Supplementary materials

Online supplemental methods (OSM)

OSM1. The experiment. The *Orbicella faveolata* elevated temperature study was described previously [1-2]. In brief, 20 colonies each from "Little Conch" (24.9465N, 80.50205W; offshore), "The Rocks" (24.95375N, -80.54806W; inshore), and "Cheeca Rocks" (24.89742N, 80.61573W; inshore) were tagged and genotyped in July 2017 (**Table 1**; discussed in Manzello et al. [3]). A subset of 5-6 were cored into 10-20 fragments with a Nemo V2 electric waterproof drill (USA) drill featuring a 4.5-cm drill bit. Half of the 360 4.5-cm diameter fragments were used in an inshore-offshore reciprocal transplant [4], while the remaining 180 were allowed to recover in the field for two days prior to transport in Yeti coolers (USA) filled with seawater to the University of Miami's Rosenstiel School of Marine and Atmospheric Sciences (RSMAS). Once at RSMAS' "Experimental Reef Laboratory," the cores were distributed randomly across four seawater tables, where they were allowed to recover for seven days (July 20-27, 2017). Fragments were randomly assigned to either 1) short-term (5-day) control temperature (30ºC), 2) long-term (31-day) control temperature (30ºC), 3) short-term (5-day) very high temperature (33ºC), or 4) long-term (31-day) high-temperature (32ºC) exposure (n=3 tanks/treatment x time) such that a spectrum of bleaching responses would hypothetically be elicited: healthy controls, high-temperature-acclimating samples, sub-lethally stressed samples, or actively bleaching fragments. These fragment health designations were properties of the individual pucks/fragments (see main text for definitions.); a second parameter, "colony health designation," was instead a property of the original colony from which fragments were made and could be "bleaching-resistant" (failure to pale at either high-temperature treatment) or "bleaching-susceptible" (paling at either of the two high-temperature treatments). The control temperature of 30ºC represented the ambient in July 2017 (see https://www.pmel.noaa.gov/co2/pco2data/cheeca/alldata/coral_cheeca-all_xco2_pres-xco2seadryair-xco2airdryair-ph-sss-sst-chl-ntu-sc_o2-sc_o2_mgl-sc_o2_umolkg-sigmatheta.dat for real-time temperature data for Cheeca Rocks.).

OSM2. Temperature data. As the mean monthly maximum at the field sites is approximately 31ºC (August), NOAA's Coral Reef Watch's algorithms would predict that corals of these reefs would bleach after 4-8 weeks of exposure to 31ºC+1ºC=32ºC (i.e., 4-8 degree-heating weeks). However, Gintert et al. [5] found that, by assuming thermal stress to only accumulate at temperatures >32ºC, bleaching likelihood and severity is underestimated since temperatures are rising by the year; instead, the temperature above which corals begin to become thermally stressed *in situ* is closer to 31.3ºC [5]. The Coral Reef Watch models, then, do not accurately predict timing of onset of bleaching, nor bleaching severity, at any of the three field sites. The "very high" (33ºC; "V" in the multivariate plots), high (32ºC; "H"), and control (30ºC; "C") treatments in the laboratory experiment consequently correspond to degree-heating weeks of 1.2 (1.7ºC x 0.7 weeks), 3.1 (0.7ºC x 4.4 weeks), and 0, respectively; degree-heating days of 8.5, 21.7, and 0, respectively; and degree-heating hours of 204, 521, and 0, respectively. As a comparison, the colonies from which fragments were made were exposed to 2.4 degree-heating weeks, 16.8 degree-heating days, and 396 degree-heating hours *in situ* over the period in which the experiment was conducted (August 2017). Please note that the latter does not simply equate to number of degree-heating days x 24 since temperature was logged every three hours *in situ*; on

many days, the temperature only rose above 31.3ºC ephemerally. In these cases, a full degree-heating day would be logged, while the number of degree-heating hours might be as low as three. Bleaching occurred in the laboratory and *in situ* at around 400 degree-heating hours in 2017. Similarly, the field test colonies monitored in 2019 bleached between 2 and 5 degree-heating weeks [6].

OSM3. Protein extraction-details. Details of the protein extraction can be found in the main text. It is worth mentioning that the liquid nitrogen+TRIzol dual-homogenization method was critical for ensuring that the durable walls of the Symbiodiniaceae cells *in hospite* were effectively ruptured. The use of gentler extraction procedures are adequate for isolating anthozoan host proteins but not those of their dinoflagellate endosymbionts (i.e., resulting in an inaccurate, overly high host:endosymbiont protein ratio). As discussed in the main text, the host:endosymbiont protein ratio using this method (~1.2:1 to 1.5:1) was actually lower than the biomass ratio (~2:1), signifying that Symbiodiniaceae proteins were effectively extracted (potentially even at the expense of host coral proteins).

OSM4. Fasta database details. The queried fasta files were derived from an RNA-Seq analysis of over 70 *O. faveolata* transcriptomes, including those 21 experimental samples (**Table 1**) from which proteins were extracted. In other words, both RNAs and proteins were extracted from the same sample, with the RNA analyzed by RNA-Seq [1] and the proteins analyzed herein (an additional 50 samples from the same experiment of Aguilar et al. [1] were analyzed by RNA-Seq & *not* by proteomics.). From these same 70 samples, a Symbiodiniaceae dinoflagellate fasta mRNA sequence database was also assembled, and this was used for the querying of the same RAW files from the mass spectrometer. Each of the RAW files was queried twice, once against the host transcriptome and once against the endosymbiont one. It is worth noting that, although genomic sequences (& even genomes in the case of the dinoflagellates) are available for these species, I opted to instead query sequences derived directly from the study organisms given that genomic sequences were obtained from different host coral genotypes. Because the 6-7 coral samples in each iTRAQ batch represented a mix of host genotypes, I did not query host transcriptomes in a genotype-by-genotype fashion (e.g., querying all proteins against the skyblue genotype, then querying all against the lightyellow genotype). If, on the other hand, all corals had been of the exact same genotype, I would have queried the contig sequence set for that genotype alone. Having queried mass spectrometry-derived spectral data from a mix of genotypes against a composite transcriptome database likely signifies that the software only considered peptides whose peptide sequences were identical across the 7-8 samples in each batch (so as not to consider sequence bias in labeling efficiency, for instance). This issue, which would also affect the endosymbiont analysis (since, like the host corals, a mix of lineages were present across samples [2].), could explain why the total number of proteins sequenced was lower than expected (see **Results & Discussion** for details.).

It should be mentioned that, unlike for transcriptomic and genetic analyses, in which contigs/reads can be confidently mapped to individual Symbiodiniaceae lineages/genera, tryptic peptides are typically too short (6-10 amino acids) to do so with confidence [7]. Indeed, in many cases sequenced peptides could not be confidently ascribed even to host or endosymbiont. For these reasons, and because *O. faveolata* hosts a large diversity of dinoflagellates in the Florida Keys [3], I did not attempt to resolve sequences into exact Symbiodiniaceae lineages. In preliminary analyses, both host and dinoflagellate endosymbiont fasta libraries were actually queried simultaneously with the same RAW file, though it was found that only a small number of

proteins (dozens) passed the false discovery rate threshold discussed above; this is because, unlike BLAST, mass spectrometry algorithms do not use exact protein sequences per se, but instead peaks that are used to infer amino acid molecular weights. For highly conserved genes/proteins, the software is unlikely to assign large numbers of peptides to the correct compartment of origin with statistical confidence ($p<0.01$ based on peptide score q-values [corrected against decoy databases]). When each compartment's transcriptome/genome is queried in isolation, however, a greater proportion of proteins (typically hundreds) can be confidently assigned. However, because I was concerned with assigning sequenced peptides to the incorrect compartment of origin, I enacted the aforementioned rule that two peptides (each >6 amino acids) mapped to the same protein. It is possible that increasing the mapping stringency even further (e.g., 3-4 peptides/reference protein) could ultimately lead to an elucidation of the exact Symbiodiniaceae species from which the protein emerged, though this would result in a far reduced number of proteins (<10).

