Community Coevolution Model: Supplementary Figures and Tables

Chaoyue Liu, Toby Kenney, Robert Beiko, and Hong Gu

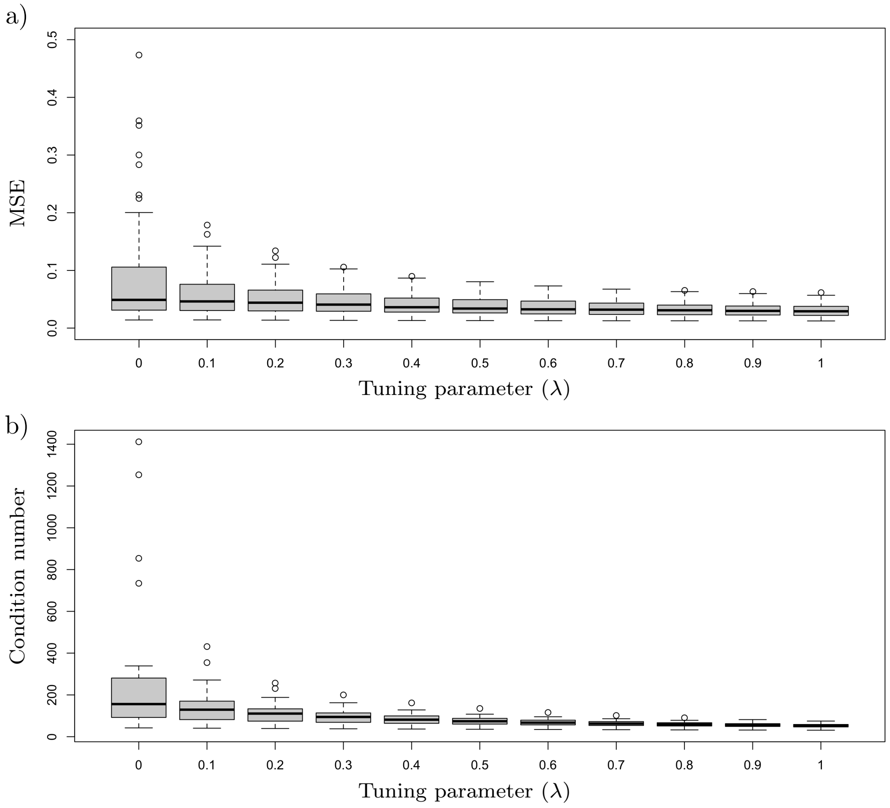


Figure S1: Evaluation of the effect of different tuning parameters: (a) MSE of the estimates for different tuning parameters. (b) The condition number of the Hessian matrix which also stands for the largest convergence rate between estimated parameters against different tuning parameters.

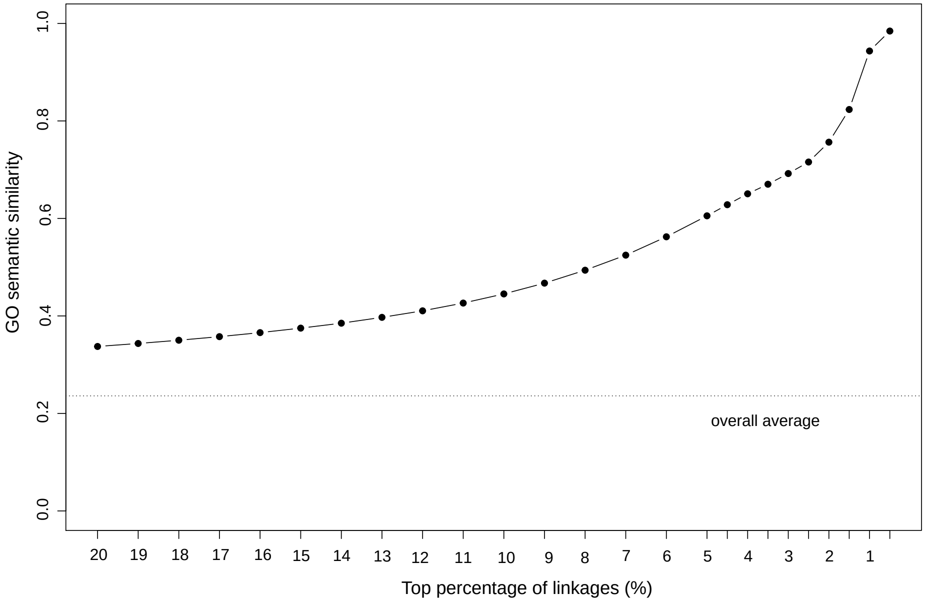


Figure S2: The association between functional similarity and the strength of the gene linkages detected by CCM. The horizontal axis shows the percentage of most significant linkages used for evaluation and the vertical axis shows the mean of the GO semantic similarity among the corresponding percentage of linkages. The horizontal line indicates the average functional similarity among all genes.

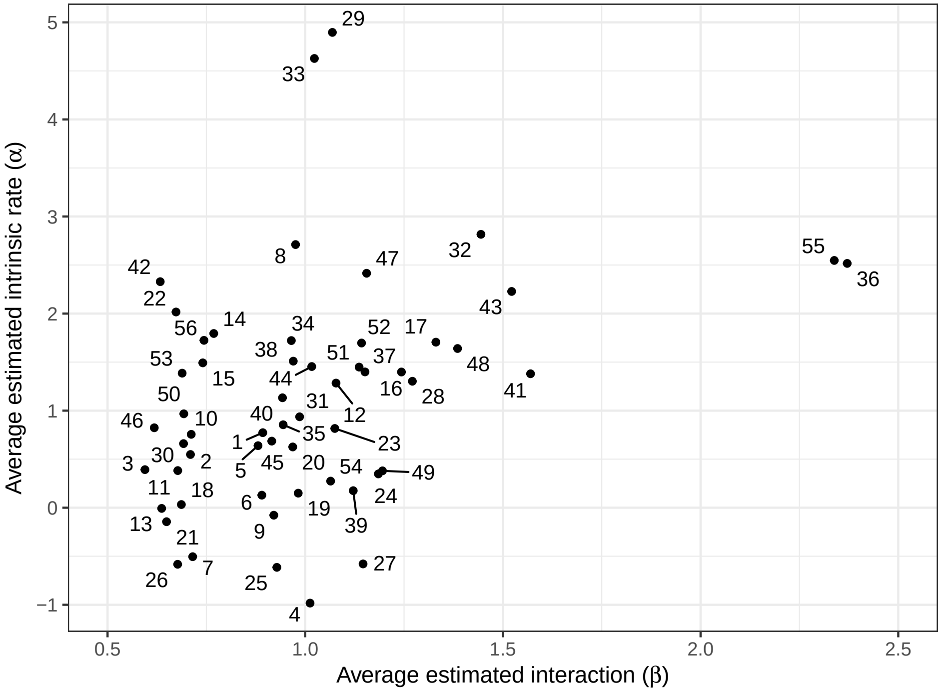
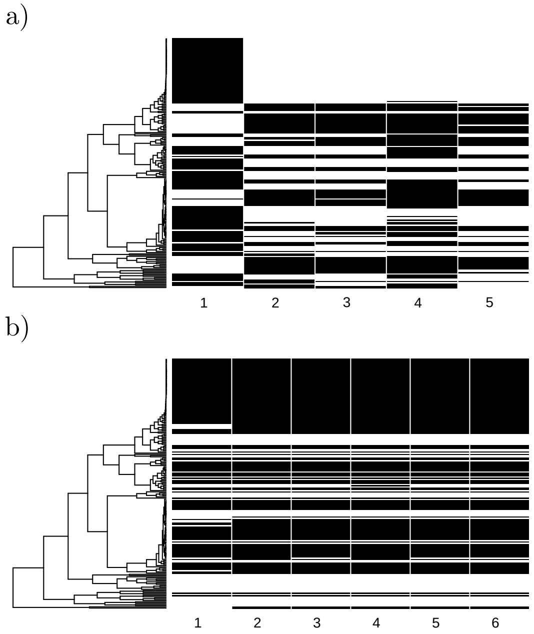


Figure S3: Averaged interaction and averaged intrinsic rates within each cluster. The cluster labels are matched to **Supplementary Table S2**.



