

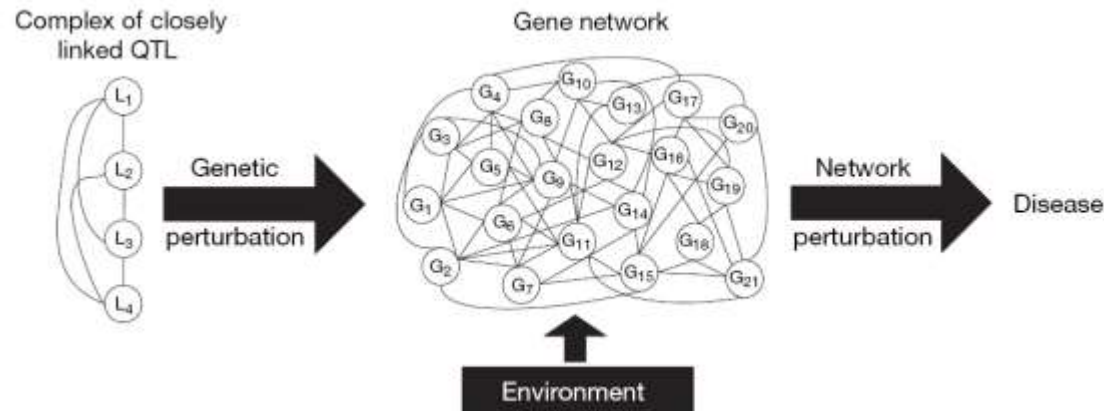
Genetical genomics of spruce

Kermit Ritland

Ilga Porth

Rebecca Dauwe

University of British Columbia



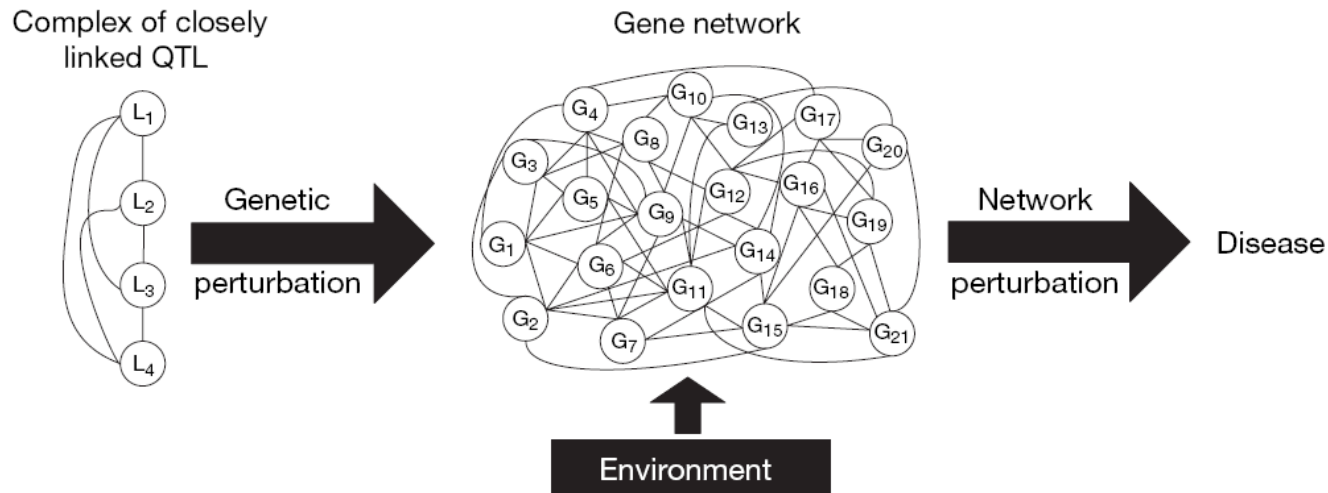
Genetical genomics

- Classical paradigm: amino acid changes underlie adaptation and evolution
- Role of gene expression gaining recognition
- **Microarrays** can infer *gene expression*

- Principle: Segregating progenies can be used to analyze genetic control of gene expression

- **Expression QTL mapping**
- **Metabolite QTL mapping**

Variation of QTLs in segregating progenies can elucidate gene networks that underlie complex traits (Chen et. al. Nature, March 2008)



Weevil and Spruce

- White pine weevil causes significant loss of growth in spruce, particularly Sitka spruce
- Weevil resistance is a quantitative genetic trait that can be addressed via QTL mapping and network analysis



Kalamalka Forestry Centre (BC Ministry of Forests)



- Interior spruce breeding center
- Controlled crosses involving trees of varying weevil resistance

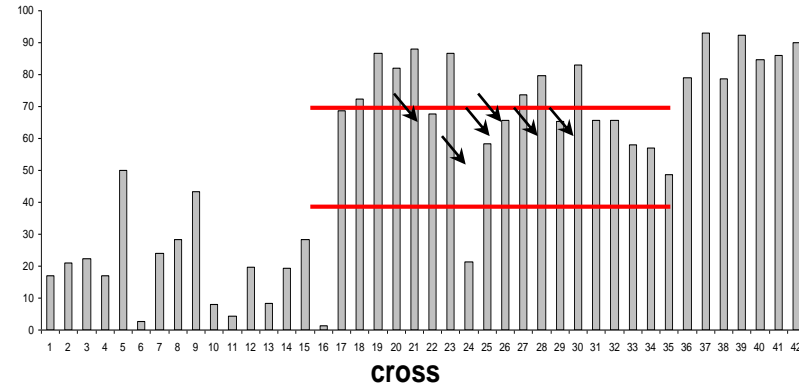
Full sib progenies at Kalamalka Research Station

Diallel crosses involving resistant and susceptible parents

(Alfaro et al. 2004, For. Ecol. Manage.)



Percentage of trees with kills for each cross



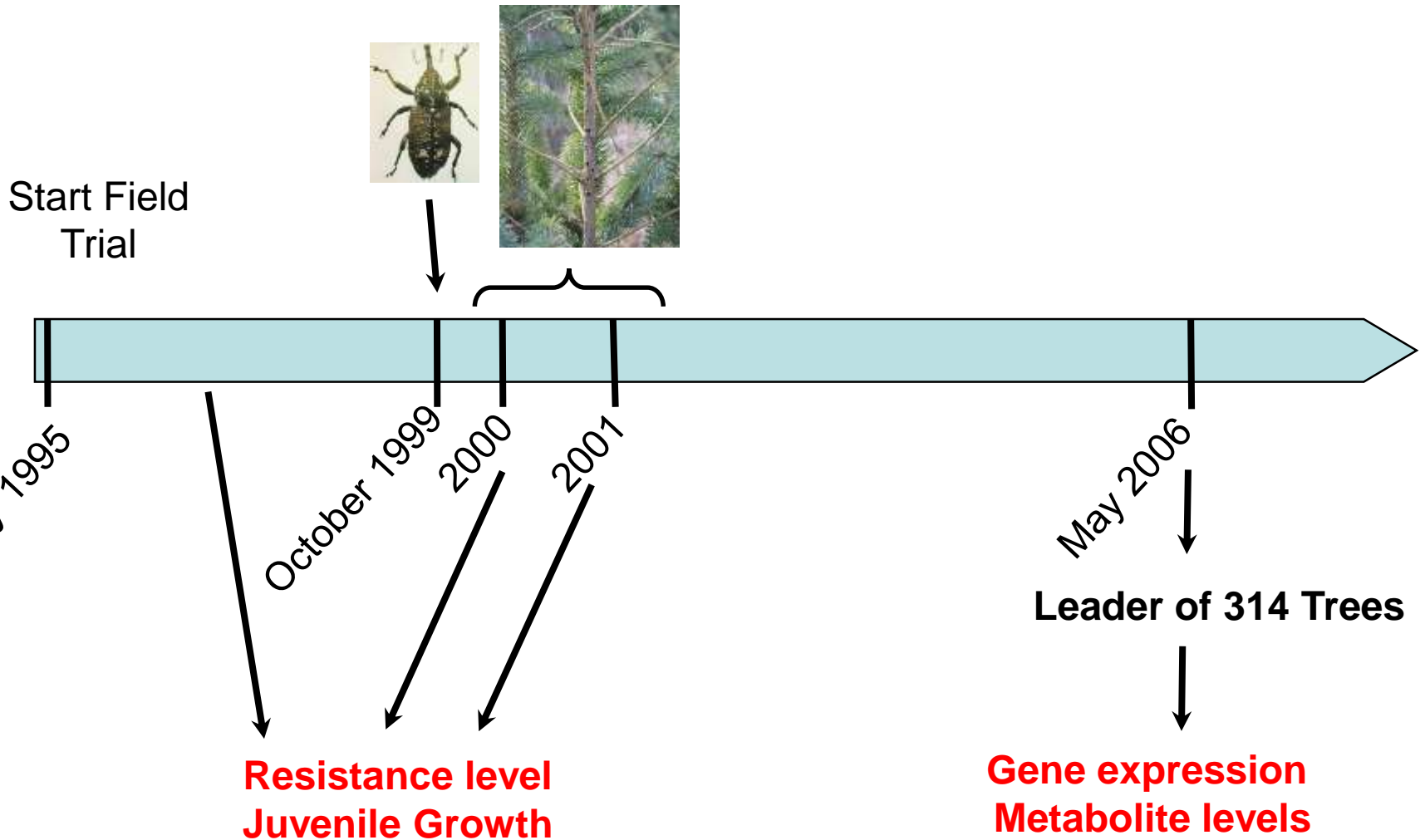
Factorial design chosen for QTL mapping:

