SUPPLEMENTAL INFORMATION

Protein signatures predict coral resilience and survival to thermal bleaching events

Brook L. Nunn[1*], Tanya Brown[2,3], Emma Timmins-Schiffman[1], Miranda C. Mudge[1], Michael Riffle[1], Jeremy B. Axworthy[2], Jenna Dilworth[4], Carly D. Kenkel[4], Jesse Zaneveld[5], Lisa J. Rodrigues[6], Jacqueline L. Padilla-Gamiño[2]

[1]Department of Genome Sciences, University of Washington, Seattle, WA, United States

[2]School of Aquatic and Fishery Sciences, University of Washington, Seattle, WA, United States

[3]Department of Biology, University of Texas, Tyler, TX, United States

[4]Department of Biological Sciences, University of Southern California, Los Angeles, CA, United States

[5]School of Science, Technology, Engineering & Mathematics, University of Washington, Bothell, WA, United States

[6]Department of Geography and the Environment, Villanova University, Villanova, PA, United States

*** Correspondence:**
Brook L. Nunn
brookh@uw.edu

This PDF file includes:

Supplemental Methods

Symbiont and Chlorophyll analyses. Chlorophyll a concentrations and dinoflagellate symbiont (Symbiodiniaceae) densities from each of the colonies were investigated (Supplementary Data 1a). Briefly, chlorophyll a was extracted with 100% acetone and absorbance was measured with a light spectrophotometer (Supplementary Data 1c). Symbionts were separated from triplicate ground coral tissue by centrifugation and symbiont pellets were homogenized prior to being counted using a hemocytometer. Chlorophyll *a* and symbiont densities were standardized to grams of ash-free dry weight (gdw) of coral tissue (Supplementary Data 1a,c). Assessments of bleaching status between T1 and T2 were completed on all colonies every week using the Coral Watch Card (Fig. S1c).

qPCR Probe Corrections. Differences in Actin probe fluorescence intensity were accounted for using known quantities of target DNA over duplicate dilution series run in duplicate reactions following Cunning et al.[1]. This method indicated that for the same template concentration, the *Cladocopium* assay amplified 0.55 cycles later than the *Durusdinium* assay. This value was subtracted from the *Cladocopium* assay C_T values to normalize these differences before subsequent analyses.

Symbiont Community Composition. Tissue samples were not available for determination of symbiont community composition at each timepoint and treatment condition for all corals in this study (Fig. S2a, b). Additionally, the symbiont concentration in many of the bleached samples at T2 fell below the detection level for our qPCR assays[2], meaning we could not confidently determine the proportion of *Cladocopium* and *Durusdinium* in such samples. However, *Montipora capitata* in Kāneo'he Bay is associated with symbionts in the genera *Cladocopium*, *Durusdinium*, or a mix of both[1]. The type of symbiont community hosted by these colonies is stable and conserved through time, even after multiple thermal stress events[3]. Thus, symbiont community composition, defined as *Cladocopium*- or *Durusdinium*-dominated (≥90% one symbiont type), or mixed, for each coral in this study was determined using the proportion of *Cladocopium* and *Durusdinium* in available samples for each coral (Table S3, Fig. S2a, b). A logistic regression analysis was conducted to assess the relationship between the proportion of *Durusdinium* symbionts pre-bleaching and coral mortality post-bleaching. Supplementary Data 1b included coral genets classified as either "resilient" (survived) or "susceptible" (died) and their respective proportions of *Durusdinium* symbionts. Rows with missing data were excluded from the analysis. The binary outcome variable (mortality) was coded as 1 for "susceptible" and 0 for "resilient." The model was fitted using the glm function in R with a binomial family to account for the binary nature of the response variable. Statistical significance of the relationship was assessed using *p*-values for the proportion of *Durusdinium* symbionts as the predictor variable. Corals for which samples throughout the entire time series of the experiment are available further support our assumption that symbiont community type is conserved through time, even after experimental thermal stress.

Proteomics. Six colonies from the resilient cohort (colony numbers R.10, R.13, R.21, R.26, R.66, R.74) and six colonies from the susceptible cohort (colony numbers S.5, S.20, S.25, S.28, S.55, S.73) were randomly selected as bioreplicates to track phenotypic differences in protein abundance through time (*i.e.,* T_1 and T_2). Proteomic

analyses were not completed on T_2 control colonies that remained at 25°C unbleached. A total of 24 samples of frozen coral (4 mm diameter x 1 mm thick, tissue plus skeletal matrix) were manually smashed using a mini mortar and pestle and resulting tissue and debris were sonicated (5 x 10 s) in 6 M urea in 50 mM NH_4HCO_3. Sample tubes were cooled between each sonication in a bath of ethanol and dry ice (30 s). Protein concentrations were measured using the Pierce BCA Protein Kit microplate assay (Thermo Scientific) following the manufacturer's protocol.

