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

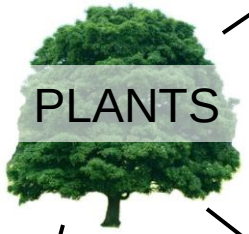
Max-Planck-Institut
für Molekulare Pflanzenphysiologie
Universität Potsdam

Metabolomics allows a new perspective on plant systems

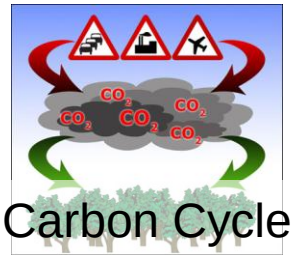
- Metabolite Quantitative Trait Loci (mQTL) enable to connect phenotype and genotype in *Arabidopsis thaliana*
  - Recombinant Inbred Line and Introgression Line populations
  - Heterosis and quantitative genetics
  - Identification and characterization of (heterotic) mQTL
  - Relation of metabolic profiles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

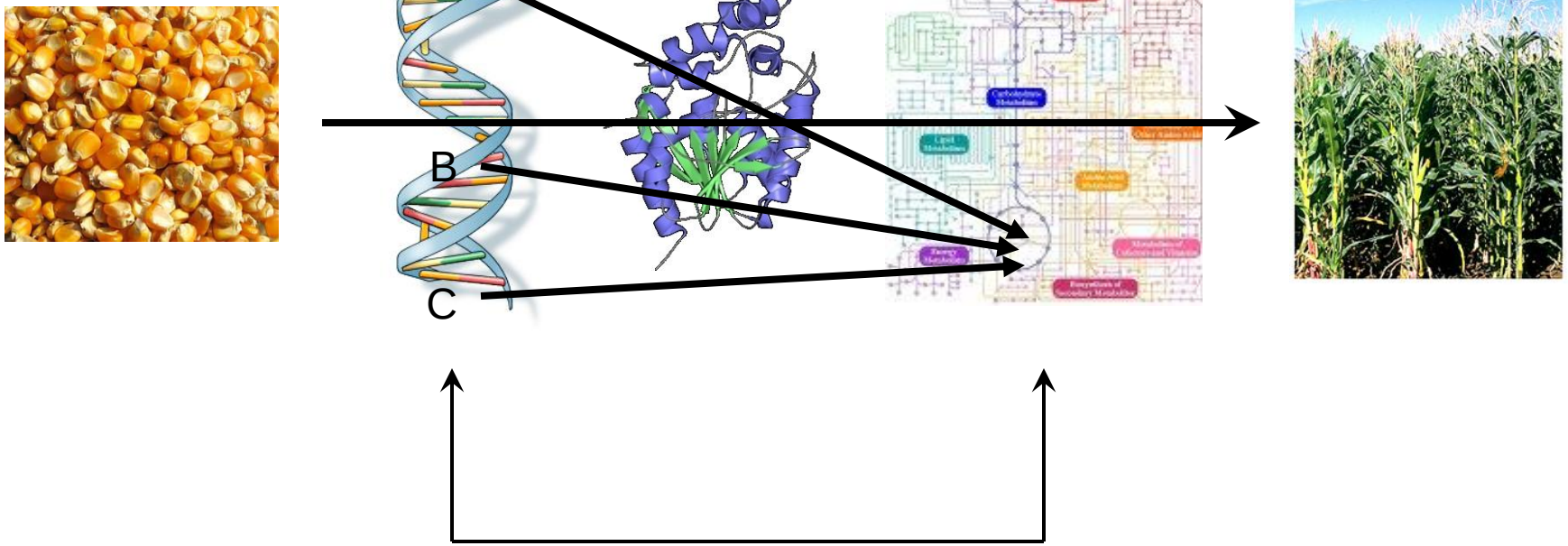
Plants are the primary producers of biomass



Renewable Resources



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

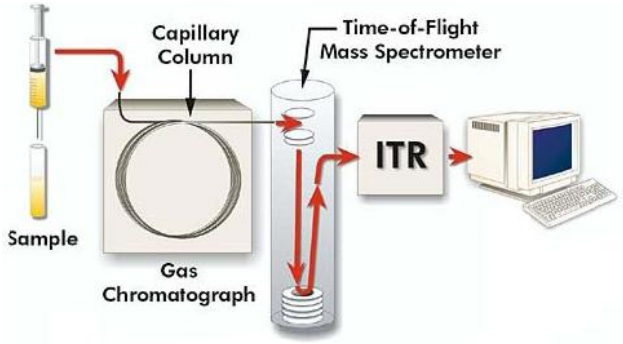
Arabidopsis thaliana

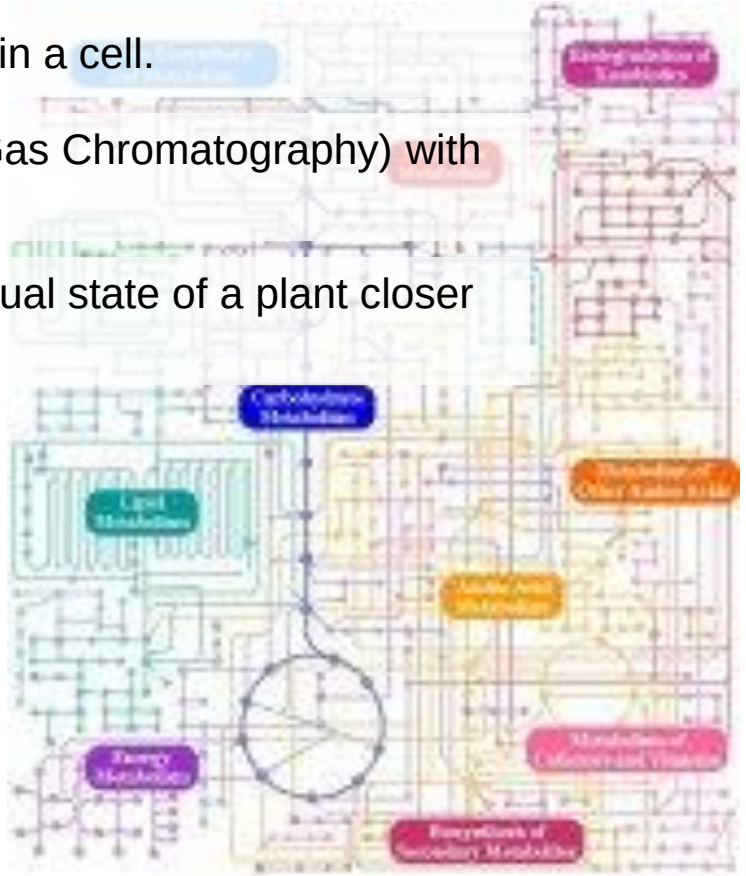
- more than 100% best-parent heterosis for yield
- heterosis effect agronomical used since more than 100 years
- of high economic importance
- difficult to cultivate in higher number in a controlled environment

Zea mais

GC-ToF-MS was used to measure metabolic profiles

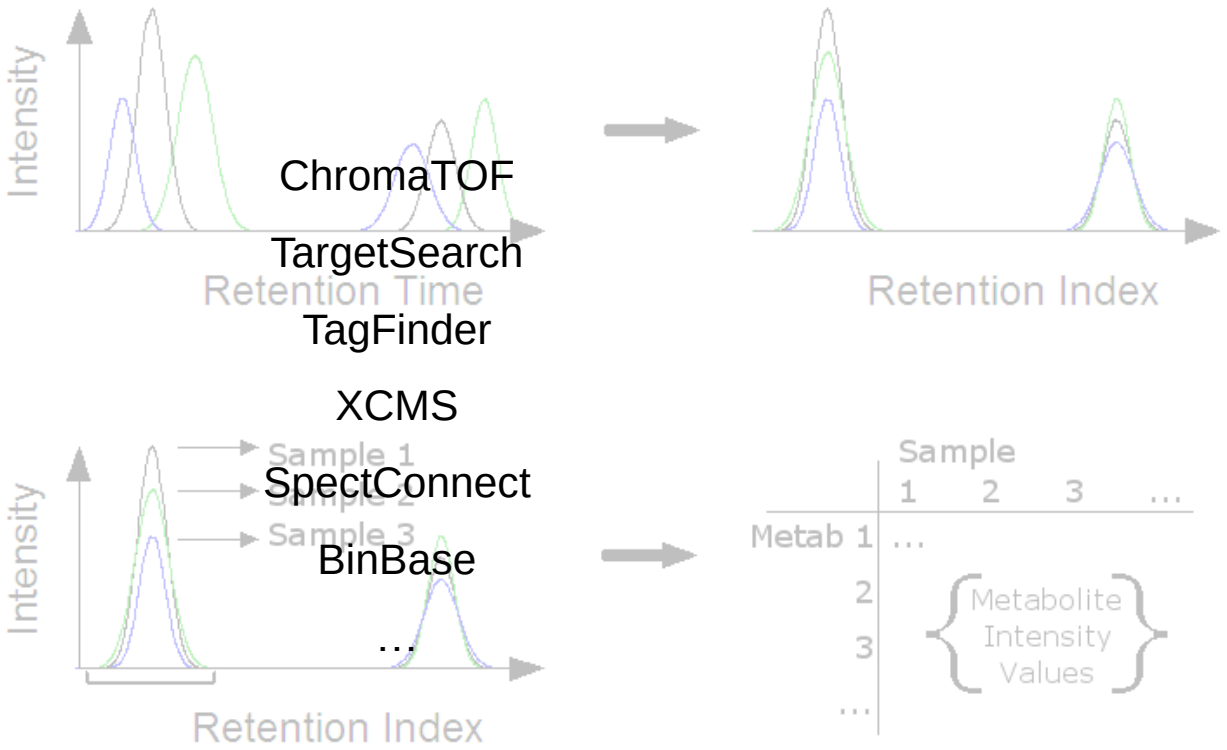
- Metabolome: The entirety of small molecules within a cell.
- GC-MS: a combination of a separation method (Gas Chromatography) with an detection system (Mass Spectrometry)
- Metabolome mirrors the biological endpoint or actual state of a plant closer than its genes.





