Gómez-Reyes Ricardo

https://orcid.org/0000-0002-6173-4986
Instituto de Investigaciones Oceanológicas (IIO), Universidad Autónoma de Baja California (UABC), Carretera Ensenada-Tijuana No. 3917, Fraccionamiento Playitas, Ensenada, Baja California CP 22860, México

Exploratory global transcriptome stress response (of *Pocillopora grandis* transplanted into different reefs: candidate genes for monitoring stress responses at holobiont level)

How the holobiont responses to abiotic factors require complex analytical methods to search important gene-related information out of large transcripts datasets. As genes work within networks, not individually, effective analytical methods to identify important drivers of biological events that focus on groups of genes may better model the mechanisms of response. Weighted Gene Correlation Network Analysis (WGCNA) is a data-driven clustering algorithm that can be used to identify clusters of genes that act similarly across individuals (Langfelder et al, 2008). In order to assess a global holobiont response to location stress, a transcriptome data-driven unsupervised clustering analysis was performed. First, the transcript expression level matrix was log-transformed (log2 + 1) and transcripts were filtered as the function of count per million > 1 and sample frequency > = 2, then the WGCNA R package (Langfelder and Horvath, 2008) was configured to calculate a weighted and signed network (soft-threshold = 29, correlation method = Pearson). A significance trait association was calculated using the first principal component (eigengenes) of the obtained modules using an asymptotic student test. Significant gene clusters (hereafter modules) were considered based on *P*-value < 0.05 (*). The resulting dissimilarity matrix was exported into ggraph (Pedersen, 2018a) and tidygraph (Pedersen, 2018b) R packages to explore candidate genes for monitoring stress. In Transcripts with high degree (percentil 95) were annotated

using gene ontology, and terms were reduced based on the semantic similarity using human Genome wide annotation in rvgo (Sergi Sayols, 2020)

Results

A dataset of 47,906 transcripts (26,619,034 reads) was splitted into 10 modules. Based on the significance trait association analysis, two modules, the turquoise (17,567 transcripts) and blue (11, 383 transcripts), were significantly associated with the translocation (between Carrizales and La Boquita) and/or the anthropogenic stress sites (Figure 1). In fact, the turquoise and blue modules grouped an important fraction of the significant transcripts up-regulated (100 %) and down-regulated (99 %) detected during the DET analysis, respectively, suggesting a regulatory turnover expression between these modules associated with the translocation stress. Besides, eight modules are summarized in Table 1. A set of 302 transcripts with high degree of interaction (percentil 95) were obtained from the turquoise (170) and blue module (132) and summarized in Supplementary Table S and Figure 3.

Table 1. Summary results

Modules	Transcripts within the modules	Reads within the modules	% Transcript	% Reads	Transcripts with Annotation	GO	DEGS	Transcripts within the network
turquoise (*)	17,567	10,919,315	36.67%	41.02%	4,957	46,725	36 (up)	13,961
blue (*)	11,383	4,964,162	23.76%	18.65%	4,667	61,016	1331 (Down)	9,133
brown	5,410	3,675,018	11.29%	13.81%	1,839	18,886		1,800
yellow	3,898	1,868,561	8.14%	7.02%	1,269	14,881		587
green	3,717	1,812,353	7.76%	6.81%	1,274	12,704		693
red	3,405	2,437,886	7.11%	9.16%	1,265	14,105	1 (Down)	901
black	1,059	249,828	2.21%	0.94%	285	3,082		76
pink	789	216,892	1.65%	0.81%	247	2,918		0