Protein digestions were conducted with 50 µg of protein per coral sample. Protein lysates were solubilized in 100 µl of 6 M urea buffer, reduced with 500 mM TCEP (1 µl), and pH was controlled with 1.5 M Tris pH 8.8 (6.6 µl) and adjusted as needed to pH > 8. Proteins were alkyated with iodoacetamide (IAA; 200 mM, 20 µl) and remaining alkylating agent was quenched with diothiothreitol (DTT; 200 mM, 20 µl). Prior to trypsin digestions, the urea was diluted with 25 mM NH_4HCO_3 (800 µl) and HPLC grade methanol (200 µl). Trypsin (modified porcine sequencing grade trypsin; Promega; 1:20 enzyme:coral protein) was added to enzymatically cleave proteins overnight at room temperature. Sample pH was then modified to $\leq$2 with 10% trifluoracetic acid (TFA) and peptides were evaporated to dryness at 4°C in a speedvac. Dried coral peptide samples were reconstituted in 5% acetonitrile and 0.1% trifluoroacetic acid and desalted on C18 macrospin columns (The NestGroup) following the manufacturer's guidelines. Eluted peptides were dried to final volume <10µl on a speedvac and reconstituted in 200 µl 2% acetonitrile and 0.1% formic acid. Peptide Retention Time Calibration Mixture (PRTC; Pierce) was added to each coral peptide sample as an internal standard to ensure consistency of peptide detection. Each sample was amended with PRTC such that 50 fmol of PRTC was analyzed with 1 µg of coral peptides for each mass spectrometry experiment.

M. capitata samples were analyzed using liquid chromatography coupled to tandem mass spectrometry (LC–MS/MS) on a Q–Exactive–HF (Thermo Scientific) in Data Dependent Acquisition (DDA) mode. An in–house analytical column (40 cm long) was packed with C18 beads (Dr. Maisch HPLC, Germany, 0.3 µm) and heated (50^{o}C) during analysis to maintain steady backpressure. Peptides were chromatographically separated on a Waters nanoAcquity UPLC using an acidified (0.01% formic acid) acetonitrile:water gradient of 2–45% over 120 minutes. MS1 was collected on 400-1200 m/z with a 70,000 resolution and AGC target of 1e6; MS2 were collected with a loop count of 20 excluding +1 and >+6 MS1 ions using a 10 s dynamic exclusion, 35,000 resolution, and AGC target of 5e4. Sample analyses were randomized, and quality controls were analyzed every 11th injection. Select peptides from QC samples were monitored using Skyline[4] to ensure that peptide peak area correlation variances were <10% through the duration of the analyses. Three bioreplicates of susceptible corals did not yield high quality mass spectrometry results due to extensive intracellular degradation (*i.e.,* cell- induced death) and were not utilized in any of the analyses (T_2 samples S.20, S.25, S.28). The raw MS data is at PRIDE accession PXD021262.

From each mass spectrometry experiment *M. capitata* peptides were identified and proteins were inferred using a proteome database derived from the *M. capitata* transcriptome[5] GSE97888_Montiporacapitata_transcriptome.fasta). The transcriptome was translated using Transdecoder v 2.0.1[6]. This final *M. capitata* proteome was concatenated with 50 common contaminants (cRAPome) and the internal standard PRTC peptides to generate our search database (Supplementary Data 6). The raw MS data were searched against the protein database using Comet v 2016.01 rev.3[7]). Comet parameters included concatenated decoy search, 20 ppm peptide mass tolerance, enzyme set at trypsin with two missed cleavages allowed, variable cysteine modification of 57 Da, and fixed methionine modification of 15.999 Da. Concatenated target–decoy

databases searches were completed and minimum protein and peptide thresholds were set at *P*> 0.95 on ProteinProphet and PeptideProphet. Protein identifications from the whole-cell lysates were accepted by ProteinProphet if the above-mentioned thresholds were passed, two or more peptides were identified, and at least one terminus was tryptic (false discovery rate <0.01)[8]. Resulting data files across all samples were analyzed with Abacus[9] to generate consistent protein inferences across replicates (*i.e.,* joining homologous protein identifications if needed) and calculating normalized spectral abundance factors (Supplementary Data 2i). Proteins were considered high confidence if two unique peptides were identified across all mass spectrometry experiments. For each treatment and time, a subset from that list was generated and included all proteins whose sum of the total spectral counts was >2 for the specific cohort (Supplementary Data 2a-d and e).

Differential relative protein abundances for resilient vs. susceptible corals were determined for each timepoint (T_1 and T_2) using the QPROT-QSPEC package[10]. QPROT-QSPEC package first normalizes each dataset and then calculates the posterior distribution of the log fold change and the associated Z-statistic[9,10](Supplementary Data 2g, h). The Z-test is similar to the T-test but accommodates for missingness of data when looking at sample size comparisons to determine significance levels to report or compare. Z-scores were converted to *p*-values using R (pvalue2sided=2*pnorm(-abs(z)) to simplify reporting. Differential abundances of proteins are reported with the following *p*-value cutoff rules: 1) $p<0.10$ if several proteins within a pathway, 2) $p<0.05$ if significance of an individual protein, or 3) $p<0.01$ if identifying a potential biomarker.

Volcano plots were made with resulting QPROT statistics in R ggplot v 3.0.4[11], and the heatmaps were generated with pheatmap v. 1.0.12[12]. NSAF (Normalized Spectral Abundance Factor) is a quantitative metric used in proteomics to estimate the relative abundance of proteins in a sample based on spectral counting. It accounts for variations in protein length and the total number of spectra in a dataset, enabling a more accurate comparison of protein abundance across samples. NSAF normalizes the spectral count of a protein by dividing it by its length, then dividing this ratio by the sum of all spectral count-to-length ratios for all identified proteins in the sample. Heatmap (Fig. 3) was generated by averaging NSAF values for each sample suite (*i.e.,* T_1S, T_1R, T_2S, T_2R) for a defined set of proteins that meet the established criteria (LFC $\geq$|1|, *p*-value<0.01) while the biomarker heatmap (Fig. 6) was generated from NSAF values for each bioreplicate to demonstrate quantitative consistency. Vertical clusters were completed with the "correlation" algorithm.