OSM5. Statistical analysis-details. Because entire coral fragments were sacrificed at each sampling time, a repeated-measures design was not employed. Instead, fragment was nested within genotype x temperature x time to ensure that intra-genotypic variability was accommodated in all statistical models outlined below and in the main text (both explanatory & predictive). In the cases when the fragment[genotype x temperature x time] term was not statistically significant, I assumed that a fragment from one genotype sacrificed after five days of laboratory treatment exposure was equivalent to an unsacrificed clonemate in the same treatment, an important assumption to be made with respect to the predictive modeling approaches outlined below and in the main text (which adhered to an "identical twin" design). This topic is discussed in detail in the main text.

OSM6. Explanatory statistics-univariate. Of the >37,000 peptides sequenced across the three normalizer and 20 experimental coral samples (1 of the 21 samples failed.), the vast majority (>99%) were filtered out (discussed in the **Results & Discussion**). The remaining 46 and 40 host coral and dinoflagellate proteins, respectively, that passed all quality control criteria were $\log_2$-transformed and used in a variety of both univariate and multivariate explanatory statistical analyses (all with the exception of permutational multivariate ANOVA [PERMANOVA] & PERMDISP [Primer, UK] were carried out with JMP® Pro 16, USA.). First, JMP's response screening platform was used to search for proteins whose concentrations were affected by temperature (30, 32, or 33ºC [df=2], as well as simply control vs. high [32+33ºC]), time (5 vs. 31 days), genotype (n=6; see **Table 1**.), shelf (inshore vs. offshore), site (The Rocks, Cheeca Rocks, & Little Conch), colony health designation, and fragment health designation at time of sampling. Briefly, response screening computes a false discovery rate-adjusted p-value to limit the possibility of generating type I statistical errors when performing numerous comparisons (>600 in this case). Proteins whose FDRlogworth values were >2 (equivalent to a false discovery rate-adjusted p-value of 0.01) were considered to represent differentially concentrated proteins. I considered a second group of "proteins of interest," which I defined as those that were included in any of the stepwise discriminant analysis models discussed below. In most cases, the differentially concentrated proteins were those most likely to feature in the stepwise discriminant analysis models; for instance, if a protein was differentially affected by temperature, it was almost certainly in the stepwise discriminant analysis model for temperature (& vice versa). In contrast, the more complex models for fragment health designation and colony health designation

outlined below did not necessarily feature the most differentially concentrated proteins, a paradox familiar to many data modelers.

OSM7. Explanatory statistics-multivariate. When dealing with large numbers of molecules, such as 'Omics datasets (in which the number of cellular targets is always greater than the number of samples), many inferential multivariate approaches cannot be used, namely multivariate ANOVA. However, exploratory-based multivariate approaches, such as principal components analysis and multi-dimensional scaling, can depict multivariate differences among samples in a semi-quantitative manner. Therefore, a principal components analysis on correlations (log_2-transformed data) was performed for the 46 host coral proteins alone, the 40 endosymbiont proteins alone, and all 86 holobiont proteins (both compartments). Since iTRAQ data are presented as ratios (to the normalizer sample found in each batch), and all were log_2-transformed prior to statistical analyses, principal components analysis and multi-dimensional scaling depict very similar information, with the latter more directly representing similarity. However, since principal component analysis biplots were not built with standardized data (i.e., normalized to account for the fact that some proteins were maintained at higher levels than others), protein concentrations were standardized prior to multi-dimensional scaling to reduce the influence of highly concentrated proteins. As such, the principal components analysis and multi-dimensional scaling biplots, while similar, depict slightly different information (**Figure 1**) since highly concentrated proteins were not down-weighted in the former.

Both principal components analysis and multi-dimensional scaling plots were created for the day-5 experiment only (30 vs. 33ºC), the 31-day experiment only (30 vs. 32ºC), and as pooled across experiments for all 86 proteins, as well as the 27 differentially concentrated proteins (including the 18 proteins of interest). Unlike multivariate ANOVA, similarity-based PERMANOVA *can* be used to uncover multivariate mean differences when the number of proteins (n=86) is larger than the number of samples (n=20), and it was used to test the effects of the same experimental factors as above with a Bray-Curtis similarity matrix calculated from standardized values of the 86 proteins as the model Y. An alpha level of 0.05 was set for this analysis. As a similar approach to PERMANOVA, a Euclidean distance-based similarity matrix was constructed from the 20 samples for the host, endosymbiont, and holobiont proteins, and the coordinates (see **Table 4**.) were instead used as the models' Y's. Since the number of dimensions (6-7, depending on suite of proteins included) was less than the number of samples, MANOVA of the effects of the various experimental parameters could then be undertaken; this approach is known as non-parametric MANOVA (NP-MANOVA). These results are not shown directly in **Table 4** and served namely to corroborate the findings of the more robust PERMANOVA and complement those of the partial least squares-based correlation loading plots (**Figures 3a** & **S5**).

Hierarchical clustering was used with both the 46- and 40- host coral and endosymbiont protein datasets, respectively, and a modified WGCNA-like analysis was performed by calculating eigenproteins from the protein contribution scores for each cluster and regressing the cluster means against the experimental parameters discussed above; only relationships with $R^2>0.4$ (typically $p<0.01$) were considered, and I hypothesized that the hierarchical clustering and stepwise discriminant analysis approaches discussed below would yield similar groupings of proteins that could be used to develop models with the predictive power to significantly partition corals by experimental parameter of interest. For this reason, these results are predominantly discussed only in the online supplemental results (**OSR**).

OSM8. Proteomic predictive modeling-overview. There were two proteomic predictive modeling goals. In the first, I sought to simply build a model capable of distinguishing corals of the 3-4 fragment health designations by using random mixes of training and validation (i.e., "holdback") samples across the 20 experimental samples. Although these models are of little use in conservation since an actively bleaching sample can be observed by eye by SCUBA divers (i.e., no need to validate with invasive sampling approaches featuring expensive 'Omics technologies), it was important to ensure that there was sufficient proteomic variation among healthy controls, high-temperature-acclimating samples, sub-lethally stressed samples, and actively bleaching corals to develop more sophisticated predictive models that could be later used with samples collected *in situ* for which no prior knowledge on bleaching susceptibility exists. This would be achieved instead by developing a predictive model for the colony health designation (see definitions above & in the main text.).

The models were constructed with colony or fragment health designation as the Y and the proteins as the predictors, and at least one sample from each of the three colony or fragment health designation levels, respectively, was held back to serve as a validation sample. It is important to note that, in the colony health designation analysis, a fragment from a bleaching-susceptible coral could actually be a healthy control, high-temperature-acclimating, sub-lethally stressed, or actively bleaching fragment; in other words, this model sought to uncover entrained proteomic differences of the colony in the presence of experimental effects. Are there proteomic signatures that manifest in fragments from bleaching-resistant colonies, for instance, regardless of which temperature they are exposed to, or, instead, are the proteomic responses more reflective of the temperature treatment itself? Prior work has shown that, although certain proteins are undoubtedly involved in the coral stress response, most canonical eukaryotic stress response proteins are maintained at high levels at all times in corals given their residence in substantially marginalized environments across the globe [2]. This means that a bleaching-resistant or bleaching-susceptible-indicative proteomic signature *could* be unveiled even in samples exposed to an array of experimental conditions (& with varying phenotypes). Models whose fragment and colony health designation misclassification rates for the validation samples were 0% (i.e., 100% accuracy) are discussed in the main text and are of likely utility to managers looking to screen for resilient colonies at their study reefs.

OSM9. Predictive modeling details. As discussed in the main text, JMP® Pro 16's "model screening" platform was run with the 86 proteins as predictors (X) and the two coral health parameters as Y: fragment health designation (the actual status of the analyzed biopsy/fragment; healthy control, high-temperature-acclimating [pooled with healthy controls for analyses listed as "FHD(3)"], sub-lethally stressed, or actively bleaching) and colony health designation (the ultimate resilience of the colony as a whole [assessed over time]; bleaching-resistant vs. bleaching-susceptible). In the initial analysis, the following modeling types were tested with 15 training and 5 validation samples (ensuring that at least one of each of the three fragment and colony health designations was represented in the validation set): bootstrap forest, discriminant analysis, generalized multivariate regression (gen-reg; using a variety of different algorithms; see **Table S1.**), k-nearest neighbors, naïve Bayes, neural networks, partial least squares, stepwise discriminant analysis (run manually as described below), stepwise regression, and support vector machines. It should be noted that, because there were more analytes (86) than samples (20), the more commonly employed ordinary least squares could not be run. The Python code for all models has been made publicly available as part of this article's **online**

supplemental data file. In certain instances, "test" data were used to further decrease chances of over-fitting.