Data Processing for large scale metabolomics was established

- R-Script 1
- R-Script 2

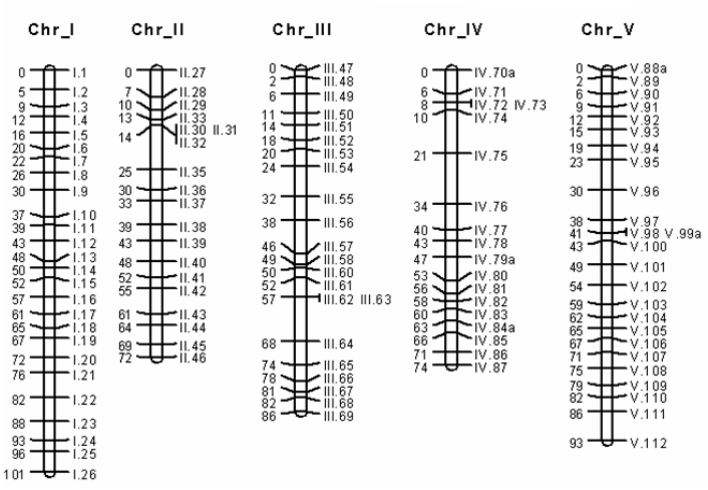


- Allowed the evaluation of >2000 samples for 181 metabolic traces (85 known)

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)

$$V_P = V_G + V_E + V_G \times V_E$$



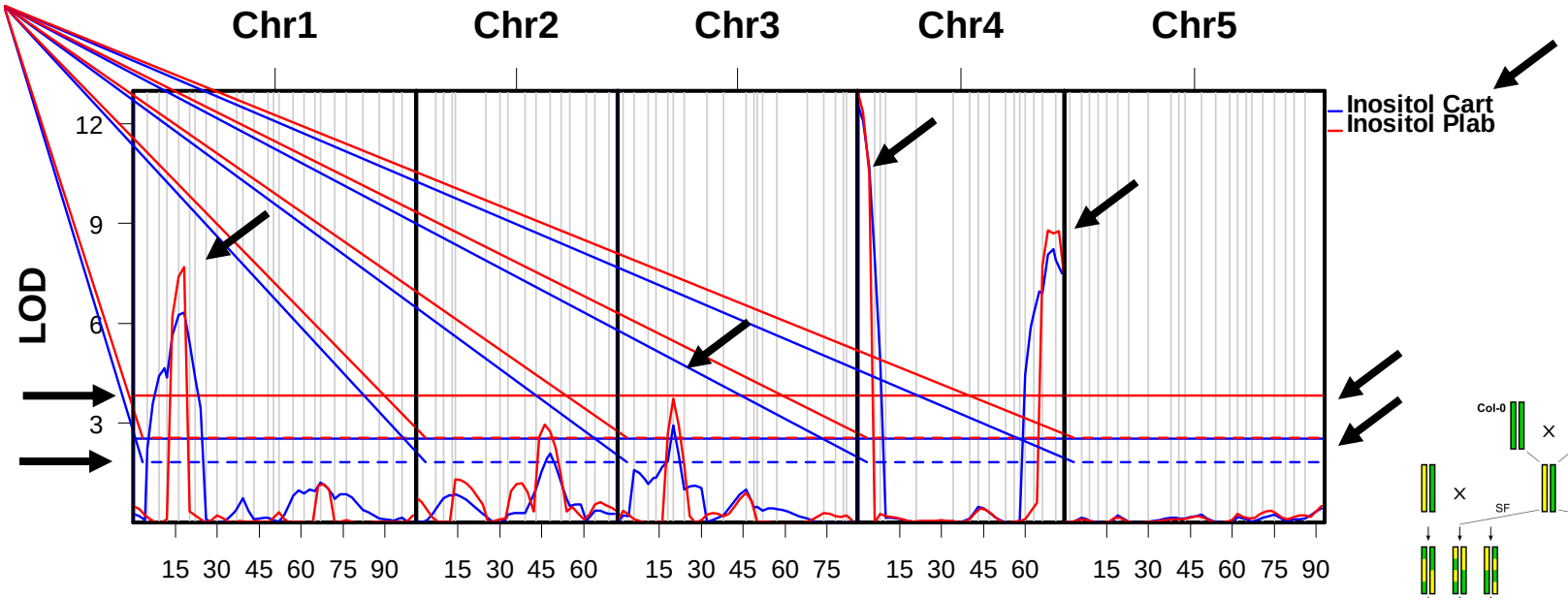
Recombinant Inbred Lines and Introgression Lines were used to identify Quantitative Trait Loci (QTL)

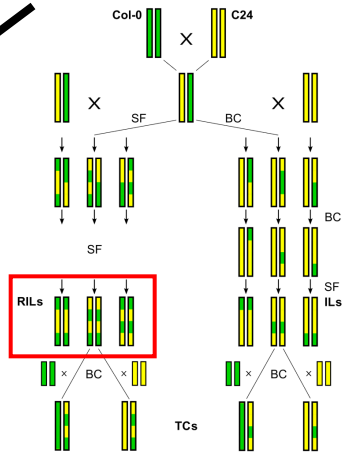
- QTL indicates a region of the genome bearing loci which influence the phenotype of an observed trait (e.g. metabolite concentration or biomass)
- quantitative traits – usually polygenic inherited

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

RIL population:
- 369 RILs (1 repl.)
- 735 RIL-TCs (1 repl.)
- $P_{1,2}$, $F_{1\text{-}a,1\text{-}b}$ (~10 repl.)