Susceptible Male
X

Resistant Female

165	165	165	117	117	117
161	87	21	161	87	21

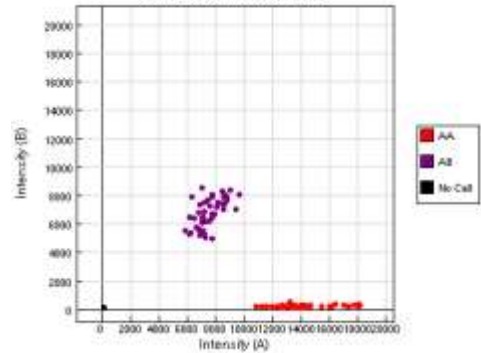
Weevil application and trait scoring timeline



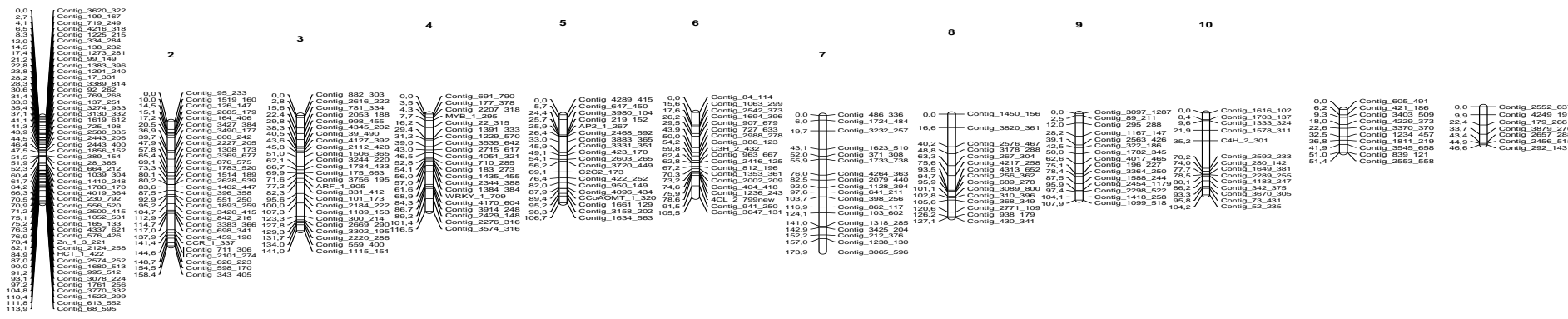
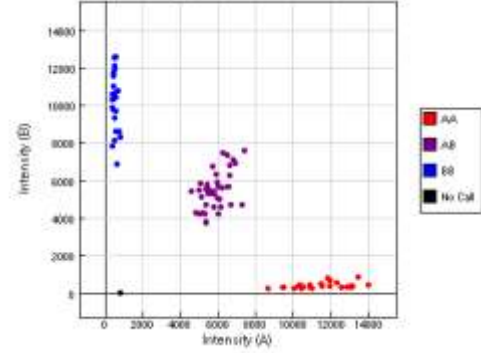
Segregation and genetic mapping in spruce

384 SNP loci genotyped in 96 well plates
About 120 segregate in any family (of the six)

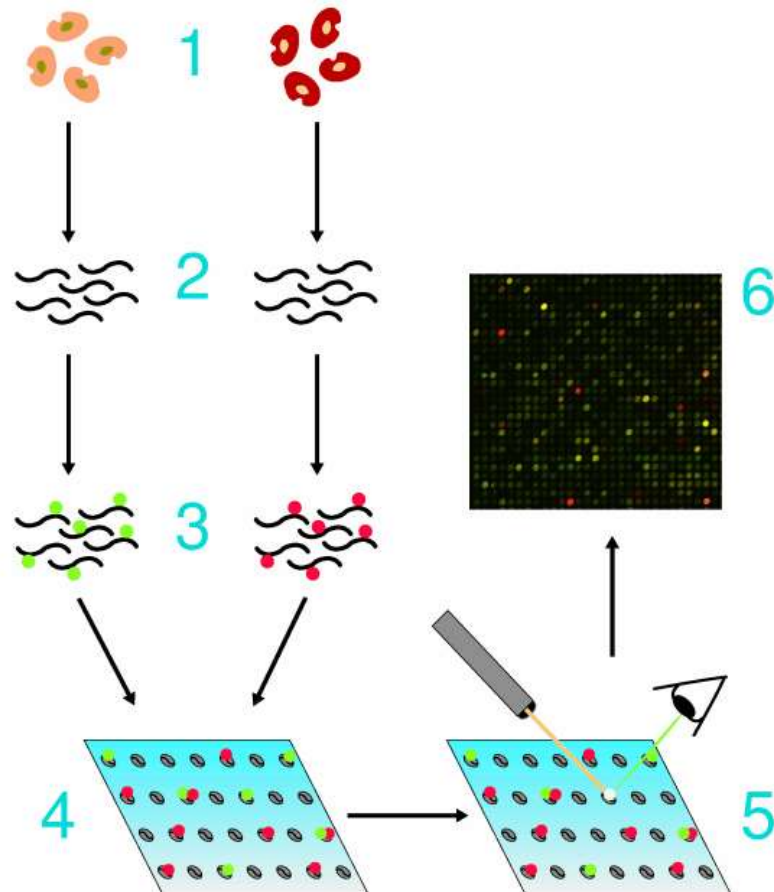
Genotype calls for locus 2152



Genotype calls for locus 5074



Comparative hybridization for assaying gene expression levels



eQTL and mQTL study

- 21843 member cDNA microarray previously developed for white spruce
- Xylem and bark samples collected June 2006 in three consecutive mornings of the same weather
- 92 hybridizations involving 184 trees (2x2 factorial)
 - ***Ilga Porth***, post-doctoral fellow, plus technical help
- 312 metabolites also assayed in xylem and bark (3x2 factorial)
 - ***Rebecca Dauwe***, post-doctoral fellow, plus technical help

A portion of genome segment for one mRNA (4 maps, one for each parent)

21843 x 4 = **87372** QTL maps!

Significance determined by randomization

SNP marker	LG	Map position	Estimation of effect by parent			
			1	2	3	4
Contig_1435_455	5	56.008	-0.036	-0.200	0.000	0.000
Contig_2344_388	5	57.017	0.000	0.230*	0.000	-0.214
Contig_1384_384	5	61.649	0.000	0.000	0.000	0.000
WRKY_1_709	5	68.852	0.000	0.000	0.000	0.000
Contig_4170_604	5	84.347	-0.107	0.047	0.000	0.000
Contig_3914_248	5	86.669	0.000	0.089	0.000	-0.089
Contig_2429_148	5	89.178	0.000	0.000	0.000	0.000
Contig_2276_316	5	101.444	0.000	0.000	0.000	0.000
Contig_3574_316	5	116.533	0.000	0.000	0.000	0.000
Contig_4289_415	6	0.000	0.000	0.000	0.000	0.000
Contig_647_450	6	5.655	0.000	0.000	0.000	0.000
Contig_3980_104	6	24.362	0.000	0.000	0.000	0.000
Contig_219_152	6	25.742	0.000	0.000	-0.084	0.000
AP2_1_267	6	25.914	0.034	0.000	0.000	0.000
Contig_2468_592	6	26.350	0.000	0.009	0.000	0.000
Contig_3883_365	6	32.999	0.000	0.000	0.000	0.000
Contig_3331_351	6	45.897	0.071	0.000	0.000	-0.083
Contig_423_170	6	49.100	0.000	0.007	0.000	0.040
Contig_2603_265	6	54.066	0.000	0.000	0.000	-0.070
Contig_3720_449	6	56.188	0.167	0.297*	0.000	0.000

