Comparative of leaf gradient transcriptomics in multiple C3 and C4 species

Zehong Ding
Bioinformatics Facility
Cornell University

Feb 2, 2015

Outline of the talk

- Motivation of the study
- Introduction of samples and workflow
- Results (e.g., C4 candidates definition)
- Main conclusion
- Pros vs. Cons
- Acknowledgement

Study background

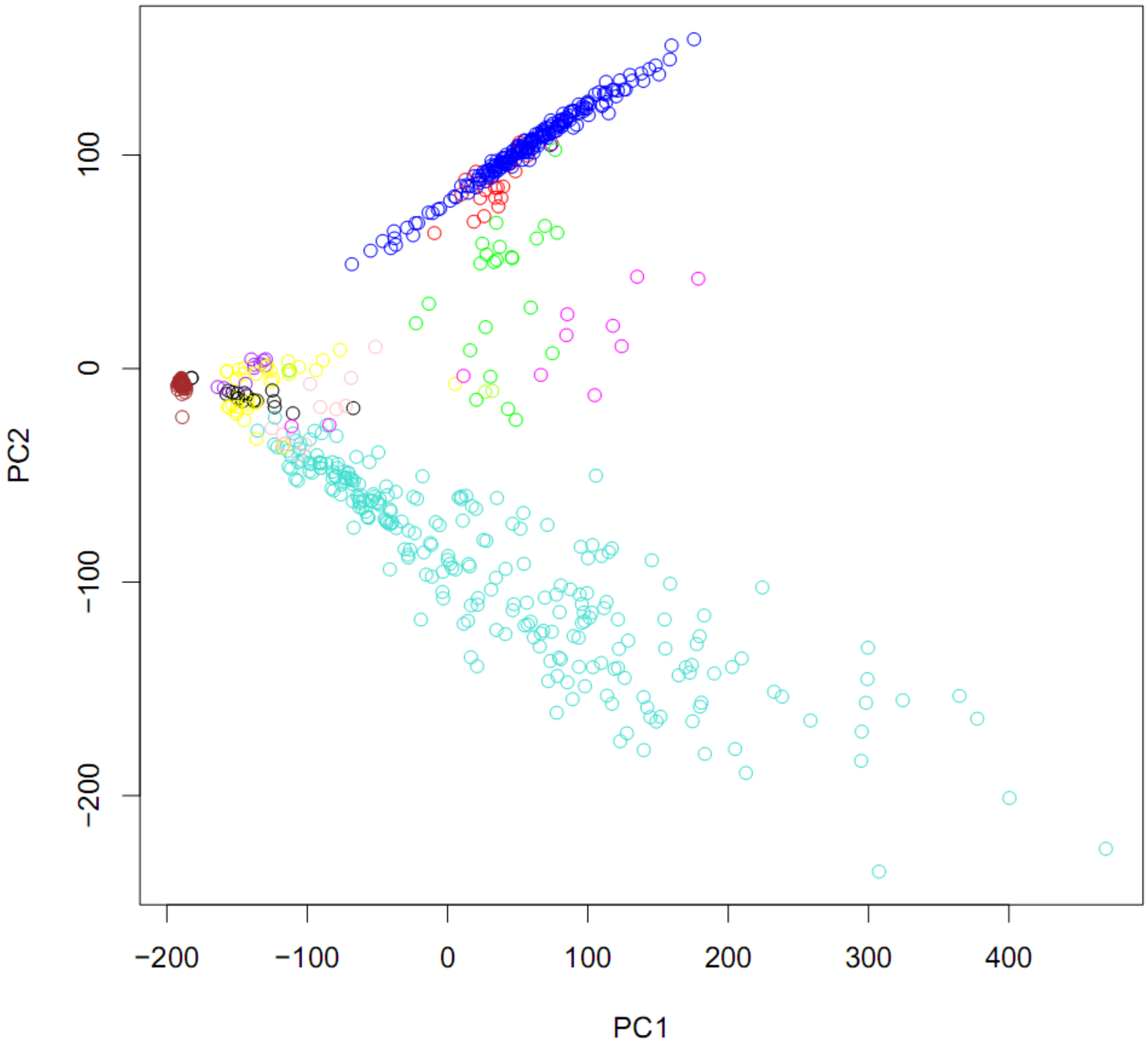
Main motivation: food crisis

- ✓Growing population,
- ✓decreasing area of agricultural land,
- ✓extreme changes of weather, e.g., drought
- ✓decreasing availability of resources, e.g., water and energy

One of solutions: introduce the C4 photosynthetic pathway into C3 crops, because C4 crops usually have higher radiation use efficiencies, water and nitrogen use efficiencies, and thus yields

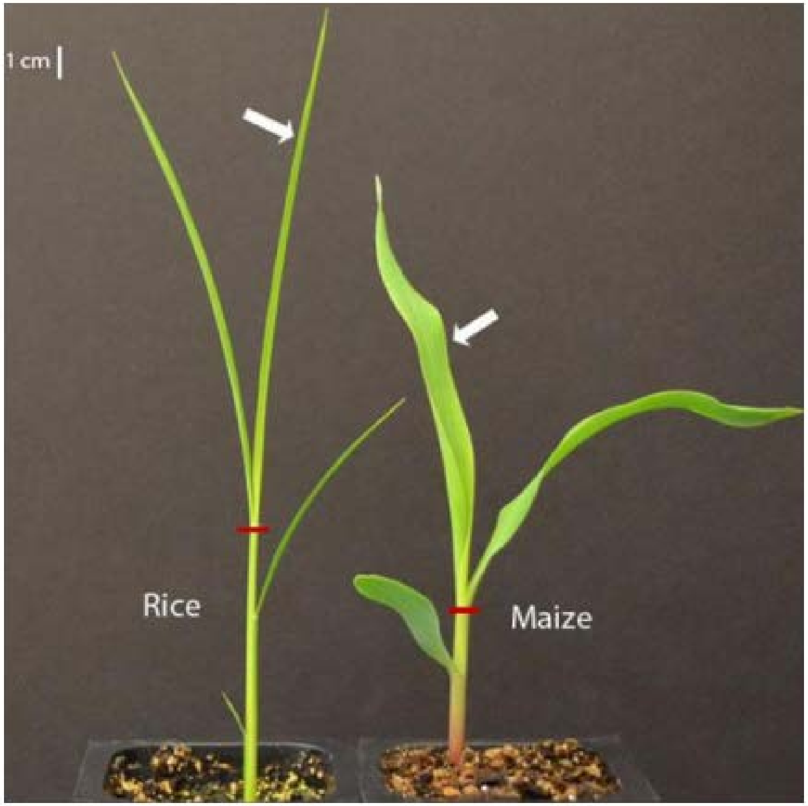
Our object: to find C4 candidate genes, which could be used for crops production improvement

Public maize data (~660 samples)

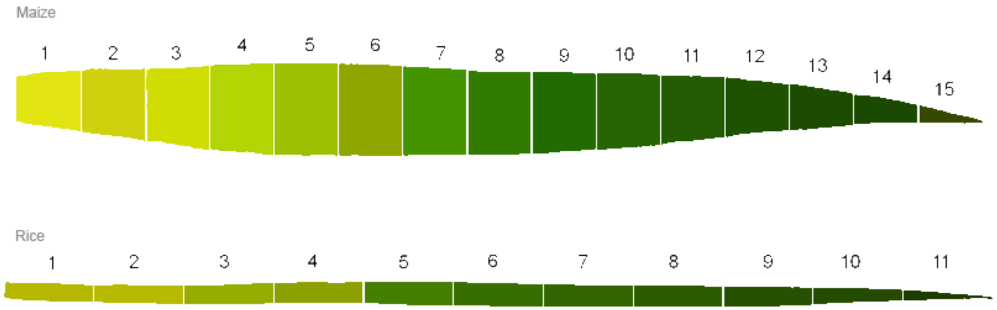


(NCBI SRA database,
2013.08)

Developmental dynamic samples

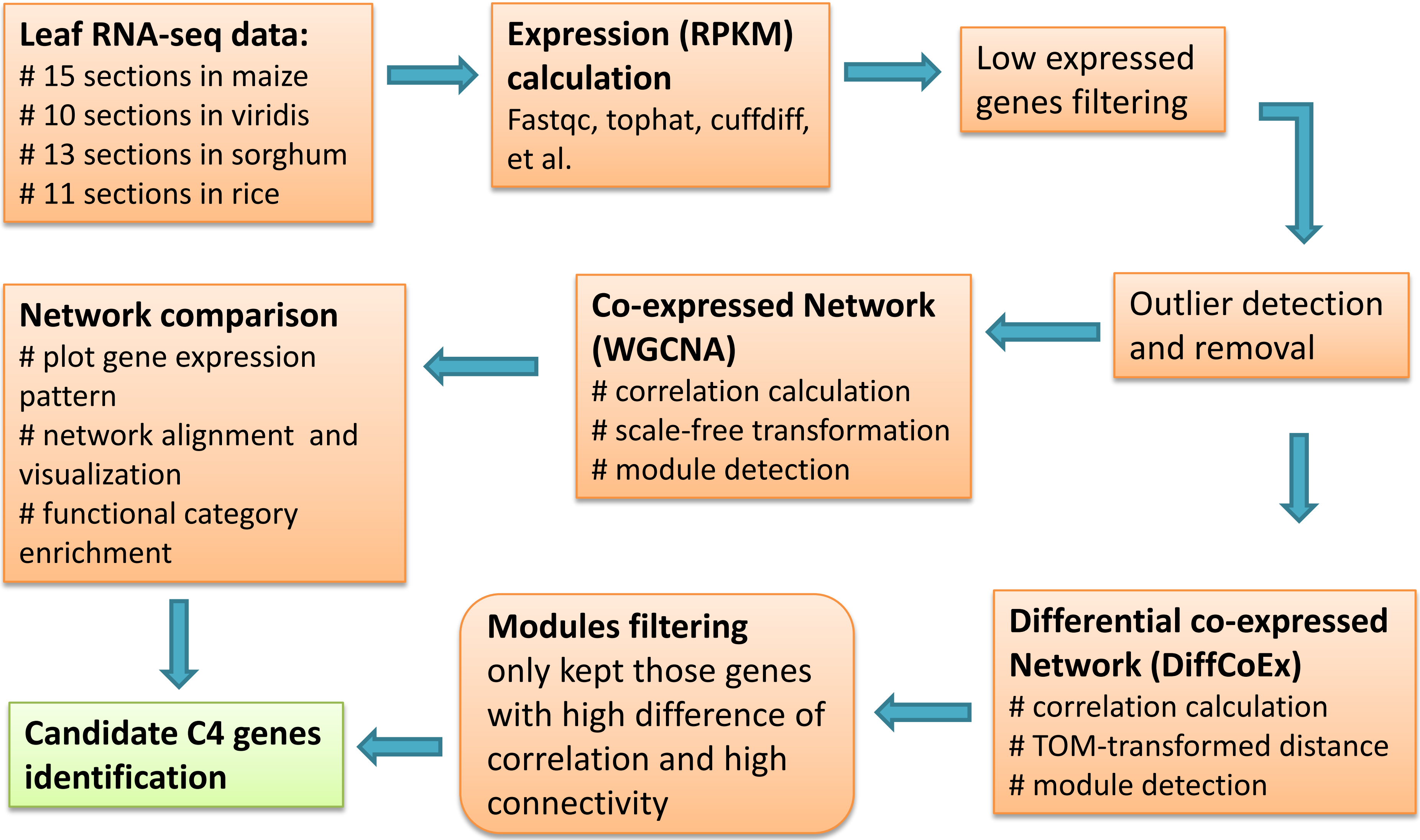


~10 days old



(Wang et al., 2014)

Workflow in this study



Leaf RNA-seq data:
15 sections in maize
10 sections in viridis
13 sections in sorghum
11 sections in rice

Expression (RPKM) calculation
Fastqc, tophat, cuffdiff, et al.