For each model, the following fit parameters were calculated: root mean square error, Akaike's information criterion (adjusted to the total number of model parameters, i.e., AICc), training R^2 (how well the model fit the training data), validation R^2 (how well the model fit the validation or test data), and model percent misclassified (i.e., the misclassification rate; how often the prediction was wrong, i.e., the inverse of accuracy). Among these parameters, I prioritized the latter; all else being equal, then, I sought the model that had the highest chance of correctly determining the bleaching status (fragment health designation) or susceptibility (colony health designation) of a coral. When two models were characterized by the same misclassification rates, the difference between the training and validation root mean square errors was the second criterion for model ranking; all else being equal, the model whose validation root mean square error was only slightly higher than that of the respective training error was prioritized.

OSM10. Neural networks. Because the artificial intelligence-(AI) based neural networks were the only modeling type capable of predicting coral bleaching susceptibility at high accuracy for the fragment and colony health designations (see **Results** for details.), ~20,000 neural network models were built in which the following input parameters were optimized (**Table S2**): number of hidden layers (1 vs. 2; i.e., no "deep" learning), numbers of each of three activation types (TanH [sigmoidal], linear [identity], &/or Gaussian [radial]; each ranging from 0 to 4), number of "boosts" (0, 1, 2, or 20), learning rate when using boosted models (0, 0.1, 0.2, or 0.5), number of tours (1, 10, 20, 50, or 100), type of penalty (absolute, squared, or weight decay), and covariate transformation (transformed vs. untransformed). That said, there was a preference for the weight decay penalty method, all else being equal, since I sought to avoid inclusion of redundantly informing predictors.

A partially factorial design was employed since not every combination of neural network input parameters was permitted. For instance, the boosting method cannot be performed with more than one hidden layer, nor is it appropriate to have a learning rate in a model that has not been boosted. Across these parameters, 16,400 and 4,000 models were built for fragment and colony health designation, respectively, across 2-4 validation approaches (**Table S2**): 15/5 training/validation (both fragment & colony health designation), 13/4/3 training/validation/test samples set *a priori* (fragment health designation only), holdback (30% for fragment health designation & 25% for colony health designation), and "exclude 1/bin" (i.e., one random sample excluded from each fragment and colony health designation). For the holdback validation approach, fragment and colony health designation samples were selected randomly (in contrast with the validation column approach, in which at least one sample from each health designation was included). For certain neural networks, kfold (5) validation was instead used due to the stochastic nature of the platform (in which running the same model twice can lead to different solutions); kfold validation increased the likelihood that the re-run model was accurate in at least one of the five folds (saving a random seed being the alternative).

OSM11. Stepwise discriminant analysis. All modeling types tested above considered all 86 proteins. This means that, were a model with high predictive capacity uncovered, it would be necessary to measure the concentrations of all 86 proteins in future field samples. It might be desirable, however, to undertake a more simplistic, bottom-up, biomarker-based approach for assessing coral bleaching likelihood; why spend $250 to sequence 1,000 proteins when the concentration of a single protein demonstrates similar predictive capacity? For this reason, a

candidate biomarker-based modeling approach known as stepwise discriminant analysis was undertaken in which JMP Pro 16 calculated the minimum number of proteins required to build a model that can correctly classify the data into the experimental treatment (e.g., temperature, time, etc.) of interest with a certain degree of confidence. I programmed JMP's stepwise discriminant analysis to create models that would predict with >85% confidence which coral sample was from which temperature, time, genotype, shelf, and fragment health designation using the minimum number of proteins. For instance, if the concentration data from two proteins could allow me to know which coral was from which temperature treatment at 80% confidence, a third protein would be added to the model provided that the model derived from all three proteins had a predictive confidence >85%. The confidence was determined by burning in random proteins, and an AI-based algorithm then simulated this random sampling several million times until the most parsimonious models were developed (i.e., the fewest number of proteins with the highest level of confidence). The AI then guessed which sample was from which treatment, and the guess was back-compared to the actual sample identity in a double-blind manner to determine correctness. Models featuring over 5-10 proteins were generally not considered.

Online supplemental results (OSR)

OSR1. Effect of Symbiodiniaceae assemblage on coral bleaching susceptibility. It is worth noting that the Symbiodiniaceae assemblages, which were inferred from mRNA contig mapping to published genomes (rather than from protein mapping), varied across host coral genotypes (see [2] & the **online supplemental data file**.), and the assemblage actually changed in response to experimental treatment for two colonies: C5 and D5; however, bleaching-resistant colonies were no more likely to host the hypothetically more thermotolerant *Durusdinium* spp. (X^2 p>0.05), and some bleaching-resistant corals were even *Breviolum* spp.-dominated (**online supplemental data file**). Similarly, some bleaching-susceptible colonies hosted primarily *Durusdinium* spp.-dominated endosymbiont communities [2]. As discussed in the shotgun proteomic analysis [2], however, mass spectrometry-derived tryptic peptides are too short (6-10 amino acids) to map confidently to one particular Symbiodiniaceae lineage (i.e., *Breviolum* vs. *Durusdinium*; see detailed treatise on this topic above.); as such, I did not consider Symbiodiniaceae assemblage as a predictor in the models discussed in the main text.

OSR2. Clustering. Hierarchical clustering generated four protein clusters for each compartment (parenthetical correlation data have been provided only when WGCNA mean eigenprotein correlation values [R^2] were >0.40 at p<0.05). For the coral host, clusters 1 (n=29 proteins), 2 (n=14), 3 (n=1), and 4 (n=2) tended to partition coral samples by fragment health designation, shelf, sampling time, and temperature, respectively, whereas endosymbiont clusters 1 (n=28 proteins), 3 (n=3 proteins), and 4 (n=2 proteins) tended to separate samples by the fragment health designation, temperature x time (R^2=0.50, p<0.01), and temperature x time (R^2=0.50, p=0.01; i.e., only 2 proteins could explain half of the variation due to temperature & time), respectively; endosymbiont cluster 2 (n=7 proteins) featured proteins unresponsive to treatment.

Although these findings are interesting for data visualization and for reducing dataset complexity, I ultimately found them inferior to stepwise discriminant analysis and neural networks given my interest in identifying small groupings of proteins that could be used to build predictive models for coral behavior. I did, however, use the clustering data to generate

eigenprotein scores that could then be regressed against experimental parameters in a manner analogous to the popular WGCNA script that attempts to quantify relationships between clusters of genes and physiological and/or experimental factors. There were significant relationships with sampling day for clusters 9 (R^2=0.47) and 4 (R^2=0.39). Cluster 9 also had the most significant influence on temperature x time (p<0.01; R^2=0.51).

OSR3. Methodological comparison: shotgun proteomics vs. iTRAQ. Of the 784 shotgun-sequenced proteins from another study with samples from the same experiment [2] that passed all quality control, only 4 were 1) sequenced herein and 2) passed all quality control imposed herein; all were of host coral origin. The first two were OFAVBQ_DN217768_c1_g1_i5 (cytoplasmic actin; top hit: XP_020600465.1 [e=0 against *O. faveolata*]) and OFAVBQ_DN222214_c0_g2_i2 (centrosome-associated protein; top hit: XP_020611767.1 [e=0 against *O. faveolata*]); both were presumed to be high-abundance housekeeping proteins in the shotgun analysis (i.e., unaffected by experimental treatment, genotype, reef site of origin, etc.), and this was confirmed with iTRAQ. Although the third and fourth common proteins, OFAVBQ_DN223482_c3_g2_i3 (myosin-10; top hit: XP_020621196.1 [e=0 against *Pocillopora damicornis*]) and DELTA-actitoxin-Ate1a-like (top hit: XP_029210457.1 [$e<10^{-84}$ against *Acropora millepora*]), were maintained at marginally higher levels in corals from Cheeca Rocks (p=0.03 & p=0.011, respectively) based on iTRAQ data, both missed the false discovery rate-adjusted differentially concentrated protein cutoff; shotgun proteomics found their concentrations to be even less variable across samples (both proteins were found in 100% of samples.).