Inferred interaction (β_{ij})

$$\begin{array}{c|ccccc} & 1 & 2 & 3 & 4 & 5 \\ \hline 1 & & -0.342 & -0.802 & -0.375 & 0.040 \\ 2 & & & 0.798 & 0.028 & 0.850 \\ 3 & & & & 0.553 & 0.655 \\ 4 & & & & & 0.499 \end{array}$$

Inferred interaction (β_{ij})

$$\begin{array}{c|cccccc} & 1 & 2 & 3 & 4 & 5 & 6 \\ \hline 1 & & 0.547 & 0.743 & 0.366 & 0.642 & 0.237 \\ 2 & & & 1.060 & 1.009 & 0.943 & 0.580 \\ 3 & & & & 0.820 & 1.151 & 0.939 \\ 4 & & & & & 0.770 & 0.249 \\ 5 & & & & & & 0.756 \end{array}$$

Figure S4: (a) The phylogenetic profiles of Cluster 49 are shown on the left. The interaction coefficients estimated by simultaneously modeling five genes as a community are shown on the right. (b) The phylogenetic profiles of Cluster 36 are shown on the left. The interaction coefficients estimated by simultaneously modeling six genes as a community are shown on the right. The gray cells indicate linkages that have p-values > 0.05.

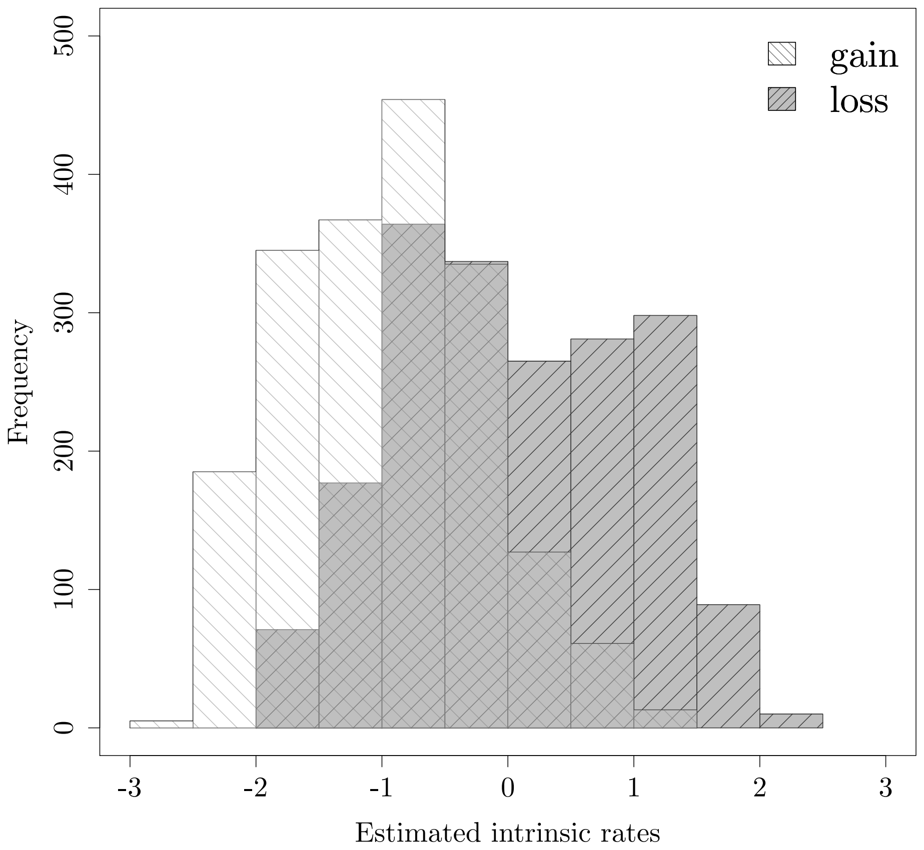


Figure S5: The estimated intrinsic gain and loss rates of the complex I genes.

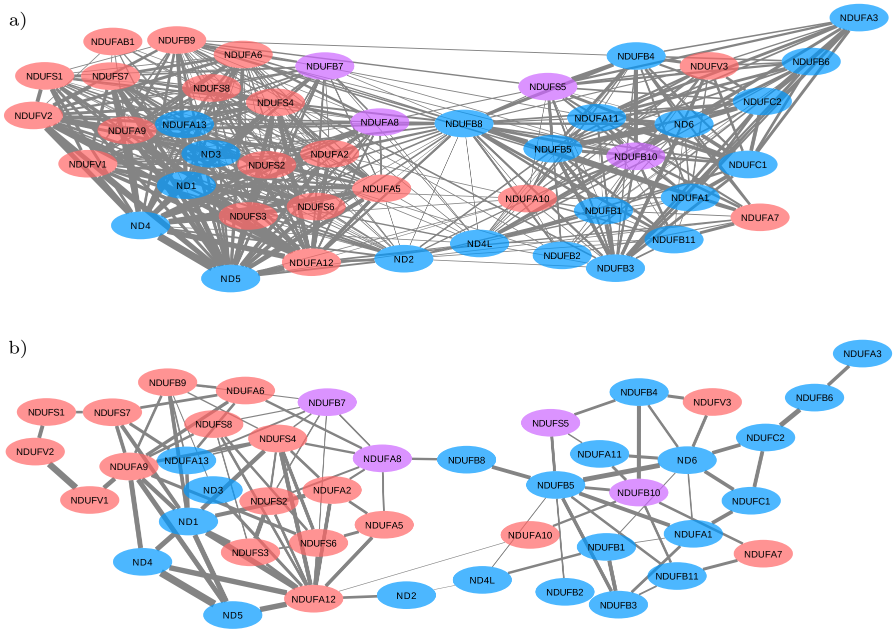


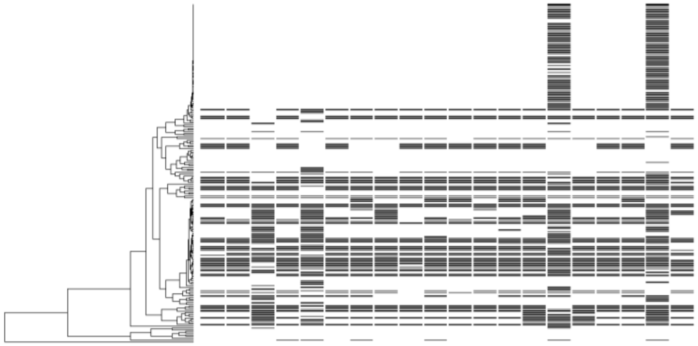
Figure S6: Network analysis of the complex I genes. (a) The original network inferred by full pairwise comparisons using CCM, which consists of 462 significant edges (p-value < 0.05). (b) The links that are significantly conditionally dependent (p-value < 0.05) in all triplets from network (a). The resulting network consists of 101 edges. The edge thickness corresponds to the estimated strength of the interaction (βij). Label colors indicate the locations of the subunits: Matrix (red), Transmembrane (blue), and Intermembrane (purple).

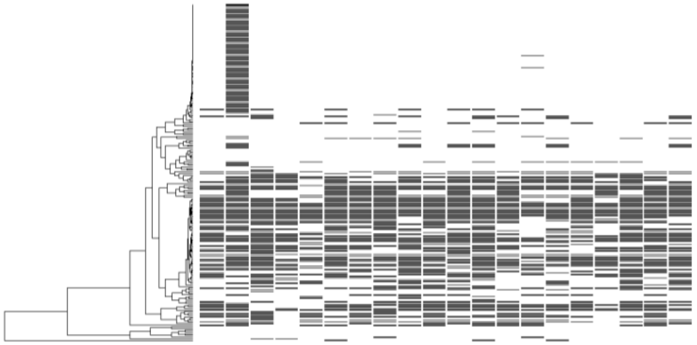
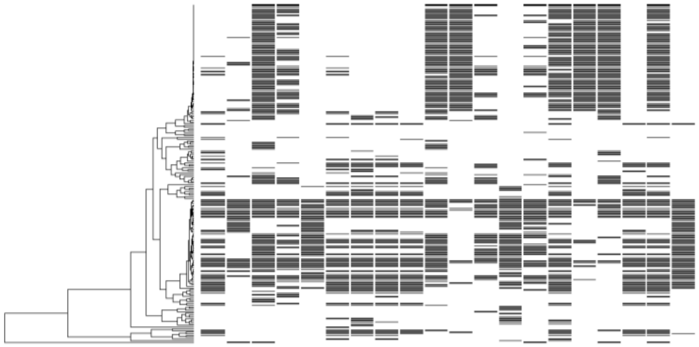
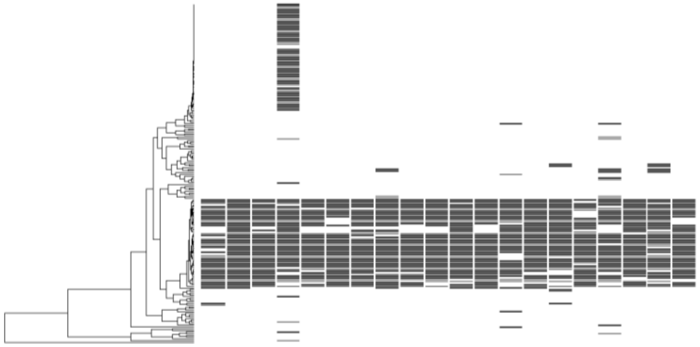
Table S1: The approximate running time of the community coevolution model for different sizes of tree and community performed on a server running Linux with 2.67 GHz CPU and 18 GB RAM. Abbreviations: s (second), m (minute) and h (hour)