Low expressed genes filtering

Outlier detection and removal

Co-expressed Network (WGCNA)
correlation calculation
scale-free transformation
module detection

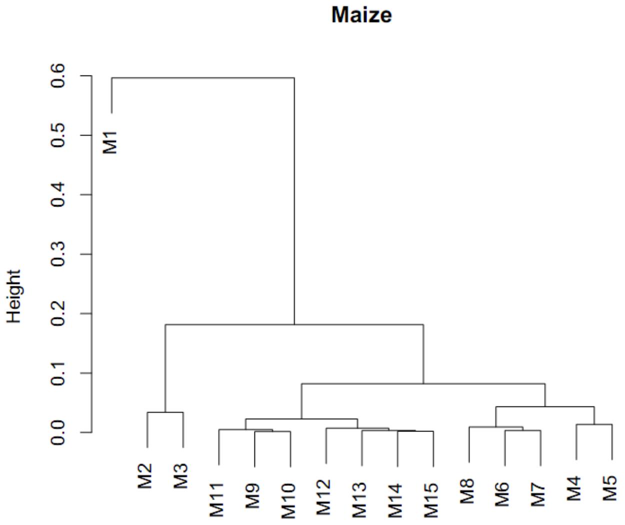
Network comparison
plot gene expression pattern
network alignment and visualization
functional category enrichment

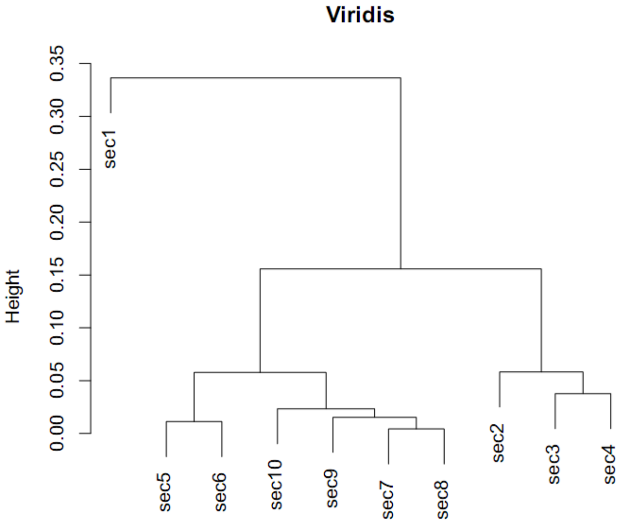
Differential co-expressed Network (DiffCoEx)
correlation calculation
TOM-transformed distance
module detection

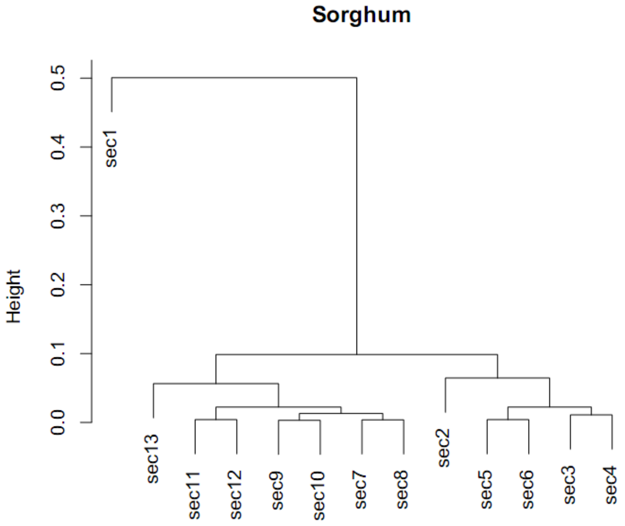
Modules filtering
only kept those genes with high difference of correlation and high connectivity