All proteins identified were analyzed with BlastKOALA[13] to find associated K numbers from the Kyoto Encyclopedia of Genes and Genomes (KEGG) and a list of metabolic pathways in which the enzymes are putatively involved. The top-scoring K numbers, KEGG descriptions, and gene names are reported. The *M. capitata* proteome was also analyzed with BLASTp against UniProtKB Swiss-Prot non-redundant protein database[14] (downloaded 10.15.2018) using the diamond algorithm[15]. Names of proteins in the manuscript text and figures are reported using the following hierarchy: 1) KEGG (gene name, description, metabolic pathways), 2) Uniprot BLAST (gene name, description), and 3) Uniprot BLAST Gene Ontology (GO Biological Process) (Supplementary Data 2f). Due to its significant LFC at both timepoints, the protein sequence for hypothetical protein m.24867 was interrogated using the protein Basic Local Alignment Search Tool (BLASTp) against the non-redundant protein database. The best-scoring descriptive annotation was "CyanoVirin-N domain containing protein" (CVNH: 95% Query coverage; e-value 2e-30; Fig. S5).

A second analysis of the mass spectrometry data was performed as a test to determine if known coral viral proteins could be detected from the whole-coral mass spectrometry

analyses. Protein sequences from five viral proteomes that were translated from viral DNA that was isolated from corals (Table S3)[16]. The resulting viral protein FASTA files were downloaded from Uniprot or NCBI (3.15.2022) and concatenated with the above mentioned *M. capitata* protein database, yielding a final database with 42,941 proteins (Supplementary Data 7). Mass spectrometry search parameters above were replicated with this database. A subset of the new ABACUS data with all viral proteins was generated with their respective spectral counts (148 unique viral proteins). Welches T-test was completed to determine if the number of viral proteins identified between T_1R and T_1S were significantly different. The same analysis was also completed on T_2R and T_2S.

MetaGOmics Biological Enrichment Analysis. To determine if categories of proteins were enriched in the resilient vs. susceptible coral cohorts at the two timepoints, a biological enrichment strategy that analyzes Gene Ontology (GO) categorical terms was used to compare sets of detected proteins[17]. Top results are reported with a cutoff E-value <1E-10. A fasta file of all *M. capitata* protein sequences confidently identified in these experiments (Supplementary Data 8) was analyzed with MetaGOmics v.0.1.1. MetaGOmics searched the truncated proteome against the UniProtKB Swiss-Prot non-redundant protein database (downloaded 10.15.2018) using BLASTp and assigned taxonomic and Gene Ontology (GO) terms for each protein sequence, and subsequently, to each peptide within the protein sequence. To determine biological enrichments of GO terms between experiments, peptides identified from each biological replicate from an experimental cohort were identified using percolator with an FDR cutoff <0.01[18]. Peptide spectral count values for each peptide that passed this threshold were averaged for the different experimental treatments (T_1S, T_1R, T_2S, T_2R). MetaGOmics generates laplace-corrected p-values to determine significance for log fold change calculations on the abundances of different GO terms. Here, we report Laplace corrected log fold change (base 2) results of all GO terms $\geq$0.5 and report a Laplace corrected Bonferroni corrected *p*-value significance <0.01 (Supplementary Data 3a, b). MetaGOmics provides taxonomic-lowest common ancestor analyses and GO-based biological enrichment analysis of functions using Uniprot identifications. Although MetaGOmics was designed to analyze microbiomes, the use of the software was modified to work with a single organism by ignoring the taxonomic enrichment analysis to instead examine functions that are significantly enriched or depleted in pairwise comparisons of coral cohorts.