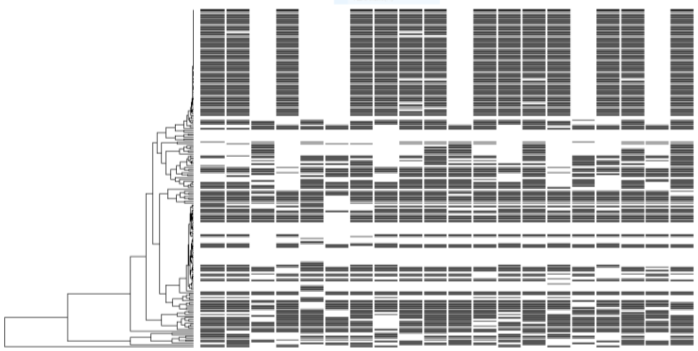
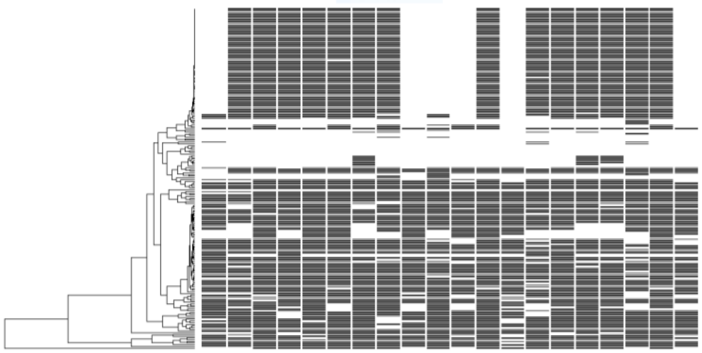
		Number of tips in tree				
		50	100	200	500	1000
Community Size	2	0.67 s	1.11 s	1.90 s	3.76 s	8.55 s
	3	2.24 s	2.90 s	5.04 s	11.22 s	22.76 s
	4	7.73 s	9.86 s	17.33 s	40.02 s	1.19 m
	5	31.04 s	52.37 s	1.60 m	4.93 m	6.60 m
	6	3.65 m	7.82 m	12.50 m	22.91 m	40.31 m
	7	15.3 m	26.43 m	37.89 m	1.25 h	2.43 h

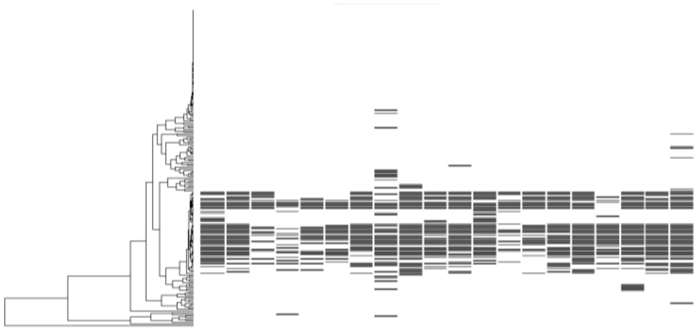
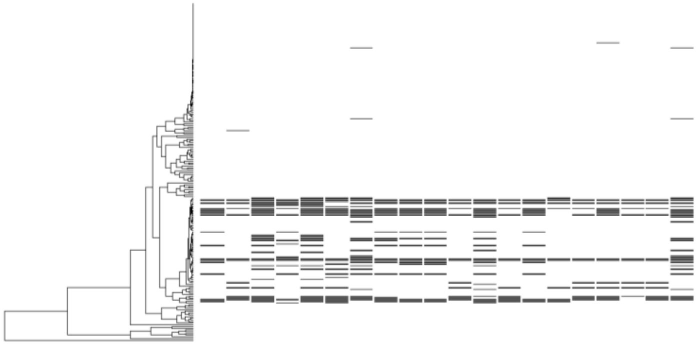
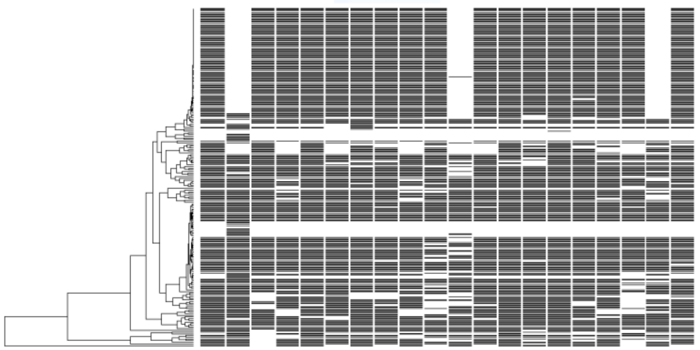
Table S2: The summary table of the gene clusters. The GO enrichment analysis was performed using the R package “TopGO” using all 1918 annotated genes out of 6505 LZ genes as the background set.

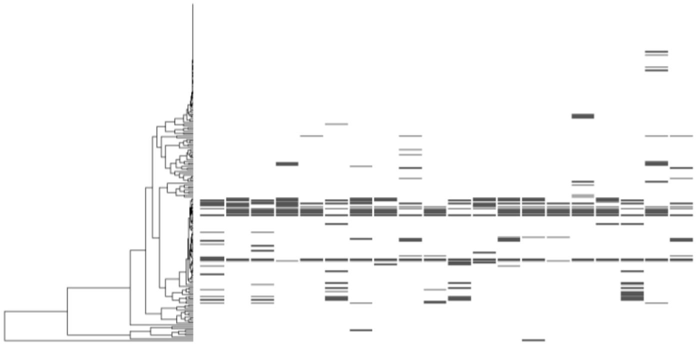
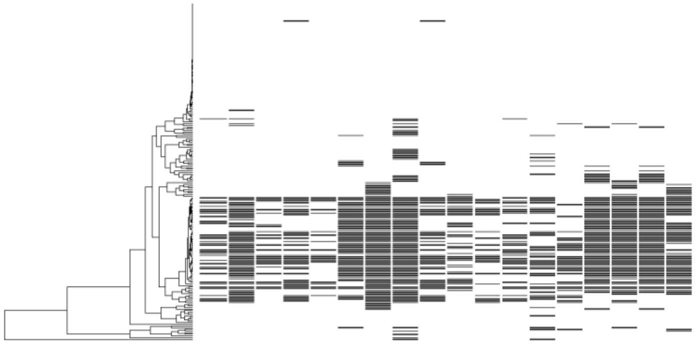
- ID: *corresponding to the clustering results and matched with the labels mentioned in the figures and main text.*
- Size: *total number of genes in the cluster.*
- Annotated: *number of genes that have GO annotations.*
- Enriched GO terms: *GO ID and GO name.*
- Frequency: *number of genes annotated with this GO; Genes can be annotated with more than one GO terms, so the sum of frequency could be greater than cluster size.*
- Expected: *expected number of this GO corresponding to cluster size.*
- P-value: *p-values of the Fisher test.*
- Phylogenetic Profiles: *due to the page limit, we only show 20 phylogenetic profiles within each cluster across a subtree with 200 tips randomly selected.*

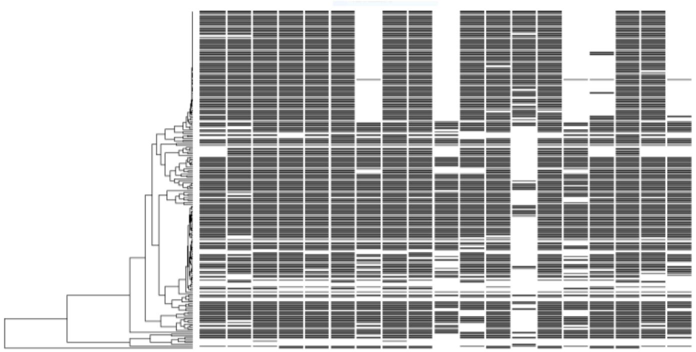
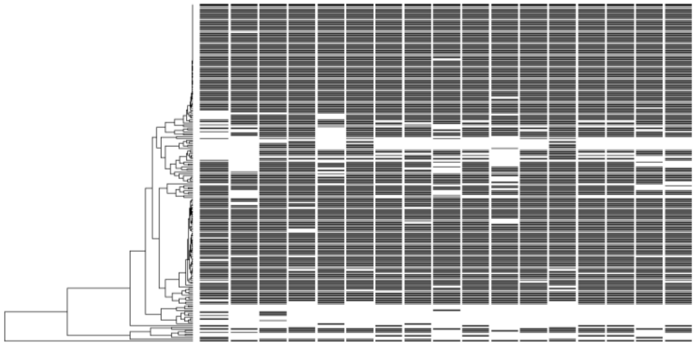
ID	Size	Annotated	Enrished GO terms	Frequency	Expected	P-value	Phylogenetic Profiles
1	307	269	GO:0055085 transmembrane transport	225	50.63	<1e-30	
			GO:0008643 carbohydrate transport	146	29.73	<1e-30	
			GO:0006013 mannose metabolic process	6	2.52	0.031	
			GO:0046373 L-arabinose metabolic process	4	1.4	0.040	