Shotgun proteomics yielded a larger number of proteins that passed all quality control criteria yet iTRAQ is nevertheless attractive in that the concentration data from the smaller number of proteins sequenced that passed all stringent quality control criteria (n=86) are truly quantitative (vs. strictly presence-absence). It is worth noting though, that, had there been less of an interest in correctly assigning the compartment of origin (host vs. dinoflagellate endosymbiont) to the sequenced proteins, a larger number would have been maintained in the dataset (hundreds). On the other hand, a multi-thousand protein dataset is currently unrealizable using iTRAQ given the low labeling efficiency [7], which normally goes unreported since it requires advanced knowledge of MS bioinformatics software (e.g., Proteome Discoverer); poor labeling efficiency is a rampant issue that surely plagues the competing technology, Thermo-Fisher Scientific's "tandem mass tags," as well, and must be confronted in the near future. No protein was found to be differentially concentrated across treatments by both shotgun proteomics and iTRAQ, though as evidenced above, this is likely due to the fact that the two methods yielded very different sets of proteins (with minimal overlap).

OSR4. Stepwise discriminant analysis-based coral biomarker profiling. Despite the fact that only 11 (9 unique) proteins differed significantly in concentration across all experimental parameters tested (**Table 5**), a larger number was identified in the stepwise discriminant analysis models: 17 and 10 for the coral host (**Figure S1**) and Symbiodiniaceae dinoflagellates (**Figure S2**), respectively. Of the 17 host proteins of interest, 3 were also deemed to be differentially concentrated by the response screening analysis (**Table 5**). When considering all 17 (**Figure 2**), 3 could not be identified (18%). Of the remaining 14, the dominant functional categories were transcription (n=3; 18%), reproduction (n=2; 12%), cell cycle (n=2; 12%), and nuclear processes (n=2; 12%); this differentially concentrated protein assemblage differs from that of a prior shotgun proteomics analysis [2], in which lipid trafficking was the most environmentally-sensitive host coral cellular pathway (X^2 p<0.01).

Of the nine endosymbiont proteins of interest (**Figure S2**), five were also found to be differentially concentrated by the response screen (**Table 5** & **Figure S3d**); the sixth and final differentially concentrated protein was not useful in stepwise discriminant analysis model building, and four proteins of interest were not differentially concentrated. When considering the 10 endosymbiont differentially concentrated proteins and proteins of interest, only half could be identified with confidence, and these proteins were involved in 1) protein processing and trafficking (n=3), translation (n=1), and metabolism (n=1); in contrast, proteins involved in photosynthesis were more likely to be differentially concentrated across treatments in the shotgun proteomic analysis [2]. Please note that, while the training model misclassification rates are low (5-15%), their validation accuracy was much lower (**Table S1**). Although the candidate biomarker approach based on discriminant analysis may, then, be adept at modeling laboratory responses, it is clearly prone to overfitting; I ultimately found this approach, then, to be better suited for describing laboratory-derived findings than for developing robust predictive models.

Online supplemental references

1. Aguilar, C., Enochs, I., Manzello, D.P., Mayfield, A.B., Kolodziej, G., Carlton, R. et al. Transcriptome profiling of thermotolerant corals of the Upper Florida Keys. *Mol Ecol* **under review**.
2. Mayfield, A.B., Aguilar, C., Enochs, I., Kolodziej, G., Manzello, D.P. Shotgun proteomics of thermally challenged Caribbean reef corals. *Front Mar Sci* **2021**, *8*, 660153.
3. Manzello, D.P., Matz, M.V., Enochs, I.C., Valentino, L., Carlton, R.D., Kolodziej, G., et al. Role of host genetics and heat-tolerant algal symbionts in sustaining populations of the endangered coral *Orbicella faveolata* in the Florida Keys with ocean warming. *Glob Change Biol* **2019**, *25*, 1016-1031.
4. Aguilar, C., Enochs, I., Manzello, D.P., Mayfield, A.B., Kolodziej, G., Carlton, R. et al. Field modulation of Caribbean reef coral thermotolerance. **in prep**.
5. Gintert, B.E., Manzello, D.P., Enochs, I.C., Kolodziej, G., Carlton, R.D., Gleason, A.C.R., et al. Marked annual coral bleaching resilience of an inshore patch reef in the Florida Keys: A nugget of hope, aberrance, or last man standing? *Coral Reefs* **2018**, *37(2)*, 533-547.
6. Mayfield, A.B. The use of neural networks to predict coral fate in the Florida Keys. **in prep**.
7. Mayfield, A.B. Proteomic signature of corals from thermodynamic reefs. *Microorganisms* **2020**, *8(8)*, 1171.

Supplemental figures

Figure S1. Host coral stepwise discriminant analysis. Centroids represent 95% confidence, and the control, high-temperature, and very high temperature samples are denoted by C, H, and V, respectively. Sample icons are colored by their 2b-RAD genotype name (a; e.g., "lightyellow" samples colored light yellow) except for genotypes black(b) and black(c), which are colored red and green, respectively. The numbers on the biplot rays correspond to the accession numbers found in the bottom-left corners of each panel; please see **Table 5** for identities of the "proteins of interest" that were also differentially concentrated or the **online supplemental data file** for identities of the proteins of interest that were *not* also differentially concentrated proteins. The model classification rates were calculated from laboratory training samples only.