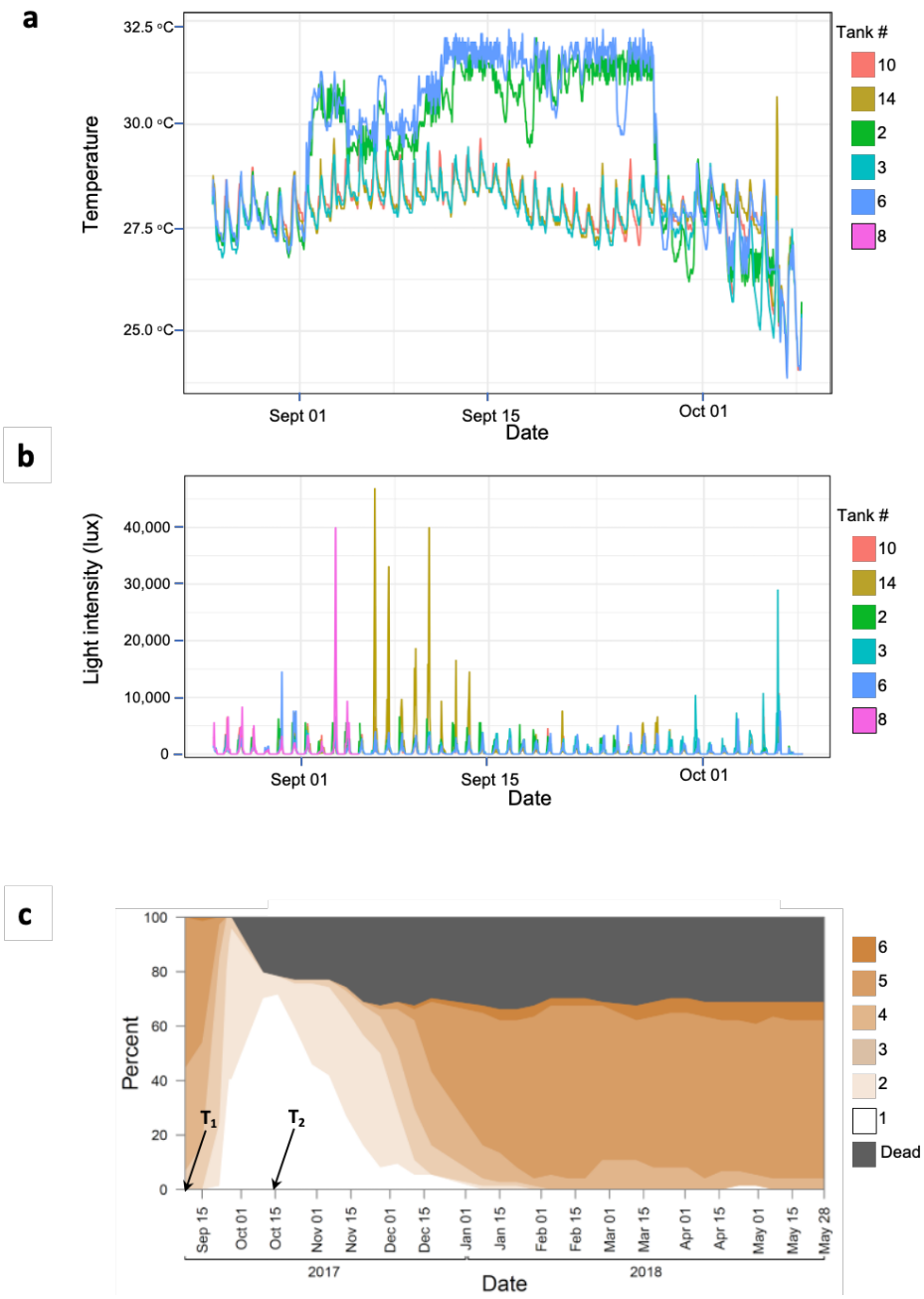

Fig. S1. a) Recorded temperature in experimental tanks and b) Recorded light intensity in experimental tanks (tank 10: coral pink, tank 14: mustard yellow, tank 2: green, tank 3: teal, tank 6: periwinkle, tank 8: pink). c) Bleaching assessments were completed on all corals weekly using the Coral Watch Card, where 1 indicates bleached, but not dead and 6 indicates the highest symbiont density (darker brown color). T1 and T2 indicate the periods when samples were collected for this study.