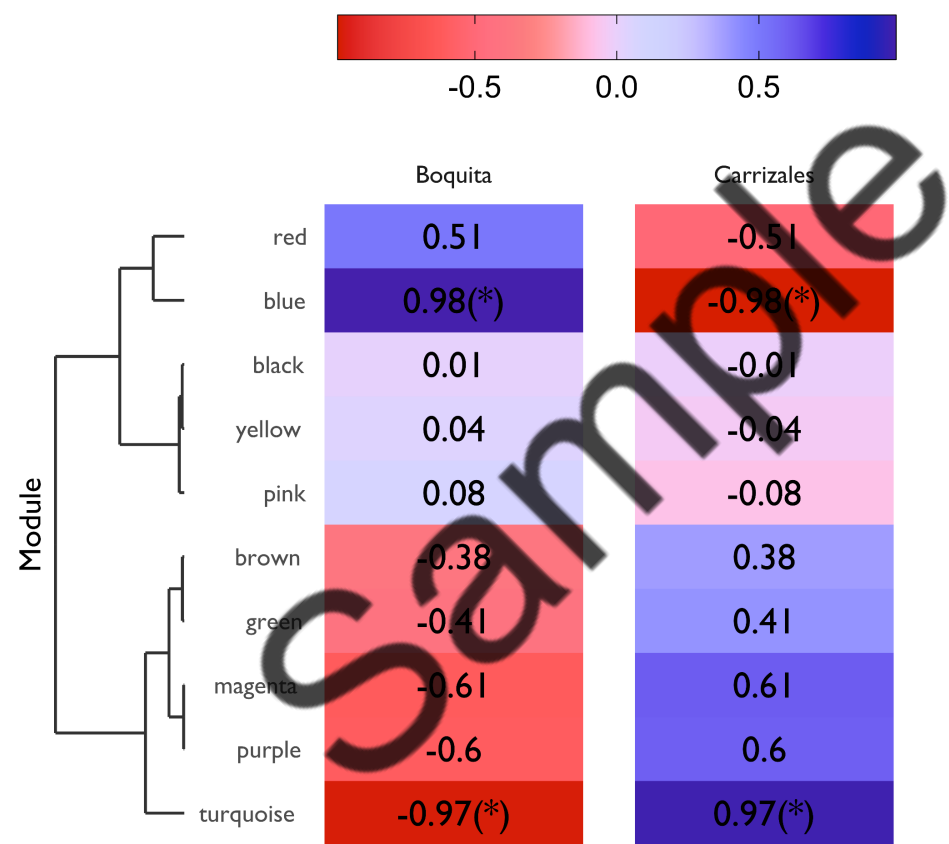|									
	magenta	377	395,294	0.79%	1.49%	130	1,660		0
	purple	301	79,725	0.63%	0.30%	79	931		101
Total	10	47,906	26,619,034	100.00%	100.00%	16,012	176,908		27,252

Figure 1. Significance trait association analysis. Two modules, the turquoise (17,567 transcripts) and blue (11, 383 transcripts), were significantly associated with the translocation (between Carrizales and La Boquita) and/or the anthropogenic stress sites.

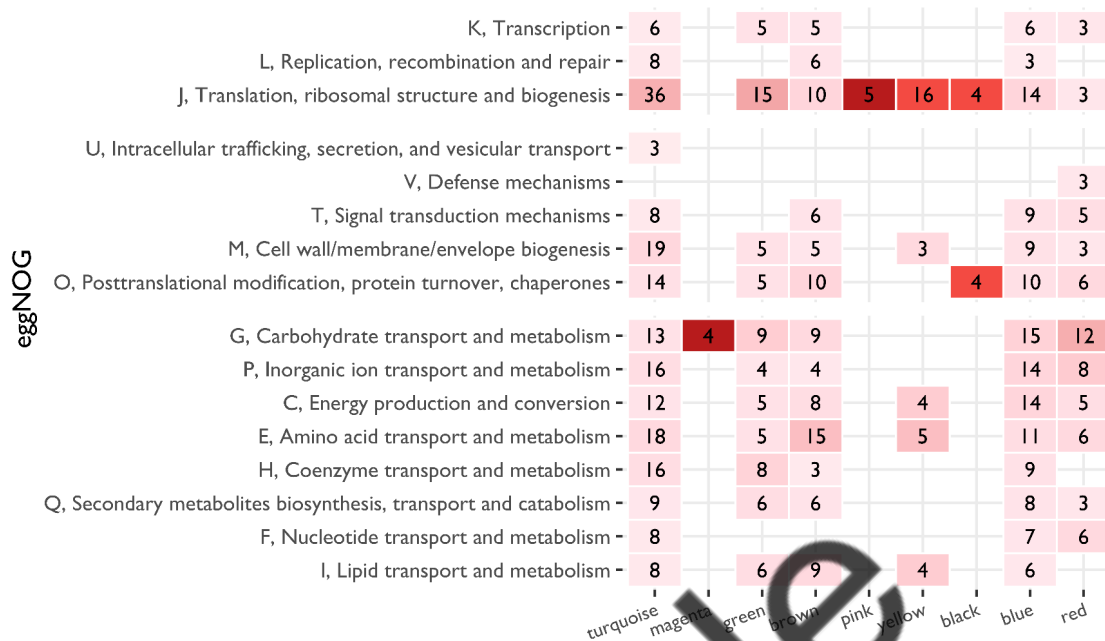| eggNOG	turquoise	magenta	green	brown	pink	yellow	black	blue	red
K, Transcription	6		5	5				6	3
L, Replication, recombination and repair	8			6				3	
J, Translation, ribosomal structure and biogenesis	36		15	10	5	16	4	14	3
U, Intracellular trafficking, secretion, and vesicular transport	3								
V, Defense mechanisms									3
T, Signal transduction mechanisms	8			6				9	5
M, Cell wall/membrane/envelope biogenesis	19		5	5		3		9	3
O, Posttranslational modification, protein turnover, chaperones	14		5	10			4	10	6
G, Carbohydrate transport and metabolism	13	4	9	9				15	12
P, Inorganic ion transport and metabolism	16		4	4				14	8
C, Energy production and conversion	12		5	8		4		14	5
E, Amino acid transport and metabolism	18		5	15		5		11	6
H, Coenzyme transport and metabolism	16		8	3				9	
Q, Secondary metabolites biosynthesis, transport and catabolism	9		6	6				8	3
F, Nucleotide transport and metabolism	8							7	6
I, Lipid transport and metabolism	8		6	9		4		6	