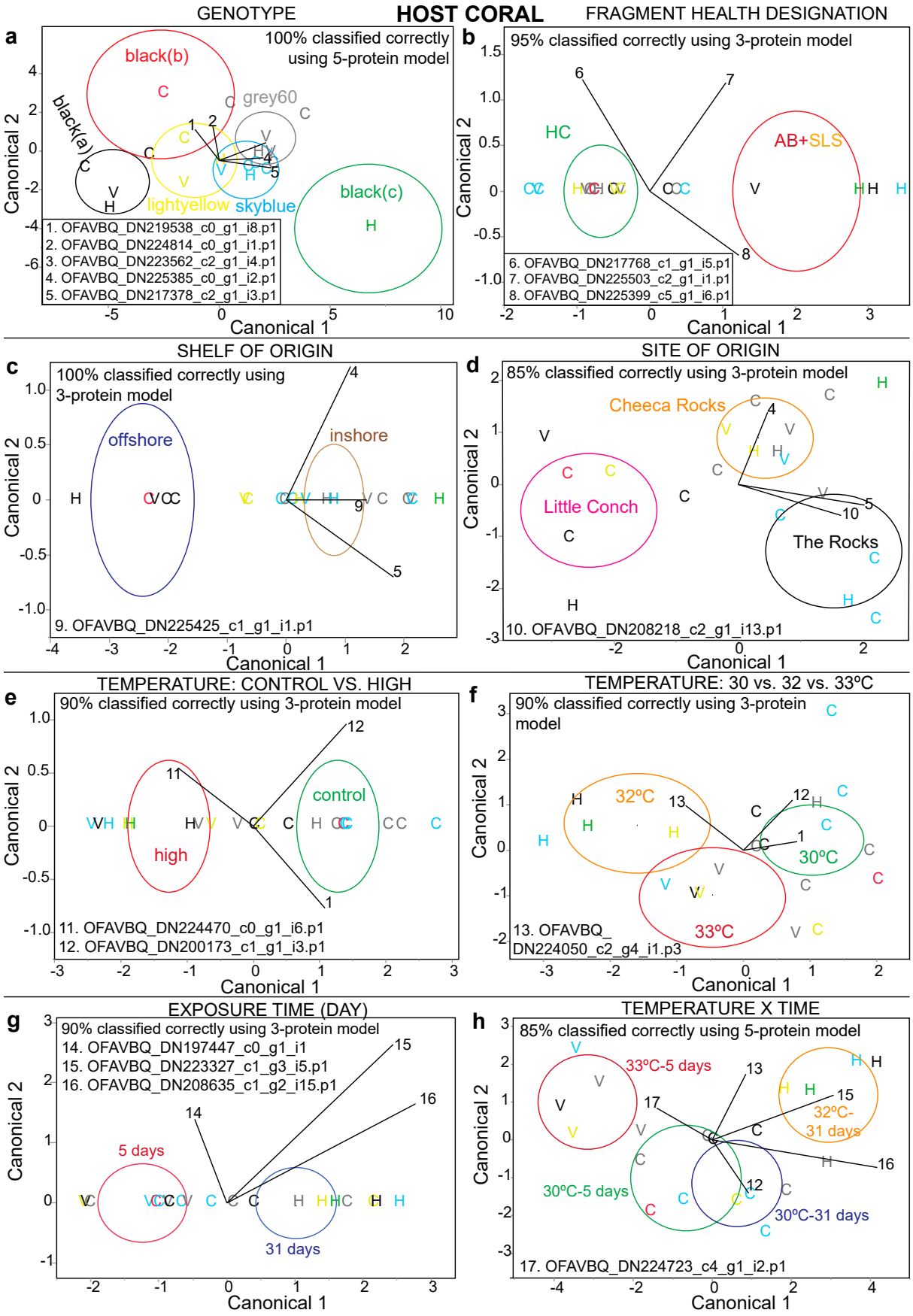
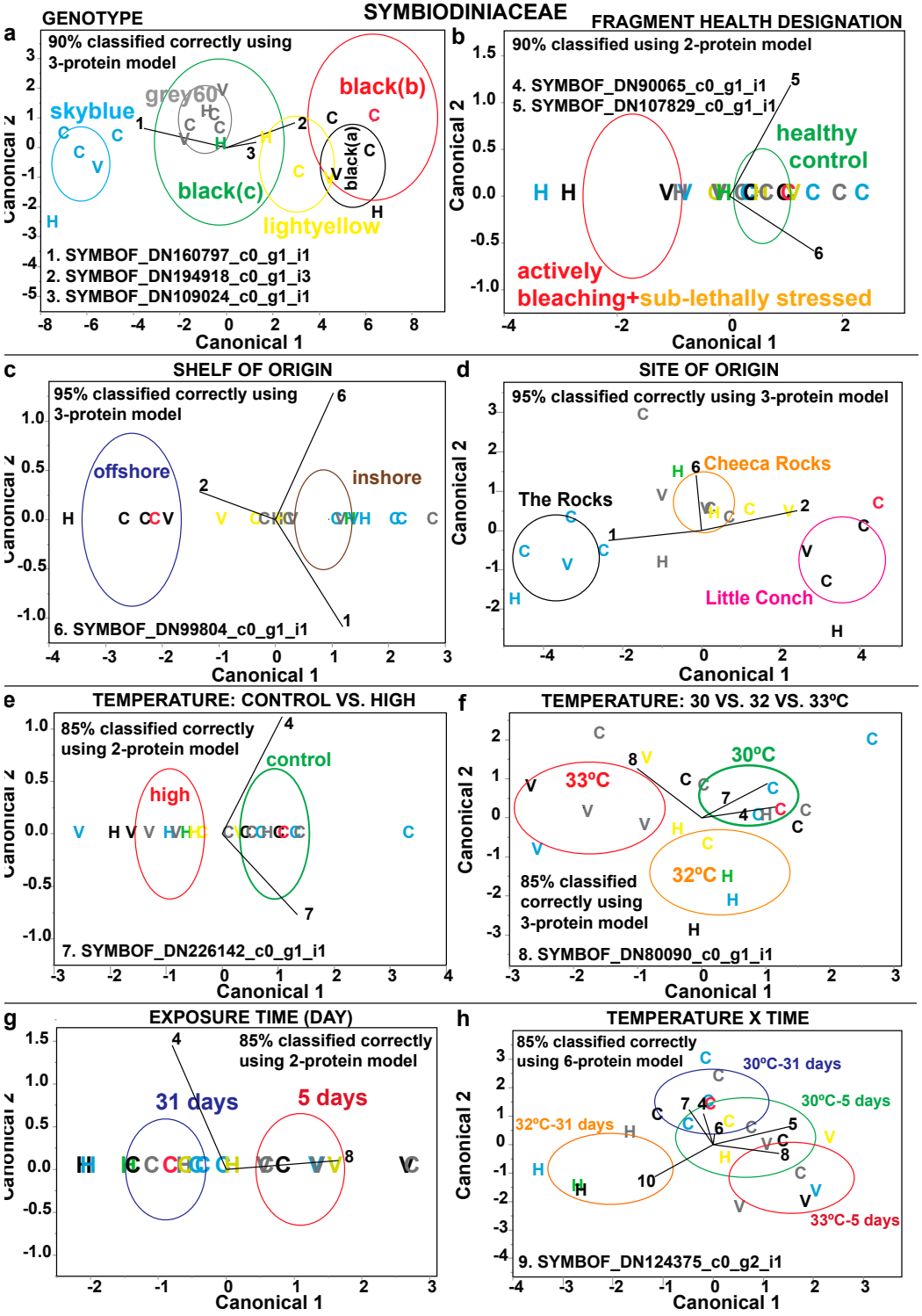

HOST CORAL
GENOTYPE
FRAGMENT HEALTH DESIGNATION
a
100% classified correctly using 5-protein model
black(b)
black(a)
grey60
lightyellow
skyblue
black(c)
1. OFAVBQ_DN219538_c0_g1_i8.p1
2. OFAVBQ_DN224814_c0_g1_i1.p1
3. OFAVBQ_DN223562_c2_g1_i4.p1
4. OFAVBQ_DN225385_c0_g1_i2.p1
5. OFAVBQ_DN217378_c2_g1_i3.p1
Canonical 1
Canonical 2
b
95% classified correctly using 3-protein model
HC
AB+SLS
6. OFAVBQ_DN217768_c1_g1_i5.p1
7. OFAVBQ_DN225503_c2_g1_i1.p1
8. OFAVBQ_DN225399_c5_g1_i6.p1
Canonical 1
Canonical 2
SHELF OF ORIGIN
c
100% classified correctly using 3-protein model
offshore
inshore
9. OFAVBQ_DN225425_c1_g1_i1.p1
Canonical 1
Canonical 2
SITE OF ORIGIN
d
85% classified correctly using 3-protein model
Cheeca Rocks
Little Conch
The Rocks
10. OFAVBQ_DN208218_c2_g1_i13.p1
Canonical 1
Canonical 2
TEMPERATURE: CONTROL VS. HIGH
e
90% classified correctly using 3-protein model
control
high
11. OFAVBQ_DN224470_c0_g1_i6.p1
12. OFAVBQ_DN200173_c1_g1_i3.p1
Canonical 1
Canonical 2
TEMPERATURE: 30 vs. 32 vs. 33ºC
f
90% classified correctly using 3-protein model
32ºC
30ºC
33ºC
13. OFAVBQ_DN224050_c2_g4_i1.p3
Canonical 1
Canonical 2
EXPOSURE TIME (DAY)
g
90% classified correctly using 3-protein model
14. OFAVBQ_DN197447_c0_g1_i1
15. OFAVBQ_DN223327_c1_g3_i5.p1
16. OFAVBQ_DN208635_c1_g2_i15.p1
5 days
31 days
Canonical 1
Canonical 2
TEMPERATURE X TIME
h
85% classified correctly using 5-protein model
33ºC-5 days
32ºC-31 days
30ºC-5 days
30ºC-31 days
17. OFAVBQ_DN224723_c4_g1_i2.p1
Canonical 1
Canonical 2

Figure S2. Symbiodiniaceae stepwise discriminant analysis. Centroids represent 95% confidence, and the control, high-temperature, and very high temperature samples are denoted by C, H, and V, respectively. Sample icons are colored by their 2b-RAD genotype name (e.g., "grey60" samples colored grey) except for genotypes black(b) and black(c), which are colored red and green, respectively. The numbers on the biplot rays correspond to the accession numbers



found in the bottom-left corners of each panel; please see **Table 5** for identities of the "proteins of interest" that were also differentially concentrated or the **online supplemental data file** for identities of the proteins of interest that were *not* also differentially concentrated proteins. The model classification rates were calculated from laboratory training samples only.