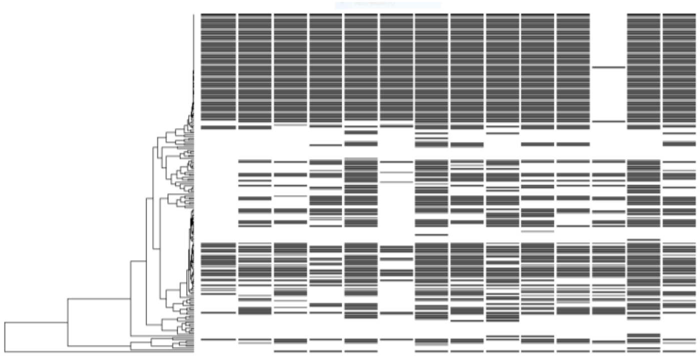
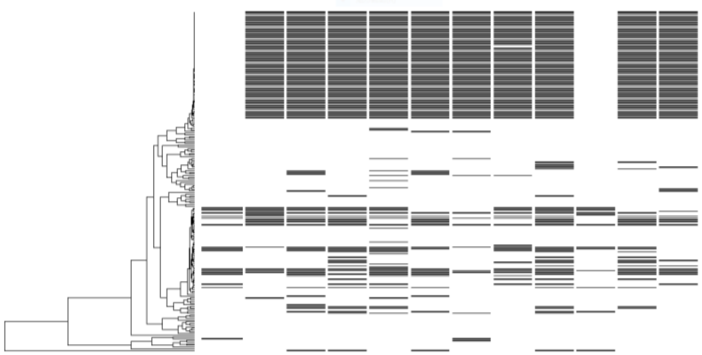
2	90	47	GO:0000160 phosphorelay signal transduction system	44	3.80	<1e-30	
3	53	8	GO:0030261 chromosome condensation	1			
			GO:0007059 chromosome segregation	1			
			GO:0015031 protein transport	1			
4	55	20	GO:0009103 lipopolysaccharide biosynthetic process	2	0.02	0.0001	
			GO:0008360 regulation of cell shape	3	0.13	0.0002	
			GO:0006412 translation	6	0.98	0.00026	

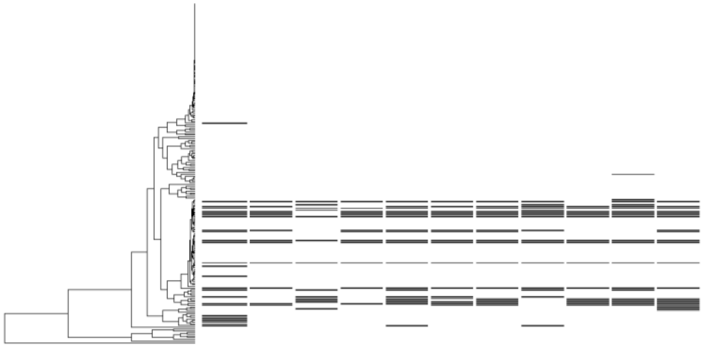
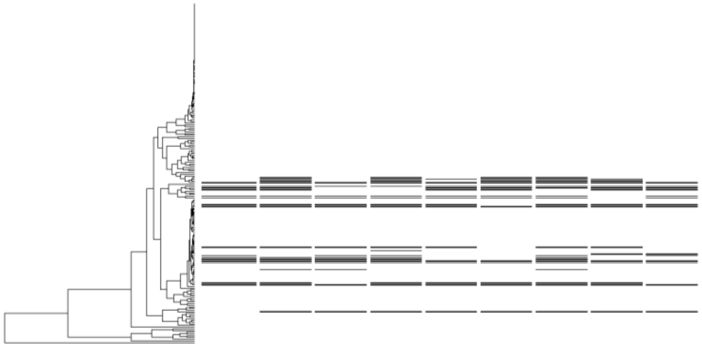
5	38	29	GO:0006935 chemotaxis	13	0.35	9.8e-20	
			GO:0071973 bacterial-type flagellum-dependent cell ...	8	0.45	4.5e-09	
			GO:0009306 protein secretion	5	0.18	4.1e-07	
6	32	26	GO:0009099 valinebiosynthetic process	5	0.12	3.7e-08	
			GO:0009097 isoleucine biosynthetic process	5	0.14	7.4e-08	
			GO:0006526 arginine biosynthetic process	5	0.11	1.7e-08	
			GO:0009098 leucine biosynthetic process	3	0.08	4.3e-05	
			GO:0000105 histidine biosynthetic process	3	0.14	0.00025	
			GO:0006541 glutamine metabolic process	4	0.19	2.4e-05	

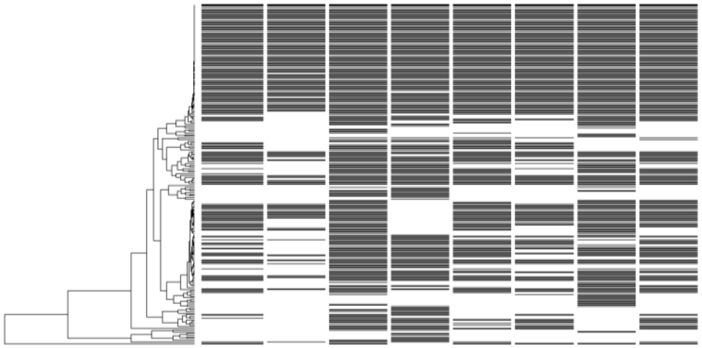
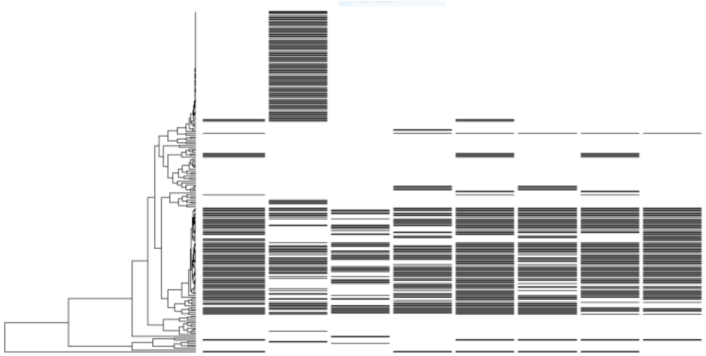
7	26	5	GO:0051301 cell division	2	0.09	0.0233	
8	26	6	GO:0015074 DNA integration GO:0006310 DNA recombination	4 4	0.08 0.12	2.8e-07 1.9e-06	
9	25	13	GO:0009847 spore germination GO:0042173 regulation of sporulation resulting in f... GO:0030435 sporulation resulting in formation of ... GO:0006352 DNA-templated transcription, initiation	2 3 4 3	0.02 0.03 0.07 0.21	0.00013 2.4e-06 2.6e-07 0.00098	

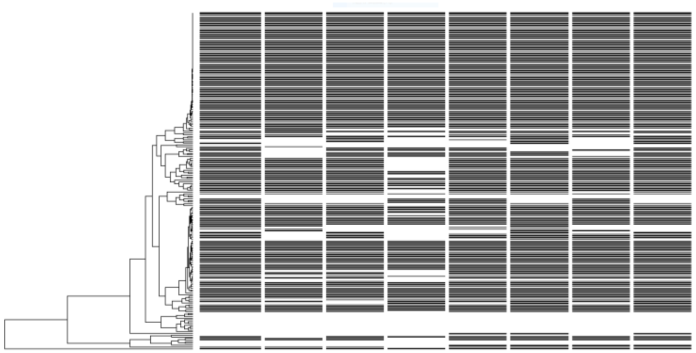
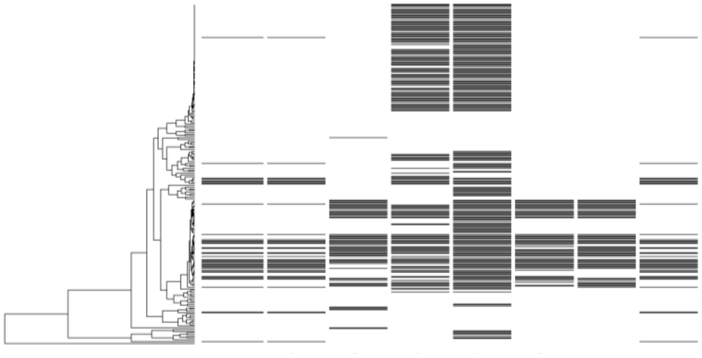
10	23	3	GO:0006139 nucleobase-containing compound metabolic...	2	0.77	0.1653	
11	18	4	GO:0009165 nucleotide biosynthetic process GO:0009116 nucleoside metabolic process	3 2	0.15 0.07	0.00021 0.00159	