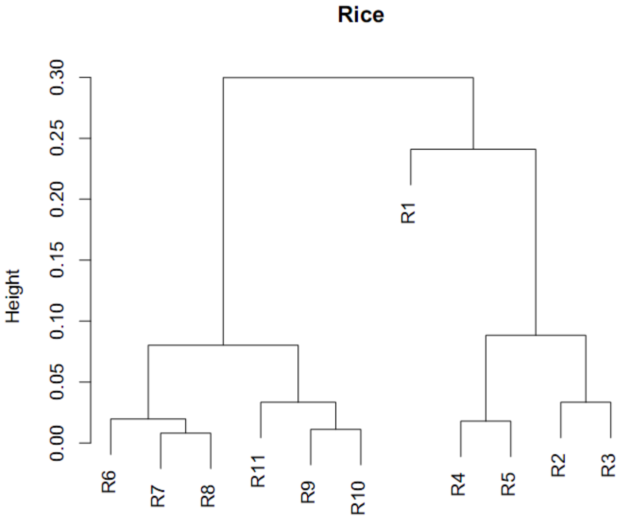
Candidate C4 genes identification

Samples clustering and outlier detection

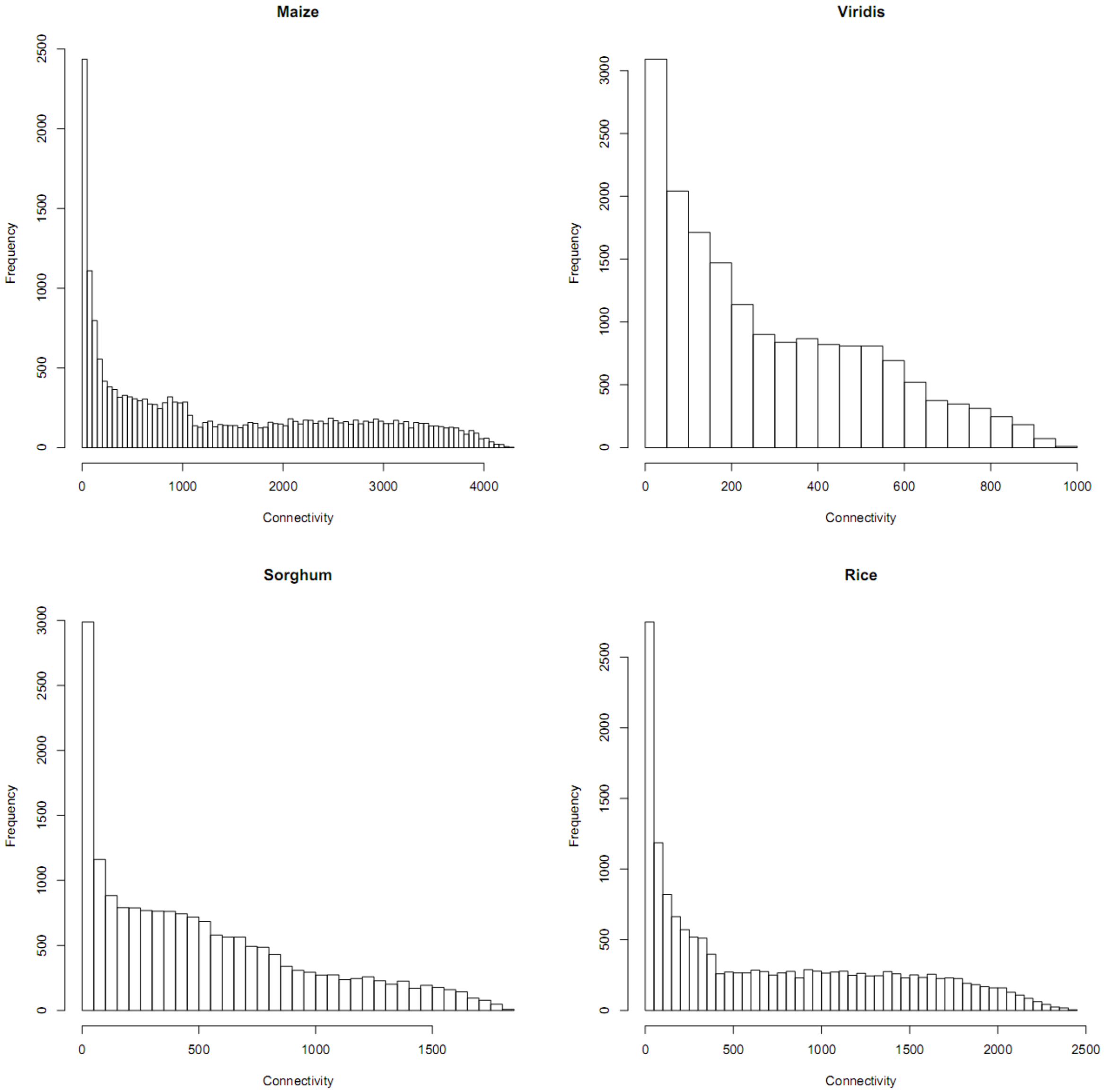

Maize


Viridis

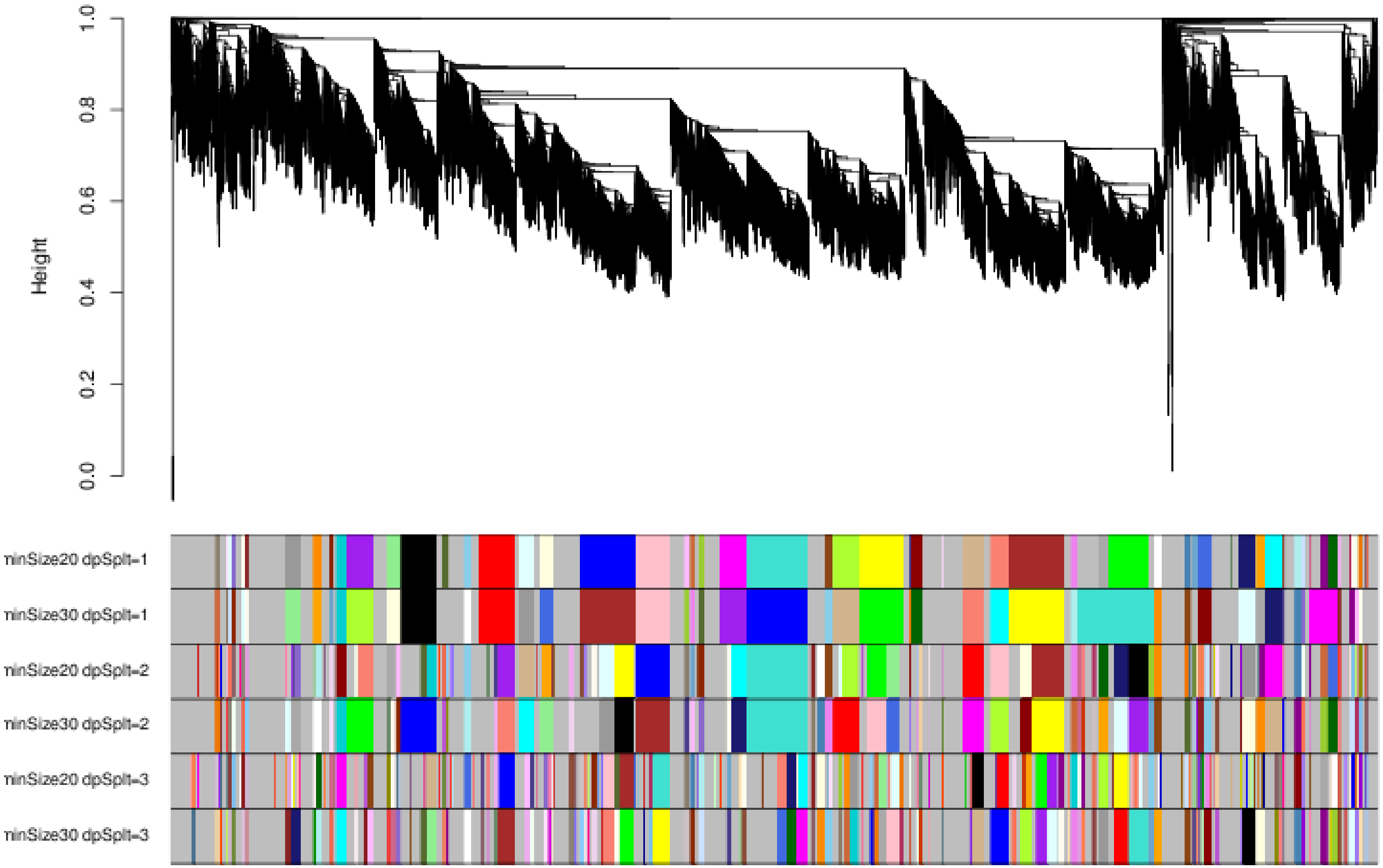

Sorghum


Rice

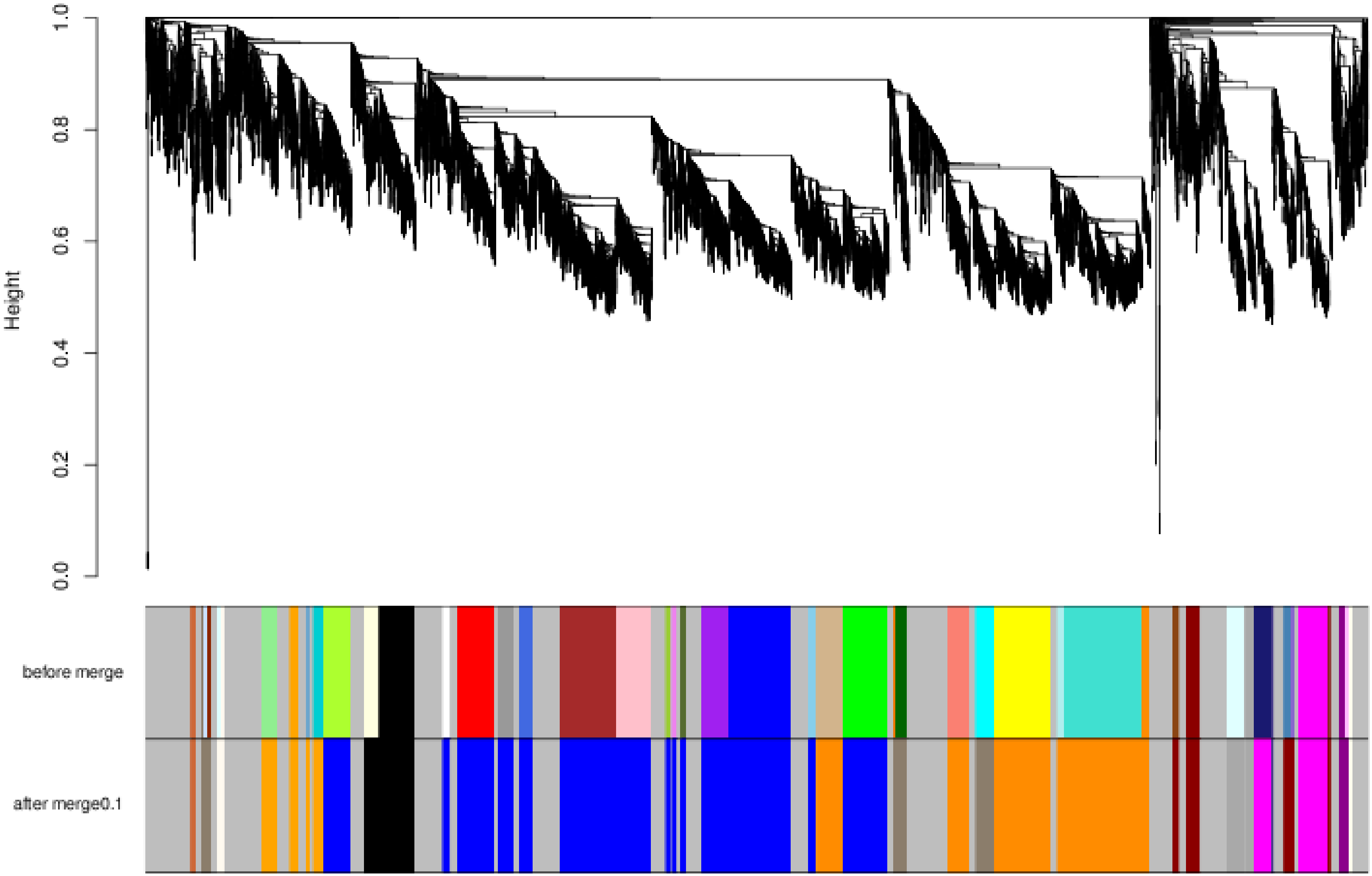
Distribution of TOM (scale-free topology transfer)