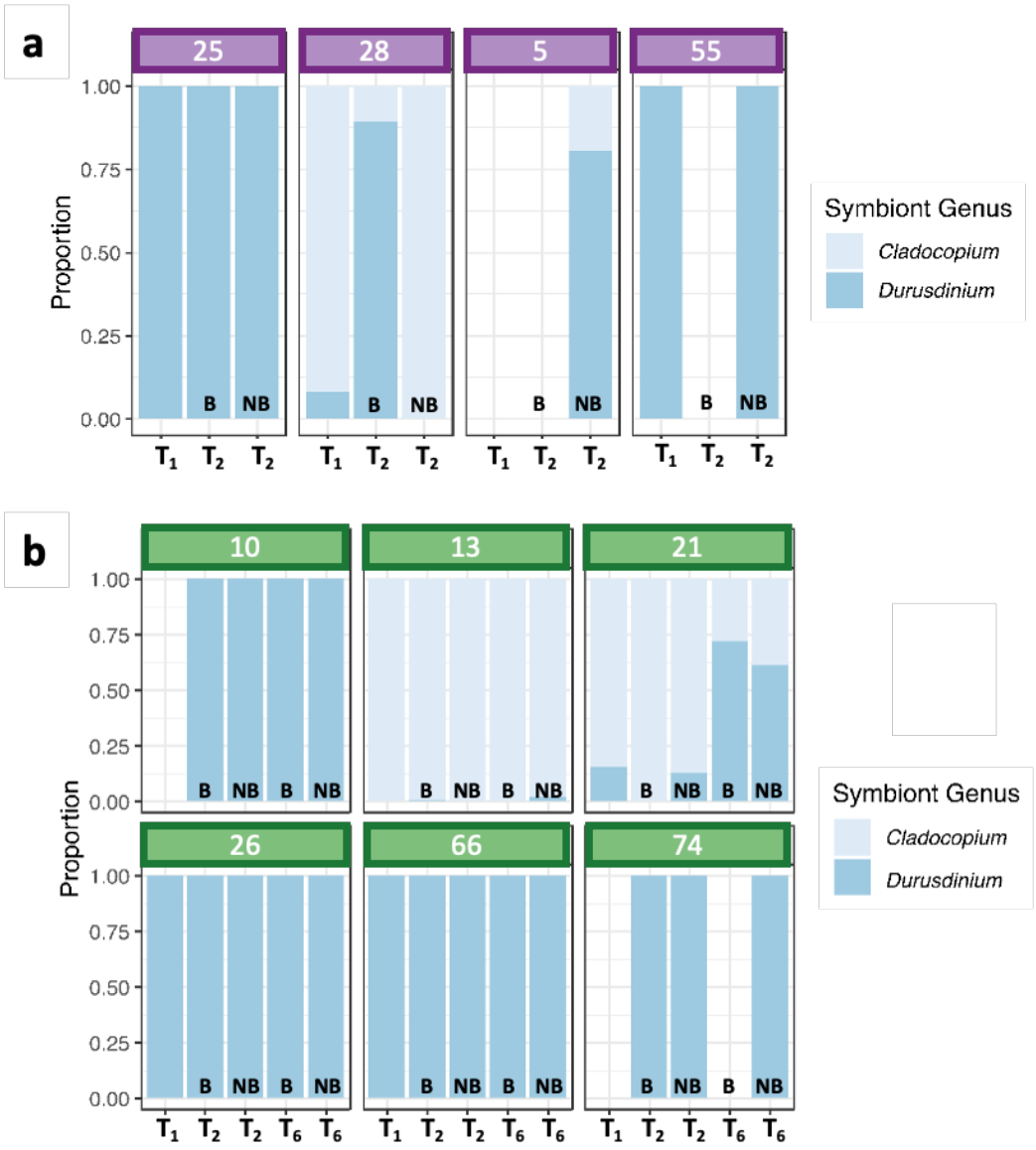

Fig. S2. Proportion of *Cladocopium* (light blue) and *Durusdinium* (dark blue) symbionts detected in a) susceptible corals (purple; *n=4*) and b) resilient corals (green; *n=6*) at time point 1 (pre-bleaching) and timepoint 2 (post bleaching, NB- nonbleached cohort, B- thermally bleached cohort). Data from time point 6 is also included when available to demonstrate consistency of symbiont community composition over time.

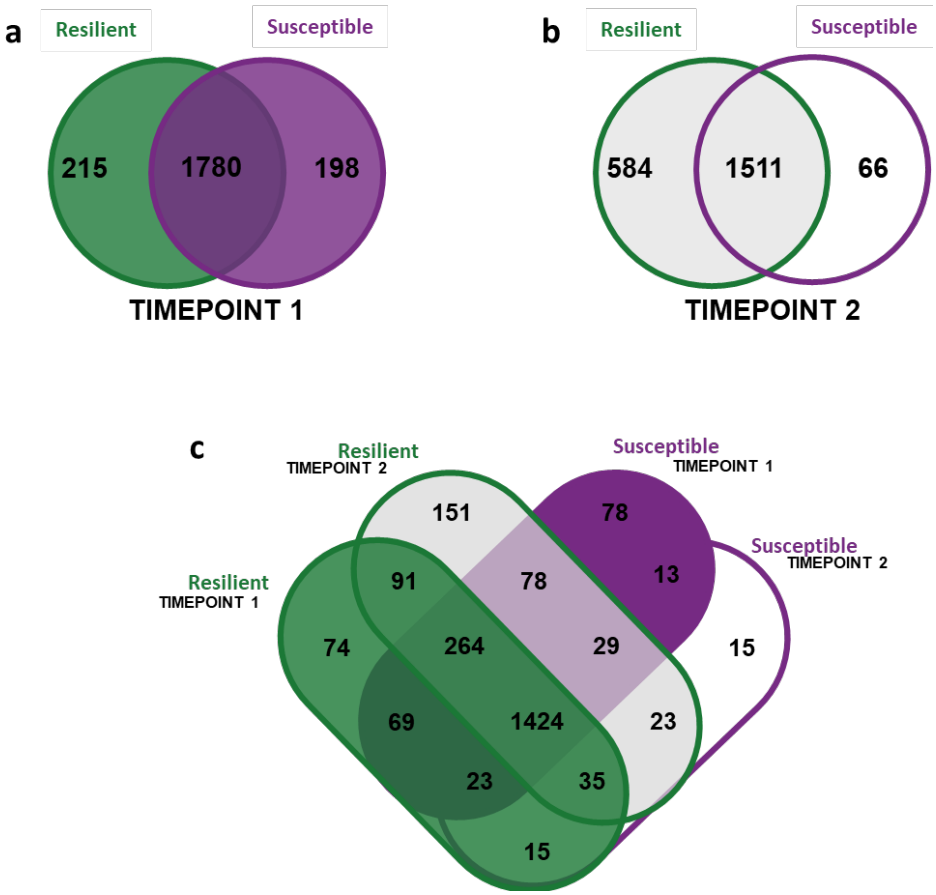

Fig. S3. Venn Diagram of proteins confidently Identified in resilient (*n=6*) and susceptible coral colonies (*n=6*) investigated at A. timepoint 1 before thermally bleached, B. timepoint 2 after thermal bleaching, and C. the overlap of all 4 treatments and timepoints.