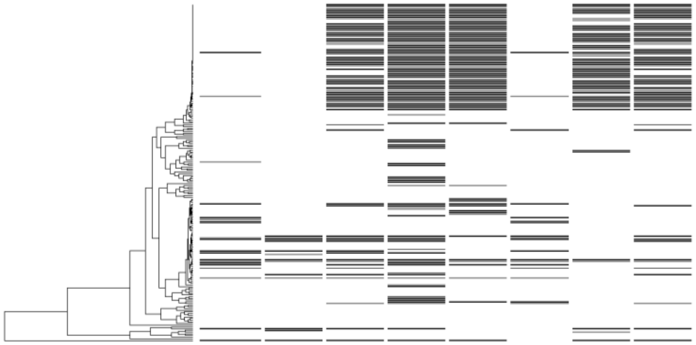
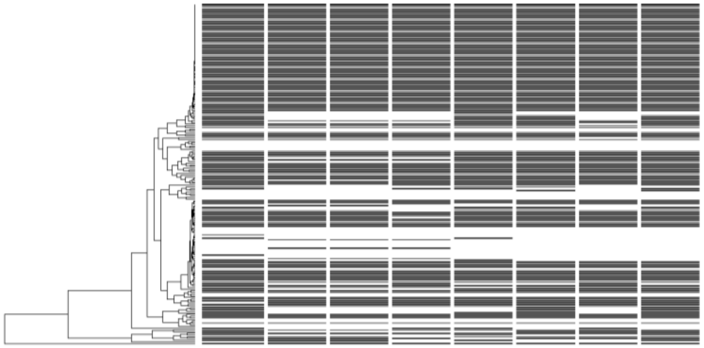
12	19	16	GO:0055085 transmembrane transport	16	3.01	1.9e-12	
13	17	10	GO:0006265 DNA topological change GO:0019281 L-methionine ... GO:0032049 cardiolipin biosynthetic GO:0009423 chorismate biosynthetic	4 1 1 1	0.06	1.8e-07	

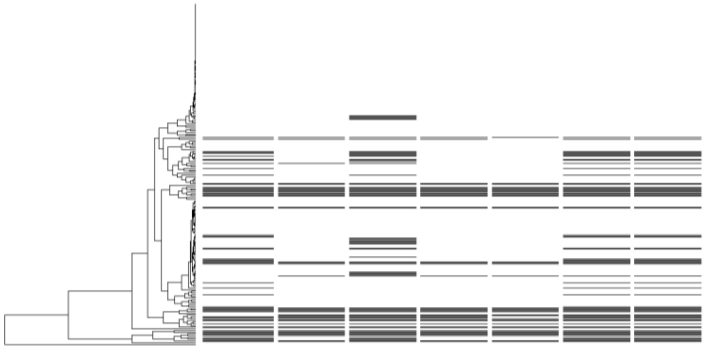
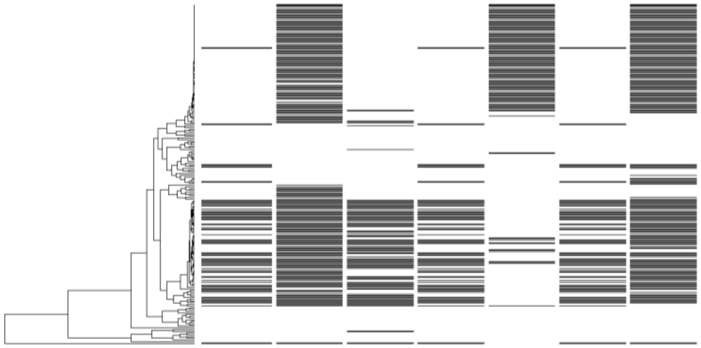
14	14	4	GO:0000160 phosphorelay signal transduction system	4	0.32	4.1e-05	
15	12	1	GO:0006352 DNA-templated transcription, initiation	1			

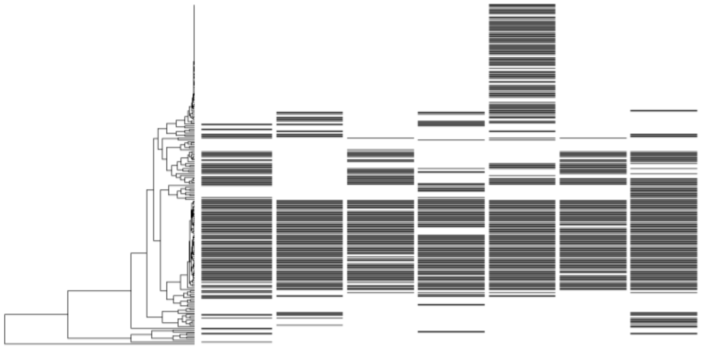
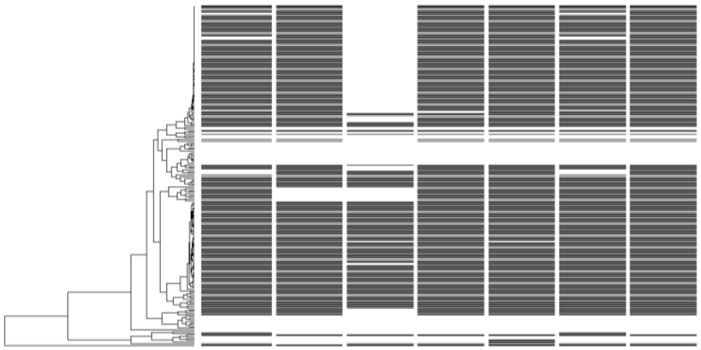
16	11	8	GO:0044780 bacterial-type flagellum assembly	3	0.07	3.2e-05	
			GO:0071973 bacterial-type flagellum-dependent cell ...	3	0.13	0.00018	
			GO:0006935 chemotaxis	1			
17	9	0	Lack of GO annotations				

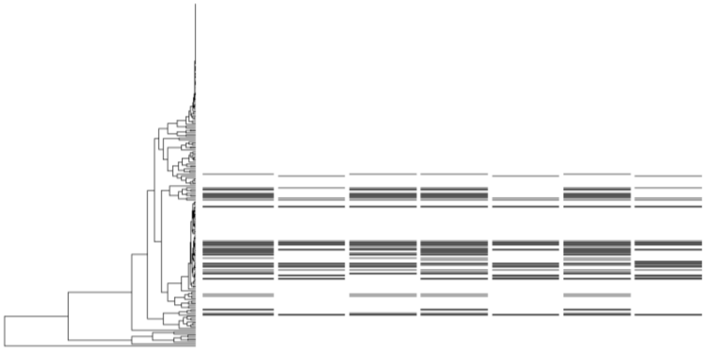
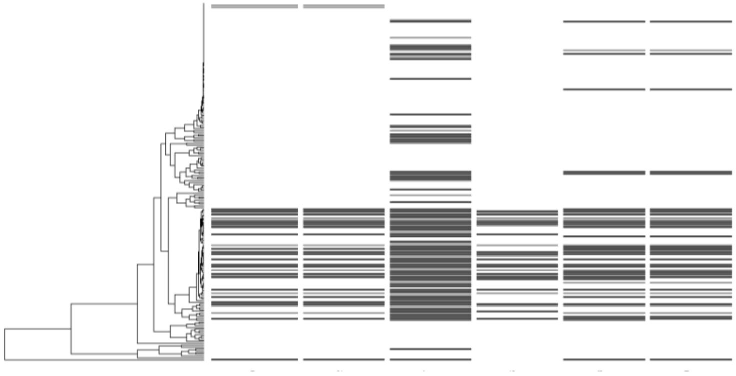
18	8	4	GO:0009401 phosphoenolpyruvate-dependent sugar phos... GO:0006352 DNA-templated transcription, initiation GO:0006355 regulation of transcription, DNA-templat...	3 1 1	0.02	5.9e-07	
19	8	1	GO:0006355 regulation of transcription, DNA-templat...	1			

