Identification of altered biochemical domains between pumpkin and tomatillo leaf metabolites
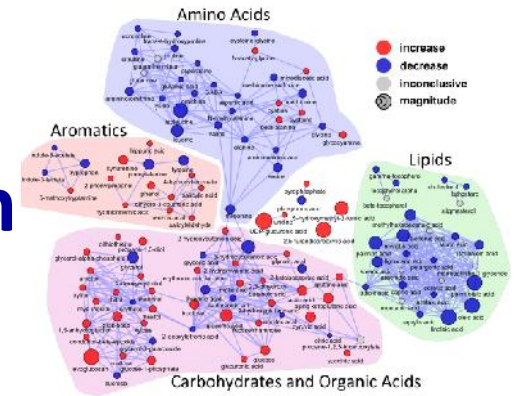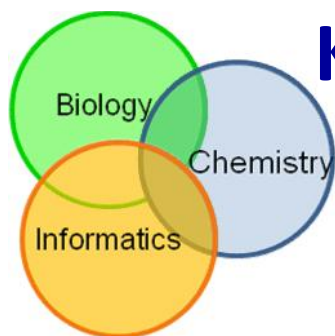
Use DATA: Pathway Enrichment data.csv

Goal: Identify significantly over represented biological pathways based on significant differences in leaf metabolites

Topics:

1. KEGG Database
2. MetaboAnalyst: Pathway enrichment analysis
3. MBrole: Over Representation Analysis
4. Hypergeometric test for enrichment

KEGG Pathway Visualization

Goals:

Use KEGG to:

1. Overview glutamate entry in KEGG (C00025)
2. Visualize a pathway of interest
3. Map metabolite of interest to pathway

- http://www.kegg.jp/dbget-bin/www_bget?C00025
- http://www.kegg.jp/kegg-bin/show_pathway?org_name=ath&mapno=00250

- Mapping example

C00064 green, black
C00025 green, black

Pathway Visualization

Pathway over representation analysis (ORA)