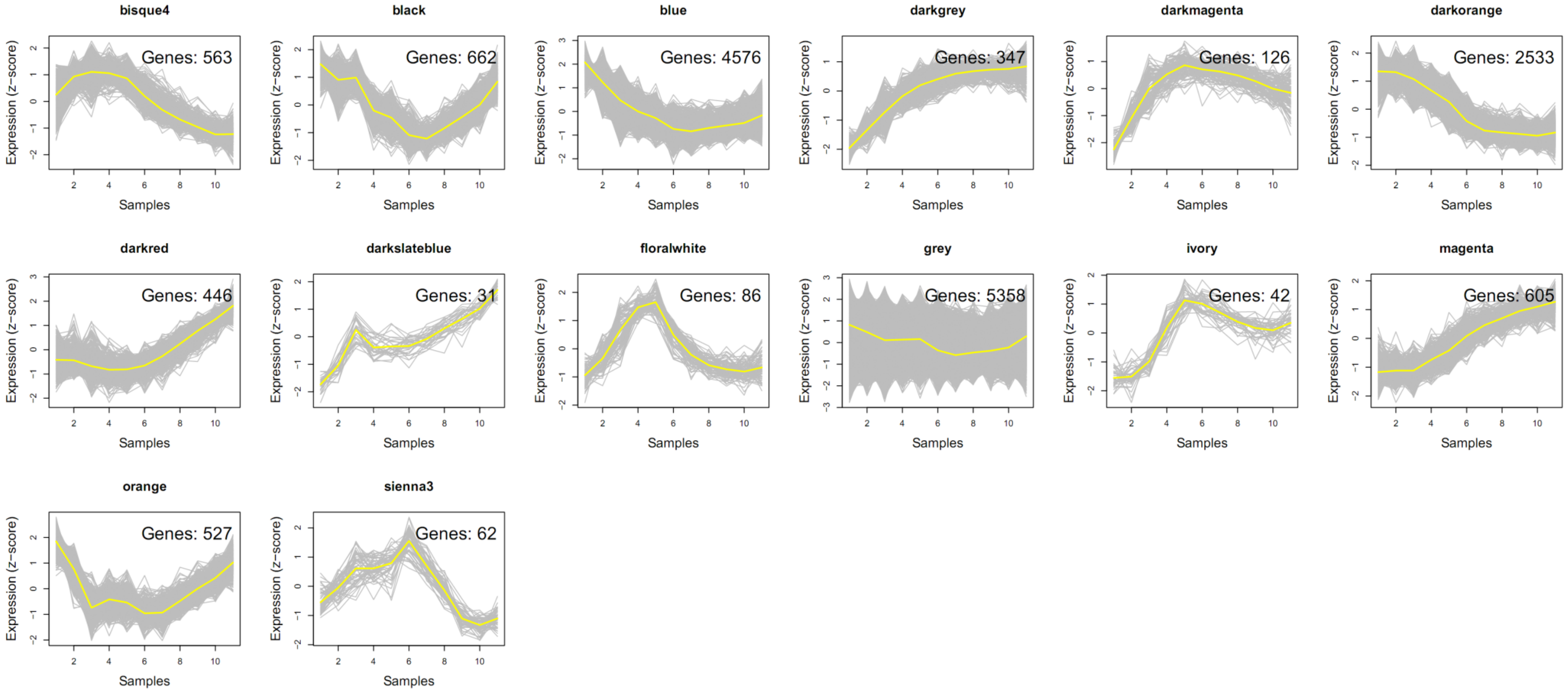
Modules detection



Merge closely related modules



Modules expression pattern



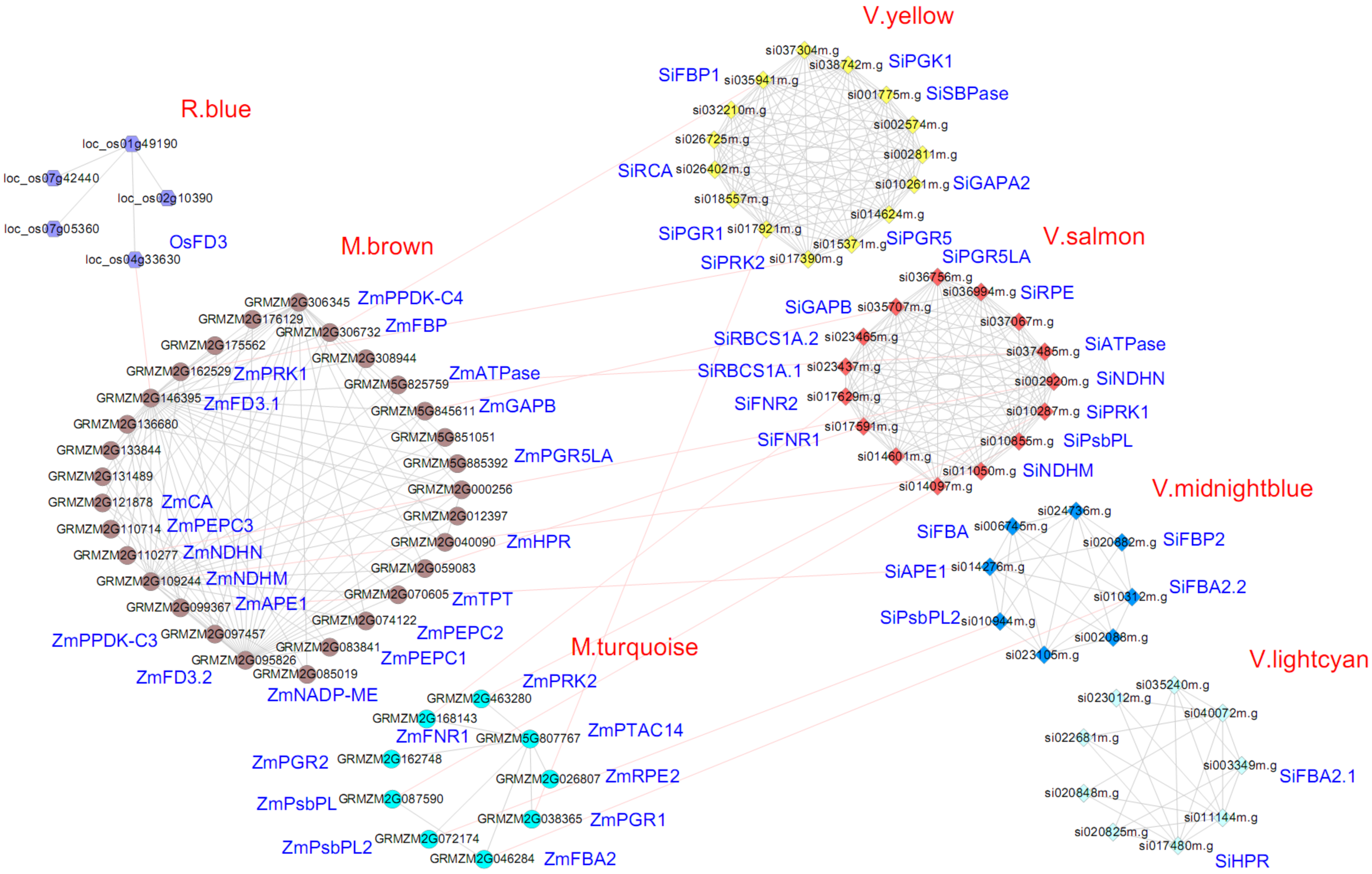
Summarize of network construction

Characteristics	Maize (M)	Viridis (V)	Sorghum (S)	Rice (R)
Loci after low expressed genes filtering	18916	17253	18119	15964
Samples after outlier removal	14	9	12	11
WGCNA selected power	10	12	18	16
Modules detected (merge at 0.1)	11	32	12	14
Percent of loci kept in modules	48%	50%	39%	34%

“Connection” among species

- Ortholog list, e.g., CoGe, Ensemble
- Syntenic orthologs build by other studies
- Network alignment, e.g., IsoRank

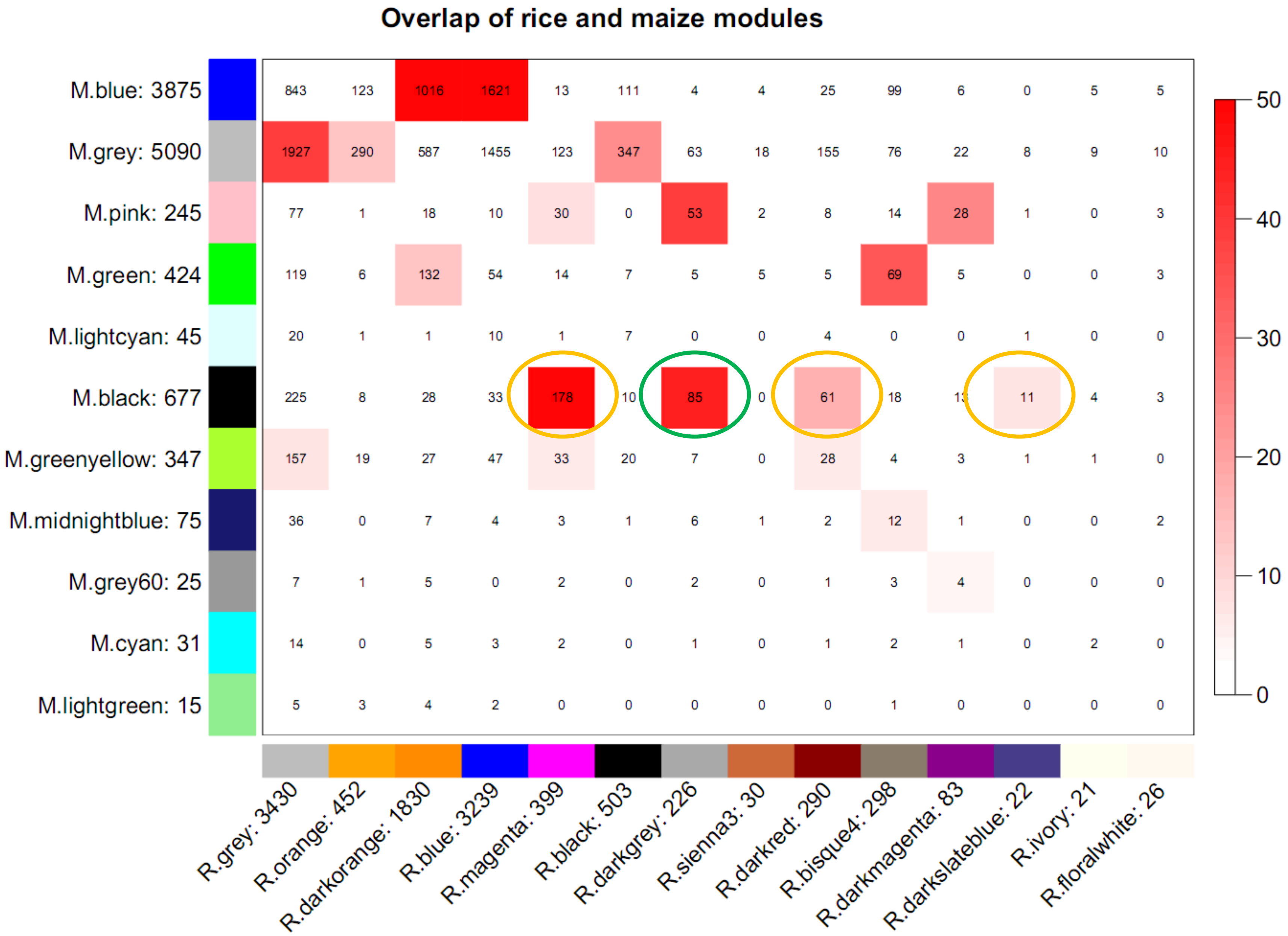
An example: networks of PS genes in three species



What kinds of rules could be used for identification of candidate C4 genes?

1. Network alignment and then compare their expression patterns based on modules
2. Case-control studies

Part 1



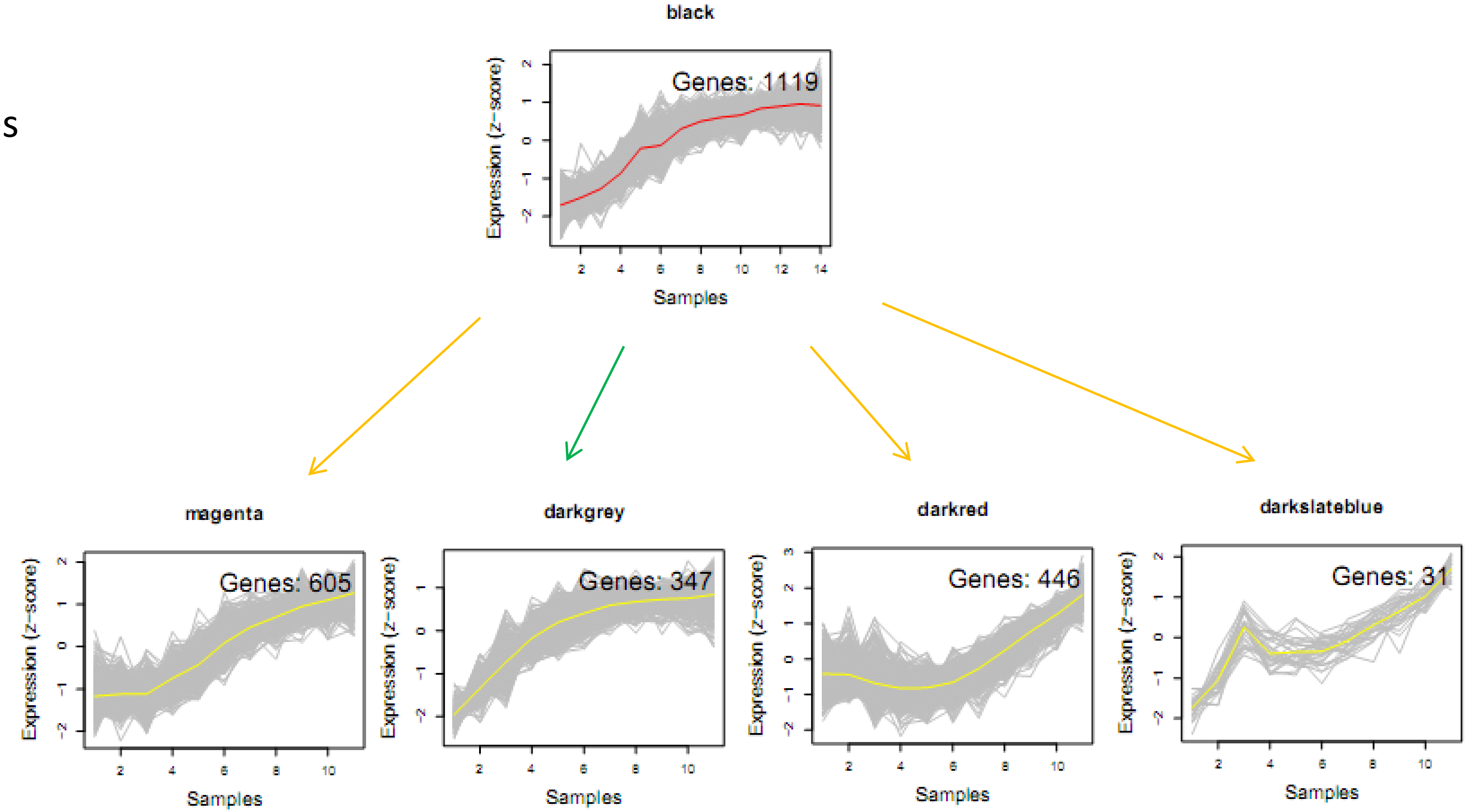
Overlap of rice and maize modules

	R.grey: 3430	R.orange: 452	R.darkorange: 1830	R.blue: 3239	R.magenta: 399	R.black: 503	R.darkgrey: 226	R.sienna3: 30	R.darkred: 290	R.bisque4: 298	R.darkmagenta: 83	R.darkslateblue: 22	R.ivory: 21	R.floralwhite: 26
M.blue: 3875	843	123	1016	1621	13	111	4	4	25	99	6	0	5	5
M.grey: 5090	1927	290	587	1455	123	347	63	18	155	76	22	8	9	10
M.pink: 245	77	1	18	10	30	0	53	2	8	14	28	1	0	3
M.green: 424	119	6	132	54	14	7	5	5	5	69	5	0	0	3
M.lightcyan: 45	20	1	1	10	1	7	0	0	4	0	0	1	0	0
M.black: 677	225	8	28	33	178	10	85	0	61	18	13	11	4	3
M.greenyellow: 347	157	19	27	47	33	20	7	0	28	4	3	1	1	0
M.midnightblue: 75	36	0	7	4	3	1	6	1	2	12	1	0	0	2
M.grey60: 25	7	1	5	0	2	0	2	0	1	3	4	0	0	0
M.cyan: 31	14	0	5	3	2	0	1	0	1	2	1	0	2	0
M.lightgreen: 15	5	3	4	2	0	0	0	0	0	1	0	0	0	0