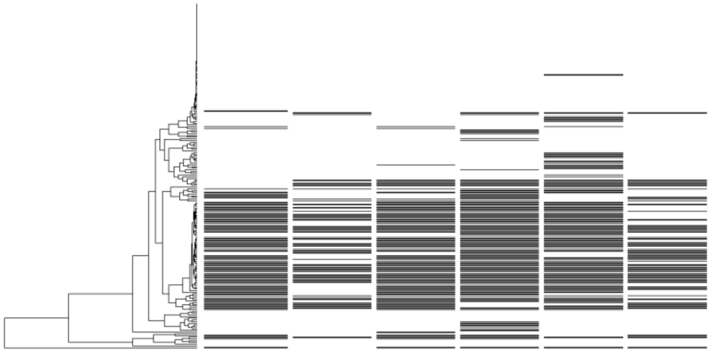
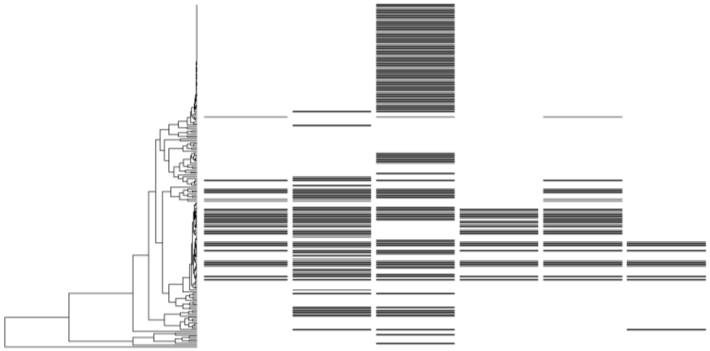
20	8	0	Lack of GO annotations		
21	8	3	GO:0009058 biosynthetic process GO:0000270 peptidoglycan metabolic process GO:0006022 aminoglycan metabolic process	1 1 1	

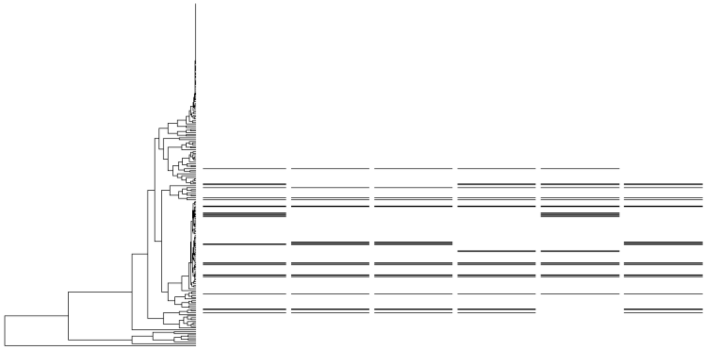
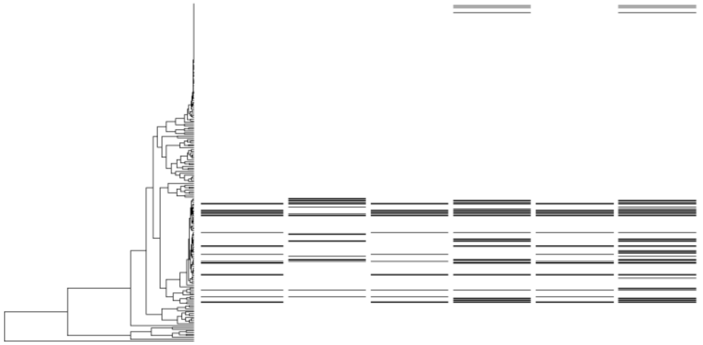
22	8	2	GO:0006355 regulation of transcription, DNA-templat... GO:0055085 transmembrane transport	1 1	
23	8	1	GO:0019605 butyrate metabolic process	1	

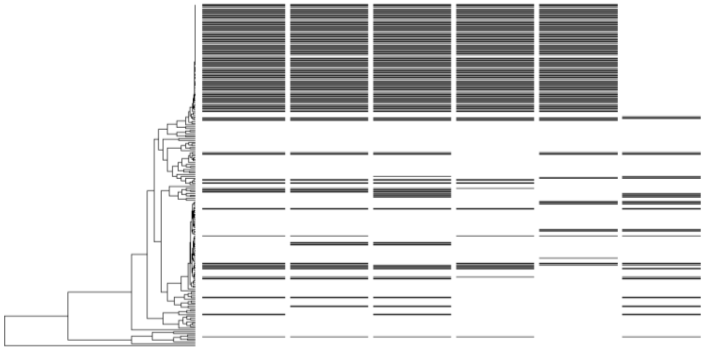
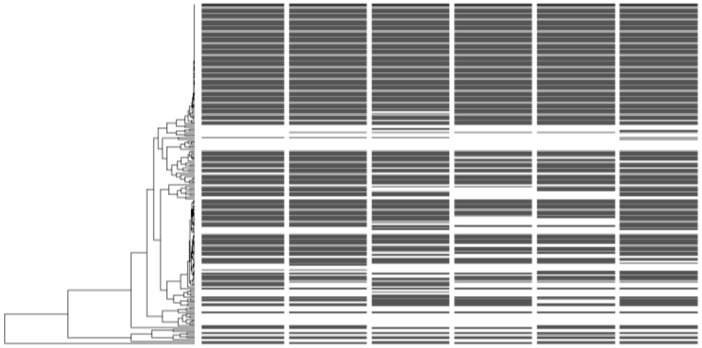
24	7	3	GO:0046944 protein carbamoylation GO:0051604 protein maturation GO:0055114 oxidation-reduction process	1 1 1	
25	7	2	GO:0009098 leucine biosynthetic process GO:0032259 methylation	1 1	

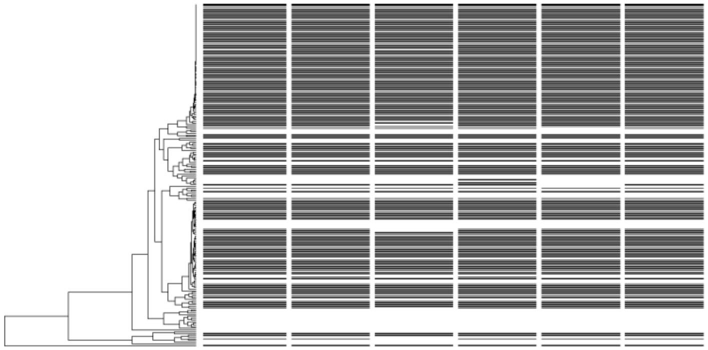
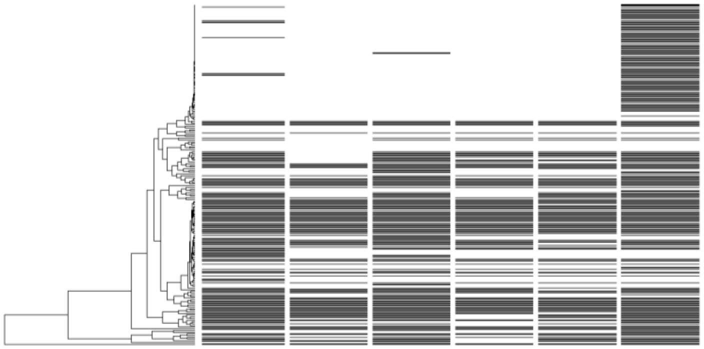
26	7	5	GO:0008654 phospholipid biosynthetic process	2	0.04	0.00064	
			GO:0015970 guanosine tetraphosphate biosynthetic pr...	1			
			GO:0006412 translation	2	0.25	0.02157	
27	7	6	GO:0005978 glycogen biosynthetic process	5	0.02	1.7e-13	
			GO:0005975 carbohydrate metabolic process	1			

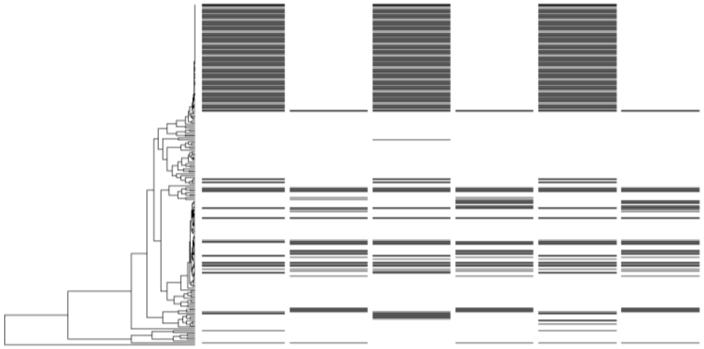
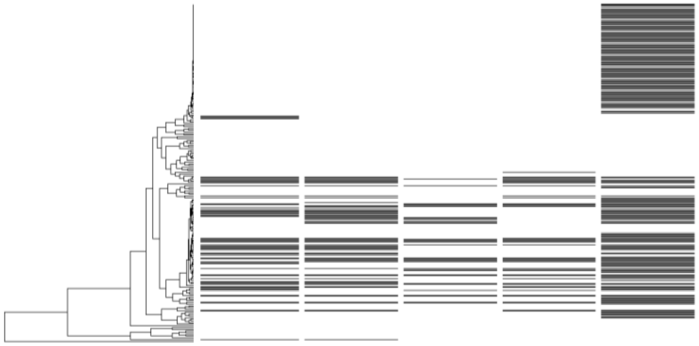
28	7	3	GO:0005975 carbohydrate metabolic process	3	0.52	0.0052	
29	6	0	Lack of GO annotations				