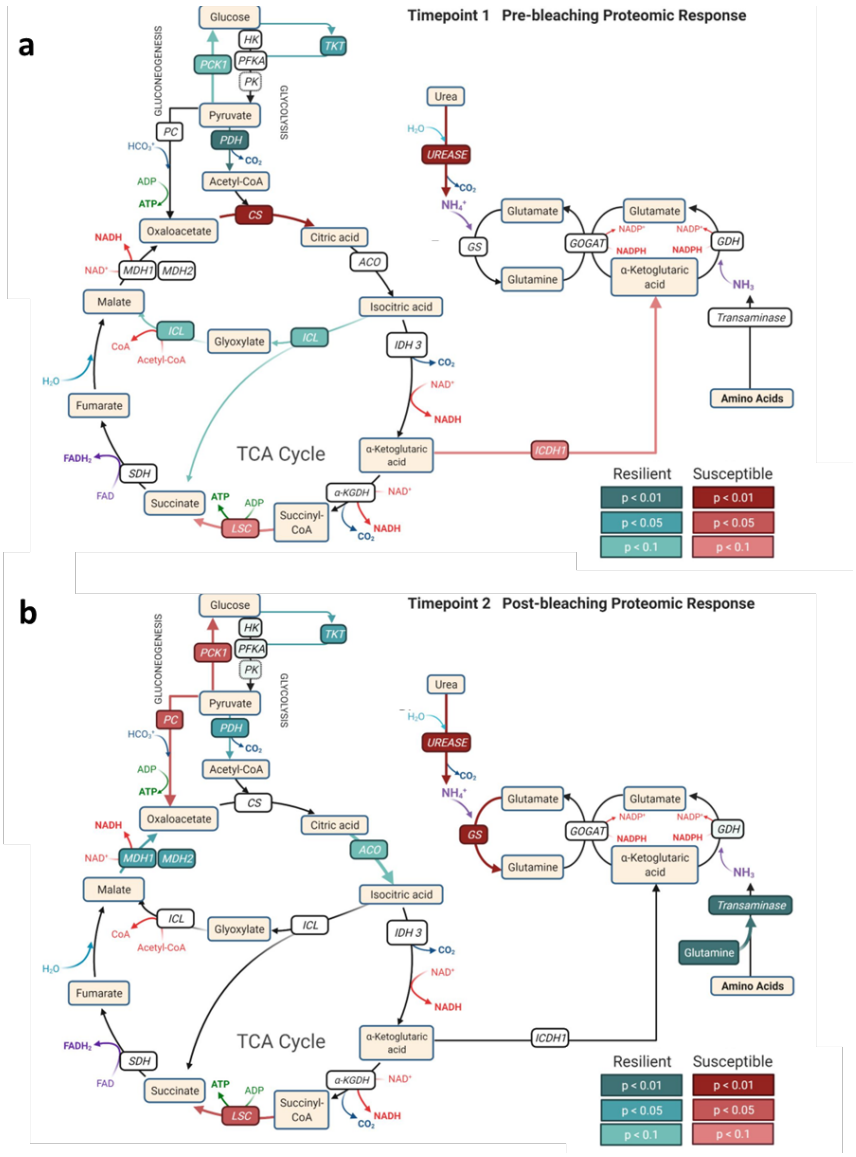

Fig. S4. Illustration of the proteins identified to be significantly increased in abundance in resilient (greens) or susceptible (reds) coral cohorts involved in the interconnected biochemical pathways of glycolysis/gluconeogenesis, the TCA cycle, urea degradation, and the glutamine synthase/glutamine oxoglutarate aminotransferase (GS/GOGAT) pathway at a. timepoint 1 before thermally bleached and b. timepoint 2 after thermal bleaching. Image was made using BioRender.

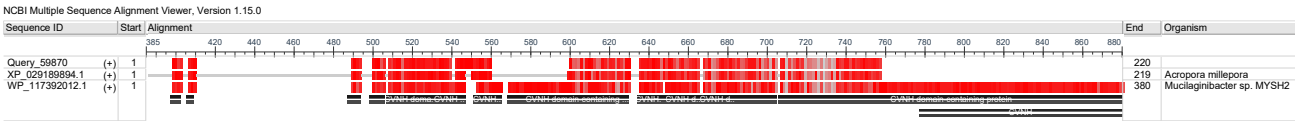

Fig. S5. Graphical illustration of a multiple sequence alignments of protein lcl|c238733_g2_i2|m.24867, noted to be significantly more abundant in the resilient colonies. The top query sequence in red (Query_59870) represents the input sequence mentioned. The next two sequences are from *Acropora millepora* and *Mucilaginbacter sp.* MYSH2 (both in red). Multiple alignment results revealed four highly conserved CyanoVirin-N domains (CVNH: 95% Query coverage; e-value 2e-30) depicted in grey.