Expression pattern: similar/different function

M- modules

R-modules



Part 2

NADP-ME type of C4 pathway

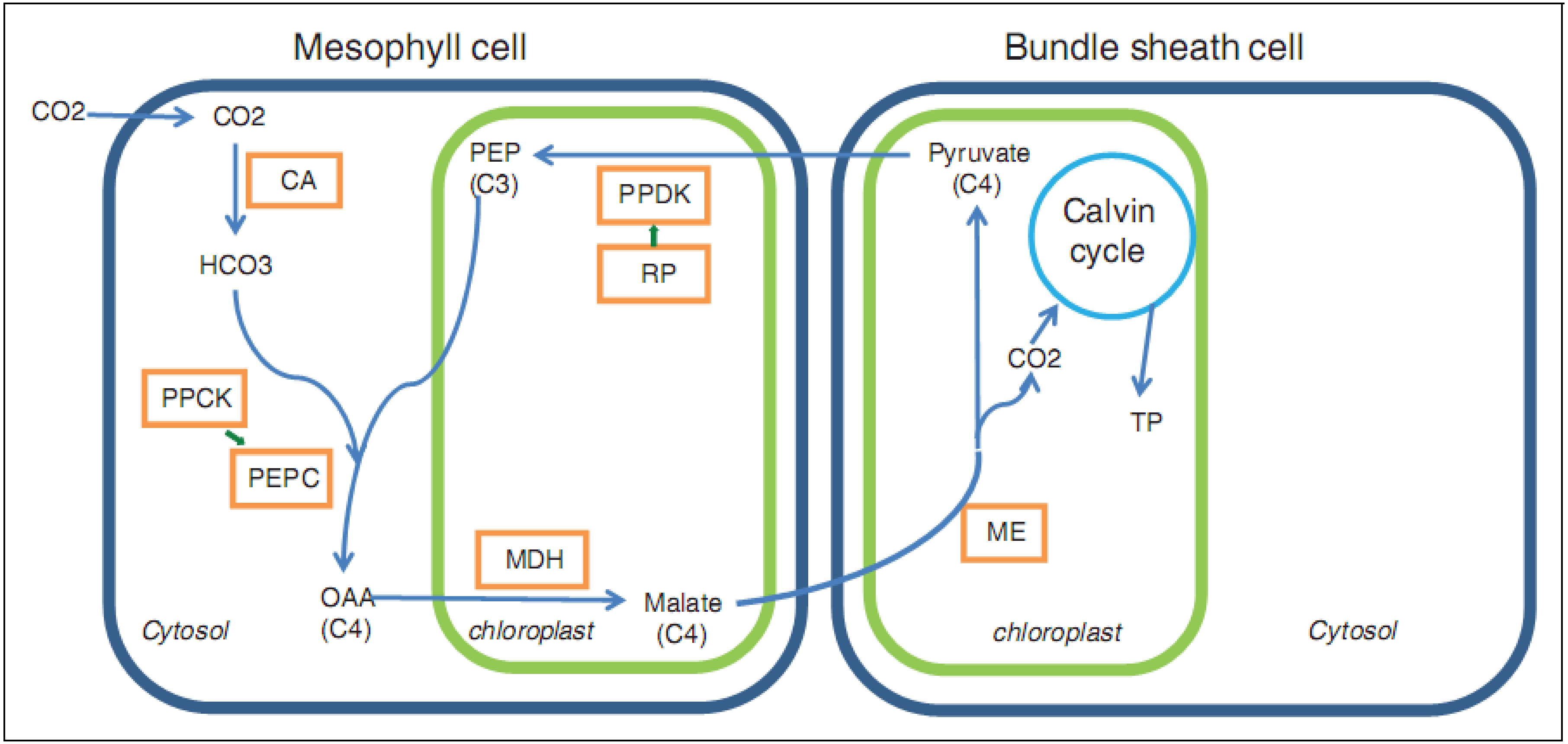
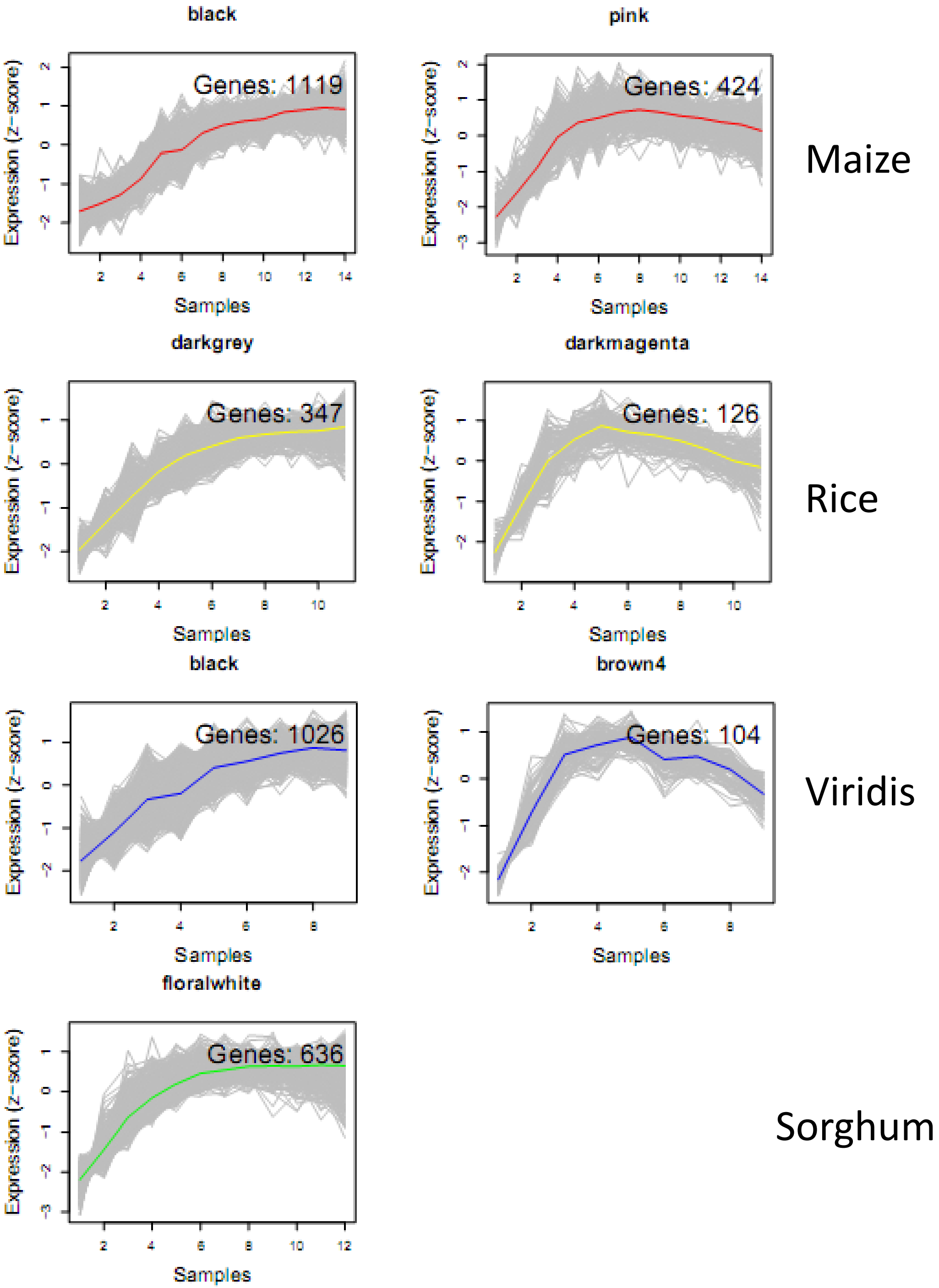


Figure 1
The NADP-ME type of C4 pathway in sorghum and maize. CA, carboxylating anhydrase; MDH, malate dehydrogenase; ME, malic enzyme; OAA, oxaloacetate; PEPC, phosphoenolpyruvate carboxylase; PPCK, PEPC kinase; PPDK, pyruvate orthophosphate dikinase; PPDK-RP, PPDK regulatory protein; TP, transit peptide.

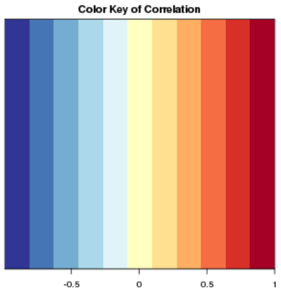
(Wang et al., 2009)

