

# **Assessing coral health in the Kingdom of Tonga with a coral health index**

Anderson B. Mayfield,<sup>1-2\*</sup> Alexandra C. Dempsey,<sup>3</sup> and Chii-Shiarng Chen<sup>4-7</sup>

<sup>1</sup>Cooperative Institute for Marine and Atmospheric Studies, University of Miami, Miami, FL 33149, USA

<sup>2</sup>Atlantic Oceanographic and Meteorological Laboratory, National Oceanic and Atmospheric Administration, Miami, FL 33149, USA

<sup>3</sup>Khaled bin Sultan Living Oceans Foundation, Annapolis, MD 21403, USA

<sup>4</sup>National Museum of Marine Biology and Aquarium, Checheng, Pingtung 944, Taiwan

<sup>5</sup>Taiwan Coral Research Center, Checheng, Pingtung 944, Taiwan

<sup>6</sup>Graduate Institute of Marine Biotechnology, National Dong Hwa University, Checheng, Pingtung 944, Taiwan

<sup>7</sup>Department of Marine Biotechnology and Resources, National Sun Yat-Sen University, Kaohsiung 804, Taiwan

\*Corresponding author. email: [andersonblairmayfield@gmail.com](mailto:andersonblairmayfield@gmail.com) or [abm64@miami.edu](mailto:abm64@miami.edu)  
tel.: 01-337-501-1976

For publication in *Platax*

## Abstract

Reef coral health is currently diagnosed retroactively; once corals bleach or become diseased, we assume they had been experiencing high stress levels. This would be akin to telling one who suffered a cardiac arrest that he/she had high blood pressure; ideally, an individual's susceptibility to a heart attack would be known in advance of such a late-stage, life-threatening event. However, sub-lethal means of assessing coral health do not yet exist, and the preferred health metric, growth, cannot be reliably measured while on research cruises, where most, if not all, reef sites are surveyed and sampled only once. Since it would be preferable to make diagnostic inferences from a single biopsy, a new parameter known as the "coral health index" (CHI; 氣) was devised herein. The CHI, which represents an amalgamation of several response variables known to scale directly with coral resilience (e.g., dinoflagellate endosymbiont density), was profiled across multiple environmental gradients in the Kingdom of Tonga. Machine learning (i.e., AI) models were made such that the CHI could be predicted from more commonly assessed environmental (e.g., salinity) and ecological (e.g., coral cover) benchmarks. Models for two pocilloporid coral species were characterized by validation  $R^2$  values approaching 1, meaning that this AI could be used to delineate relative levels of coral resilience on a pre-bleaching timescale.

**key words:** artificial intelligence, coral reefs, diagnostics, dinoflagellates, environmental stress, global climate change, machine learning, ocean health

## Introduction

In the face of the global climate change (GCC) crisis (Grottoli et al., 2021), marine biologists have missed the opportunity to characterize the physiology of “pristine” corals whose habitats have not become extensively marginalized by human impact (Cruz-Garcia et al., 2020); although this does not signify that no corals will acclimatize or even adapt to GCC scenarios (Enochs et al., 2020; Rubin et al., 2021; Demerlis et al., [under review](#)), it *does* mean that coral health diagnostics are inherently complicated (Mayfield and Chen, 2020) since corals now constitutively exhibit the hallmarks of a cellular stress response (Mayfield et al., 2021). Concentrations of canonical eukaryotic stress marker genes/ proteins, then, do not provide direct, easily interpretable insight into coral health when measured on their own. Without watching a coral colony first bleach or succumb to disease, it is not currently possible to predict its resilience in a proactive manner (Mayfield, [under review](#)). To that end, we developed a “coral health index” (CHI; 氣) that represents a combination of several response variables whose absolute or relative levels/concentrations may reflect deviation from normal coral behavior that is diagnostic of underlying stress (beyond the aforementioned, high-baseline coral stress levels).

To then understand how coral stress loads, as approximated by the CHI, vary over space and time at sub-regional scales, we re-explored a published dataset (Mayfield et al., 2017a) that emerged from the Khaled bin Sultan Living Oceans Foundation’s (LOF) 2013 expedition to the Kingdom of Tonga (**Figure 1**) as part of their “Global Reef Expedition” (GRE). The target species were the model corals for research (Putnam et al., 2013; Mayfield et al., 2018a-b; McRae et al., 2021): *Pocillopora acuta* and its closely related sister species *Pocillopora damicornis* (Mayfield et al., 2013a-b).

Both were sampled across a number of environmental (ENV) gradients, with ecological (ECO) transect surveys undertaken nearby (depths ranging from ~5 to 30 m for both surveys and coral sampling). This provided us the opportunity to look at relationships among coral physiological data (CPD; e.g., the aforementioned CHI [discussed in detail below]), ENV, and ECO, and we hypothesized that we could use these data to make inferences about coral health and coral reef ecology (Rodriguez-Troncoso et al., 2019) with machine learning (i.e., artificial intelligence [AI]); the ultimate goal was to develop analytical tools that will allow us to identify where resilient reef-building corals and coral reefs, respectively, will be found (*sensu* Mayfield et al., 2019a-b; Mayfield, 2020b).