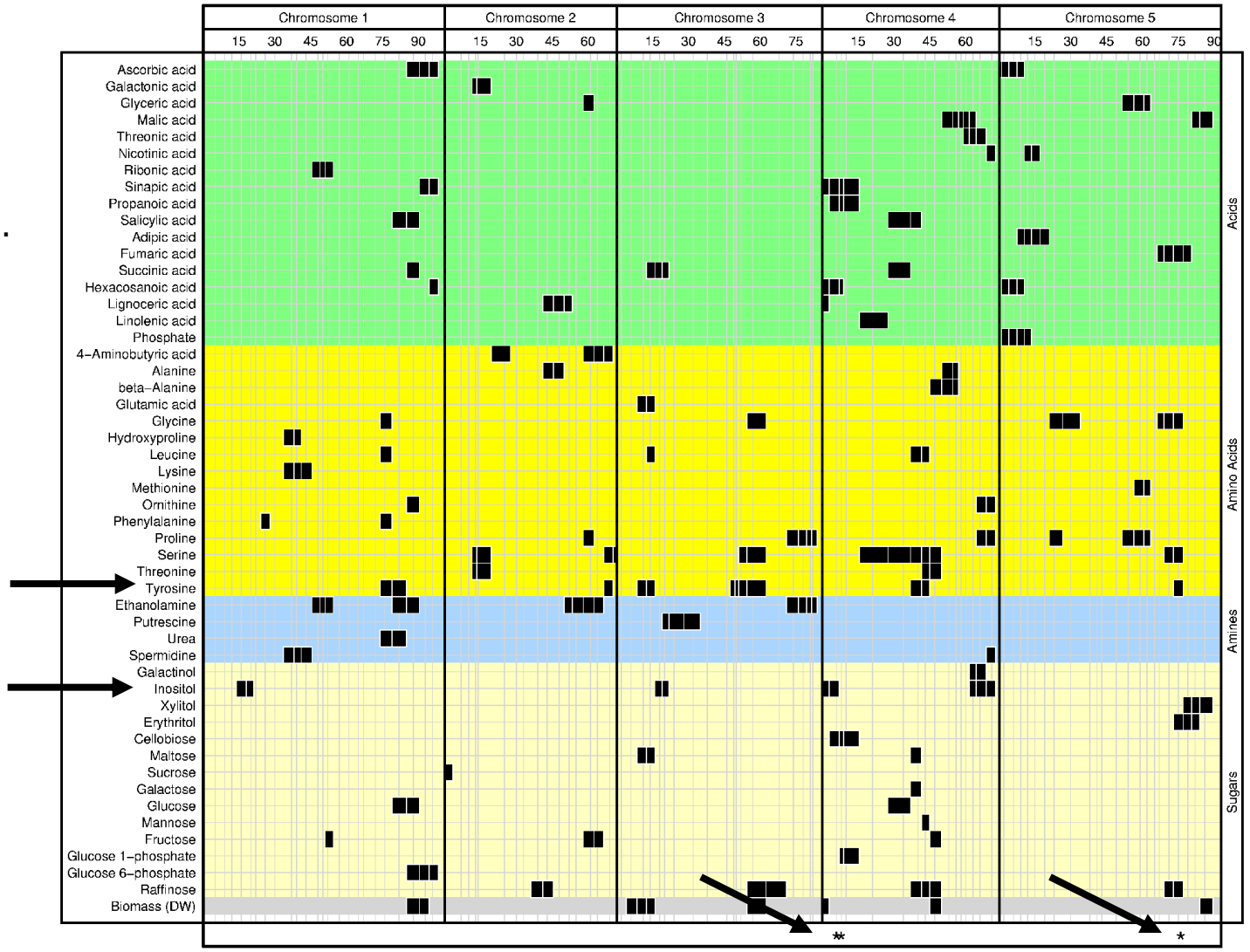
A case study: *myo*-Inositol has 4 QTL



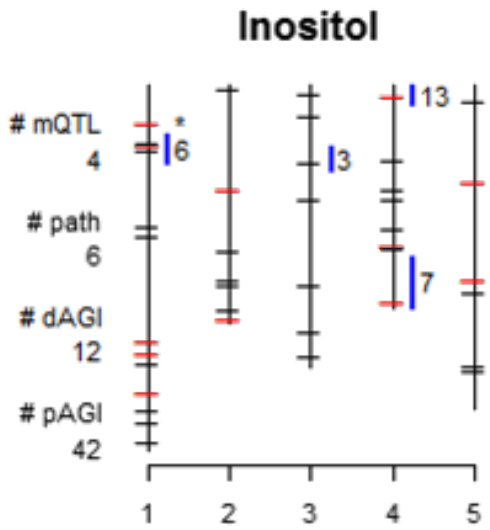


A birds view on RIL QTL for known metabolites

- 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



Candidate genes for metabolic QTL were identified using AraCyc 3.5 database



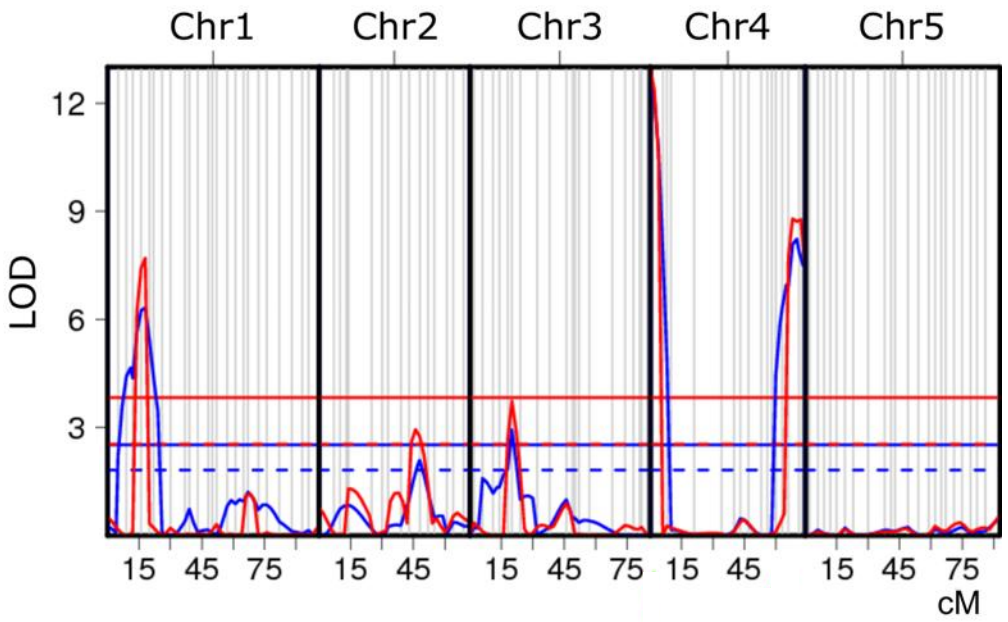
mQTL

direct AGI

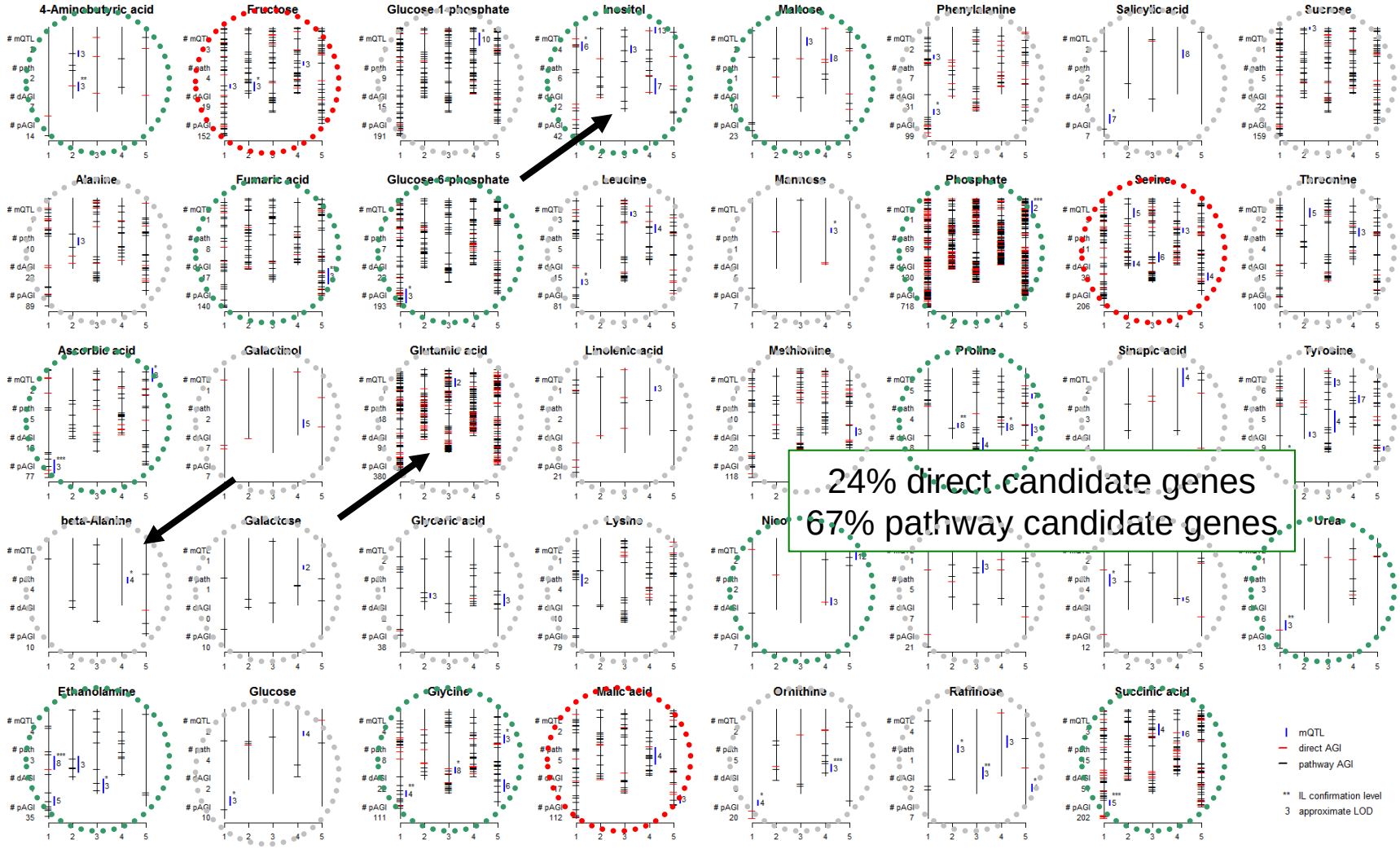
pathway AGI

** IL confirmation level

3 approximate LOD



Metabolic QTL are enriched for enzyme candidate genes



Max-Planck-Institut für Molekulare Pflanzenphysiologie

Universität Potsdam

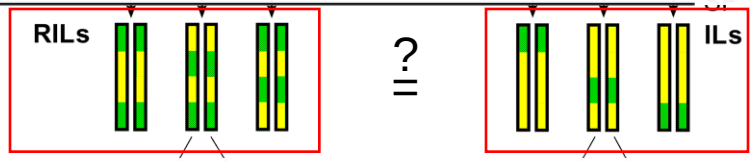
lisec@mpimp-golm.mpg.de

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

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



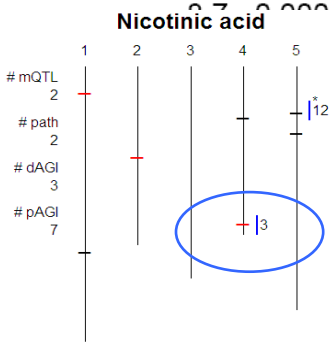
Direct candidate gene	IL confirmation rate			
	>0.05	0.05 *	0.01 **	0.001 ***
no	40	18	4	2
yes	10	4	3	3

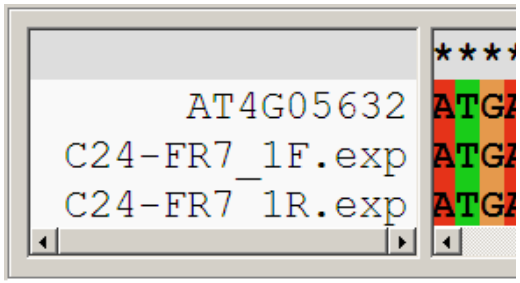
Resequencing of 8 mQTL candidate genes reveals amino acid substitutions

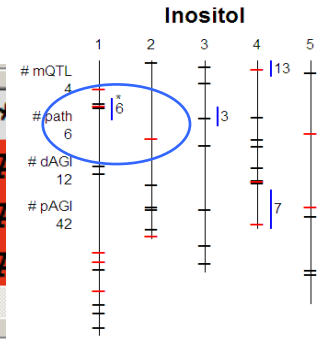
Candidate Genes to test selected according to:

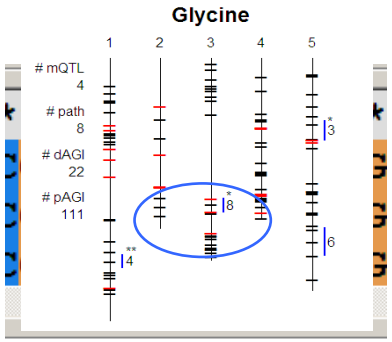
- LOD value
- Confirmation level in IL-population
- Number of known genes in AraCyc 8.5
- Catalyzed reaction

AGI	Substitution	Synonymus		Intron	bp	Metabolite	Gene function	LOD_{RIL}	P_{IL}
AT1G14520 →	3	6		7	2169	Inositol	MIOX1 (myo-inositol oxygenase)	6.5	0.047
AT5G53970 →	1	1		0	2240	Tyrosine	tyrosine aminotransferase	9.6	0.054
AT1G43710	0	1	?	1	2064	Ethanolamine	glutamate decarboxylase		
AT3G44740 →	1	0		0	1446	Glycine	glycyl-tRNA synthetase		
AT4G15210	0	0	?	4	3132	Maltose	beta-amylase activity		
AT2G38400	0	0	?	2	2660	4-Aminobutyric acid	glyoxylate aminotransfera		
AT4G05632	0	0	?	4	747	Glucose 1-phosphate	unknown (G3P DH)		
AT5G15600 →	1	0		0	891	Nicotinic acid	unknown (Nitrilase)		