MBR🌸LE
Metabolites Biological Role

Biochemical Enrichment Analysis
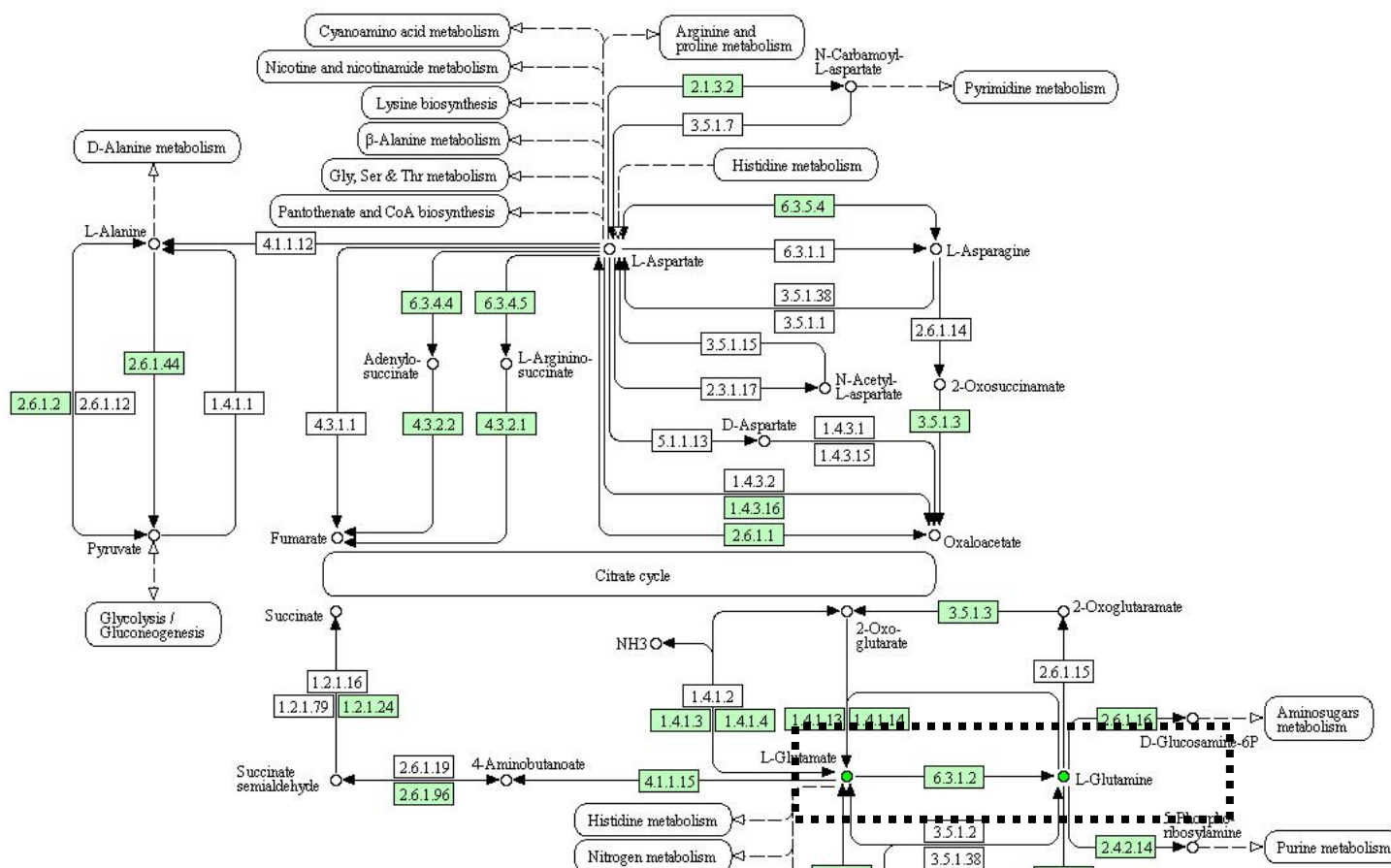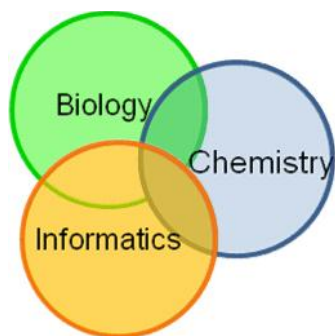
Steps:

1. Use MBrole to conduct:
   - Pathway over representation analysis
   - url: http://csbg.cnb.csic.es/mbrole/

ORA:

- is used to evaluate whether a particular set of metabolites is represented more than expected by chance within a given compound list [doi: 10.1093/nar/gkq329].

- p-value is calculated using hypergeometric or Fisher's exact test
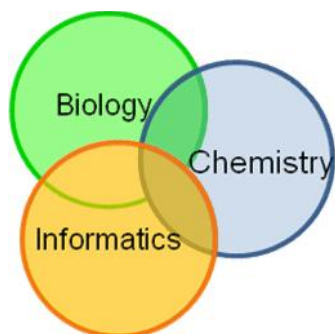
MBRole:
Pathway Over Representation Analysis (ORA)

Goal: Identify an over represented pathway and visualize it in KEGG

Biochemical Enrichment Analysis

MBR⬤LE
Metabolites Biological Role

Home | Analysis | ID conversion | Help

MBRole results for:

Pathways (73)

Enzyme interactions (223)

Other interactions (88)

Biological role (51)

Chemical groups (28)

Pathways

☐ Show compounds with no annotations (15)
☑ Show compounds in results table
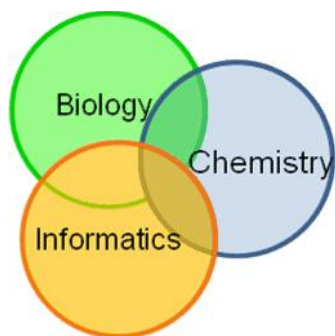
Show annotations with p-value ≤ 0.05   [Set filter]

