thead>
<tr>
<th>Inbred Parent 1</th>
<th>F1 Hybrid</th>
<th>Inbred Parent 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genotype</td>
<td>Phenotype</td>
<td>Genotype</td>
</tr>
<tr>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>B</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td>C</td>
<td>C</td>
<td>C</td>
</tr>
</tbody>
</table>
Heterosis for metabolic traits was analyzed

Calculate Metabolite Heterosis for C24 x Col-0 and Col-0 x C24 hybrids

Do we observe a heterosis effect for metabolites in *Arabidopsis thaliana*?

\[
MPH = 100 \cdot \left( \bar{F}_1 - \bar{P} \right) / \bar{P}
\]

\[
\bar{F}_1 = \left( F_{1-a} + F_{1-b} \right) / 2
\]

\[
\bar{P} = \left( P_1 + P_2 \right) / 2
\]
Identification of heterotic metabolic QTL in RILs and ILs

QTL mapping of heterotic effects

• in RILs and RIL-TCs:

\[
AMPH_{P_1} = TC_{R_1,i} - 0.5(RIL_i + P_1)
\]

\[
AMPH_{P_2} = TC_{P_2,i} - 0.5(RIL_i + P_2)
\]

\[
Z_1 = TC_{P_1,i} + TC_{P_2,i}
\]

\[
Z_2 = TC_{P_1,i} - TC_{P_2,i}
\]

• in ILs and IL-TCs:
  - t-test and mean value comparison

lisec@mpimp-golm.mpg.de

São Paulo, 07.12.2009
Distribution of heterotic metabolic QTL identified in RILs/RIL-TCs

- 232 hmQTL (AMPH_{Col}, AMPH_{C24}, Z_2)
- PVE: 4.8% (1.3-18.5%)
- Overlap with mQTL is 14%
Mode of Inheritance for metabolic QTL identified in ILs/IL-TCs reveals mainly dominant effects

- 634 significant effects (at a FDR of 5%)
- Mode of Inheritance:
  - 21% additive
  - 67% dominant/recessive
  - 12% overdominant
- at comparable significance levels the confirmation of RIL QTL in ILs is lower for heterotic than for non-heterotic effects

Lisec et al. (2009, Plant J)
Biomass Prediction using Metabolic Profiles

- Canonical Correlation Analysis (CCA) finds the optimal combination of metabolites to predict biomass
- 1144 measurements ($P_{1,2}, F_{1-a,1-b}, RIL_i, TC_i$)

<table>
<thead>
<tr>
<th>METABOLITE</th>
<th>COR</th>
<th>PV</th>
</tr>
</thead>
<tbody>
<tr>
<td>unknown_038</td>
<td>0.23298</td>
<td>1.55E-15</td>
</tr>
<tr>
<td>unknown_086</td>
<td>-0.21139</td>
<td>5.06E-13</td>
</tr>
<tr>
<td>unknown_035</td>
<td>0.20551</td>
<td>2.25E-12</td>
</tr>
<tr>
<td>Citric acid</td>
<td>-0.18815</td>
<td>1.41E-10</td>
</tr>
<tr>
<td>Ethanolamine</td>
<td>0.18662</td>
<td>2.00E-10</td>
</tr>
<tr>
<td>Fructose 6-phosphate</td>
<td>-0.18193</td>
<td>5.69E-10</td>
</tr>
<tr>
<td>Raffinose</td>
<td>-0.17577</td>
<td>2.16E-09</td>
</tr>
<tr>
<td>Glucose 6-phosphate</td>
<td>-0.16496</td>
<td>2.00E-08</td>
</tr>
<tr>
<td>Glutamine</td>
<td>-0.16277</td>
<td>3.09E-08</td>
</tr>
<tr>
<td>Succinic acid</td>
<td>-0.15043</td>
<td>3.19E-07</td>
</tr>
<tr>
<td>Sinapic acid (cis)</td>
<td>-0.1443</td>
<td>9.53E-07</td>
</tr>
<tr>
<td>Salicylic acid</td>
<td>-0.13687</td>
<td>3.38E-06</td>
</tr>
<tr>
<td>unknown_061</td>
<td>0.13612</td>
<td>3.83E-06</td>
</tr>
<tr>
<td>unknown_078</td>
<td>0.13446</td>
<td>5.03E-06</td>
</tr>
<tr>
<td>Tyrosine</td>
<td>-0.13087</td>
<td>8.97E-06</td>
</tr>
<tr>
<td>Glycerol-3-phosphate</td>
<td>-0.12249</td>
<td>3.26E-05</td>
</tr>
<tr>
<td>Spermidine (major)</td>
<td>0.11964</td>
<td>4.97E-05</td>
</tr>
<tr>
<td>Ornithine</td>
<td>0.11326</td>
<td>1.23E-04</td>
</tr>
<tr>
<td>Malic acid</td>
<td>-0.10976</td>
<td>2.00E-04</td>
</tr>
</tbody>
</table>

What is the relation between biomass and metabolic profiles?

Correlation R=0.73  Prediction R=0.58

Summary of the *Arabidopsis thaliana* analyses

- Data analysis of metabolic profiles from large scale experiments has been improved. (RI correction, deconvolution errors, peak detection)

- Metabolic QTL and heterotic metabolic QTL were identified in two independent experiments (RILs, RIL-TCs and ILs, IL-TCs) and further characterized.

- Enzyme candidate genes for QTL were suggested and amino acid substitutions have been found in sequenced candidates.

- Heterosis effects for metabolites were found ($P_{1,2}$, $F_{1-a,1-b}$) and characterized as mainly dominant (ILs).

- A comparison of RIL/IL for mQTL and hmQTL suggests both approaches to be complementary, especially for heterotic effects.

- It was shown that a complex trait (biomass) can be predicted based on a metabolic profile. (CCA, $R = 0.58$)

lisec@mpimp-golm.mpg.de

São Paulo, 07.12.2009
Maize roots were analyzed for 112 metabolites using GC-MS (in 6 replicates of 6 parental and 14 hybrid genotypes)
Major variance in root metabolites separates parents from hybrids (PC1) and European from American lines (PC2)

More analyses were applied:
- metabolite correlation networks (heterosis in network properties)
- metabolite-phenotype correlations
- metabolite heterosis

Lisec et al. (2009, submitted)
Polar plots reveal differences in metabolite heterosis pattern between reciprocal hybrids, thus raising more questions than providing answers.
Lower variability in metabolite levels of hybrids compared to parents suggests a metabolic optimum which could explain heterosis.
Two representative examples for metabolite variance in parental and hybrid genotypes show additive and overdominant heterosis pattern.
A paternal effect is observed for the network property Average Degree only in the American lines.
The correlation between Average Degree and root fresh weight is overlaid by a paternal effect.
Summary of the *Zea mais* heterosis analyses

- Maize metabolome allows to distinguish ecotypes and homozygous from heterozygous lines.

- A metabolic optimum for given environmental conditions is suggested. Evidence for the metabolic optimum is the lower variation of metabolite levels in hybrids and a significant correlation of a deviation from the supposed optimum with biomass.

- Metabolite correlation networks differentiating homozygous and heterozygous genotypes could be calculated.

- A paternal effect for network properties of metabolite correlation networks was observed.
Acknowledgements

AG Altmann
AG Willmitzer
AG Selbig/Walther
AG Kopka
AG Fernie

Barbara Kusterer
(Uni Hohenheim)
Lilla Roemisch-Margl
(TU München)
Achim Walter
(FZ Jülich)
David Riewe
(IPK Gatersleben)

São Paulo, 07.12.2009