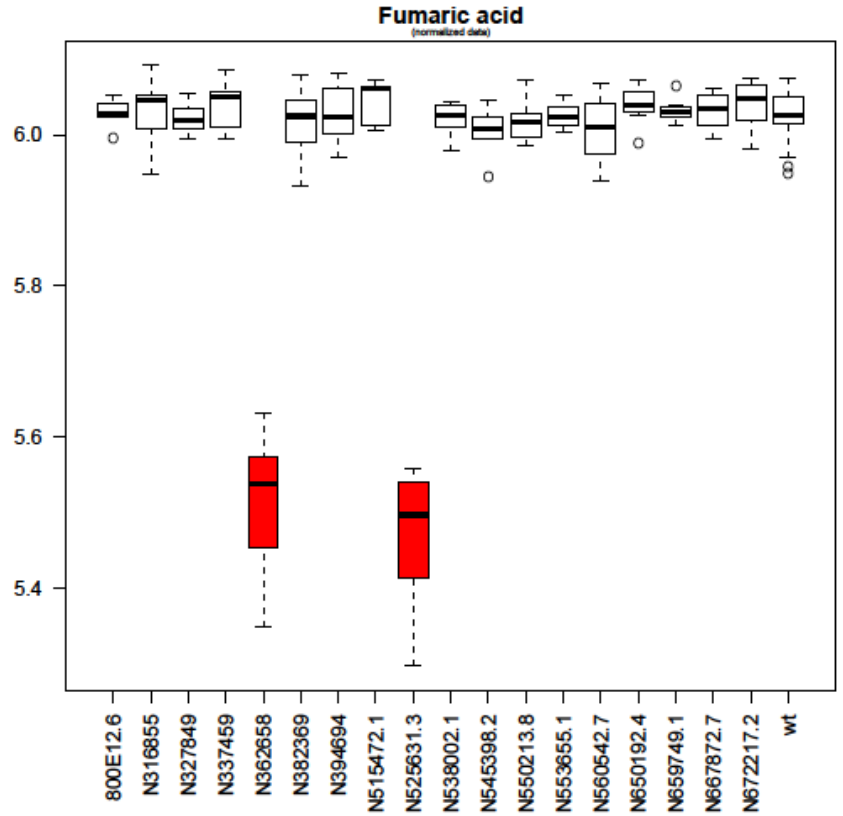


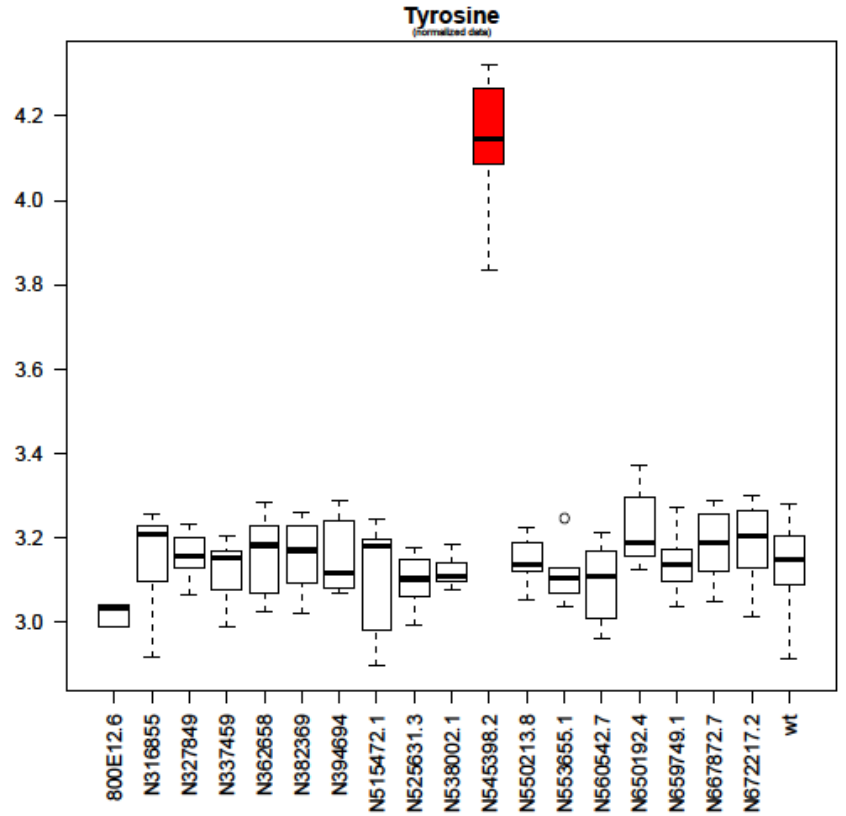






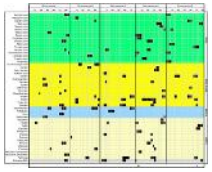
Metabolite QTL Candidate genes can be confirmed using knock-out lines



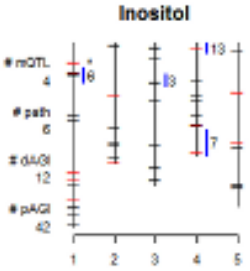


Summary for metabolic QTL

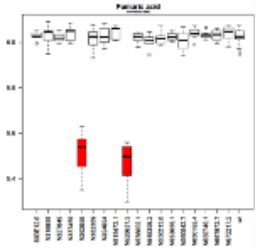
- In total 157 metabolic QTL (for 84 out of 181 metabolites) were found using Recombinant Inbred Lines.
- Introgression Lines allowed the independent confirmation for 11-55% of all mQTL and revealed 160 additional effects at p ≤ 0.001.
- Candidate genes (using AraCyc 3.5) were identified for 24-67% of 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.



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



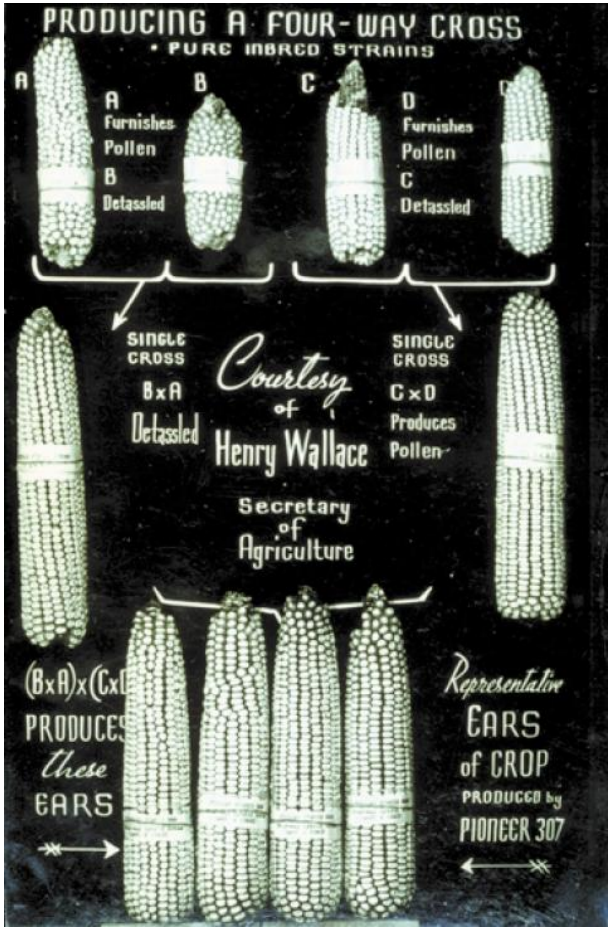
AGI	Substitution	Synonymus	Intron	bp	Metabolite	Gene function	LOD_{RIL}	P_{IL}
AT1G14520	3	6	7	2169	Inositol	MIOX1 (myo-inositol oxygenase)	6.5	0.047
AT5G53970	1	1	0	2240	Tyrosine	tyrosine aminotransferase	9.6	0.054
AT1G43710	0	1	1	2064	Ethanolamine	glutamate decarboxylase	8.7	0.000
AT3G44740	1	0	0	1446	Glycine	glycyl-tRNA synthetase	8.0	0.015
AT4G15210	0	0	4	3132	Maltose	beta-amylase activity	9.9	-
AT2G38400	0	0	2	2660	4-Aminobutyric acid	glyoxylate aminotransferase	3.6	0.004
AT4G05632	0	0	4	747	Glucose 1-phosphate	unknown (G3P DH)	10.7	0.013
AT5G15600	1	0	0	891	Nicotinic acid	unknown (Nitrilase)	13.2	0.013



Heterosis – the hybrid vigour

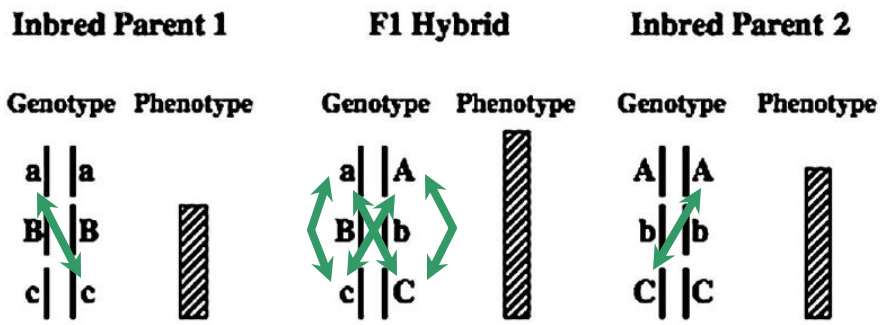
- Heterosis: increased fitness of a hybrid cross compared to its homozygous parents
- ‘fitness’: biomass, size, yield, fertility, speed of development and stress resistance

- Best-Parent-Heterosis: the hybrid increase over the better performing parent
- Mid-Parent-Heterosis: the hybrid deviation from the parental mean

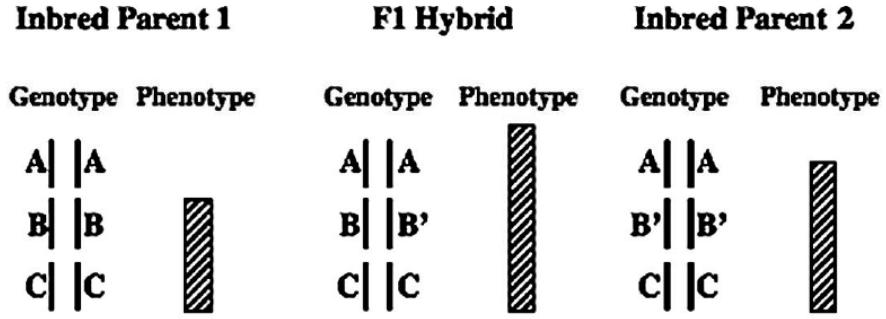


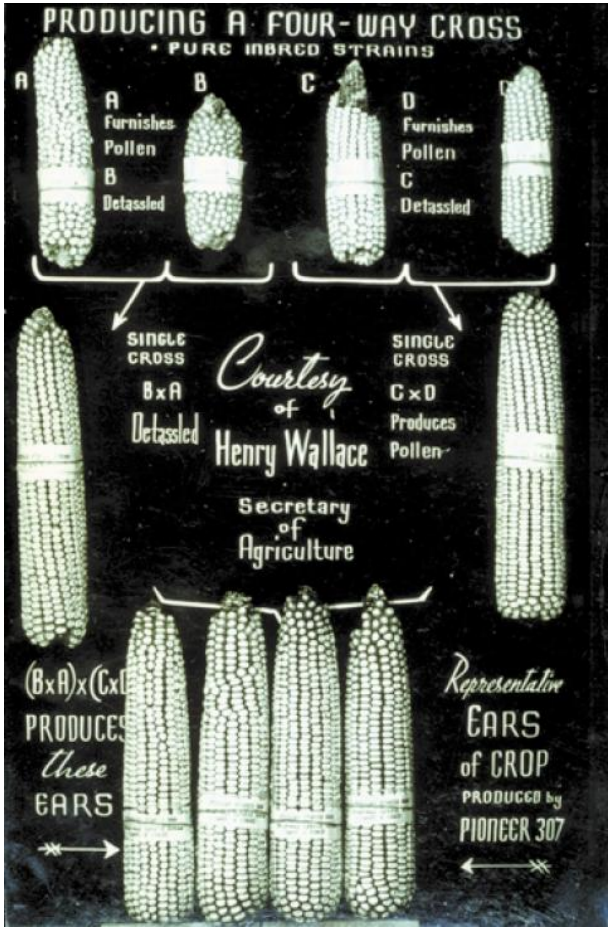
Two classic heterosis theories

- the dominance theory:
  a complementation of deleterious alleles



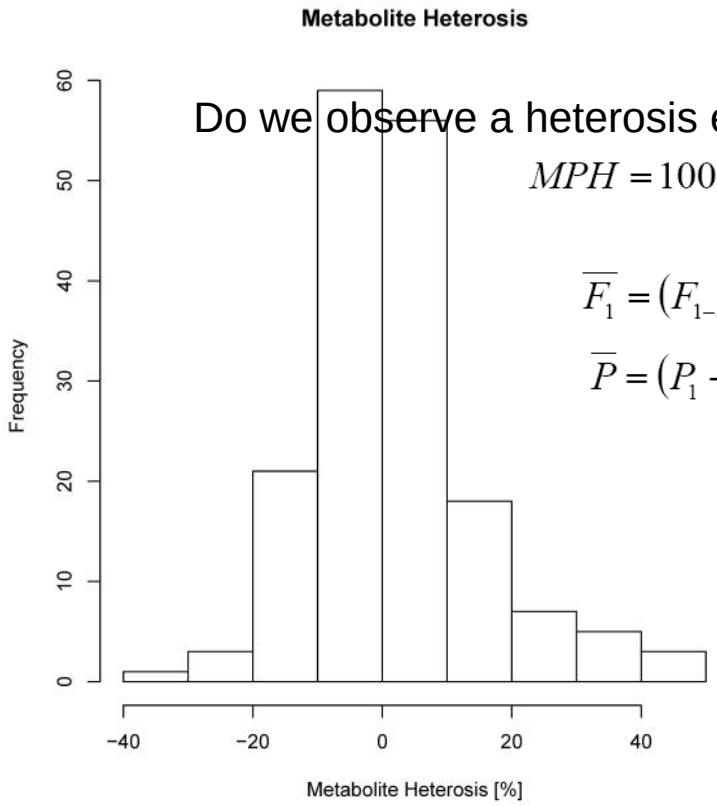
- the overdominance theory:
  an interaction of two alleles