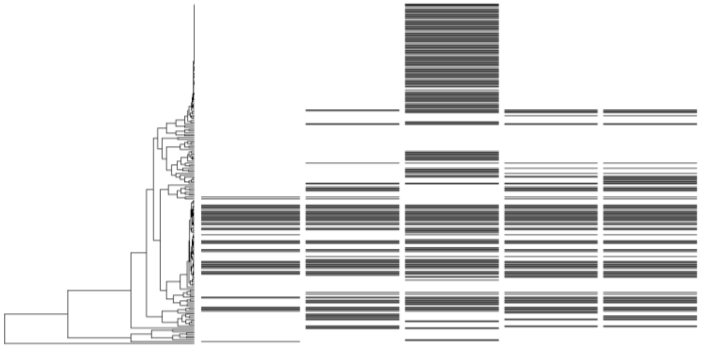
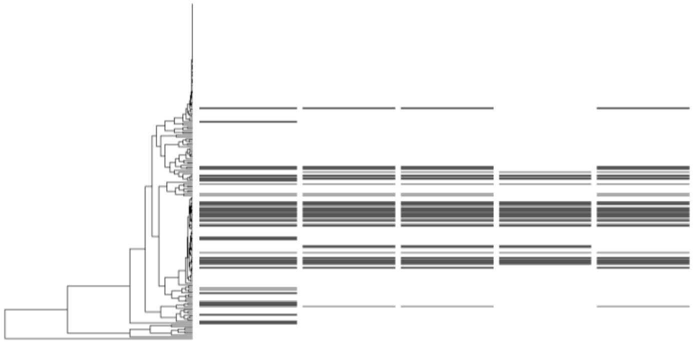
30	6	3	GO:0006564 L-serine biosynthetic process GO:0006807 nitrogen compound metabolic process	2 3	0.00 1.22	1.6e-06 0.06659	
31	6	3	GO:0005996 monosaccharide metabolic process GO:0000160 phosphorelay signal transduction system	2 1	0.13	0.0057	

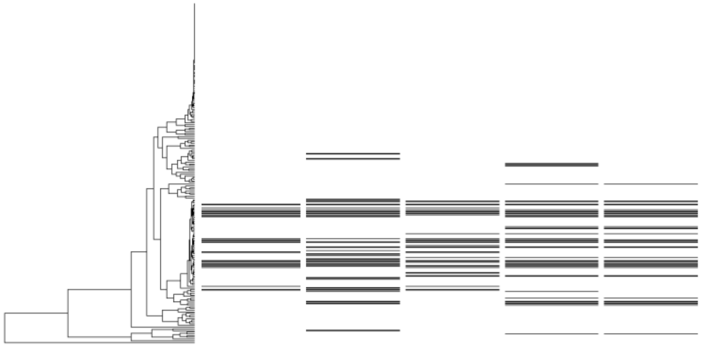
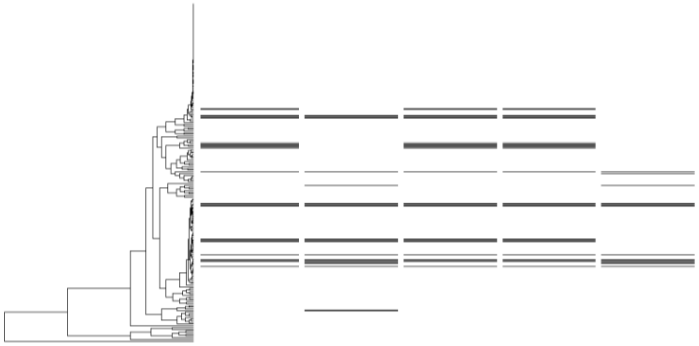
32	6	0	Lack of GO annotations				
33	6	2	GO:0015074 DNA integration	2	0.03	0.00015	
			GO:0006310 DNA recombination	2	0.04	0.00038	

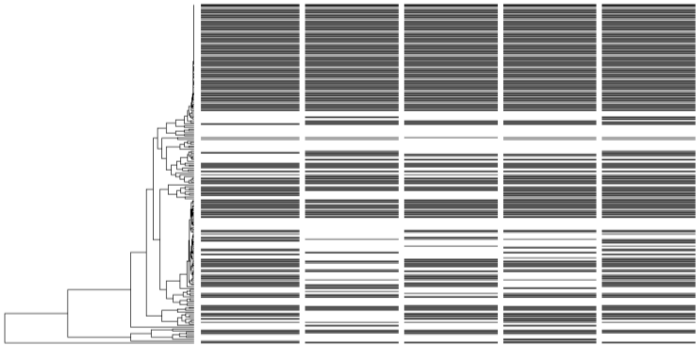
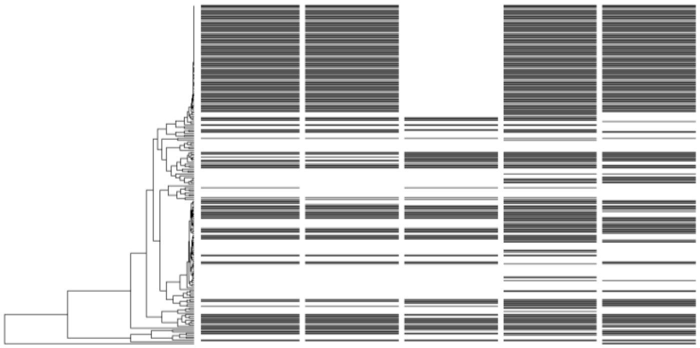
34	6	0	Lack of GO annotations	
35	6	0	Lack of GO annotations	

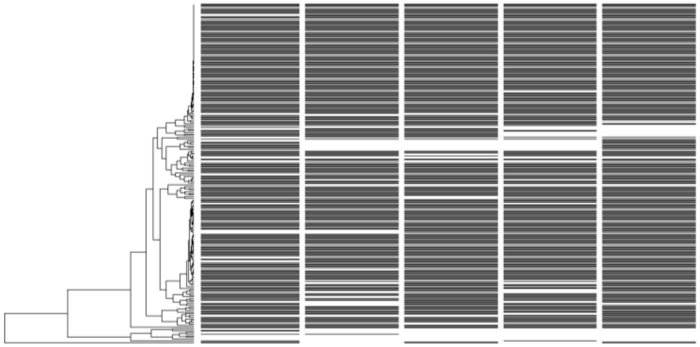
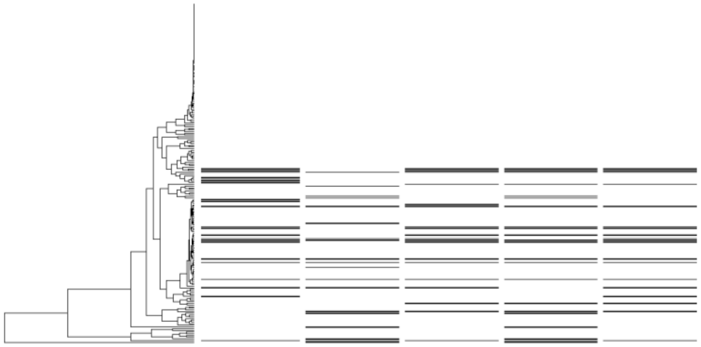
36	6	6	GO:0042121 alginic acid biosynthetic process	6	0.03	4.1e-16	
37	6	0	Lack of GO annotations				