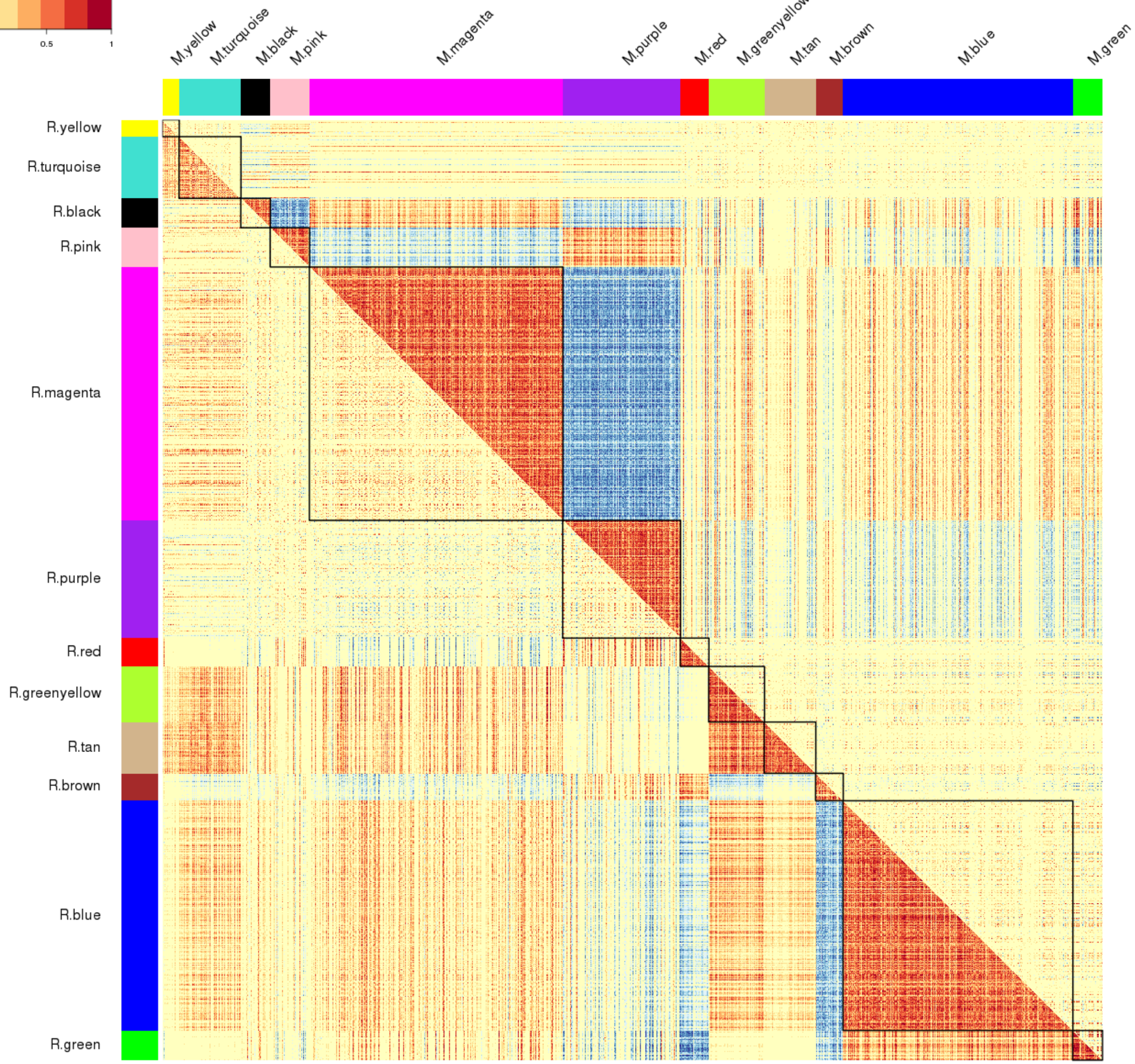
Focus on "C4 modules", which are both photosynthesis (PS)-enriched and showing PS-expression pattern



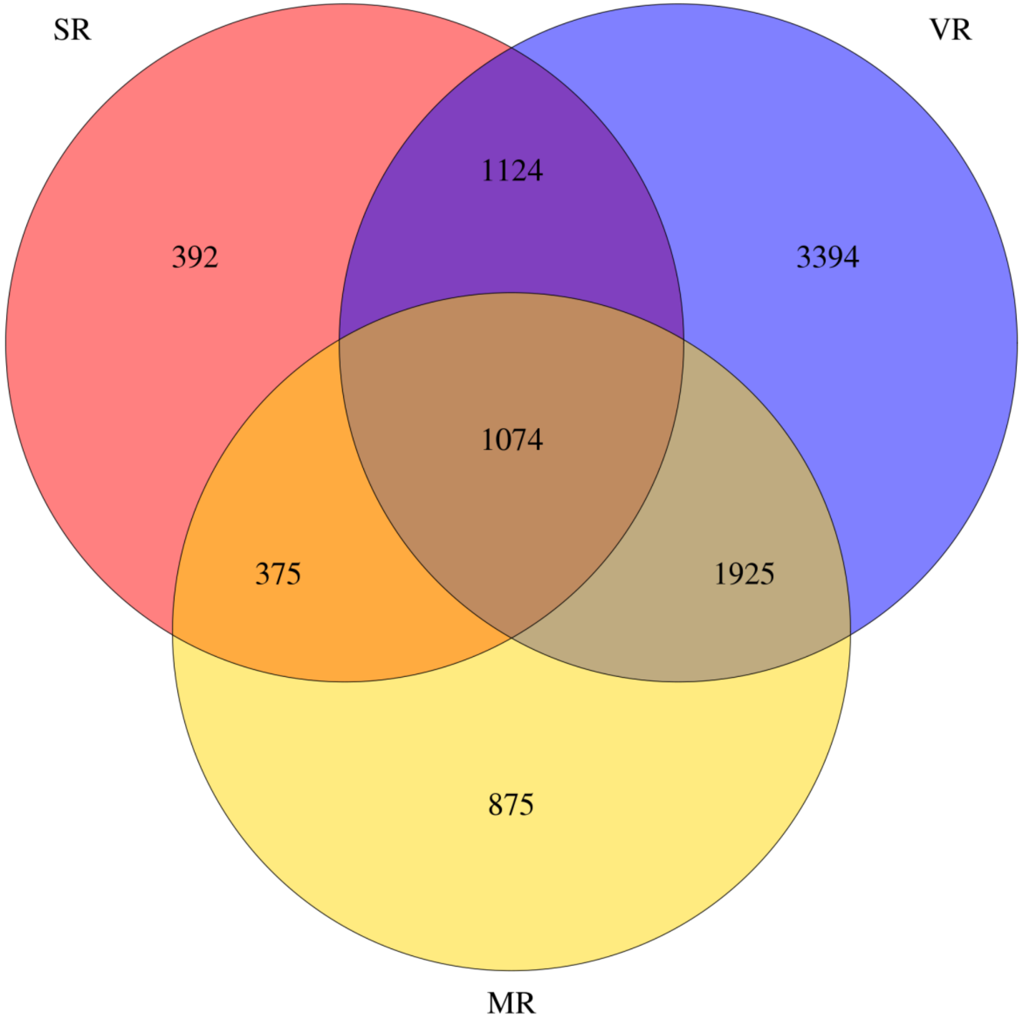
Differential expressed modules detected between maize (M) and rice (R).

Based on the color keys, blue represent negative correlation while brown represent positive correlation.