Figure S3. Venn diagrams depicting relative influence of various environmental predictors on numbers of differentially concentrated proteins (DCPs) and "proteins of interest" (POIs). Please note that the three host coral (**a-b**) and six endosymbiont (**c-d**) response screening (RSA)-identified proteins represent the true DCPs, with the remaining 14 and 4 proteins, respectively, instead representing POIs that were useful in stepwise discriminant analysis (SDA) model building (but not necessarily statistically significant at the false discovery rate-governed *p*-value). Also note that "space" includes both reef site and shelf (inshore vs. offshore). The 9 DCPs and 18 POIs have also been shown as pooled across both compartments of the mutualism (**e-f**).

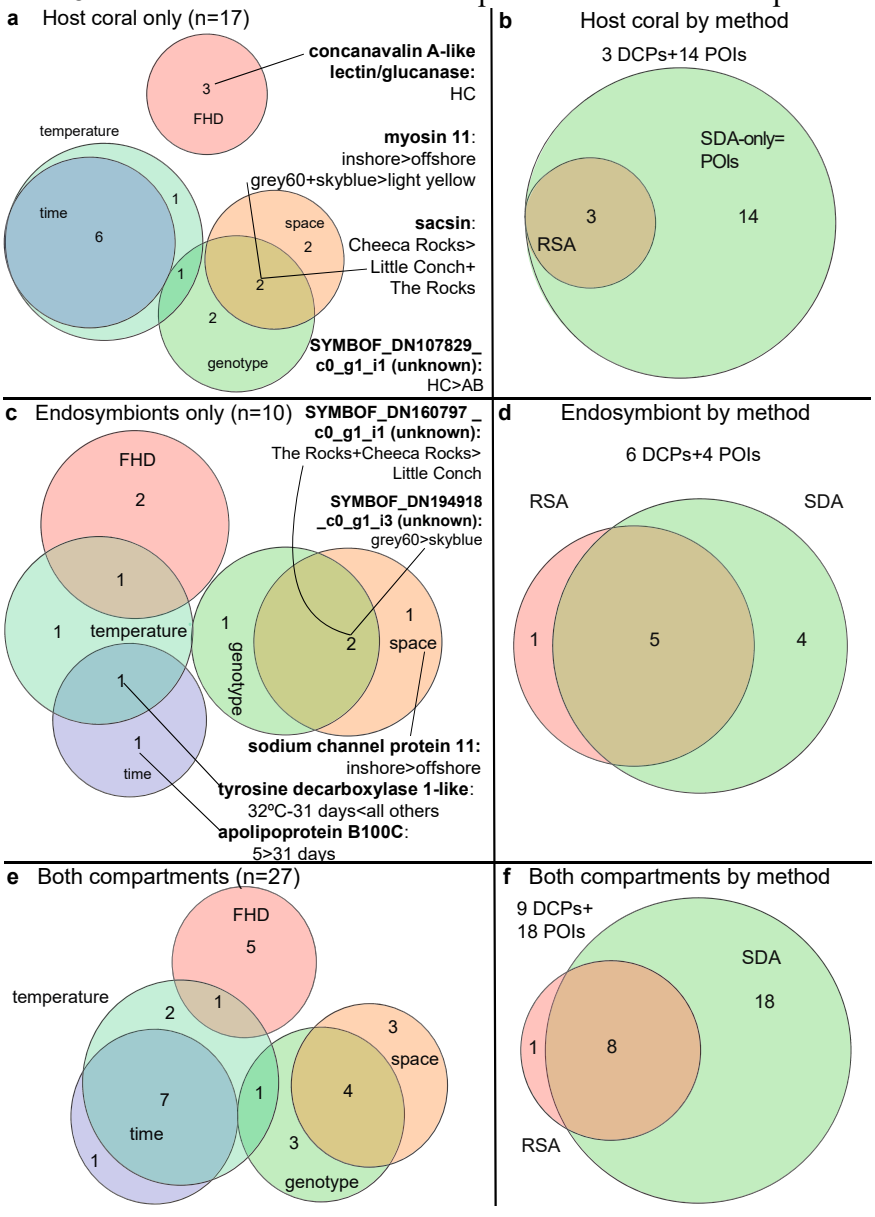


Figure S4. T^2 plot of multivariate variability for the healthy control (HC), sub-lethally stressed (SLS), and actively bleaching (AB) samples. The upper control/confidence limit (UCL) has been presented as a hatched blue line. Using the default settings of JMP Pro 16, the two AB samples would both be considered multivariate outliers.

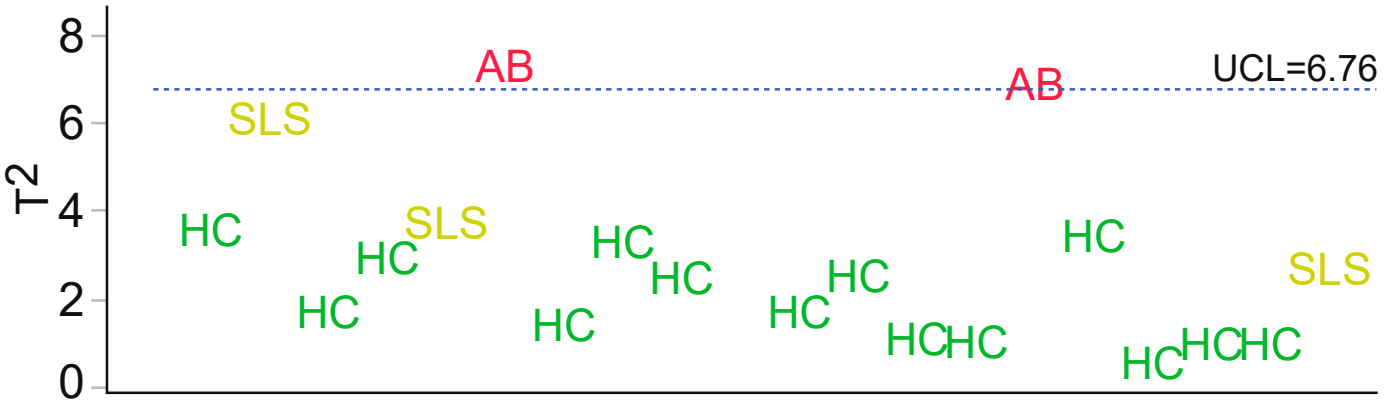
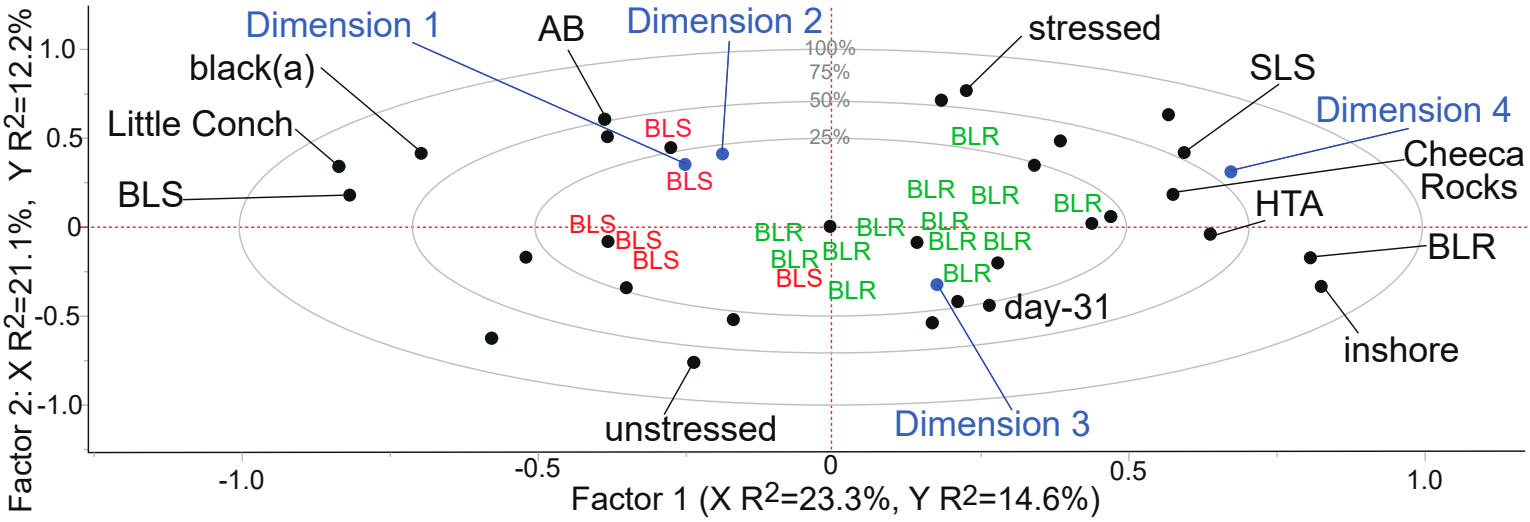