Pathways (set: 72 background: 3358)

id	label	p-val	adjusted p-val	in bckgnd	in set	%	Compounds
ath01100	Metabolic pathways	1.72E-6	1.79E-5	1455	51	70.8	C00474 C00407 C00149 C00751 C00085 C00041 C00093 C05422 C00037 C00026 C00334 C00103 C00864 C05437 C00064 C06427 C00315 C00049 C00493 C00031 C00078 C00009 C00097 C00042 C02291 C00134 C00072 C00446 C00047 C00092 C00333 C00811 C00327 C00082 C02477 C00095 C00073 C00158 C01419 C00065 C00187 C00160 C00514 C00025 C00263 C01724 C00189 C00152 C00209 C00079 C00198

BIOSYNTHESIS OF PLANT HORMONES
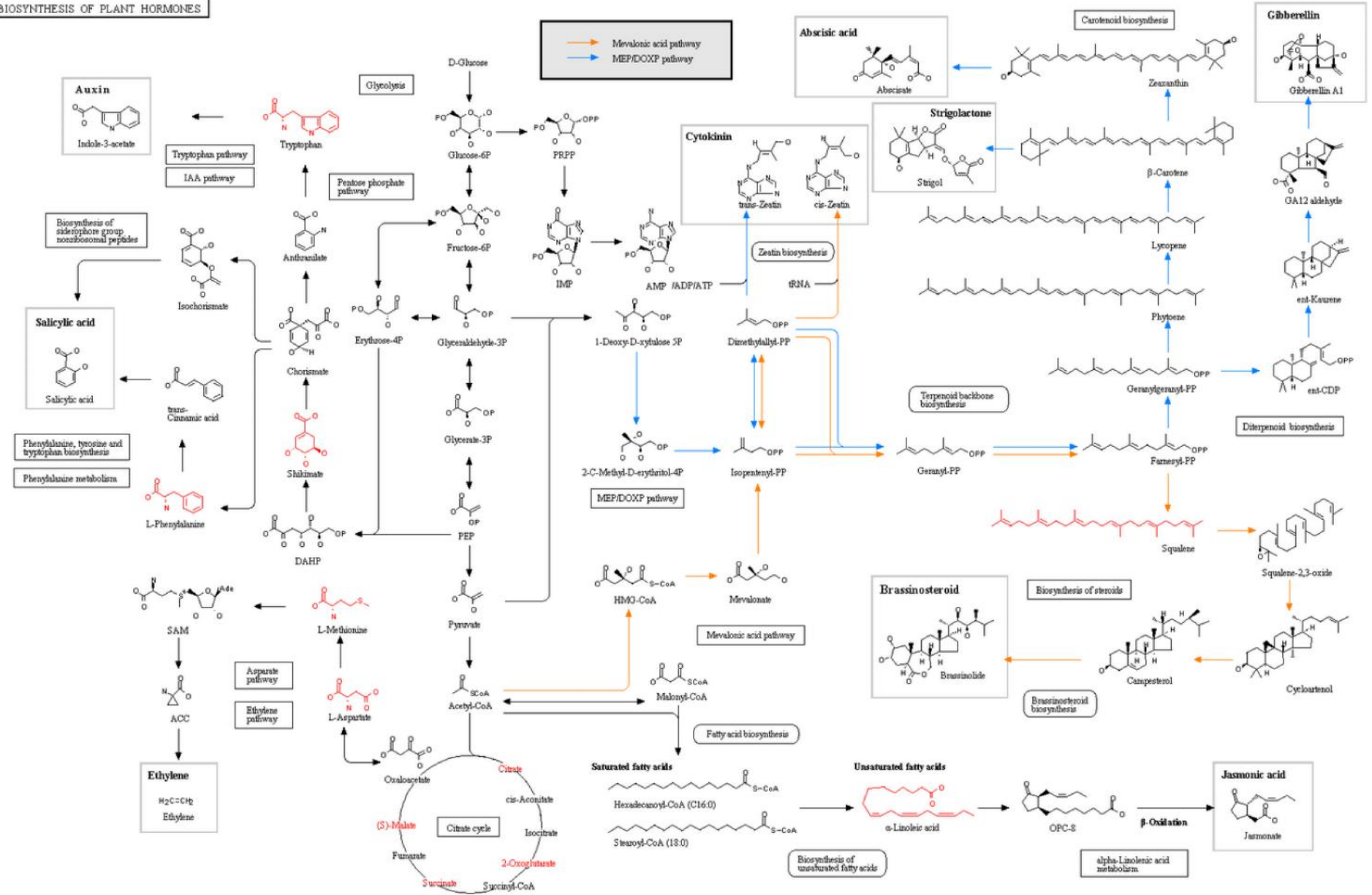
http://www.genome.jp/kegg-
bin/show_pathway?map01070+C06427+C00158+C00049+C00493+C00079+C00026+C00042+C00751+C00149+C00078+C00073+