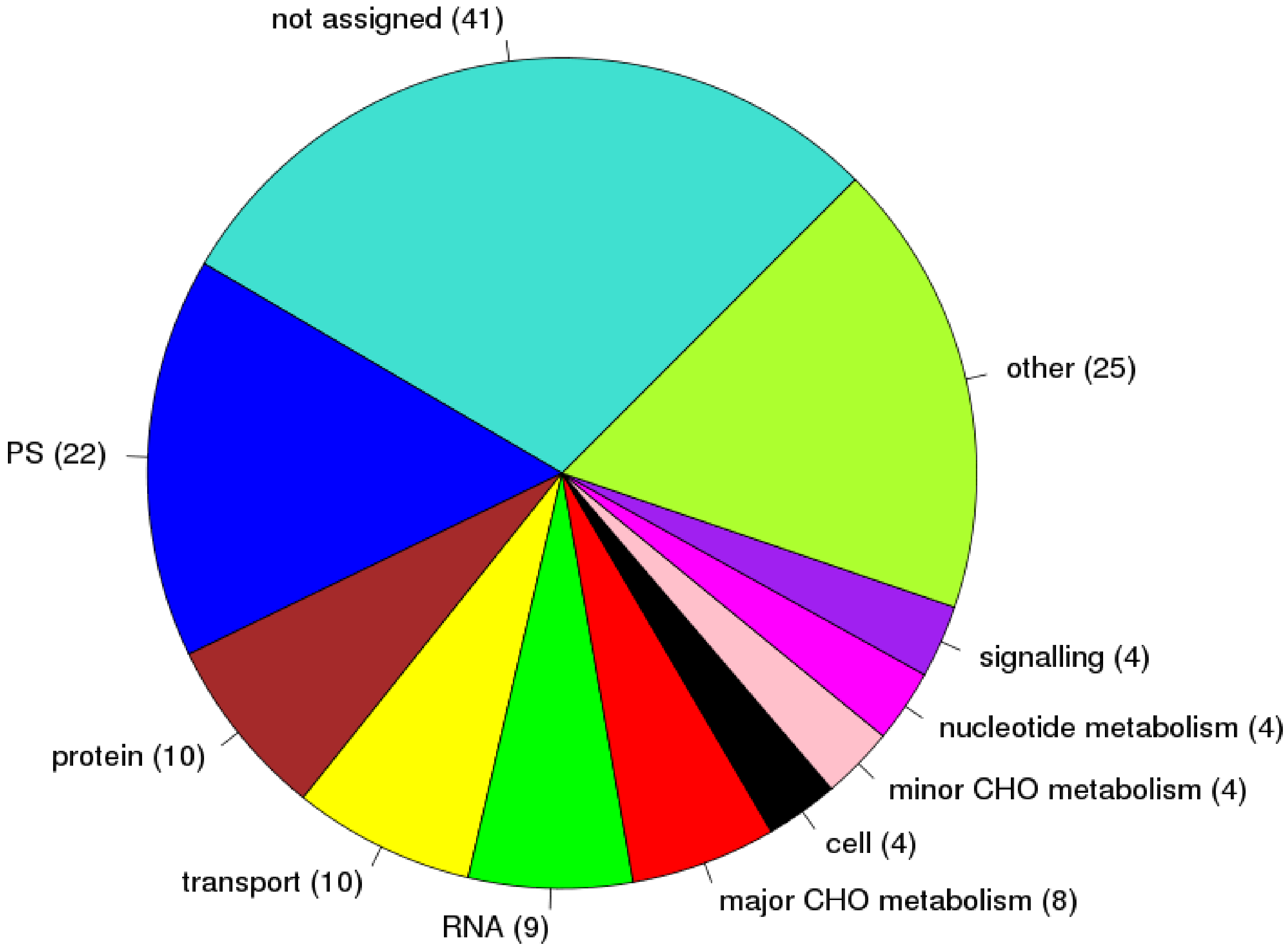




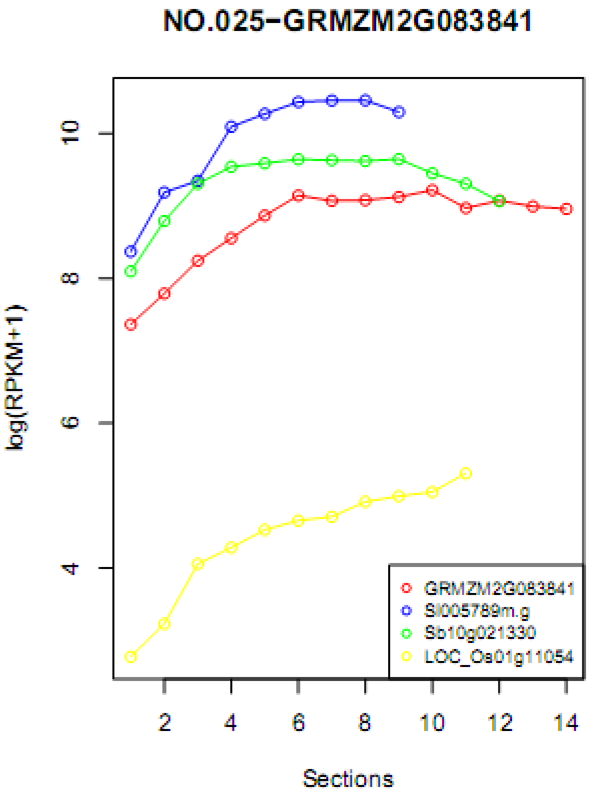
Venny diagram of diffCoEx results

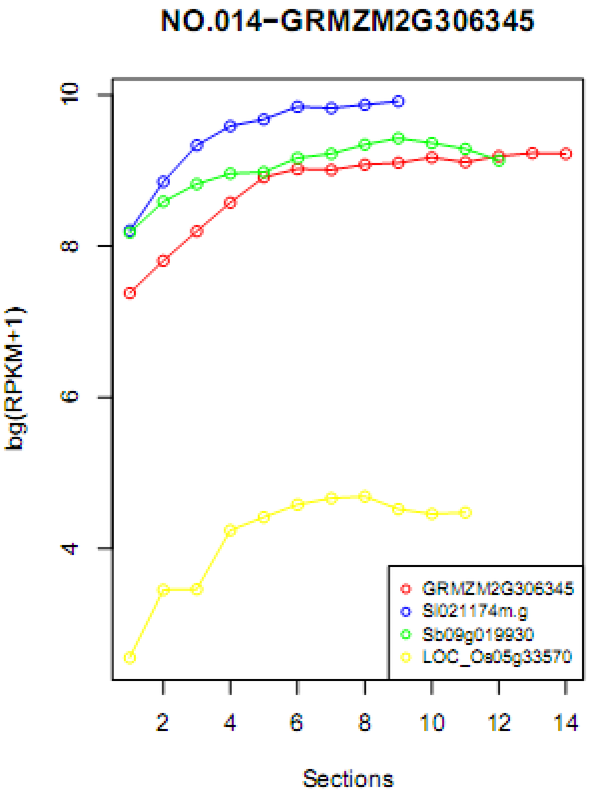


Functional distribution of C4 candidate genes

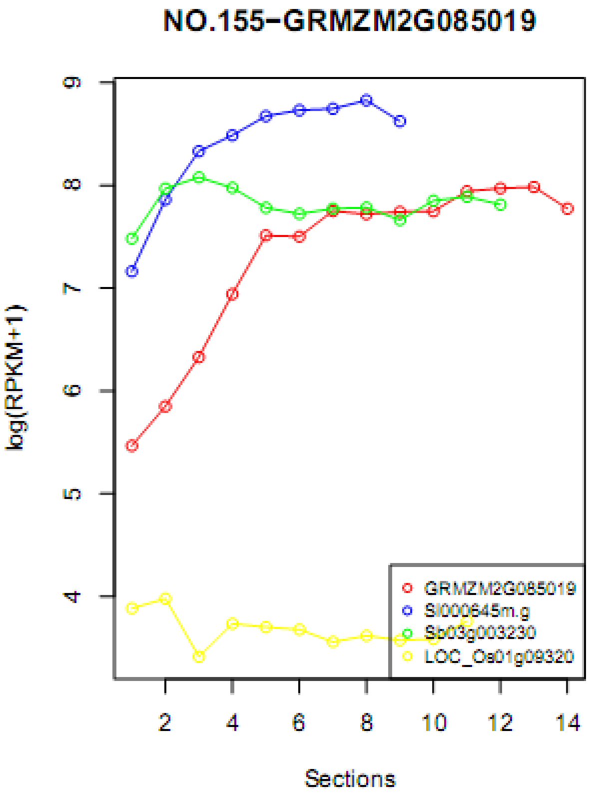


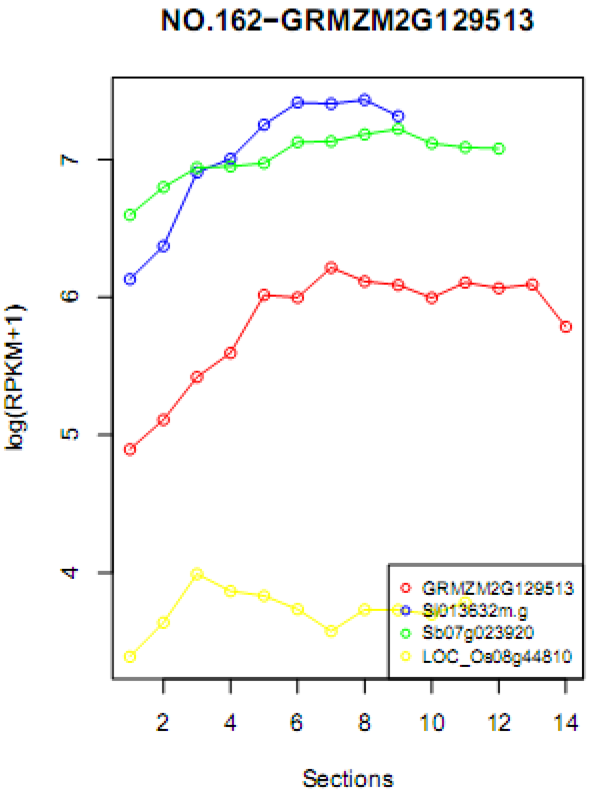
Examples: type1

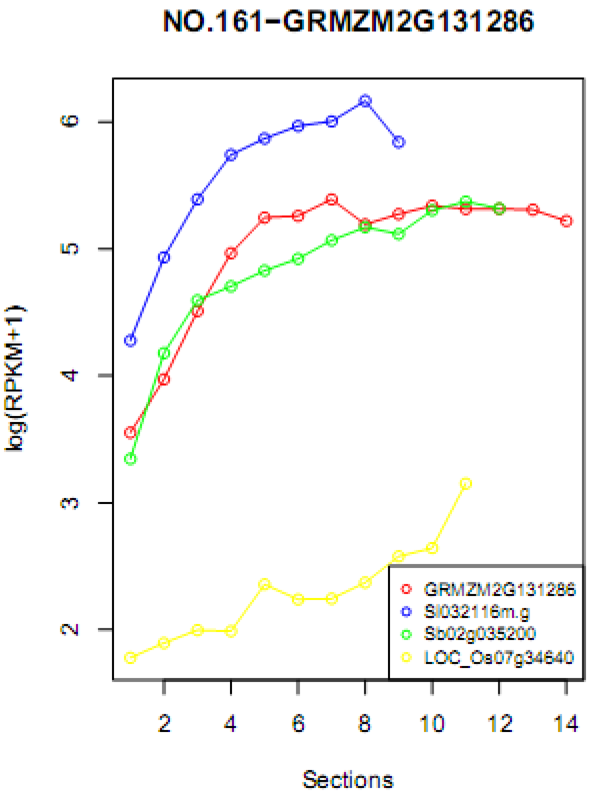




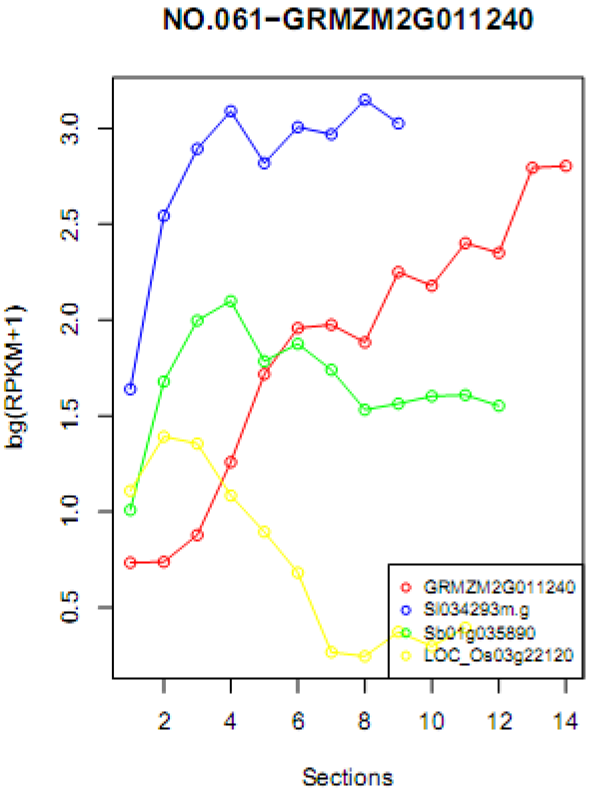
Examples: type2


NO.155−GRMZM2G085019

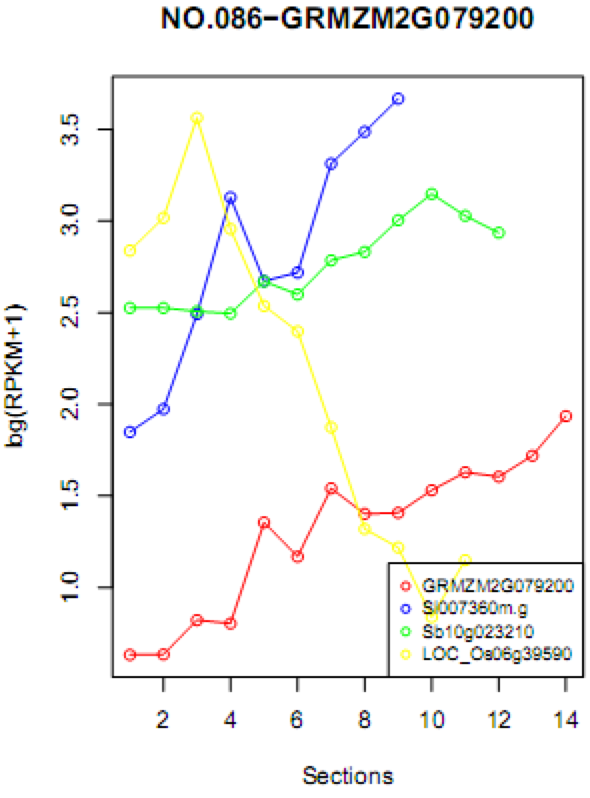

NO.162−GRMZM2G129513


NO.161−GRMZM2G131286

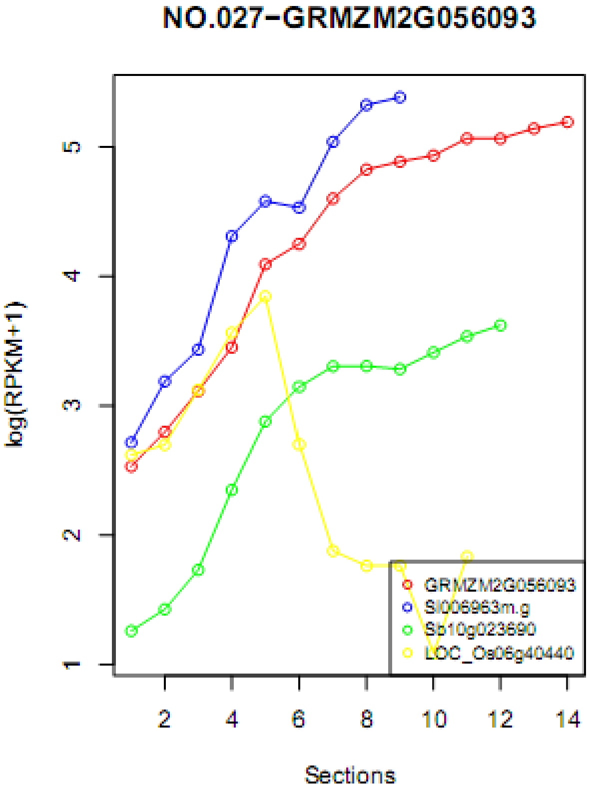
Examples: type2



Sucrose synthase 3

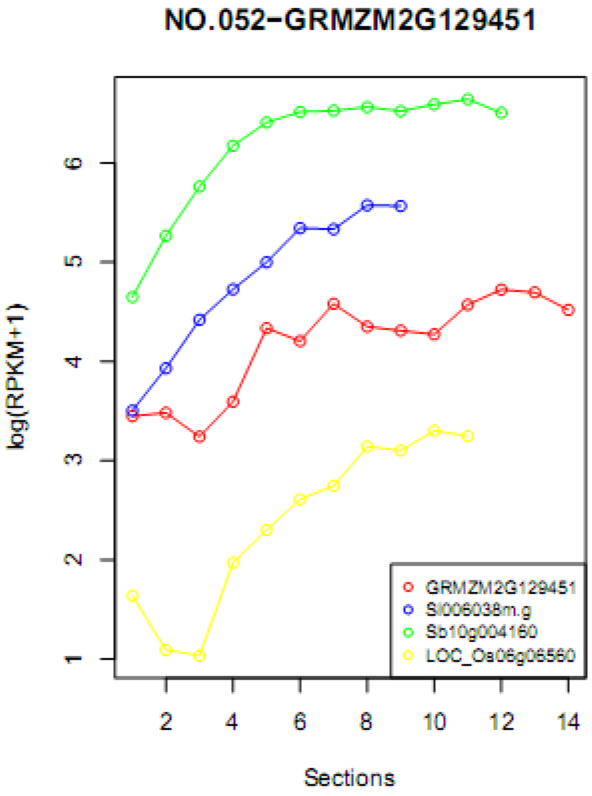


SHORT HYPOCOTYL 2

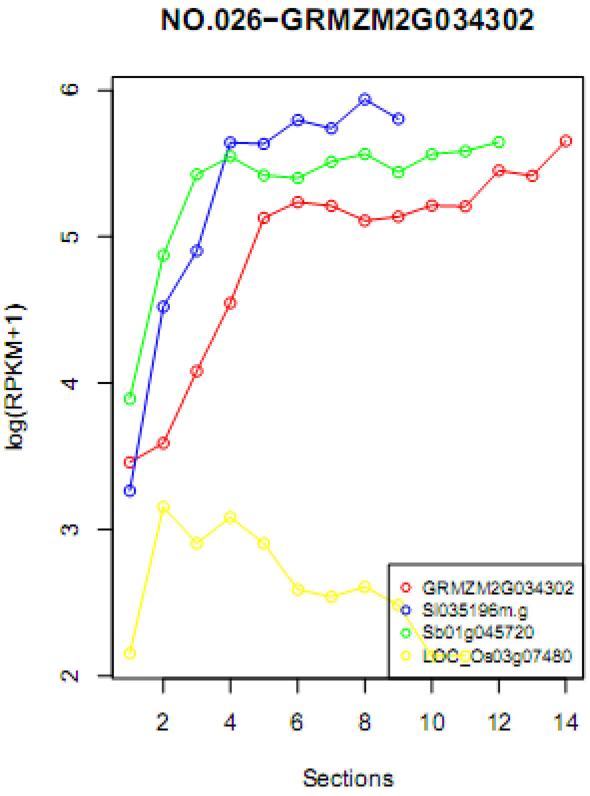


un-known genes

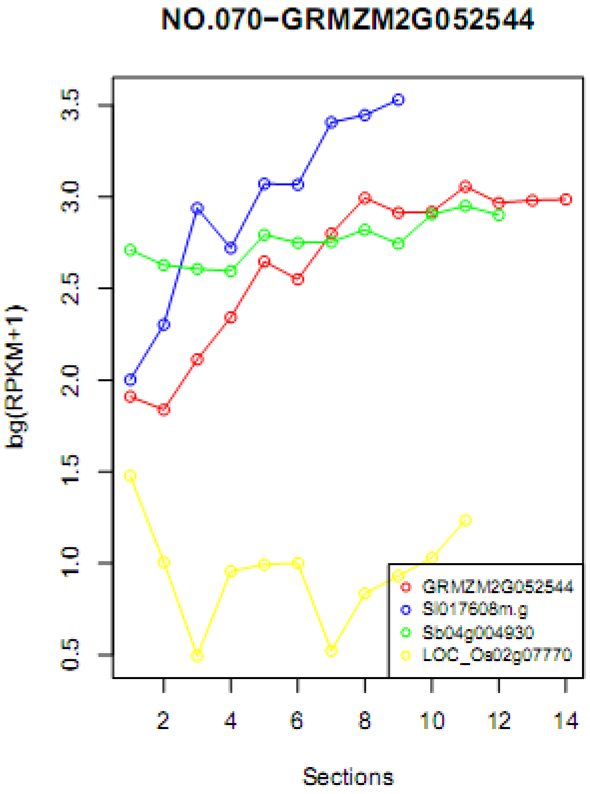
Examples: non-classic C4 genes



Starch pathway genes (SSI)



Transporters (SUT2)



Transcription factors (APL)

Conclusion

✓A method, which was mainly based on gene co-expression or diffCoEx patterns and synteny orthologs, to identify candidate C4 genes ;

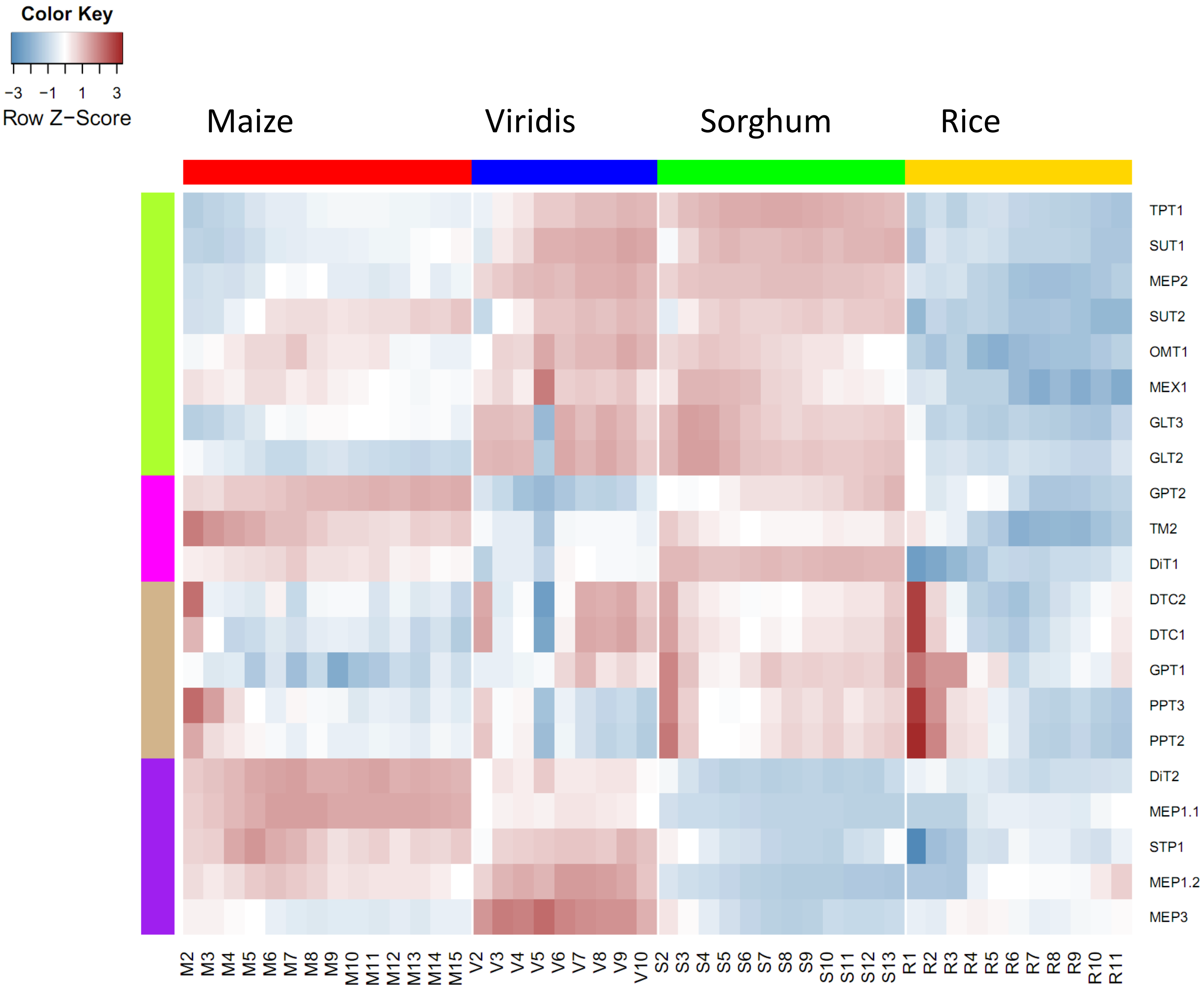
✓Besides the classic C4 genes, a lot of non-classic C4 genes were also identified; they were associated with starch metabolism, transcription factors and transporters;

✓These candidate C4 genes could be classified into three groups, which may corresponding to different evolutionary processes of C4 genes,

Pros vs. Cons

- ■ A method to dissect gene relationships between and within C3 and C4 species at systematic and multi-species level