Figure S5. Partial least squares-based correlation loading plot of the 20-sample dataset. The concentration data from the 86 proteins were standardized and used to construct a four-dimensional multi-dimensional scaling plot; these coordinates were used as the model Y's to reduce dataset complexity (blue). Black circles denote model predictors, many of which having been labeled. The samples (n=19) are denoted by their colony health designation: bleaching-susceptible (BLS; red) vs. bleaching-resistant (BLR; green). One sample, D6-6, was excluded since its ultimate bleaching resilience could not be determined (too few fragments to fate-track over course of entire 31-day experiment). AB=actively bleaching. HTA=high-temperature-acclimating. SLS=sub-lethally stressed.



Online supplemental tables

Table S1. Additional statistical approaches with log_2-transformed protein concentrations. Only the approaches that yielded statistically significant results, identified differentially concentrated proteins, or correctly predicted coral bleaching susceptibility with 100% accuracy are shown in **Table 2**. Please note that the number of predictors corresponds to the number used in model screening; the ultimate number in the final model that passed all quality control criteria could feature fewer predictors (see **Table 6**.). BIC=Bayesian information criterion. CHD=colony health designation. EP=experimental parameter (e.g., temperature, time, genotype, etc.). FDR=false discovery rate. FHD=fragment health designation. MDS=multi-dimensional scaling. NA=not applicable. OSDF=online supplemental data file. RV=response variable (Y). tMPM=training model percent misclassified. vMPM=validation model percent misclassified.

Analytical goal Approach (abbreviation)	**RV (Y)/predictor (X)**	**Acceptance criterion**	**Primary finding(s)**	**Data location(s)**
Uncover multivariate treatment effects (explanatory)				
Permutational multivariate ANOVA	86 proteins/all EP	alpha=0.05	Effect of FHD on host proteome	**Table 4**
Non-parametric MANOVA	MDS coordinates/all EP	alpha=0.05		**Table 4**
Identify differentially concentrated proteins				
Response screening analysis	86 proteins/all EP	FDR-p<0.01	9 differentially concentrated proteins identified	**Table 5** & **Figure 2**
Stepwise discriminant analysis-manual	86 proteins/all EP	MPM<10%	18 "proteins of interest" identified	**Table 5** & **Figures S1-2**
Predict bleaching susceptibility				
Fragment health designation (FHD; 15/5 training/validation samples)			***Pass/fail quality control: reason***	
Bootstrap forest (100 trees)	FHD/86 proteins	MPM<10%	Failed: vMPM=40%	Data not shown
Discriminant analysis-automated	FHD/86 proteins	MPM<10%	Failed: vMPM=40%	Data not shown
Generalized multivariate regression (GMR)-lasso	FHD/86 proteins	MPM<10%	Failed: vMPM=40%	Data not shown
GMR-adaptive lasso	FHD/86 proteins	MPM<10%	Failed: vMPM=40%	Data not shown
GMR-elastic net	FHD/86 proteins	MPM<10%	Failed: vMPM=40%	Data not shown
GMR-adaptive elastic net	FHD/86 proteins	MPM<10%	Failed: vMPM=40%	Data not shown
GMR-pruned forward	FHD/86 proteins	MPM<10%	Failed: vMPM=40%	Data not shown
k-nearest neighbors	FHD/86 proteins	MPM<10%	Failed: vMPM=40%	Data not shown
Naïve Bayes	FHD/86 proteins	MPM<10%	Failed: vMPM=60%	Data not shown
Neural network	FHD/86 proteins	MPM<10%	**Passed: vMPM=0%**	**Tables 2 & 6**
Nominal logistic	FHD/86 proteins	MPM<10%	Failed: vMPM=60%	Data not shown
Partial least squares (NIPALS)	FHD/86 proteins	MPM<10%	Failed: vMPM=40%	Data not shown
Stepwise discriminant analysis-manual	FHD/86 proteins	MPM<10%	Failed: vMPM=40%	**Figures S1b & S2b**

Analytical goal Approach (abbreviation)	RV (Y)/predictor (X)	Acceptance criterion	Primary finding(s)	Data location(s)
Stepwise regression	FHD/86 proteins	MPM<10%	Failed: vMPM=80%	Data not shown
Support vector machine	FHD/86 proteins	MPM<10%	Failed: vMPM=60%	Data not shown
XGBoost	FHD/86 proteins	MPM<10%	Failed: vMPM=60%	Data not shown
Colony health designation (CHD; 15/5 training/validation samples)				
Bootstrap forest (100 trees)	CHD/86 proteins	MPM<10%	Failed: vMPM=40%	Data not shown
Discriminant analysis-automated	CHD/86 proteins	MPM<10%	Failed: vMPM=40%	Data not shown
GMR-lasso	CHD/86 proteins	MPM<10%	**Passed: vMPM=0%**	**Tables 2 & 6**
GMR-adaptive lasso	CHD/86 proteins	MPM<10%	Failed: vMPM=40%	Data not shown
GMR-elastic net	CHD/86 proteins	MPM<10%	Failed: MPM=40%	Data not shown
GMR-adaptive elastic net	CHD/86 proteins	MPM<10%	Failed: vMPM=40%	Data not shown
GMR-pruned forward	CHD/86 proteins	MPM<10%	**Passed: vMPM=0%**	**Tables 2 & 6**
k-nearest neighbors	CHD/86 proteins	MPM<10%	Failed: vMPM=40%	Data not shown
Naïve Bayes	CHD/86 proteins	MPM<10%	Failed: vMPM=60%	Data not shown
Nominal logistic	CHD/86 proteins	MPM<10%	Failed: vMPM=40%	Data not shown
Neural network	CHD/86 proteins	MPM<10%	**Passed: vMPM=0%**	**Tables 6 & S2-3**
Partial least squares (NIPALS)	CHD/86 proteins	MPM<10%	Failed: MPM=33%	Data not shown
Stepwise discriminant analysis-manual	CHD/86 proteins	MPM<10%	Failed: MPM=33%	Data not shown
Stepwise regression (minimum BIC-forward)	CHD/86 proteins	MPM<10%	Failed: MPM=67%	Data not shown
Support vector machine	CHD/86 proteins	MPM<10%	Failed: MPM=67%	Data not shown
XGBoost	CHD/86 proteins	MPM<10%	Failed: vMPM=60%	Data not shown

Table S2. Neural networks. The protein predictors (n=86) were log_2-transformed prior to screening. In the "validation type" column, the first, second, and, when present, third values reflect the number of training, validation, and test samples, respectively; when the 25 or 30% holdback approach was used to validate, samples were chosen randomly. For the fragment health designation (FHD) analyses, models were run in which either three ("FHD (3)") or four ("FHD (4)") outcomes were possible. The common three were healthy control, sub-lethally stressed, and actively bleaching. In the four-outcome analysis, high-temperature acclimating samples were also included (considered "health controls" in FHD (3)). When the "exclude 1/bin" validation method was used, one randomly selected sample from each group was excluded (three & four holdback samples for the three- & four-outcome analyses, respectively). CHD=colony health designation. HL=hidden layer. QC=quality control (validation &/or test model misclassification rate[s]=0).