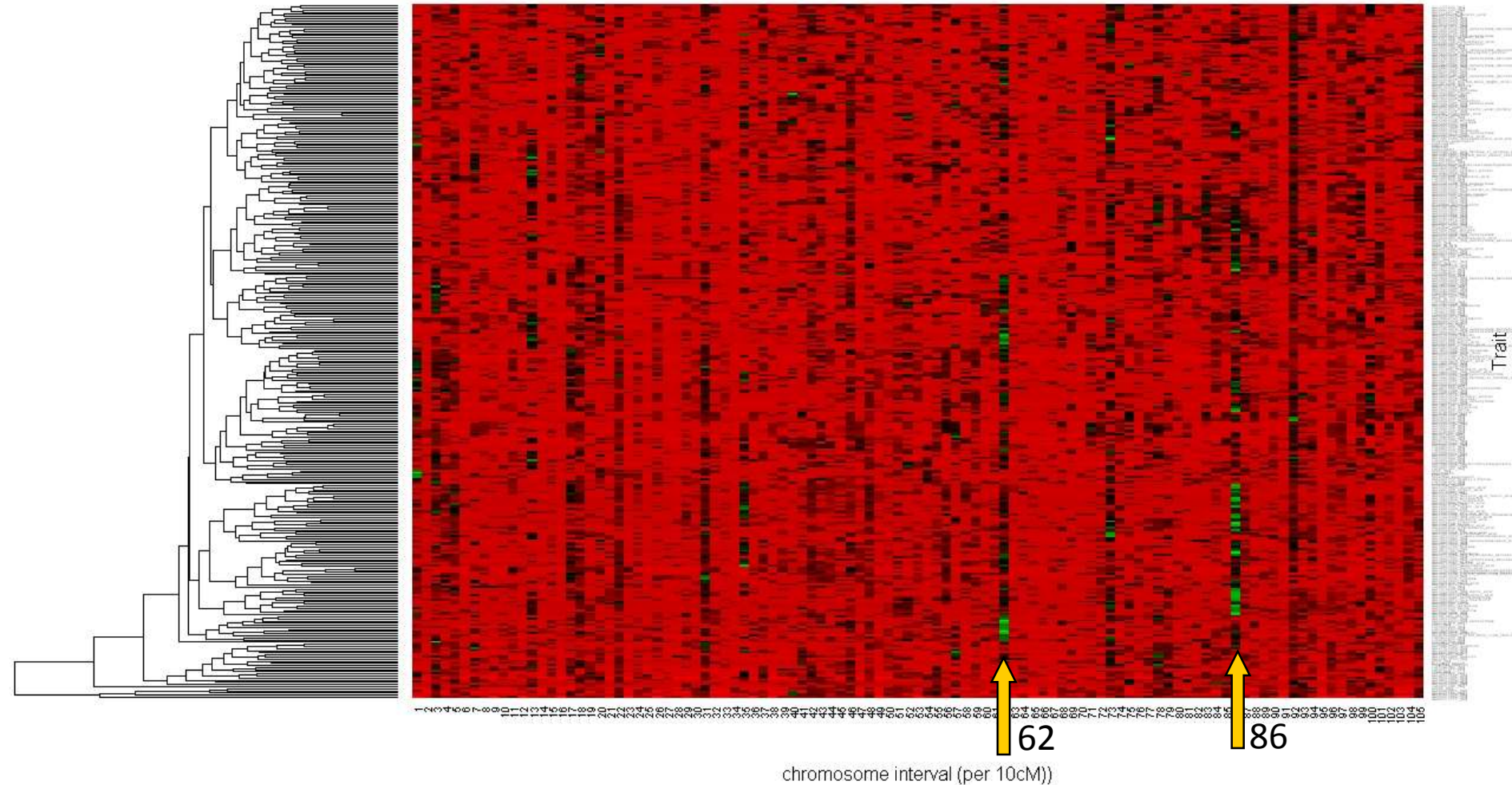
...

*p<.05 (effect estimates of zero are due to non-informative sites)

QTL co-localization:

- If an eQTL is in same spot as an QTL for growth/resistance, then the mRNA might control growth/resistance
- Therefore we can identify candidate genes for molecular breeding

Co-localizations of metabolite QTL along the genetic map of spruce



- Traits are grouped based on the hierarchical clustering of mapping patterns
- red - green: QTL effect; map position 62 and 86 have hot spots

QTL co-localization: two QTLs for different mRNAs map to the same position

QTL co-localization indicates shared pathway

In the following, an example for secondary metabolism pathways important for insect defense (total of 519 mRNAs):

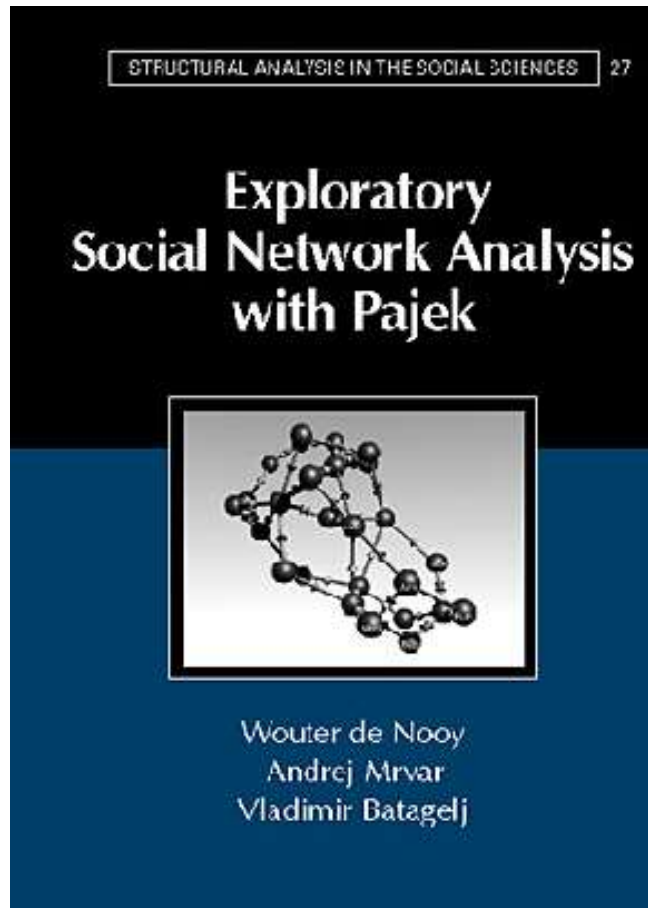
- Transcription factors
 - MYB (n=54)
 - AP2 (n=37)
- Secondary compound pathways
 - terpene synthase (n=60)
 - phenylpropanoids (n=277)
 - cytochrome P450 (n=89)

Numbers of mRNAs that are co-located with resistance/growth QTLS ($p < .05$)

(of 21843 mRNAs)

- Resistance traits 1355
- Growth traits 672

Network inference



- The computer program "Pajek" (Slovene word for spider)
- Network of vertices and edges
- Pioneered by **Rebecca Dauwe**, post-doctoral fellow (metabolomics)

All pathways

Dimension: 517

The lowest number of neighbours: 2

The highest number of neighbours: 98

All genes are connected to the network by at least 2 connections with p-value < 0.05