- ■ Annotation for function unknown genes

- ◆ Limited resolution (only tens of samples)
- ◆ Genes important in other studies but not identified in our study, mainly because they did not expressed in our sample or they did not show a PS-similar expression pattern


Color Key
−3 −1 1 3
Row Z−Score
Maize
Viridis
Sorghum
Rice
TPT1
SUT1
MEP2
SUT2
OMT1
MEX1
GLT3
GLT2
GPT2
TM2
DiT1
DTC2
DTC1
GPT1
PPT3
PPT2
DiT2
MEP1.1
STP1
MEP1.2
MEP3
M2 M3 M4 M5 M6 M7 M8 M9 M10 M11 M12 M13 M14 M15
V2 V3 V4 V5 V6 V7 V8 V9 V10
S2 S3 S4 S5 S6 S7 S8 S9 S10 S11 S12 S13
R1 R2 R3 R4 R5 R6 R7 R8 R9 R10 R11

Pros vs. Cons

- A method to dissect gene relationships between and within C3 and C4 species at systematic and multi-species level
- Annotation for function unknown genes

- Limited resolution (only tens of samples)
- Genes important in other studies but not identified in our study, mainly because they did not expressed in our sample or they did not show a PS-similar expression pattern
- No BS and M info was provided. But fortunately, most (94%) of the candidate C4 genes were either BS- or ME-enriched, based on previous cell-specific studies

Acknowledgement

Cornell university
Qi Sun
Minghui Wang
Klaas van Wijk
Christopher Myers
Lei Huang

Donald Danforth Plant Science Center
Thomas P Brutnell
Sarit Weissmann
Anthony Studer (Tony)

Shandong Agricultural University
Pinghua Li

University of Nebraska–Lincoln
James Schnable

Thank you!