Supplemental Figure 6

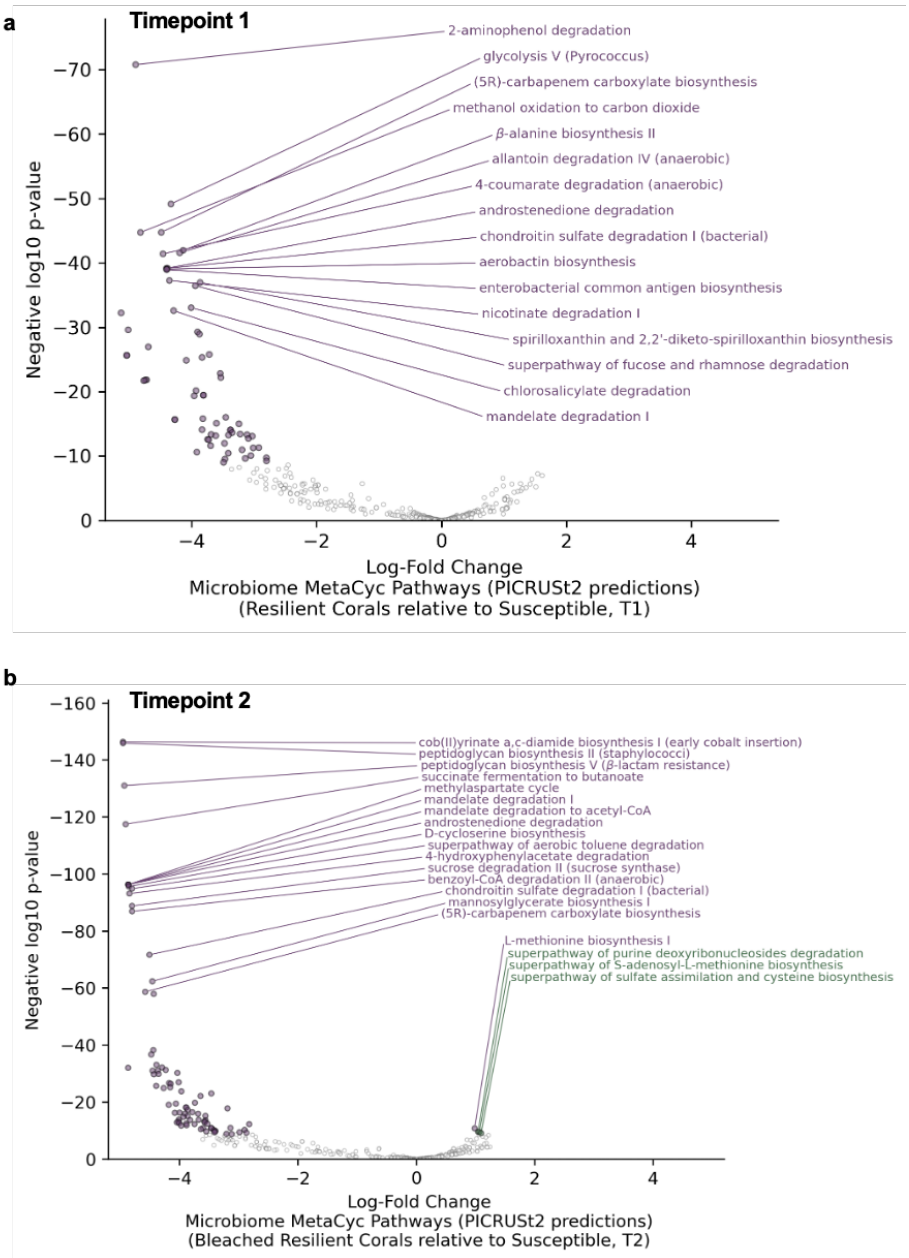

Fig. S6. Differentially abundant MetaCyc pathways in microbiomes of resilient corals relative to susceptible corals at a) Timepoint 1 and b) Timepoint 2-post bleaching. Volcano plot displaying predicted MetaCyc pathways from microbiome functional profiles generated using PICRUSt2. The x-axis represents log-fold changes in predicted pathway abundance, with positive values indicating higher abundance in resilient corals and negative values indicating higher abundance in susceptible corals. The y-axis shows the negative log10-transformed p-values from ANCOM-BC analysis, with more significant differentials appearing higher on the plot.

Table S1: Symbiont community composition classification of corals in this study.

Coral Genet ID	Susceptibility	Timepoint	Proportion *Cladocopium*	Proportion *Durusdinium*	Community Type
5	S	T2-NB	0.196	0.804	Mixed
20	S	T2-NB	0.142	0.858	Mixed
25	S	T1	0	1	*Durusdinium*-dominated
28	S	T1	0.081	0.919	*Cladocopium*-dominated
55	S	T1	0	1	*Durusdinium*-dominated
73	S				NO DATA
10	R	T2-NB	0	1	*Durusdinium*-dominated
13	R	T1	1	0	*Cladocopium*-dominated
21	R	T1	0.843	0.157	Mixed
26	R	T1	0	1	*Durusdinium*-dominated
66	R	T1	0	1	*Durusdinium*-dominated
74	R	T2-NB	0	1	*Durusdinium*-dominated

Table S2. Resulting table from MaAsLin2 analysis.[27].

TAXONOMY	FAMILY	METADATA	VALUE	COEF	STDERR	N	N.NOT.0	PVAL	QVAL
D_0__Bacteria.D_1__Proteobacteria.D_2__Gammaproteobacteria.D_3__Pseudomonadales.D_4__Moraxellaceae	Moraxellaceae	TimePoint	TimePoint	-0.156	0.035	36	6	0.0001	0.005
D_0__Bacteria.D_1__Proteobacteria.D_2__Alphaproteobacteria.D_3__Caulobacterales.D_4__Caulobacteraceae	Caulobacteraceae	Resilience	S	-0.411	0.089	36	6	0.0006	0.012
D_0__Bacteria.D_1__Proteobacteria.D_2__Gammaproteobacteria.D_3__Pseudomonadales.D_4__Moraxellaceae	Moraxellaceae	Resilience	S	-0.233	0.062	36	6	0.0007	0.012
D_0__Bacteria.D_1__Proteobacteria.D_2__Alphaproteobacteria.D_3__Sphingomonadales.D_4__Sphingomonadaceae	Sphingomonadaceae	Resilience	S	-0.181	0.062	36	7	0.0120	0.144
D_0__Bacteria.D_1__Firmicutes.D_2__Bacilli.D_3__Bacillales.D_4__Bacillaceae	Bacillaceae	Resilience	S	-0.134	0.054	36	6	0.0194	0.186

Table S3. Table of coral viral proteomes selected, location the database was found, and the number of proteins downloaded.