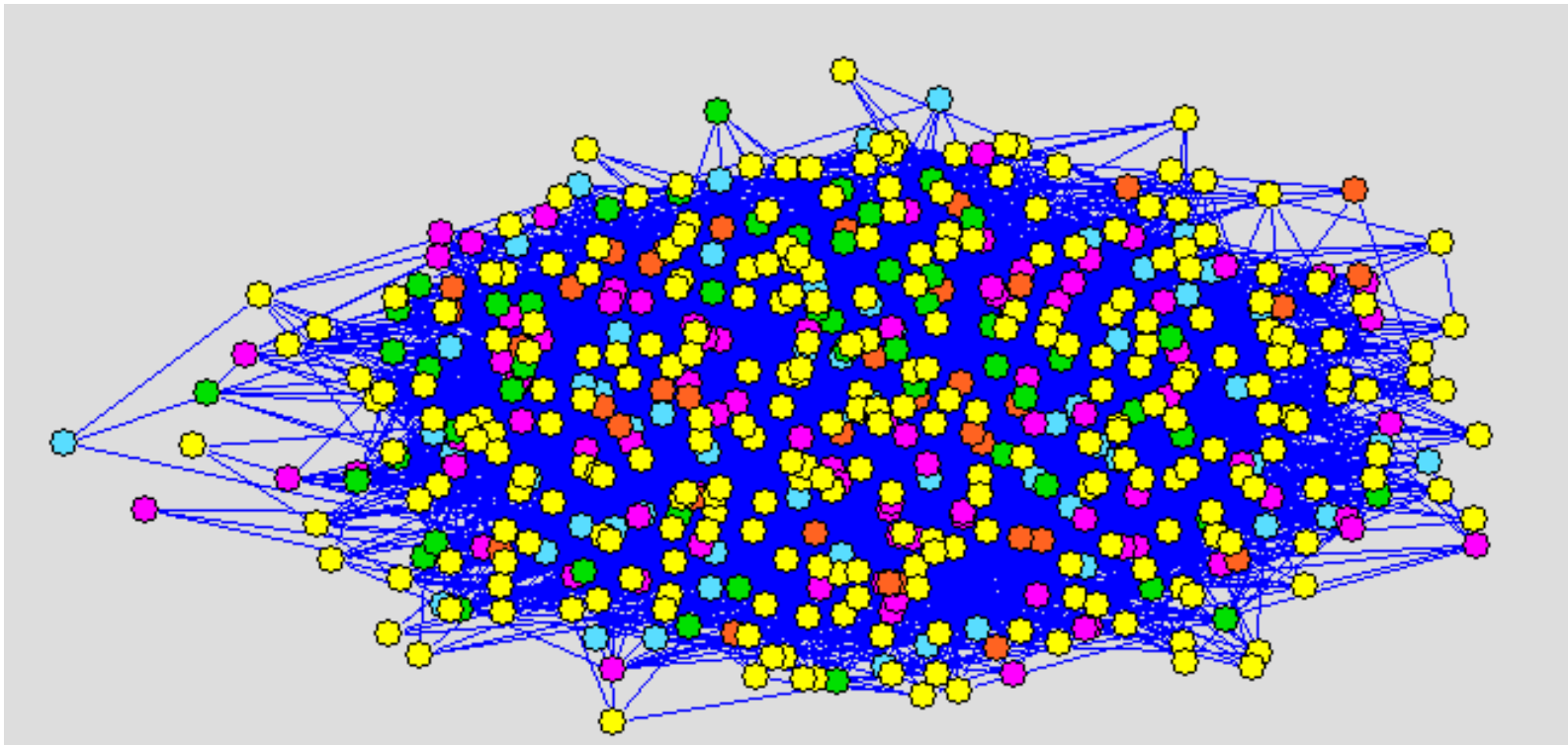
Phenylpropanoid

P450

Terpene

MYB

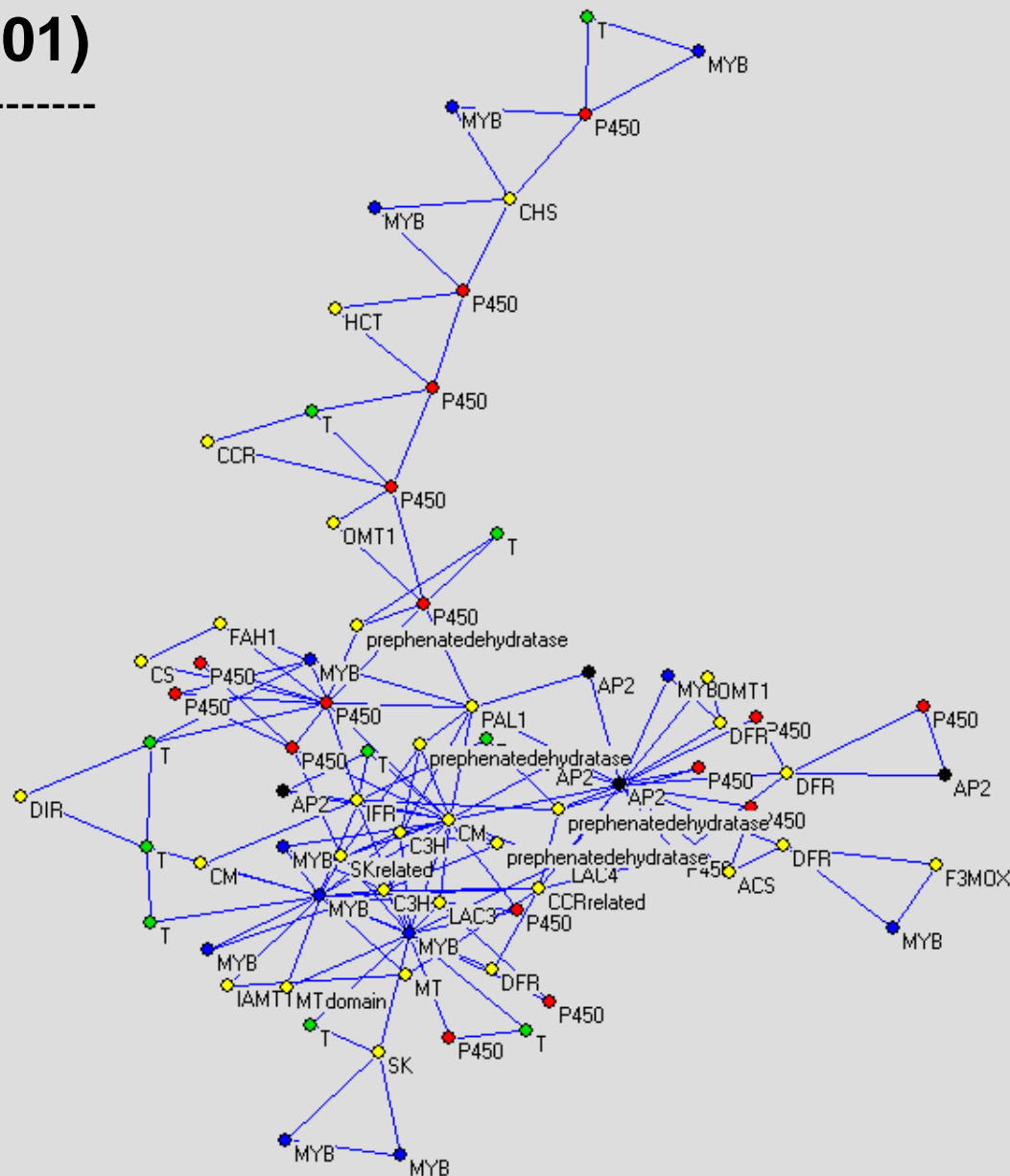
AP2



All pathways (p-value < 0.01)

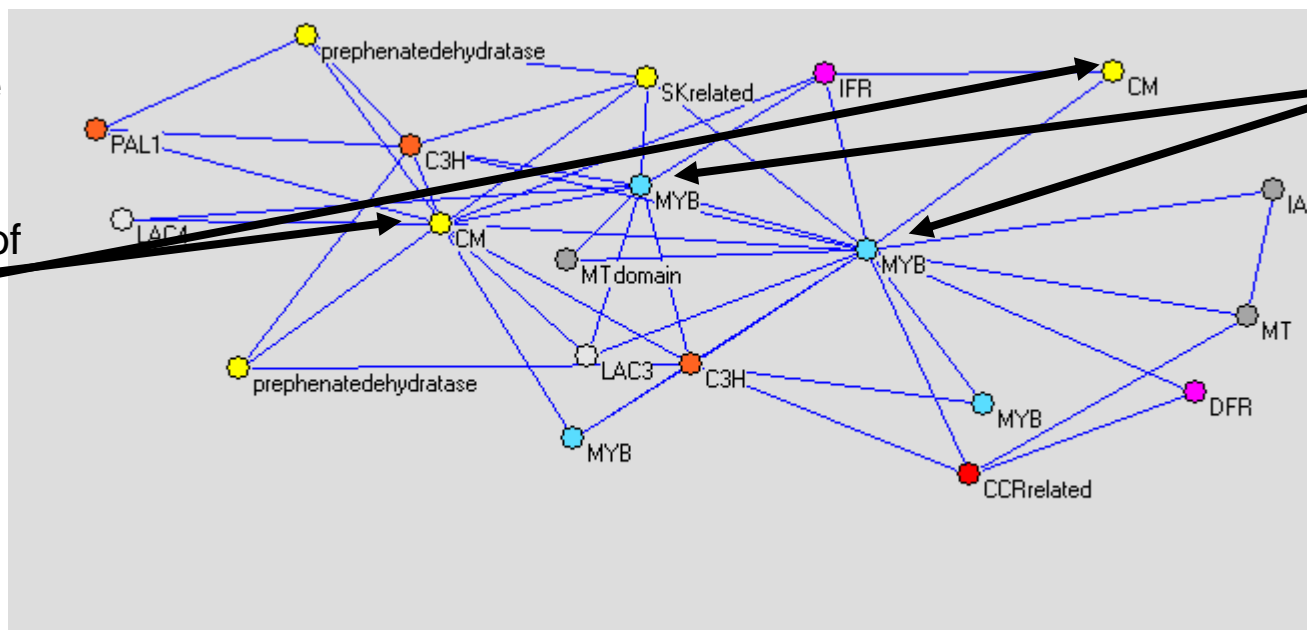
Note positions of MYBs, AP2 (stress induced transcription factor), and chorismate mutase