Heterosis for metabolic traits was analyzed

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

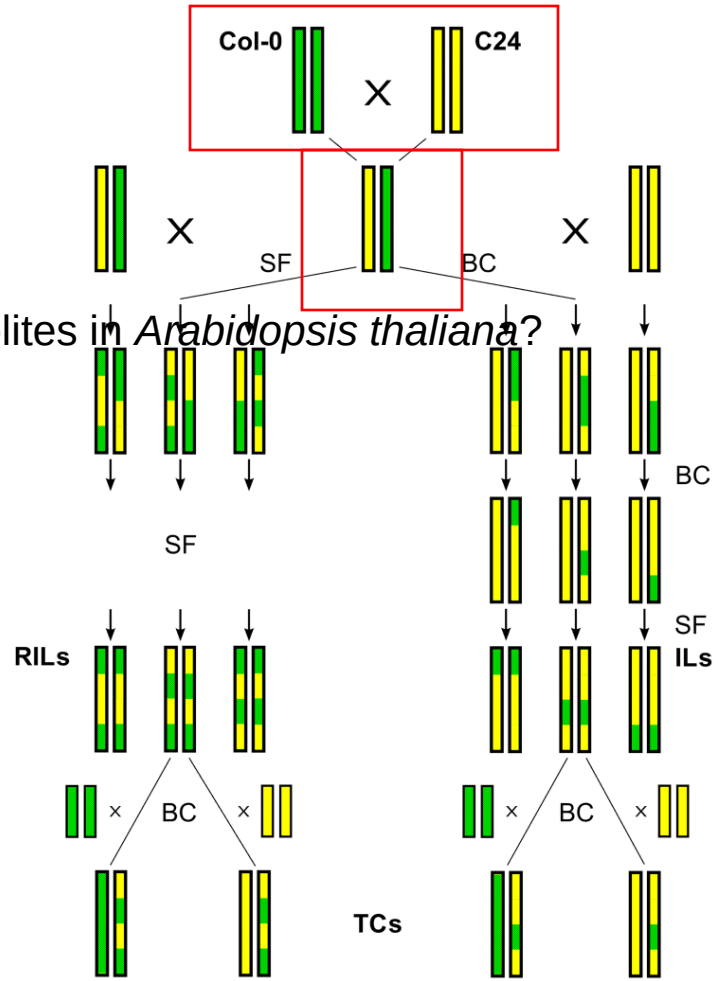


Do we observe a heterosis effect for metabolites in *Arabidopsis thaliana*?

$$MPH = 100 \cdot \left(\overline{F_1} - \overline{P}\right) / \overline{P}$$

$$\overline{F_1} = \left(F_{1-a} + F_{1-b}\right) / 2$$

$$\overline{P} = \left(P_1 + P_2\right) / 2$$



Max-Planck-Institut für Molekulare Pflanzenphysiologie
Universität Potsdam
lisec@mpimp-golm.mpg.de

Identification of heterotic metabolic QTL in RILs and ILs

QTL mapping of heterotic effects

- in RILs and RIL-TCs:

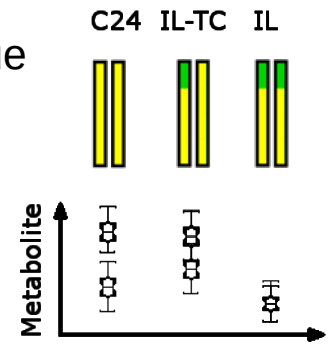
$$AMPH_{P_1} = TC_{P_1,i} - 0.5\left(RIL_i + \overline{P_1}\right)$$

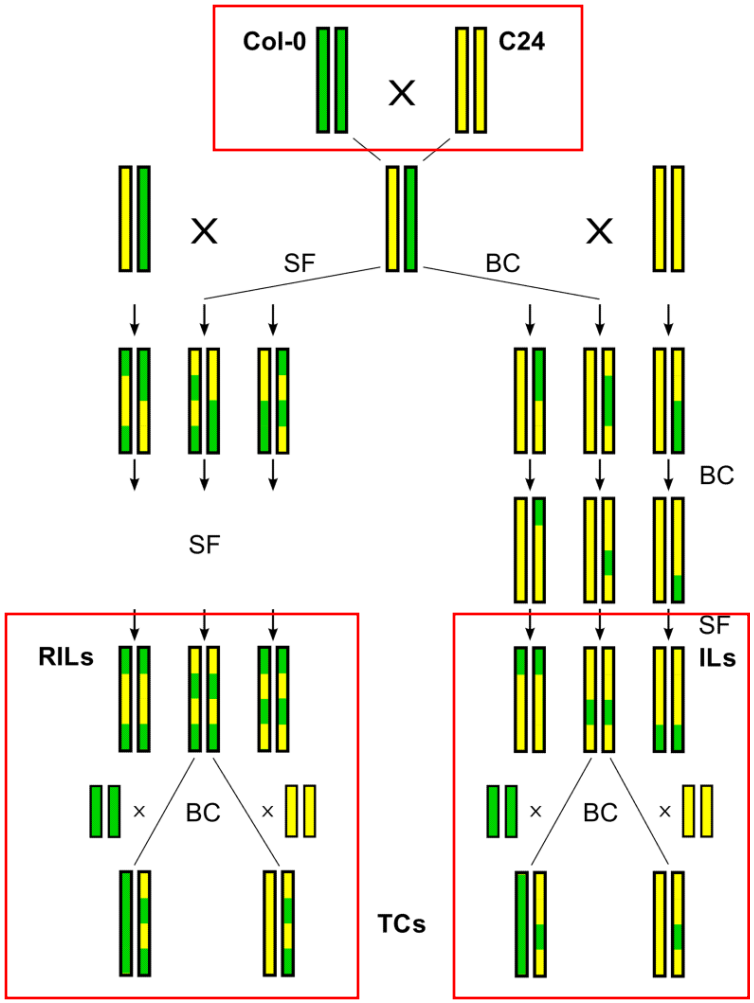
$$AMPH_{P_2} = TC_{P_2,i} - 0.5\left(RIL_i + \overline{P_2}\right)$$

$$Z_1 = TC_{P_1,i} + TC_{P_2,i}$$

$$Z_2 = TC_{P_1,i} - TC_{P_2,i}$$

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





Distribution of heterotic metabolic QTL identified in RILs/RIL-TCs

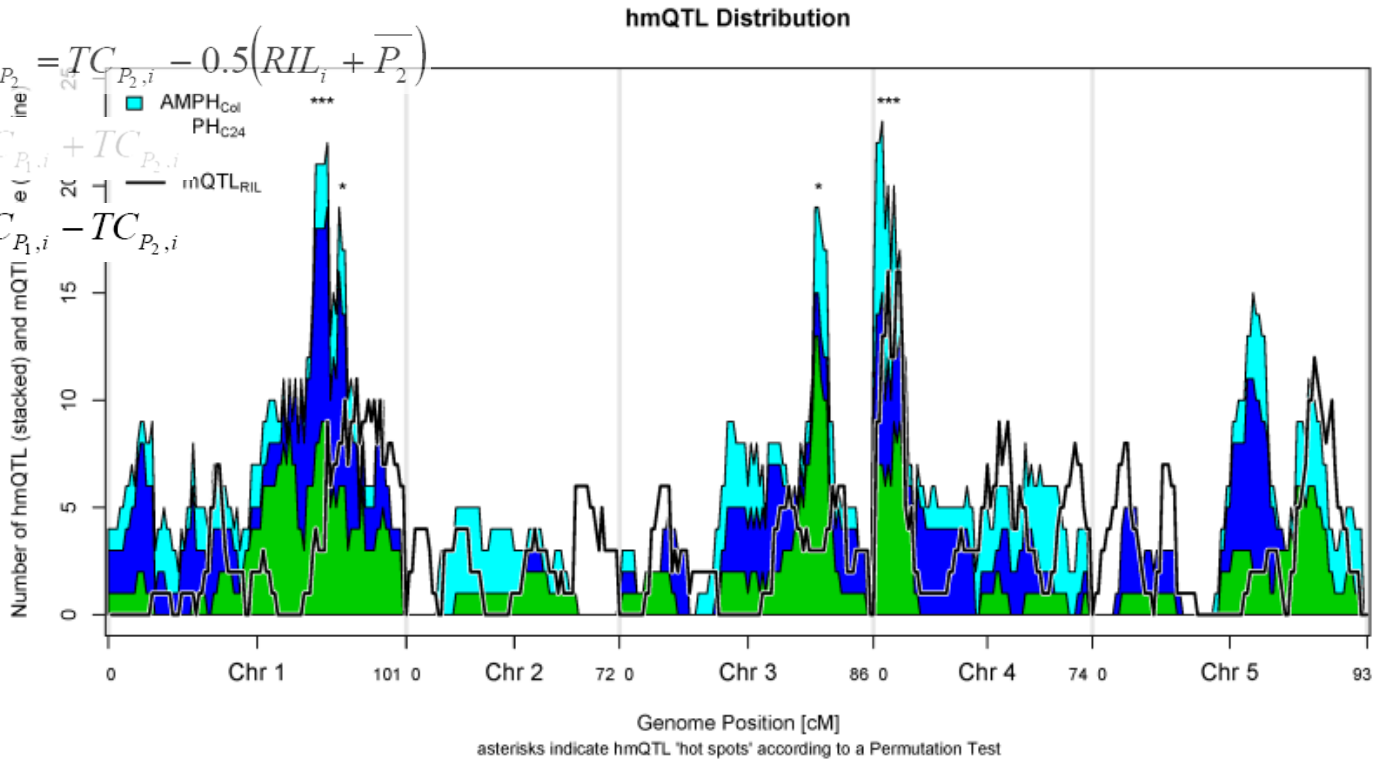
- 232 hmQTL ($AMPH_{Col}$, $AMPH_{C24}$, Z_2)
- PVE: 4.8% (1.3-18.5%)
- Overlap with mQTL is 14%