Test for significance: Hypergeometric Test

Biochemical Enrichment Analysis

Pathways (set: 72 background: 3358)

id	label	p-val	adjusted p-val	in bckgnd	in set	%
ath01100	Metabolic pathways	1.72E-6	1.79E-5	1455	51	70.8

How to calculate statistics to determine network enrichment?
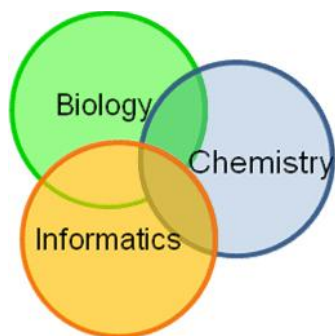
hit.num = 51 # number of significantly changed pathway metabolites
set.num = 1455 # number of metabolites in pathway
full = 3358 # all possible metabolites in organism
q.size = 72 # number of significantly changed metabolites

phyper(hit.num-1, set.num, full-set.num, q.size, lower.tail=F)
= 1.717553e-06

MetaboAnalyst:
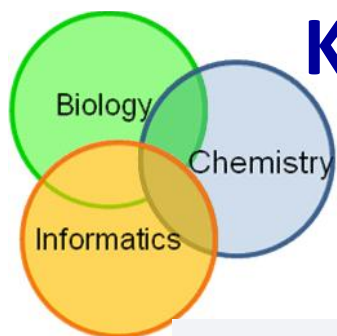# Pathway Enrichment Analysis (PEA)

Use MetaboAnalyst to conduct:

- Pathway enrichment Analysis
- url: http://www.metaboanalyst.ca/MetaboAnalyst/faces/UploadView.jsp

PEA:

- is an advanced form of over representation analysis (ORA) which takes into account pathway topology and is based on gene set enrichment analysis (GSEA) [doi:10.1093/bioinformatics/btq418]