- Phenylpropanoid
- P450
- Terpene
- MYB
- AP2

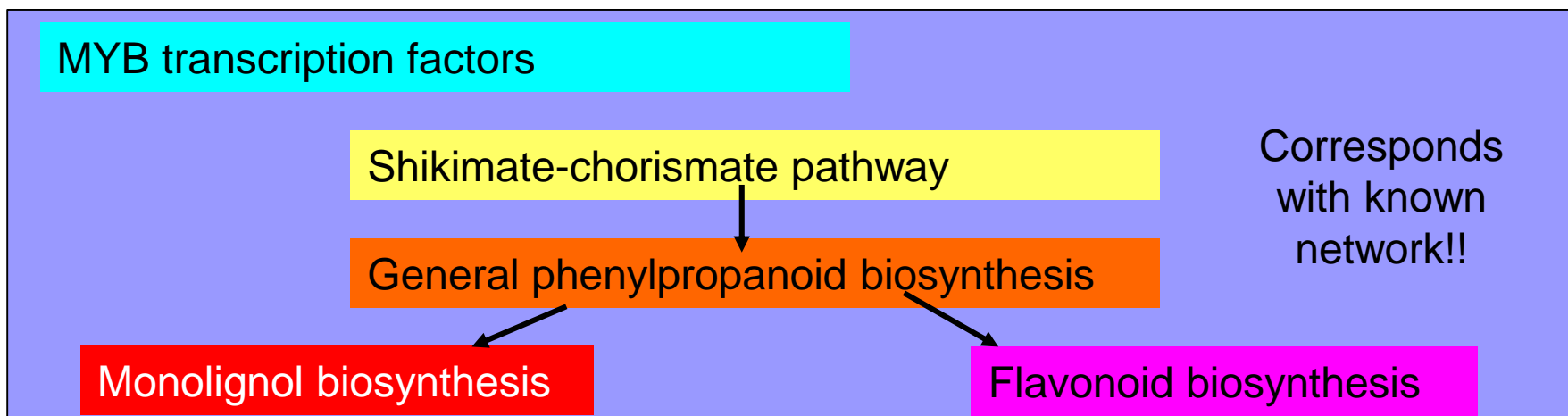


Network including only phenylpropanoid and MYB (terpene, P450, AP2 excluded) (P-value<0.01)

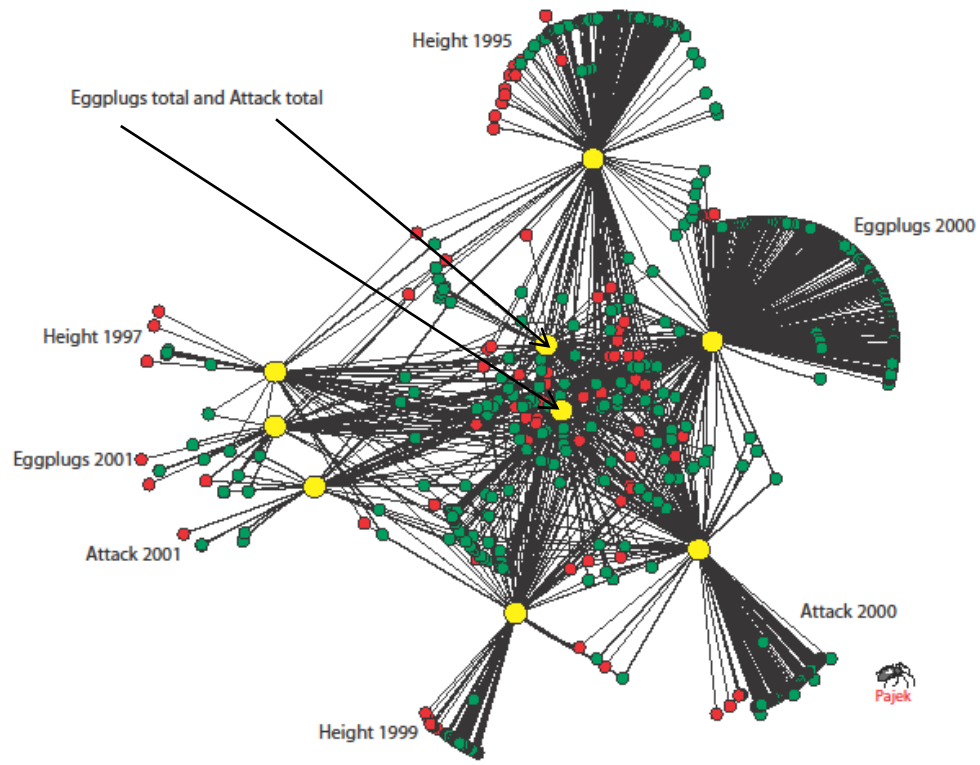
Network indicates the central regulatory importance of Chorismate Mutase in phenylpropanoid metabolism



Network indicates the 2 main MYB factors involved in phenylpropanoid metabolism

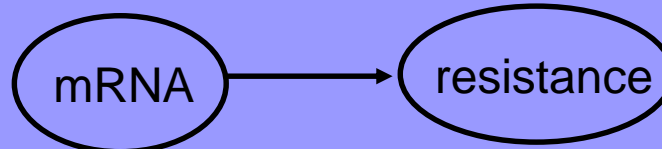


Co-localization of mRNAs with resistance and growth traits (preliminary global analysis)

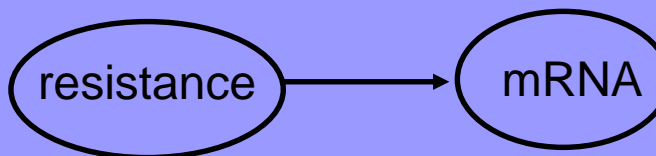


Next step: Causality can be inferred using the fact that random segregation allows directionality of gene action to be inferred between co-located QTLs (Schadt et al. Nature Genetics, July 2005)

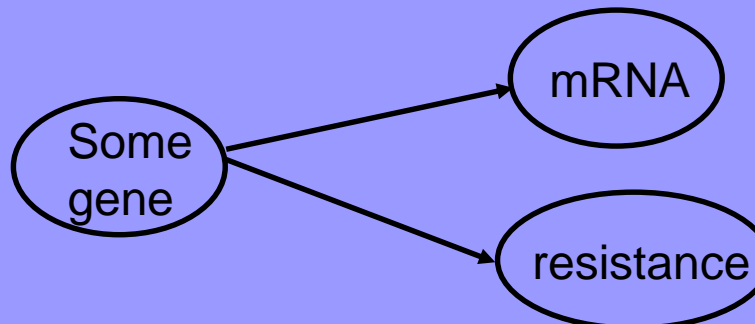
Causal model
(here, mRNA is a true candidate gene)



Reactive model



Independent model



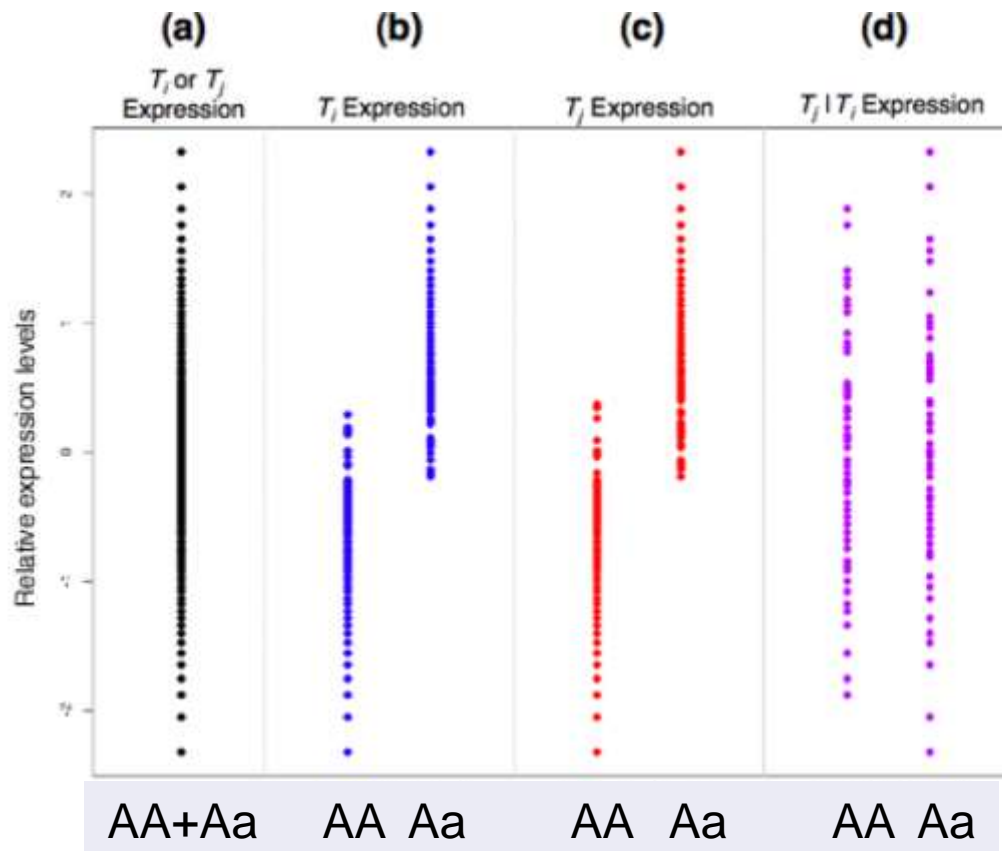
Harnessing naturally randomized transcription to infer regulatory relationships among genes