$$AMPH_{P_1} = TC_{P_1,i} - 0.5\left(RIL_i + \overline{P_1}\right)$$

$$AMPH_{P_2} = TC_{P_2,i} - 0.5\left(RIL_i + \overline{P_2}\right)$$

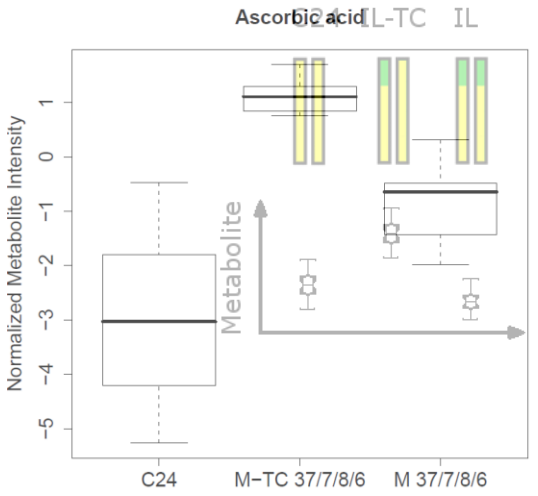
$$Z_1 = TC_{P_1,i} + TC_{P_2,i}$$

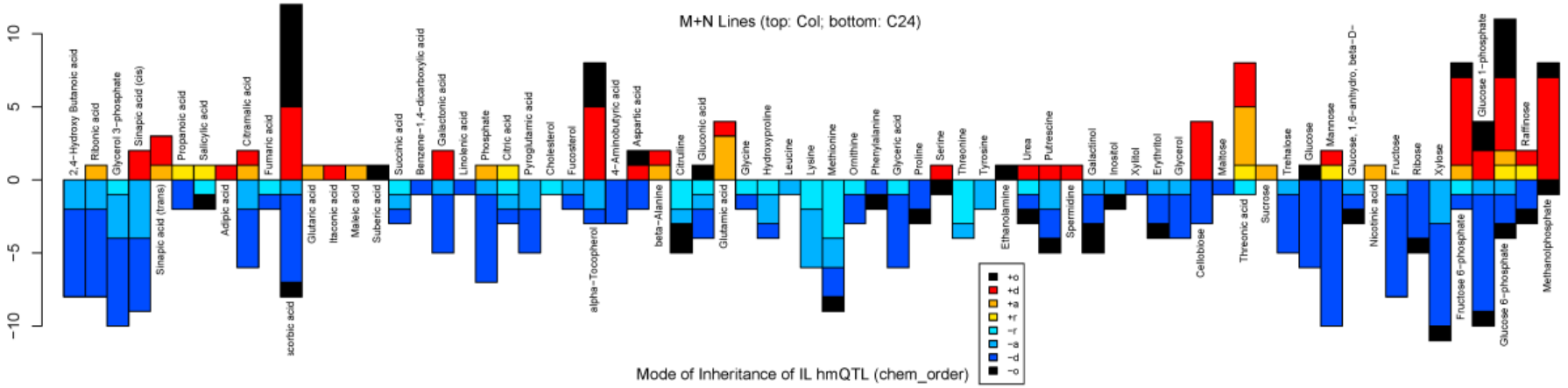
$$Z_2 = TC_{P_1,i} - TC_{P_2,i}$$



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





Max-Planck-Institut für Molekulare Pflanzenphysiologie
Universität Potsdam
lisec@mpimp-golm.mpg.de

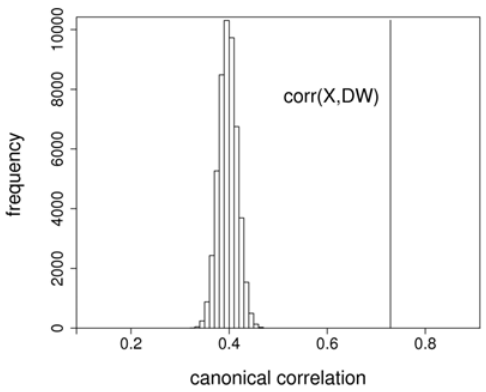
Biomass Prediction using Metabolic Profiles

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

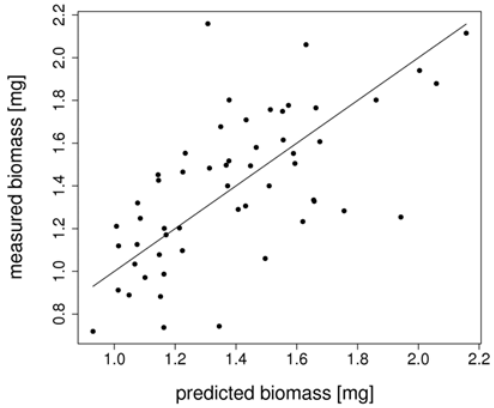
- Canonical Correlation Analysis (CCA) finds the optimal combination of metabolites to predict biomass
- 1144 measurements ($P_{1,2}$, $F_{1\text{-}a,1\text{-}b}$, RIL_i, TC_i)

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 ($P_{1,2}$, $F_{1\text{-}a,1\text{-}b}$) 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)

Maize roots were analyzed for 112 metabolites using GC-MS (in 6 replicates of 6 parental and 14 hybrid genotypes)

A

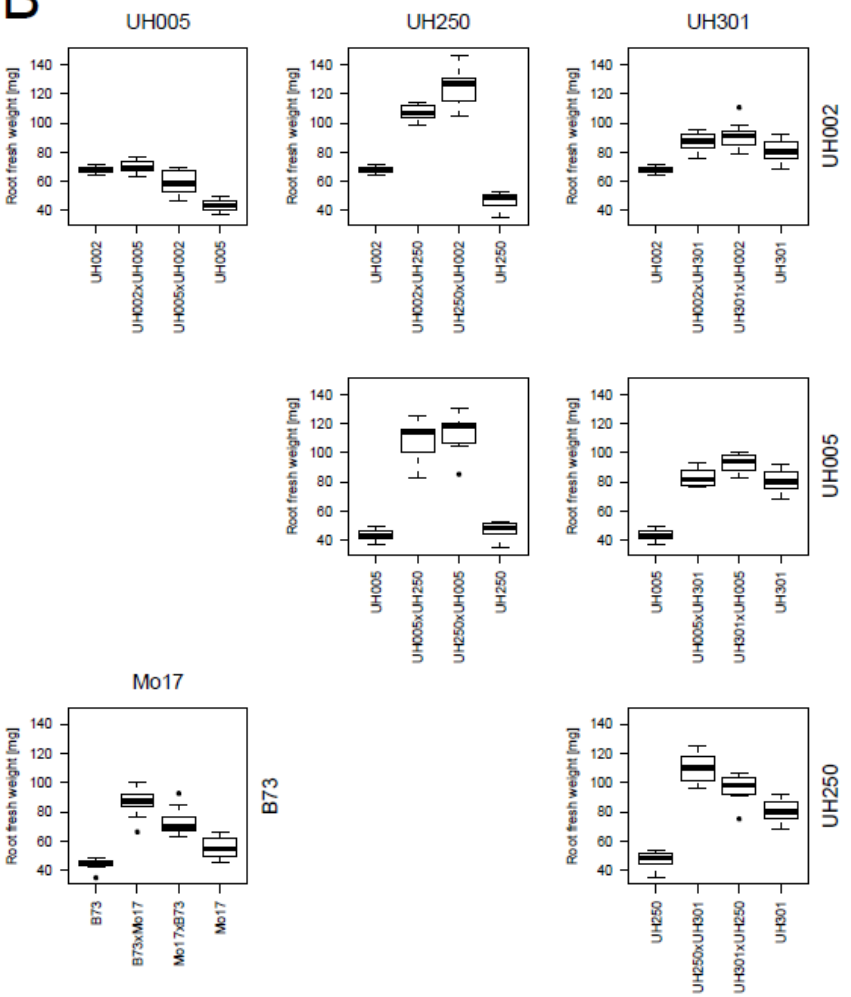
European Lines
(Flint-Poot: UH002, UH005;
Dent-Pool: UH250, UH301)

	UH002	UH005	UH250	UH301
UH002	UH002 x UH002	UH002 x UH005	UH002 x UH250	UH002 x UH301
UH005	UH005 x UH002	UH005 x UH005	UH005 x UH250	UH005 x UH301
UH250	UH250 x UH002	UH250 x UH005	UH250 x UH250	UH250 x UH301
UH301	UH301 x UH002	UH301 x UH005	UH301 x UH250	UH301 x UH301

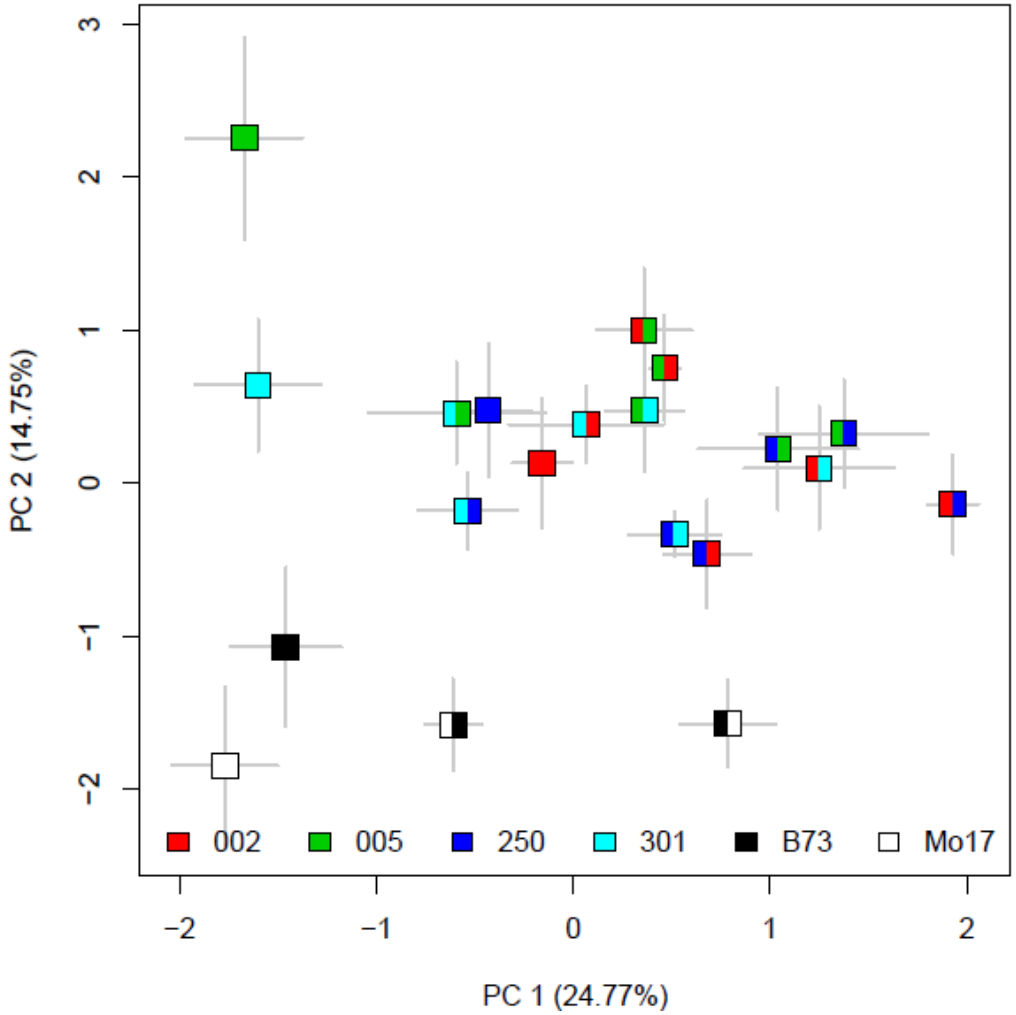
American Lines

	B73	Mo17
B73	B73 x B73	B73 x Mo17
Mo17	Mo17 x B73	Mo17 x Mo17

B

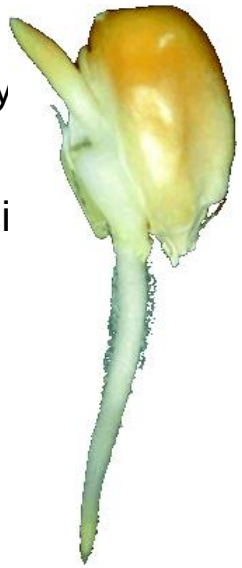


Major variance in root metabolites separates parents from hybrids (PC1) and European from American lines (PC2)