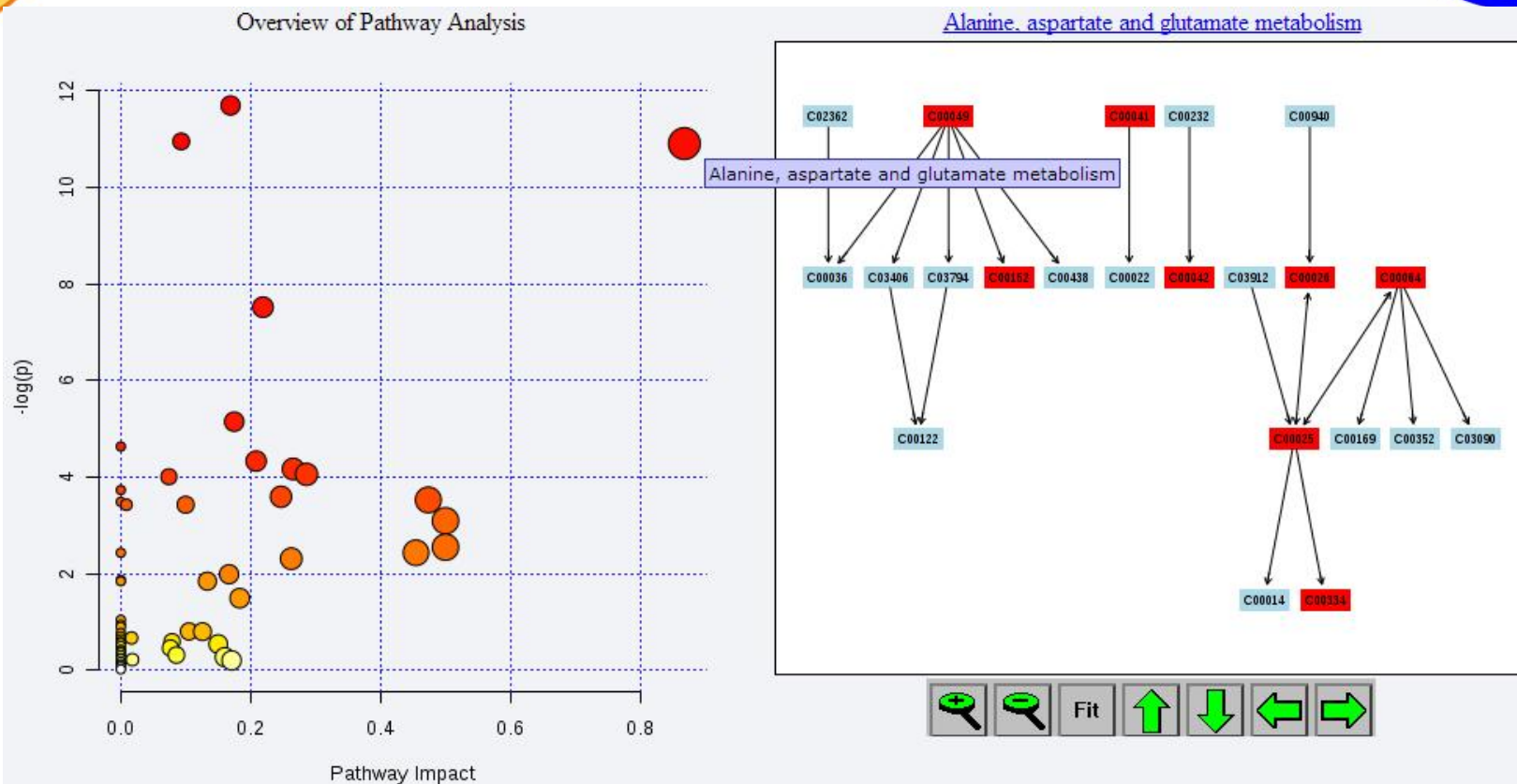- p-value is calculated using hypergeometric or Fisher's exact test

Questions:

1. What pathway is the most important based on ORA and topology?

KEGG Pathway Enrichment

Biochemical Enrichment Analysis

Pathway Name	Total	Hits	p	-log(p)	Holm p	FDR	Impact	Details
Glutathione metabolism	26	9	8.3684E-6	11.691	7.2805E-4	5.3375E-4	0.16897	KEGG
Aminoacyl-tRNA biosynthesis	67	14	1.7588E-5	10.948	0.0015126	5.3375E-4	0.09302	KEGG
Alanine, aspartate and glutamate metabolism	22	8	1.8405E-5	10.903	0.0015645	5.3375E-4	0.86781	KEGG
Galactose metabolism	26	7	5.44E-4	7.5166	0.045696	0.011832	0.21886	KEGG
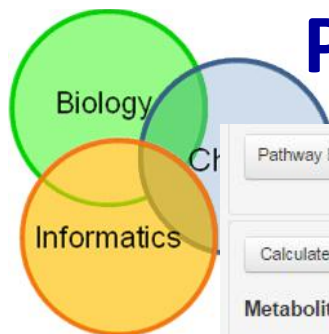
PathViewR Visualize changes in metabolites for a pathway/organism of interest

increased or decreased in tomatillo vs. pumpkin

Biochemical Enrichment Analysis