Lin S Chen*, Frank Emmert-Streib*† and John D Storey*†

Genome Biology 2007, 8:R219

1. Segregating SNPs randomizes putative upstream genes
2. Causality of effect can be disentangled
3. **Need shared QTL**

Principle:



Conclusion:
 T_j depends on T_i
(ideally, mRNA \Rightarrow resistance)

Inferring Causal Phenotype Networks From Segregating Populations

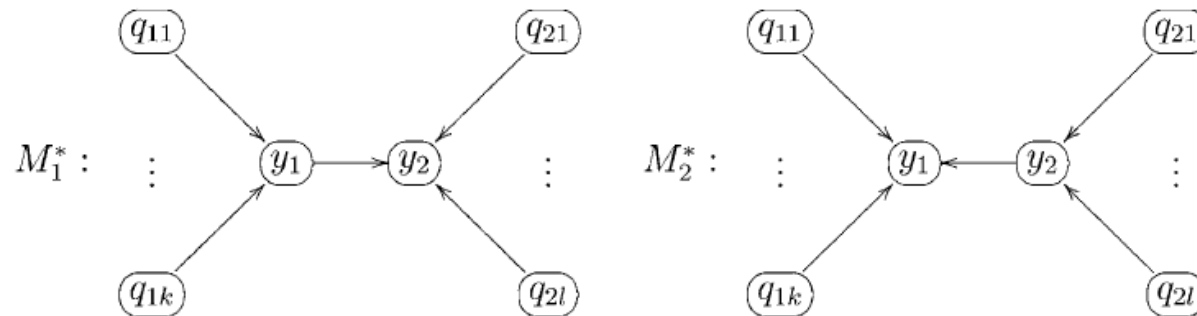
Elias Chaibub Neto,* Christine T. Ferrara,^{1,2} Alan D. Attie² and Brian S. Yandell^{*8,1}

^{*}Department of Statistics, ²Biochemistry Department and ⁸Department of Horticulture, University of Wisconsin, Madison, Wisconsin 53706 and ¹Sarah W. Stedman Nutrition and Metabolism Center and Department of Pharmacology and Cancer Biology, Duke University Medical Center, Durham, North Carolina 27704

Manuscript received November 28, 2007

Accepted for publication April 6, 2008

Maybe we don't need shared QTLs between mRNAs to infer direction



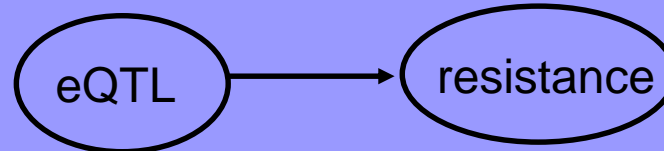
Y_1, Y_2 are different mRNAs, q 's are QTL

$$f(y_1 | \mathbf{q}_1) f(y_2 | y_1, \mathbf{q}_2).$$

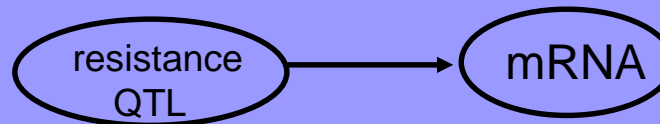
$$f(y_2 | \mathbf{q}_2) f(y_1 | y_2, \mathbf{q}_1).$$

Causality is inferred from random segregation between *co-located* QTLs (Schadt et al. Nature Genetics, July 2005)

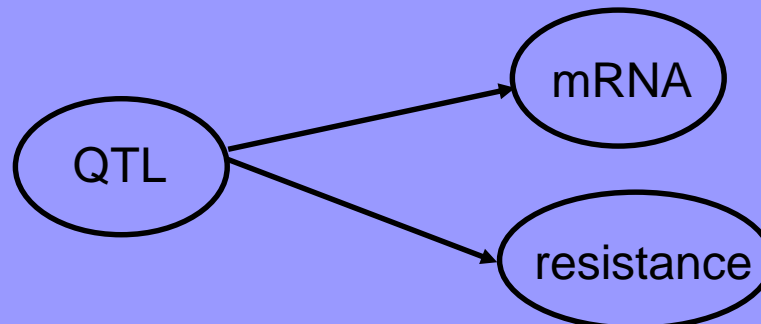
Causal model
(here, mRNA is a
true candidate gene)



Reactive model



Independent model



Procedure for identifying causation

Model 1. Marker \rightarrow trait X \rightarrow trait Y

- Predictor of X: U_A for trait X if marker is “AA”, U_a for trait X if marker is “Aa”.
- Predictor of Y: (predictor of X) \times (regression of Y on X)
- These are estimated using all individuals in a pedigree \rightarrow variance explained

Model 2. Marker \rightarrow trait Y \rightarrow trait X

Similar calculation, find % variance explained

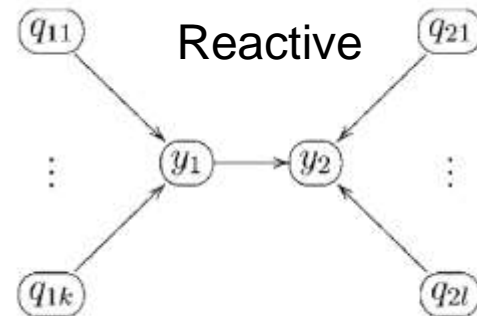
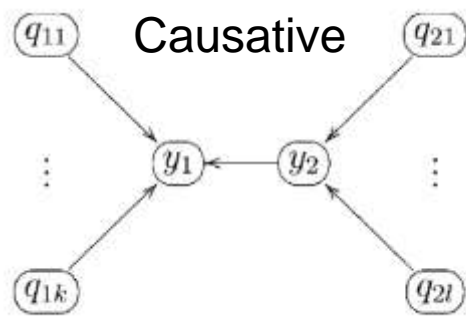
\rightarrow X

Model 3. Marker

\rightarrow Y

Both traits X and Y have means conditioned on marker genotype “AA” vs. “Aa”

The model that explains the most variance is picked

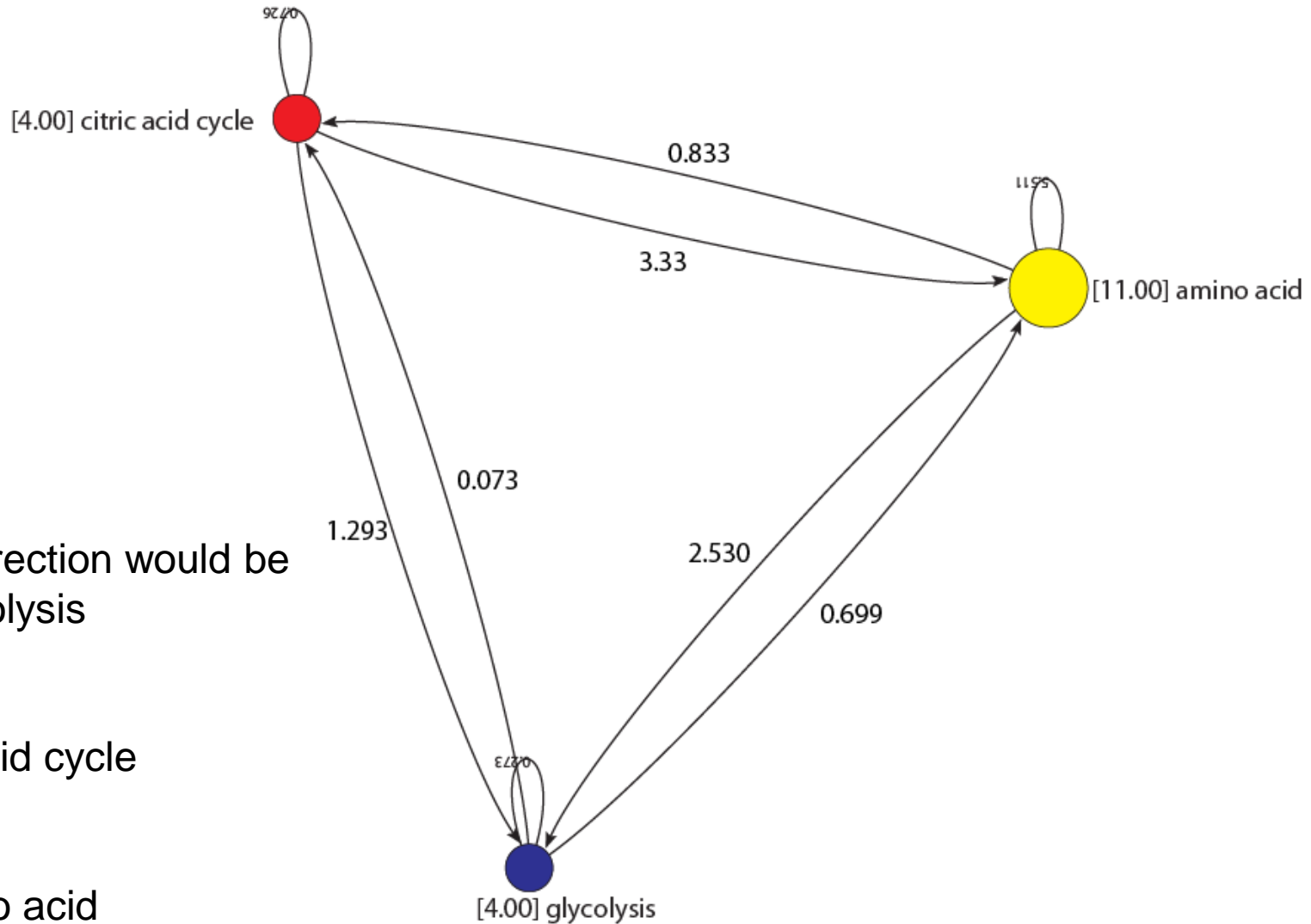


Numbers of cases where growth/resistance traits are causative vs. reactive, as inferred by Neto et. al