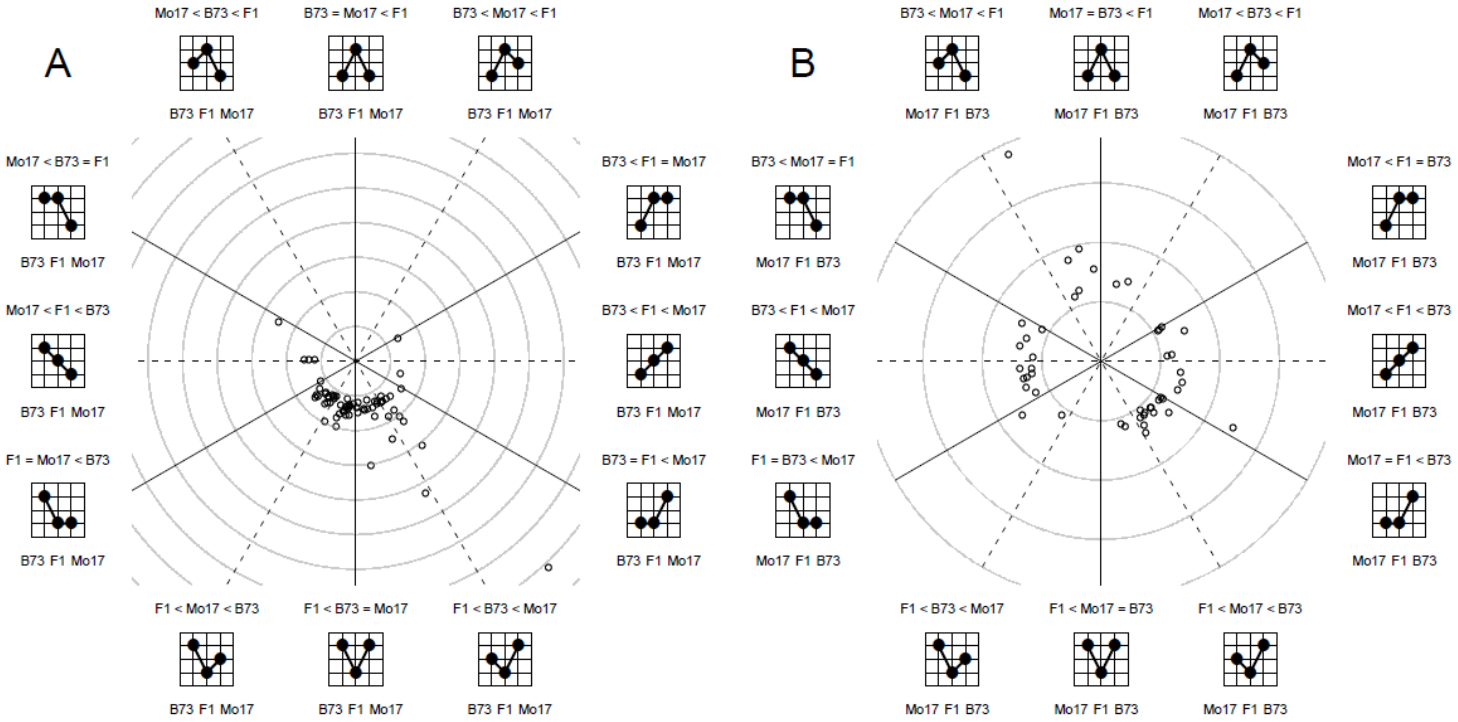
More analyses were applied:

- metabolite correlation networks (heterosis in network properties)
- metabolite-phenoty correlations
- metabolite heterosi

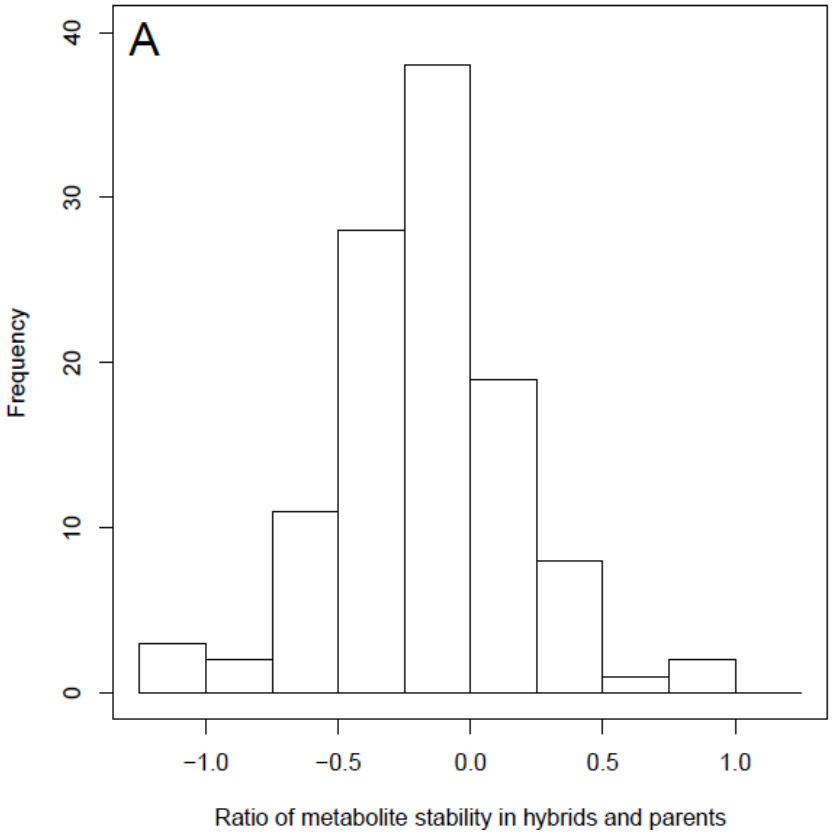


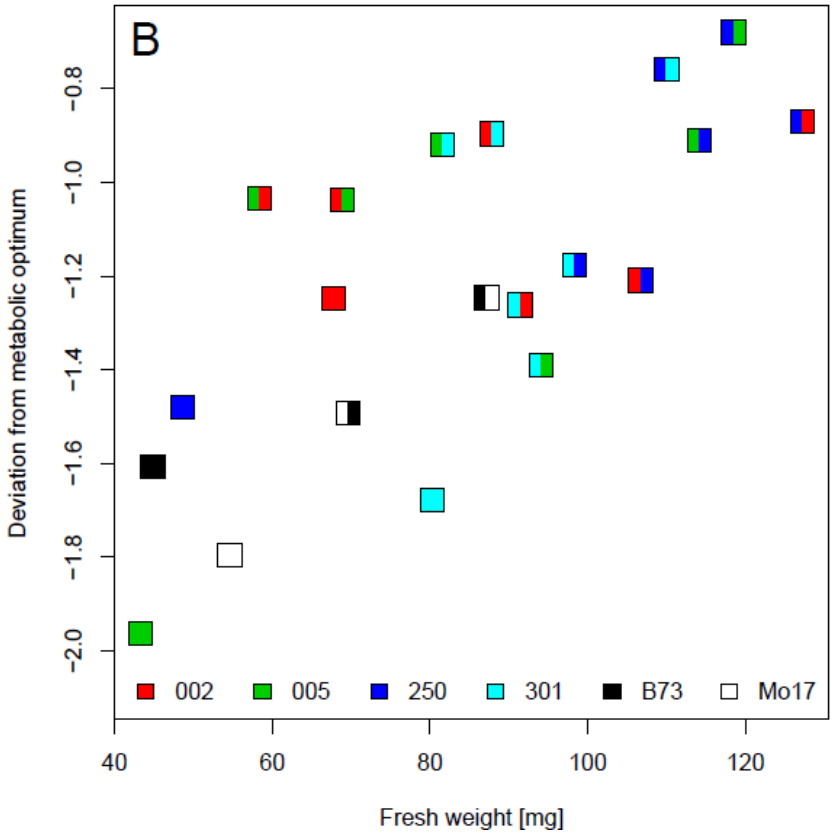
Max-Planck-Institut für Molekulare Pflanzenphysiologie
Universität Potsdam
lisec@mpimp-golm.mpg.de

Polar plots reveal differences in metabolite heterosis pattern between reciprocal hybrids, thus raising more questions than providing answers