Figure 2. The transcriptome annotated against the eggNOG database were splitted based on the modules detected during the co-expression analysis. The pathways are categorized in Genetic information processing (K, L and J), Cellular processes (U, V, T, M and O), and metabolism (G, P, C, E, H, Q, F and I) .Transcripts withs known pathway classification are showed.
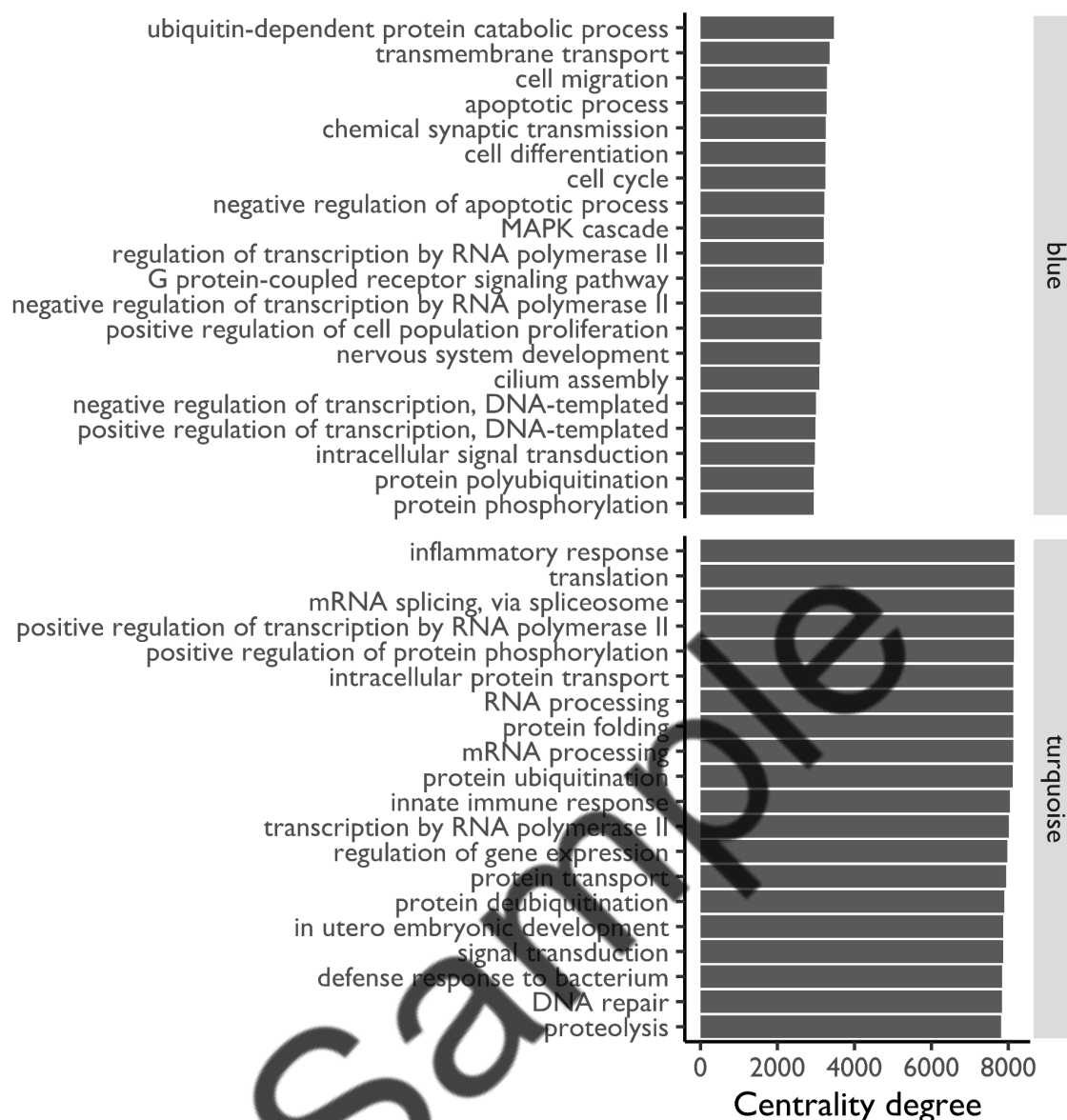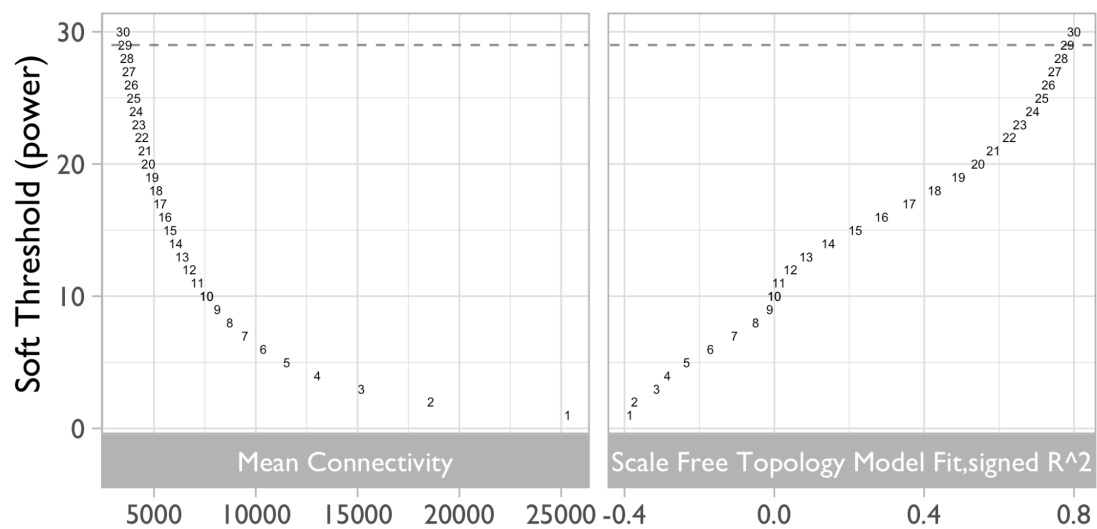
Figure 3. Transcripts with high degree (percentil 95) were annotated using gene ontology, and terms were reduced based on the semantic similarity using human Genome wide annotation.