Growth/resist. traits	metabolites		genes on microarrays	
	<i>reactive</i>	<i>causative</i>	<i>reactive</i>	<i>causative</i>
Leader length 1999	155	167	52	466
Height 1995	202	120	125	393
Height 1999	285	37	51	467
Attack 2000	204	118	327	191
Eggplugs 2000	224	98	312	206

Metabolites: directed connections inferred from Pajek and that recent paper:

xylem metabolites and 3 connected pathways do NOT correspond to previous knowledge

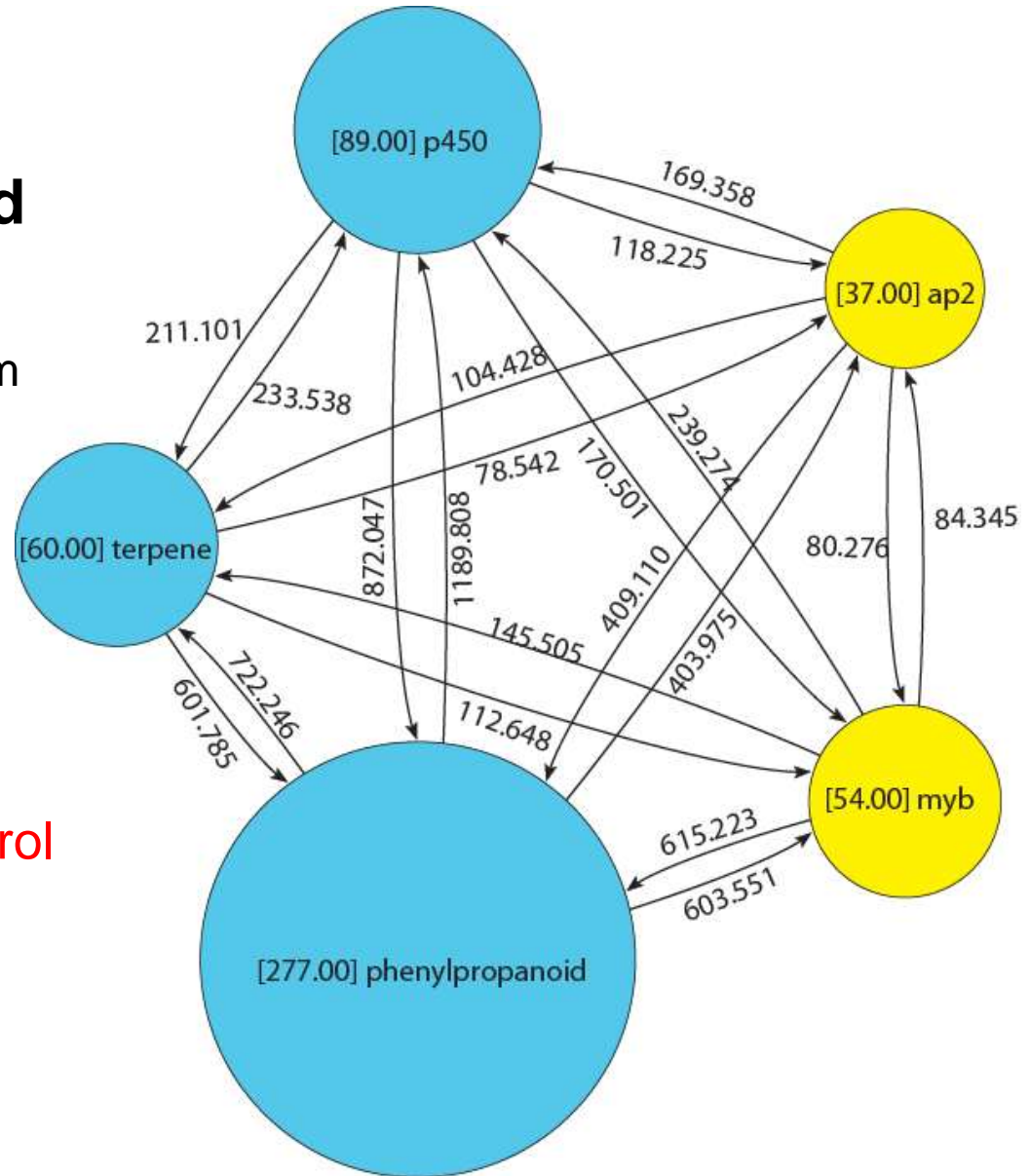


The expected direction would be
glycolysis
↓
citric acid cycle
↓
amino acid

517 transcripts (five gene families putatively involved in defense)

The direction is expected from transcription factors (ap2, myb) towards three gene families (p450, terpene, phenylpropanoid)

1. Myb Ap2 controls terpenes, p450
2. Phenylpropanoids control terpenes, p450





Genetical genomics II.

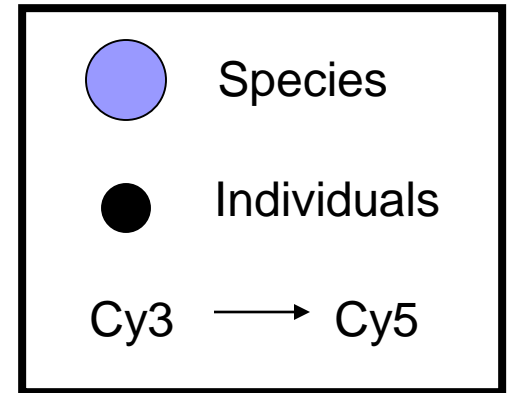
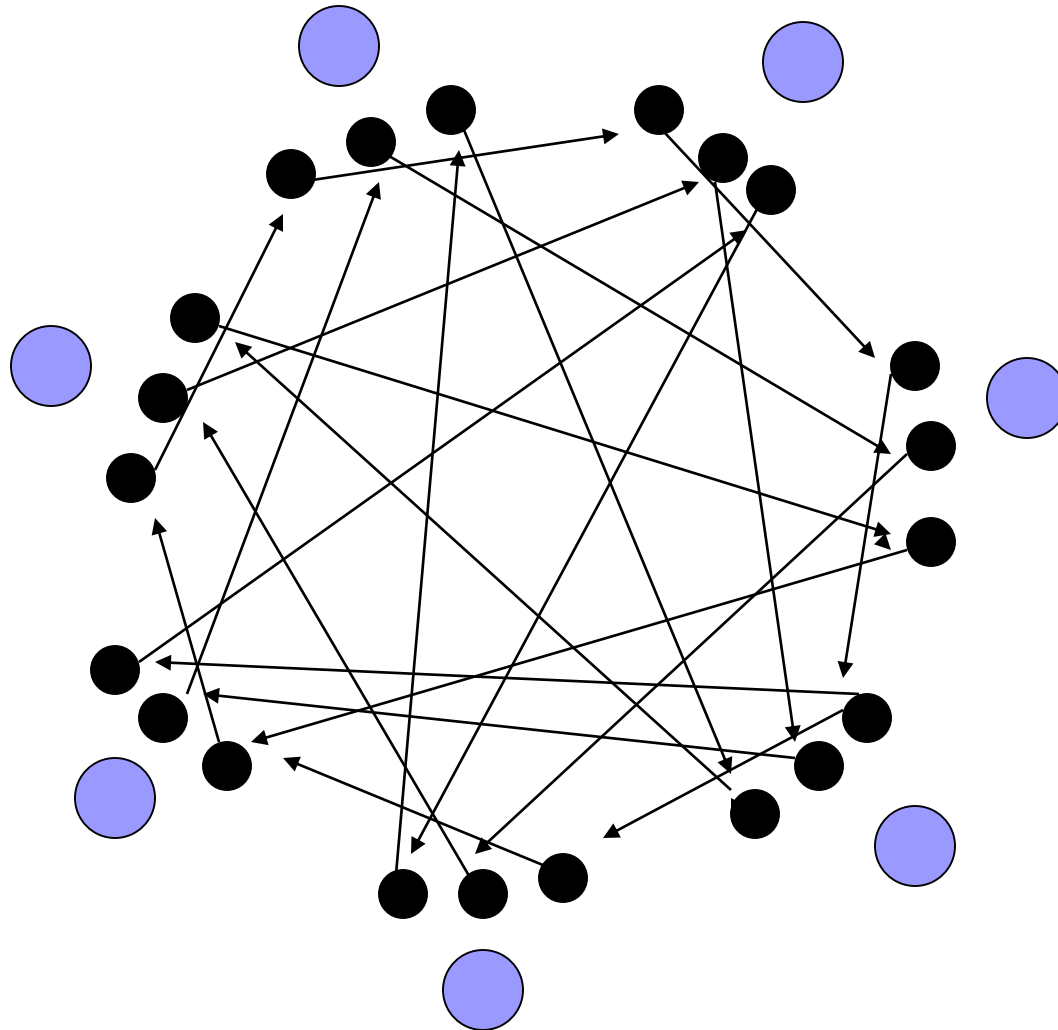
Gene expression evolution