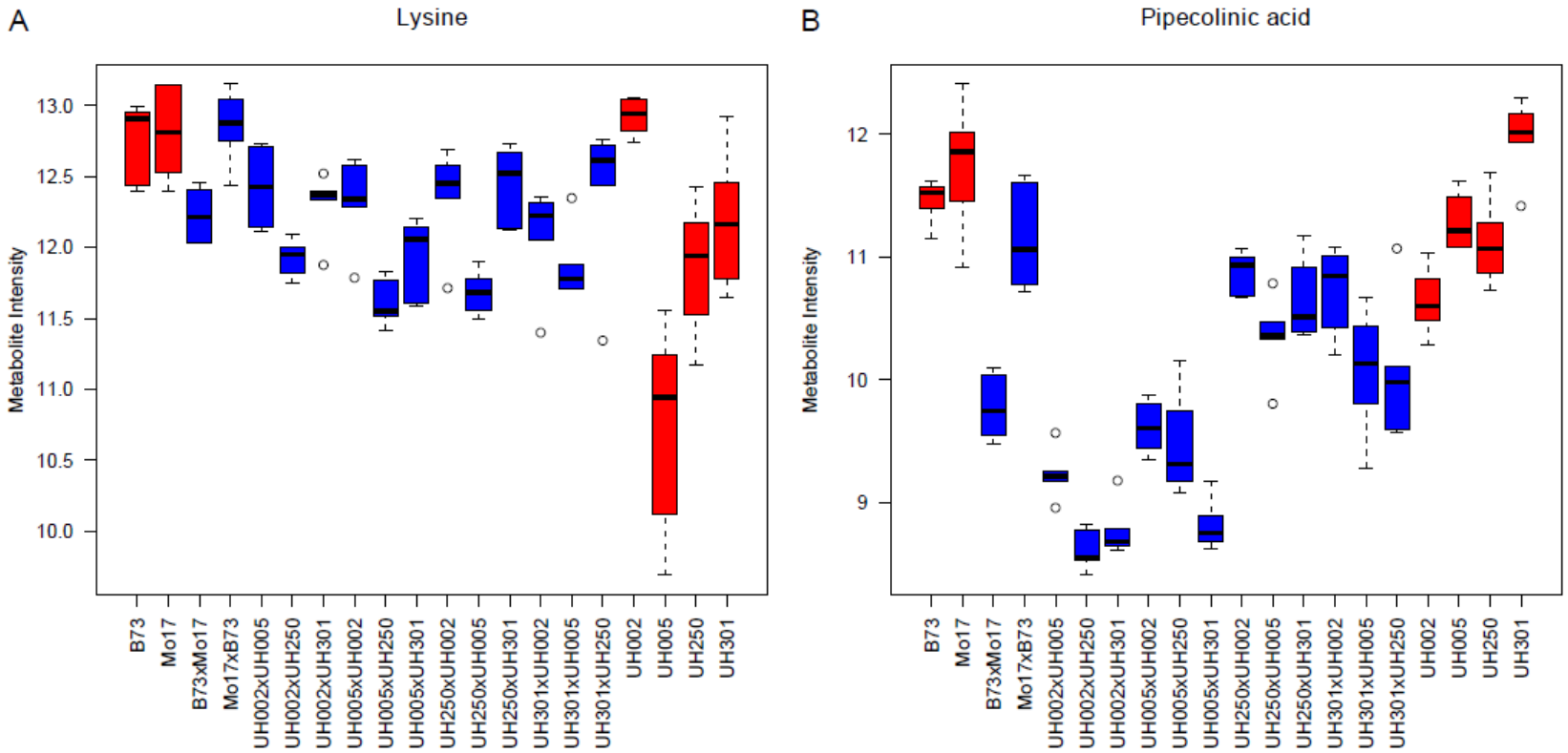
Lower variability in metabolite levels of hybrids compared to parents suggests a metabolic optimum which could explain heterosis.



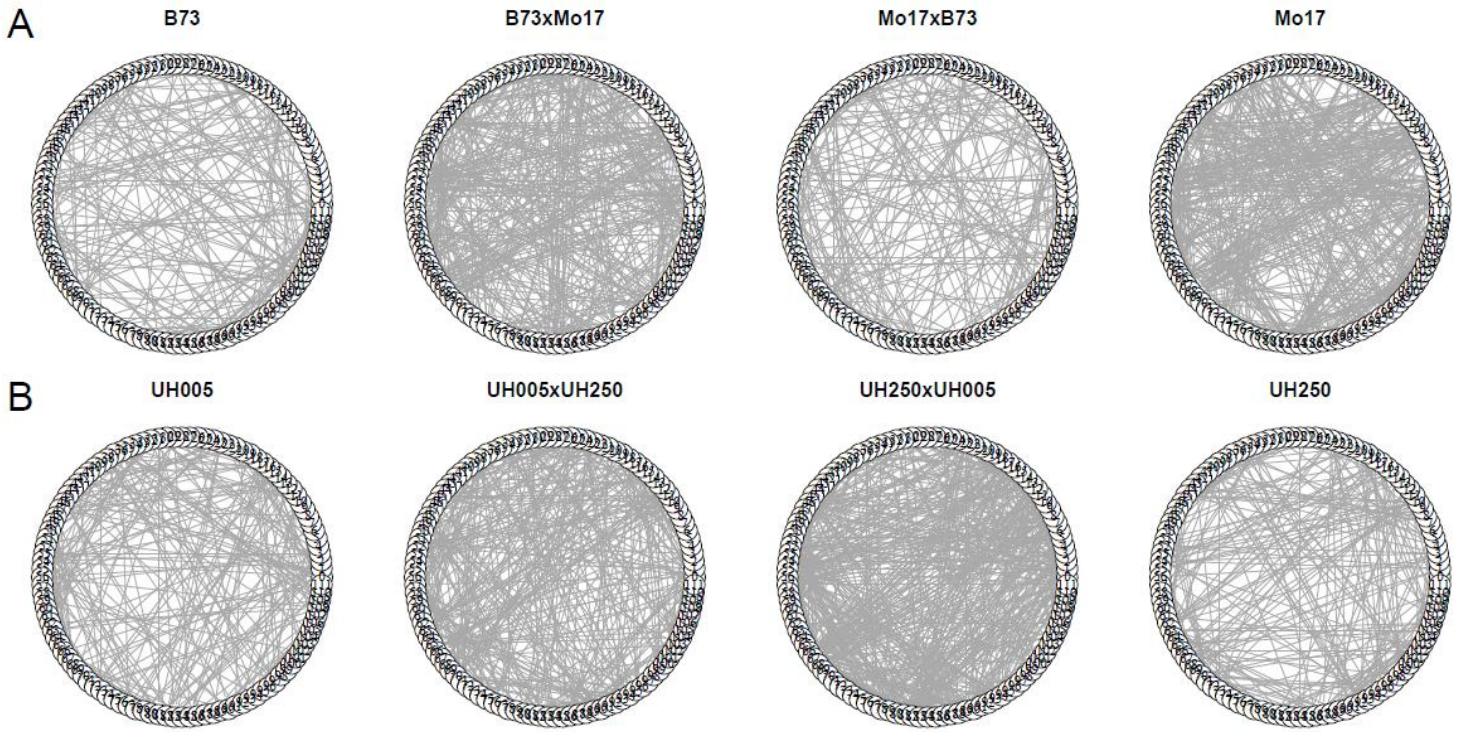


Max-Planck-Institut für Molekulare Pflanzenphysiologie
Universität Potsdam
lisec@mpimp-golm.mpg.de

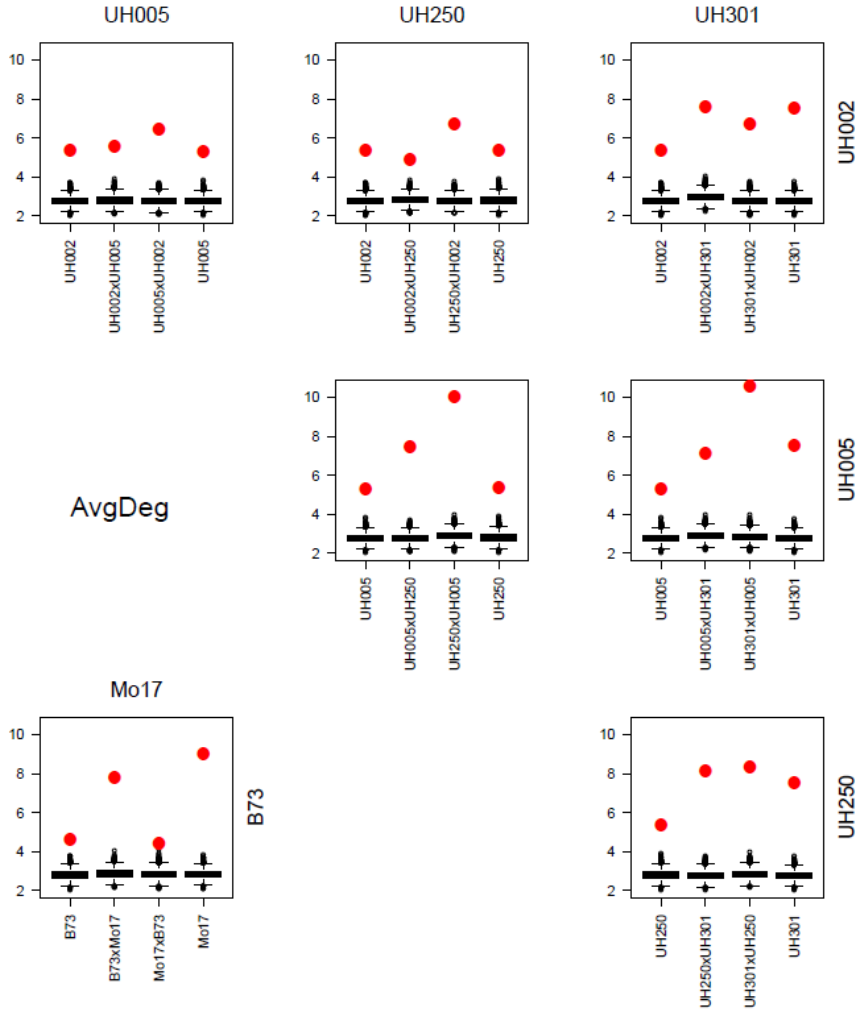
Two representative examples for metabolite variance in parental and hybrid genotypes show additive and overdominant heterosis pattern.



A paternal effect is observed for the network property Average Degree only in the American lines.

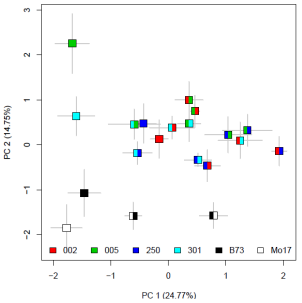


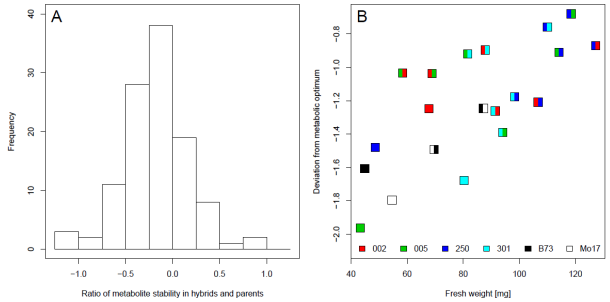
The correlation between Average Degree and root fresh weight is overlaid by a paternal effect.

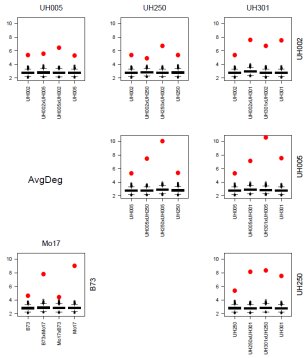


Summary of the *Zea mais* heterosis analyses

- Maize metabolome allows to distinguish ecotypes and homzygous 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.







Max-Planck-Institut für Molekulare Pflanzenphysiologie
Universität Potsdam
lisec@mpimp-golm.mpg.de

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)