Response	Validation type	# HL	#TanH nodes	#Linear nodes	#Gaussian nodes	#Boosts	Learning rate	#Tours	# models	#passed QC (%)
FHD (4)	15/5	1 or 2	0, 1, 2, 3, or 4			0, 1, 2, or 20	0, 0.1, 0.2, or 0.5	1, 10, 20, 50, or 100	2,100	257 (12.2%)

FHD (4)	holdback (30%)	1 or 2	0, 1, 2, 3, or 4	0, 1, 2, or 20	0, 0.1, 0.2, or 0.5	1, 10, 20, 50, or 100	2,100	278 (13.2%)
FHD (4)	exclude 1/ bin	1 or 2	0, 1, 2, 3, or 4	0, 1, 2, or 20	0, 0.1, 0.2, or 0.5	1, 10, 20, 50, or 100	2,000	34 (1.7%)
FHD (4)	13/4/3	1 or 2	0, 1, 2, 3, or 4	0, 1, 2, or 20	0, 0.1, 0.2, or 0.5	1, 10, 20, 50, or 100	2,000	12 (0.6%)
FHD (3)	15/5	1 or 2	0, 1, 2, 3, or 4	0, 1, 2, or 20	0, 0.1, 0.2, or 0.5	1, 10, 20, 50, or 100	2,100	613 (29.1%)
FHD (3)	holdback (30%)	1 or 2	0, 1, 2, 3, or 4	0, 1, 2, or 20	0, 0.1, 0.2, or 0.5	1, 10, 20, 50, or 100	2,100	271 (12.9%)
FHD (3)	exclude 1/ bin	1 or 2	0, 1, 2, 3, or 4	0, 1, 2, or 20	0, 0.1, 0.2, or 0.5	1, 10, 20, 50, or 100	1,000	154 (15.4%)
FHD (3)	13/4/3	1 or 2	0, 1, 2, 3, or 4	0, 1, 2, or 20	0, 0.1, 0.2, or 0.5	1, 10, 20, 50, or 100	3,000	44 (1.5%)
CHD	15/5	1 or 2	0, 1, 2, 3, or 4	0, 1, 2, or 20	0, 0.1, 0.2, or 0.5	1, 10, 20, 50, or 100	2,000	122 (6.1%)
CHD	holdback (25%)	1 or 2	0, 1, 2, 3, or 4	0, 1, 2, or 20	0, 0.1, 0.2, or 0.5	1, 10, 20, 50, or 100	2,000	480 (24%)

Table S3. Additional neural network models whose sample misclassification rates were 0%. A maximum of four models have been shown (see **Table S2** for totals.), all featured a weight decay penalty with untransformed covariates unless noted otherwise. FHD=fragment health designation: either three (healthy controls, sub-lethally stressed, & actively bleaching; "FHD (3)") or four (the former three plus high-temperature-acclimating; "FHD (4)"). On average, a model that was characterized by 100% accuracy possessed one hidden layer (HL), four each of TanH, linear, and Gaussian activation nodes, no boosts, absolute penalty method, untransformed covariates, and 100 tours. That said, only numbers of boosts and tours significantly influenced the misclassification rate (independent uniform input total effect sizes of 0.85 & 0.20, respectively), with the former negatively influencing accuracy; accuracy instead increased with increasing tours (no more than 100 were tested.). MPM=model percent misclassified. RMSE=root mean square error.

Model #	**#HL**	**# TanH activation nodes**	**# linear activation nodes**	**# Gaussian activation nodes**	**Total activation nodes**	**# Boosts (learning rate)**	**#tours**	**Training/validation/test MPM (RMSE)**
			1st &, when present, 2nd value=HL 1 & 2, respectively					
FHD (4): 15/5 training/validation (n=4/257)								
1	1			1	1	20 (0.1)	50	0/0% (0.00/0.00)
2[b]	2	1/4	2/4	4/3	18	0	50	0/0% (0.00/0.00)
3	1			3	3	20 (0.2)	50	0/0% (0.00/0.00)
4[a]	2	3/0	4/1	2/0	10	0	50	0/0% (0.00/0.00)
FHD (4): 30% random holdback (n=4/278)								
5[a]	2	1/1	2/3	4/2	13	0	50	0/0% (0.00/0.00)
6[b]	2	2/4	0/1	0/4	11	0	100	0/0% (0.00/0.00)
7[b]	2	2/1	4/3	4/0	14	0	100	0/0% (0.00/0.00)
8	1	4		3	7	20 (0.2)	100	0/0% (0.00/0.00)

Model #	#HL	# TanH activation nodes	# linear activation nodes	# Gaussian activation nodes	Total activation nodes	# Boosts (learning rate)	#tours	Training/validation/test MPM (RMSE)
FHD (4): exclude 1/bin (n=4/34)								
9	1			1	1	20 (0.1)	100	0/0% (0.00/0.00)
10	1			2	2	20 (0.2)	100	0/0% (0.00/0.00)
11[b]	1			2	2	20 (0.5)	100	0/0% (0.00/0.00)
12[b]	2	4/0	0/1	2/0	7	0	50	0/0% (0.01/0.02)
FHD (4): 13/4/3 training/validation/test (n=4/12)								
13[c]	2	1/1	2/3	2/0	9	0	50	0/0.25/0% (0.18/0.40/0.35)
14[b]	2	2/3	0/0	4/0	9	0	1	0.69/0.25/0% (0.73/0.64/0.42)
15[a,c]	1	4	0	1	5	2 (0.1)	100	0/0/0.33% (0.03/0.13/0.53)
16[a,c]	1	4	0	4	8	20 (0.5)	100	0/0/0.33% (0.00/0.03/0.58)
FHD (3): 15/5 training/validation (n=4/613)								
17[c]	2	3/1	4/3	4/3	18	0	20	0/0% (0.00/0.00)
18	2	4/2	0/1	3/0	10	0	100	0/0% (0.00/0.00)
19[b]	1	3	1	3	7	20 (0/5)	100	0/0% (0.00/0.00)
20[c]	2	2/3	3/2	0/3	13	0	100	0/0% (0.00/0.00)
FHD (3): 30% random holdback (n=4/271)								
21	2	2/4	0/1	3/3	13	0	50	0/0% (0.02/0.10)
22[a-b]	1	4	2	1	7	2 (0.5)	10	0/0% (0.01/0.03)
23[a-b]	2	4/3	1/2	4/4	18	0	50	0/0% (0.00/0.05)
24[a-b]	1	1	4	1	6	20 (0.5)	50	0/0% (0.00/0.05)
FHD (3): 13/4/3 training/validation/test (n=4/44)								
25[a]	2	4/0	1/1	0/2	8	0	20	0/0% (0.00/0.00)
26[a,c]	2	3/3	1/4	2/3	16	0	100	0/0% (0.00/0.00)
27[c]	1	2	4	1	7	20 (0.1)	20	0/0% (0.00/0.00)
28	2	4/3	3/2	1/4	17	0	50	0/0% (0.00/0.00)
CHD: 15/5 training/validation (n=4/122)								
29[a-b]	1	3	0	1	4	20 (0)	20	0/0% (0.18/0.00)
30[a]	2	4/4	4/3	4/4	23	0	50	0/0% (0.00/0.00)
31[b]	1	1	0	4	5	2 (0.5)	100	0/0% (0.04/0.00)

Model #	**#HL**	**# TanH activation nodes**	**# linear activation nodes**	**# Gaussian activation nodes**	**Total activation nodes**	**# Boosts (learning rate)**	**#tours**	**Training/validation/test MPM (RMSE)**
32[a]	2	0/0	2/2	4/2	10	0	100	0/0% (0.00/0.00)
CHD: 25% random holdback (n=4/480)								
33[a]	2	2/4	3/3	3/4	19	0	50	0/0% (0.00/0.00)
34[b]	2	1/1	2/1	1/4	10	0	100	0/0% (0.00/0.00)
35	1	1	2	2	5	0	20	0/0% (0.00/0.00)
36[b]	2	4/1	3/4	1/3	16	0	100	0/0% (0.00/0.00)
mode	2	4/0	0/0	1 & 4/0	8	0	100	

[a]Covariates transformed. [b]squared penalty method. [c]absolute penalty method.