- Divergent species illustrate the patterns of natural selection upon gene expression
- Ca. 30 spruce species
- 7 chosen for pilot study
- Project of **Rockney Albouyeh**, PhD student

Microarray experiment design

7 spruce species, 3 individuals for each species

21 hybridizations total



Inference of lineage-specific expression differences

-relative to mean over all species

-”star” phylogeny assumed

Clone	Lineage a	Lineage b	Lineage c	Lineage d	Lineage e	Lineage f	Lineage g	Species variance	Error variance	Species /total	chi-square (6 df)
WS0262_L03	-0.03	-0.01	0.18	-0.24	-0.12	0.19	0.03	0.024	0.001	0.972	206.558
WS0047_B04	-0.10	0.17	0.09	-0.17	0.27	0.17	0.02	0.04	0.001	0.971	206.558
WS01027_A02	0.05	0.05	-0.22	-0.04	0.15	0.15	-0.13	0.020	0.001	0.971	203.785
WS00723_G05	0.05	-0.44	0.02	-0.02	0.06	-0.17	0.50	0.080	0.002	0.971	203.785
...
WS0061_B08	0.02	-0.08	0.08	0.07	0.22	-0.17	-0.13	0.019	0.019	0.497	5.917
WS0053_E12	0.13	-0.20	0.21	0.05	-0.32	-0.12	0.26	0.048	0.049	0.496	5.917
WS0035_A17	0.21	-0.05	-0.17	0.02	0.12	0.19	0.08	0.021	0.022	0.496	5.916
WS0038_D14	-0.22	0.04	0.17	0.05	0.46	-0.20	-0.30	0.071	0.072	0.496	5.913
WS0093_M07	0.05	0.02	0.06	-0.20	-0.08	0.09	0.05	0.011	0.011	0.496	5.909
...
WS00937_G12	0.09	0.01	-0.01	0.06	-0.25	0.23	-0.12	0.023	0.721	0.031	0.195
WS01010_O12	-0.52	0.04	-0.03	-0.14	0.55	-0.03	0.13	0.101	3.128	0.031	0.194
WS00122_B05	0.78	-0.19	0.30	-0.21	0.2	-0.30	0.69	0.132	4.129	0.031	0.191
WS01028_G20	0.15	0.05	-0.01	-0.01	0.00	0.16	-0.07	0.013	0.425	0.031	0.190
WS01014_I04	-0.01	-0.07	0.07	0.07	-0.03	-0.01	-0.01	0.003	0.088	0.031	0.189

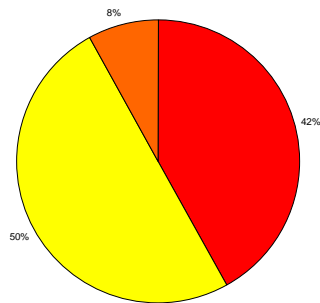
Diversifying or “Darwinian” selection

No selection

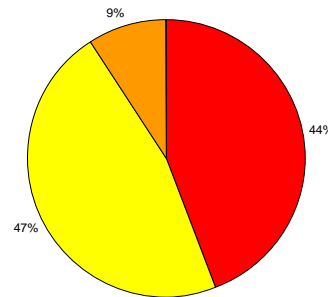
Stabilizing selection

Evolution of gene expression in gene families involved with resistance

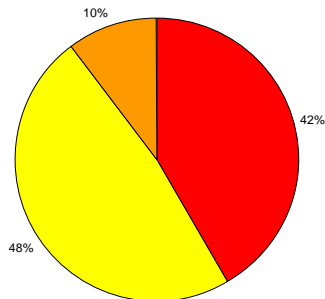
All clones (n=21,334)



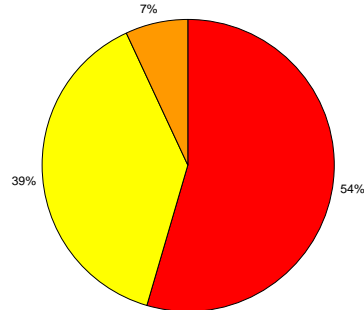
Phenylpropanoid genes (n=329)

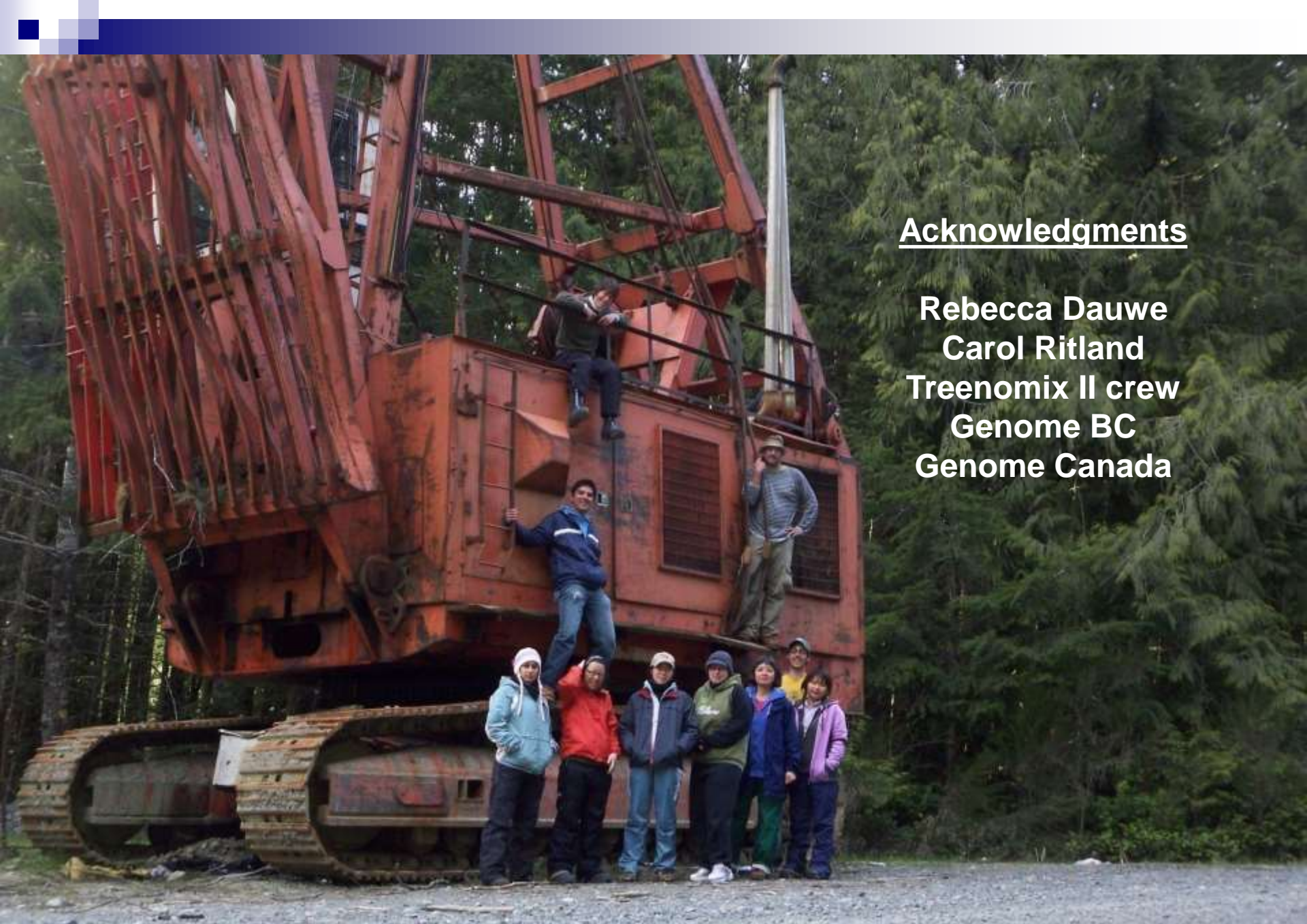


P450 genes (n=96)



Terpenoid genes (n=88)





Acknowledgments

Rebecca Dauwe
Carol Ritland
Treenomix II crew
Genome BC
Genome Canada