Lowest power for which the scale free topology index reaches the 77.5 %

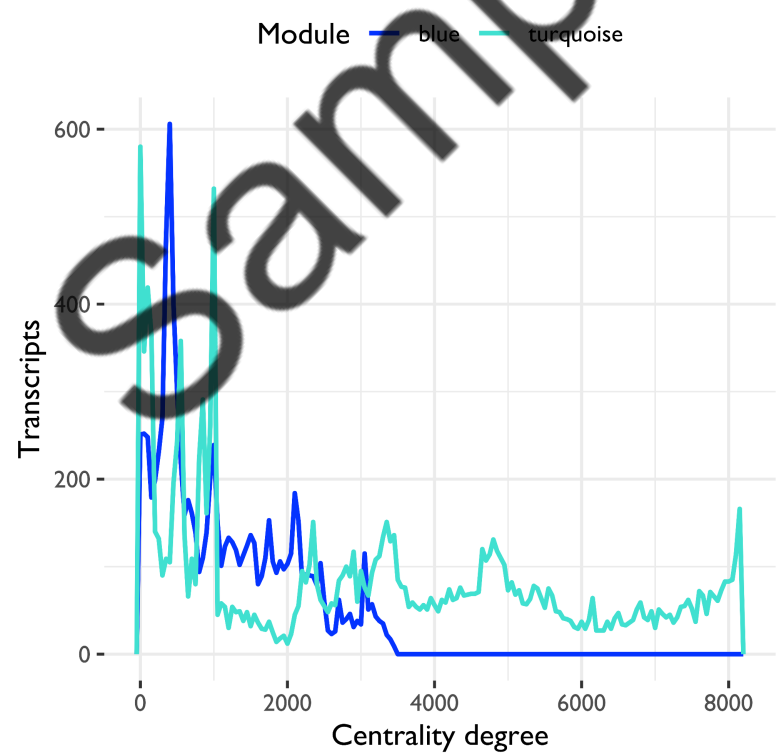Fig. S1. Analysis of scale free topology for multiple soft thresholding powers

Figure 2S. Transcripts degree centrality shows an important number of interactions within the
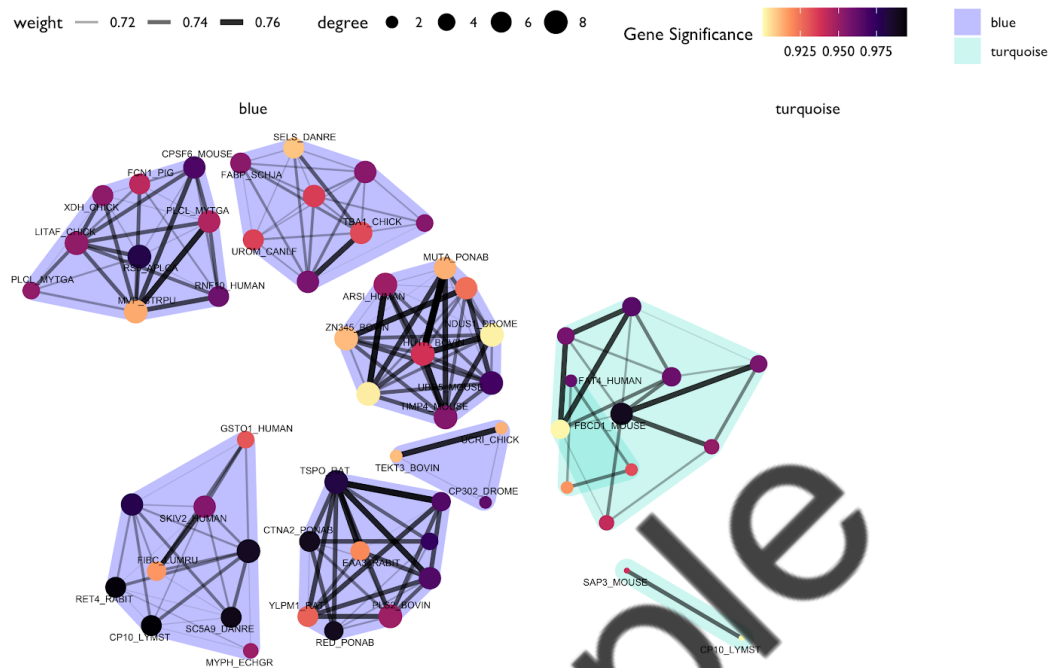
transcriptome

Fig 8. Example of Interconnectivity visualization of enriched modules. Gene significance to each module is colored, showing high gene significance (> 0.97) into the central nodes. Non-annotated but significant transcripts are included in the regulatory network in both modules.

Cite:

Langfelder, P., & Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. BMC bioinformatics, 9(1), 1-13.

Pedersen TL. Tidygraph: a tidy API for graph manipulation. R package version 1.1.1. 2018. (https://CRAN.R-project.org/package=tidygraph)

Pedersen TL, Ggraph: an implementation of grammar of graphics for graphs and networks. R package version 1.0.2. 2018. (https://CRAN.R-project.org/package=ggraph)

Sergi Sayols (2020). rrvgo: a Bioconductor package to reduce and visualize Gene Ontology terms (https://ssayols.github.io/rrvgo.)

Pipeline: https://github.com/RJEGR/CORAL_PAPER