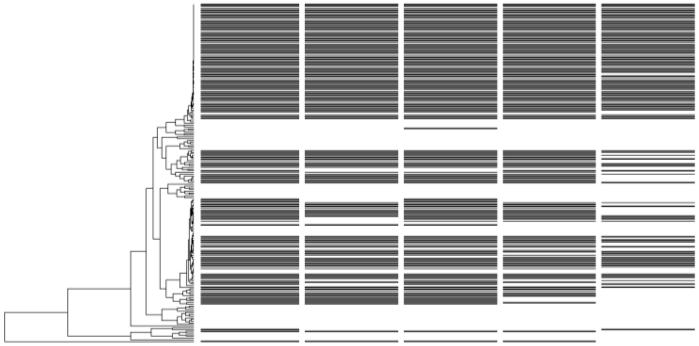
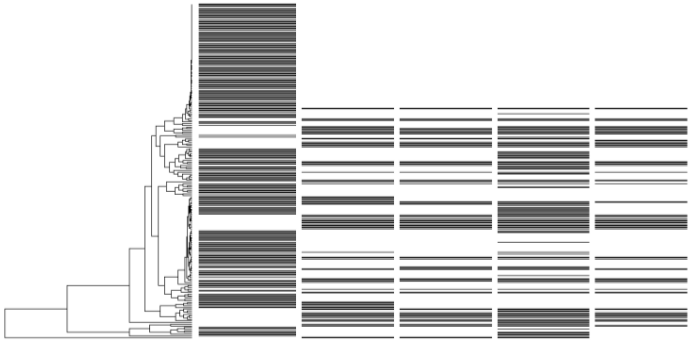
38	6	3	GO:0005975 carbohydrate metabolic process	3	0.52	0.0052	
39	5	0	Lack of GO annotations				

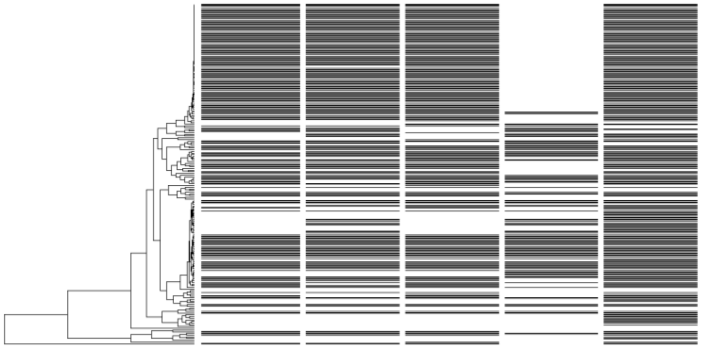
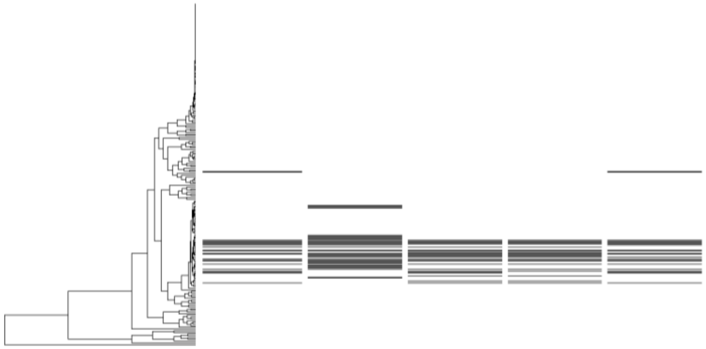
40	5	3	GO:0006064 glucuronate catabolic process GO:0006355 regulation of transcription, DNA-templat...	2 1	0.01	9.8e-06	
41	5	1	GO:0045490 pectin catabolic process	1			

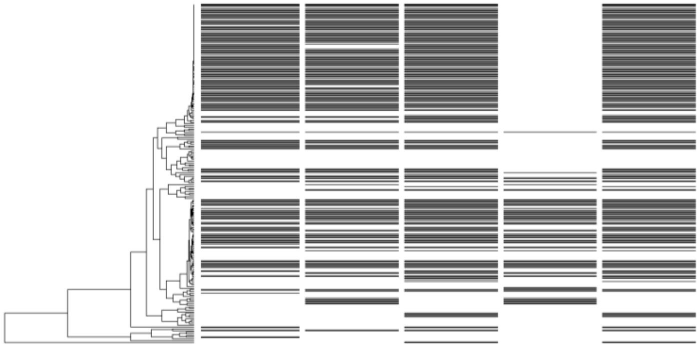
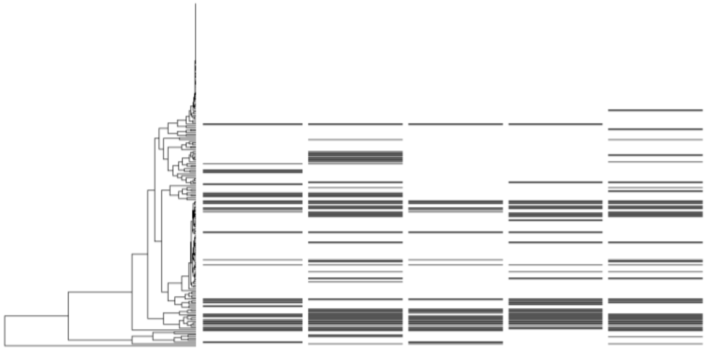
42	5	0	Lack of GO annotations	
43	5	0	Lack of GO annotations	

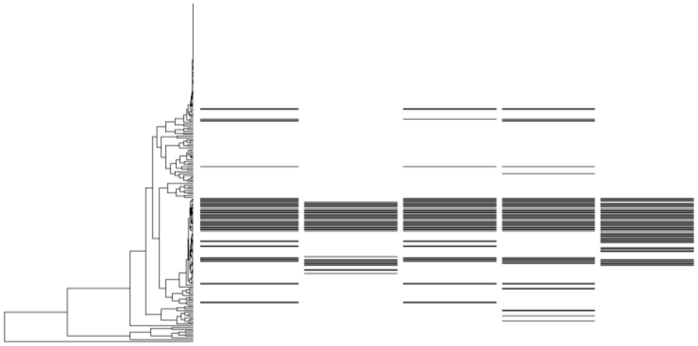
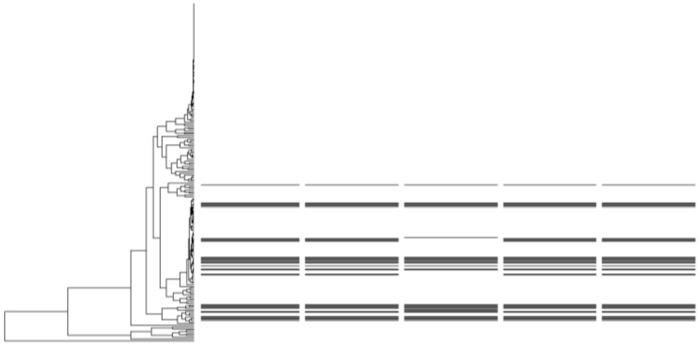
44	5	0	Lack of GO annotations				
45	5	5	GO:0019464 glycine decarboxylation via glycine clea... GO:0045454 cell redox homeostasis	3 2	0.01 0.02	8.5e-09 0.00015	

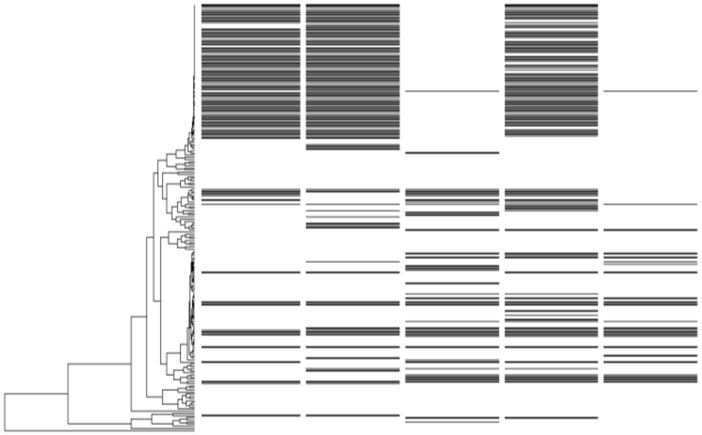
46	5	3	GO:0019877 diaminopimelate biosynthetic process GO:0006400 tRNA modification GO:0032259 methylation	1 1 1			
47	5	2	GO:0019310 inositol catabolic process	2	0.00	1.6e-06	

48	5	0	Lack of GO annotations				
49	5	4	GO:0016226 iron-sulfur cluster assembly	3	0.01	6.8e-08	
			GO:0006534 cysteine metabolic process	1			
			GO:0006790 sulfur compound metabolic process	4	0.06	3.1e-08	

50	5	4	GO:0000160 phosphorelay signal transduction system GO:0055085 transmembrane transport	3 1	0.32	0.0020	
51	5	1	GO:0036376 sodium ion export across plasma membrane	1			

52	5	2	GO:0006051 N-acetylmannosamine metabolic process	2	0.00	5.4e-07	
53	5	2	GO:0006355 regulation of transcription, DNA-templat...	2	0.14	0.0052	

54	5	1	GO:1901135 carbohydrate derivative metabolic proces...	1			
55	5	5	GO:0005975 carbohydrate metabolic process	5	0.87	0.00016	

56	5	1	GO:0000160 phosphorelay signal transduction system	1	

The clusters of size at least 5 are reported.