DESCRIPTION	DATABASE LOCATION	IDENTIFIER	# OF PROTEINS DOWNLOADED
Cafeteria roenbergensis virus	NCBI	NC_014637.1	1099
Paramecium bursaria Chlorella virus	NCBI	NC_000852.5	1911
Lymphocystis disease virus	NCBI	NC_005902.1	1488
Acanthamoeba polyphaga mimivirus	UNIPROT	NC_014649.1	8727
Megavirus Chiliensis	UNIPROT	NC_016072.1	4514

Supplemental References

1 Cunning, R. & Baker, A. C. Excess algal symbionts increase the susceptibility of reef corals to bleaching. *Nat Clim Change* **3**, 259-262 (2013).

2 Matz, M. V. Bayesian analysis of qRT-PCR data using MCMC. qpcr package: a tutorial. (2013).

3 Dilworth, J., Caruso, C., Kahkejian, V. A., Baker, A. C. & Drury, C. Host genotype and stable differences in algal symbiont communities explain patterns of thermal stress response of Montipora capitata following thermal pre-exposure and across multiple bleaching events. *Coral Reefs* **40**, 151-163 (2021).

4 Pino, L. K. *et al.* The Skyline ecosystem: Informatics for quantitative mass spectrometry proteomics. *Mass Spectrom Rev*, doi:10.1002/mas.21540 (2017).

5 Frazier, M., Helmkampf, M., Bellinger, M. R., Geib, S. M. & Takabayashi, M. De novo metatranscriptome assembly and coral gene expression profile of Montipora capitata with growth anomaly. *BMC genomics* **18**, 1-11 (2017).

6 Haas, B. J. *et al.* De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc* **8**, 1494-1512, doi:10.1038/nprot.2013.084 (2013).

7 Eng, J. K., Jahan, T. A. & Hoopmann, M. R. Comet: An open-source MS/MS sequence database search tool. *Proteomics* **13**, 22-24 (2013).

8 Keller, A., Eng, J., Zhang, N., Li, X. J. & Aebersold, R. A uniform proteomics MS/MS analysis platform utilizing open XML file formats. *Molecular systems biology* **1**, 2005 0017, doi:10.1038/msb4100024 (2005).

9 Fermin, D., Basrur, V., Yocum, A. K. & Nesvizhskii, A. I. Abacus: a computational tool for extracting and pre-processing spectral count data for label-free quantitative proteomic analysis. *Proteomics* **11**, 1340-1345, doi:10.1002/pmic.201000650 (2011).

10 Choi, H., Kim, S., Fermin, D., Tsou, C. C. & Nesvizhskii, A. I. QPROT: Statistical method for testing differential expression using protein-level intensity data in label-free quantitative proteomics. *J Proteomics* **129**, 121-126, doi:10.1016/j.jprot.2015.07.036 (2015).

11 The ggplot package. r-project (2007).

12 Package 'pheatmap'. R package (2015).

13 Kanehisa, M., Sato, Y. & Morishima, K. BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. *J Mol Biol* **428**, 726-731, doi:10.1016/j.jmb.2015.11.006 (2016).

14 UniProt, C. UniProt: a hub for protein information. *Nucleic Acids Res* **43**, D204-212, doi:10.1093/nar/gku989 (2015).

15 Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using DIAMOND. *Nat Methods* **12**, 59-60, doi:10.1038/nmeth.3176 (2015).

16 Messyasz, A. *et al.* Coral bleaching phenotypes associated with differential abundances of nucleocytoplasmic large DNA viruses. *Frontiers in Marine Science*, 789 (2020).

17 Riffle, M. *et al.* MetaGOmics: A Web-Based Tool for Peptide-Centric Functional and Taxonomic Analysis of Metaproteomics Data. *Proteomes* **6**, doi:10.3390/proteomes6010002 (2017).

18 Kall, L., Canterbury, J. D., Weston, J., Noble, W. S. & MacCoss, M. J. Semi-supervised learning for peptide identification from shotgun proteomics datasets. *Nat Methods* **4**, 923-925, doi:10.1038/nmeth1113 (2007).

19 Caporaso, J. G. *et al.* Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proceedings of the national academy of sciences* **108**, 4516-4522 (2011).
20 Bolyen, E. *et al.* Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol* **37**, 852-857, doi:10.1038/s41587-019-0209-9 (2019).
21 Callahan, B. J. *et al.* DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods* **13**, 581-583, doi:10.1038/nmeth.3869 (2016).
22 Brown, T., Sonett, D., Zaneveld, J. R. & Padilla-Gamiño, J. L. Characterization of the microbiome and immune response in corals with chronic Montipora white syndrome. *Mol Ecol* **30**, 2591-2606 (2021).
23 Quast, C. *et al.* The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* **41**, D590-596, doi:10.1093/nar/gks1219 (2013).
24 Yilmaz, P. *et al.* The SILVA and "All-species Living Tree Project (LTP)" taxonomic frameworks. *Nucleic Acids Res* **42**, D643-648, doi:10.1093/nar/gkt1209 (2014).
25 Weiss, S. *et al.* Normalization and microbial differential abundance strategies depend upon data characteristics. *Microbiome* **5**, 27, doi:10.1186/s40168-017-0237-y (2017).
26 Anderson, M. J. A new method for non-parametric multivariate analysis of variance. *Austral ecology* **26**, 32-46 (2001).
27 Mallick, H. *et al.* Multivariable association discovery in population-scale meta-omics studies. *PLoS Comput Biol* **17**, e1009442, doi:10.1371/journal.pcbi.1009442 (2021).