## **Materials and Methods**

### ***Overview and coral stress metrics***

Results from the Khaled bin Sultan LOF research cruise to the Kingdom of Tonga in 2013 were first described in a field report from Purkis et al. (2017) followed thereafter by our attempt to describe the molecular eco-physiology of a subset of 115 pocilloporid corals that were sampled at diverse sites across two archipelagos within the nation (Mayfield et al., 2017a): Ha'apai and Va'vau. From the sampled corals, which spanned four species (*Pocillopora acuta*, *Pocillopora damicornis*, *Pocillopora verrucosa*, and *Pocillopora meandrina*), 18 molecular-physiological measurements were made (**Table 1**); only the results of *P. acuta* and *P. damicornis* are portrayed herein given small sample sizes for the other species. Colony color was scored qualitatively (normal, pale, very pale, or bleached) while underwater, so we first recoded these values to 5, 3.5, 2, and 1, respectively; we then converted the recoded data to z-scores (i.e., standardization). Since a more quantitative proxy for dinoflagellate endosymbiont (family

Symbiodiniaceae) concentration was derived from the DNAs extracted from the small (~50-100 mg) coral tissue+skeleton biopsies (the *Symbiodiniaceae* “genome copy proportion” [GCP]; Mayfield et al., 2011), we also standardized the endosymbiont GCP data and then averaged these z scores with the standardized color z scores to create a response variable called the “mean color score.”

The mean color score was factored into the CHI (discussed below) because of the well-studied link between coral pigmentation and health (Mayfield and Gates, 2007); all reef-building corals require gastrodermal dinoflagellate photosynthesis to survive and calcify (Mayfield et al., 2009; Peng et al., 2011), and, although pigmentation can change across light levels/depths (Mayfield et al., 2015), corals with low symbiont densities and white or off-white appearances will, all else being equal, physiologically underperform on account of malnourishment (Mayfield et al., 2010, 2012; Mayfield, 2016; Peng et al., 2020). Since the qualitative, diver-derived color score was not used in the statistical models outlined below given its inherent redundancy with the mean color score, 17 CPD model terms (the 19 listed in **Table 1** minus qualitative color scores and the CHI) were used in the majority of the multivariate analyses, though the two categorical ones, polyp extension (yes vs. no) and Symbiodiniaceae assemblage (*Cladocopium* spp. only vs. *Symbiodinium* spp.+*Cladocopium* spp.), were omitted from modeling types that cannot accommodate categorical data (e.g., multi-dimensional scaling [MDS]).

### ***Coral health index***

Although all 18 CPD were hypothesized to be important for coral health diagnostics, we sought to distill this multivariate trait space into a single value that could be used as a proxy for health: the aforementioned CHI. This value was calculated by

taking the mean of the standardized values of four coral response variables: the mean color score, the Mahalanobis distance, the “heat map score,” and the variability index (the latter three are described in Mayfield et al., 2017a-b.). For the latter three CPD, the z-scores were multiplied by -1 since high values are reflective of aberrant, potentially stress-indicative behavior, and the CHI was designed to where high values (5) are associated with healthier corals (with values of 0 given to the most stressed corals). The Mahalanobis distance is the multivariate outlier metric (inter-sample aberrancy), with high values reflecting significant deviation from normal behavior; this was hypothesized to signify stress since it could be evidence for loss of control of homeostasis.

The heat map score is the summed total of all response variables that demonstrated aberrant behavior, defined as being beyond two standard deviations above or below the local mean. For instance, if a coral host’s stress genes’ z scores were -3 (+1), -1 (+0), 1 (+0), and 2.5 (+1), it would be given a heat map score of 2 since two of the four genes’ z scores were beyond the two-standard deviation threshold. The variability index is simply the standard deviation of the z scores across all CPD, with high values of this univariate-derived parameter hypothesized to be associated with loss of control of homeostasis (intra-sample aberrancy; Mayfield, 2020a). Upon multiplying the Mahalanobis distance, heat map score, and variability index z-scores by -1, the resulting values were averaged with the mean color score (for which *high* values are instead reflective of healthy corals), and the resulting means of the standardized values of the four CPDs were re-scaled to percentiles (1-100%) and then re-scaled again to follow a 0-5 distribution. Finally, values were rounded to the nearest whole integer.

### ***Analytical approaches-benthic ecology***

There were two overarching goals of this work (summarized in **Table 2**). First, we sought to determine the ENV that contributed most significantly to variation in the benthic assemblage. In these analyses, we considered 14 ENV (**Table 1**), which included both continuous (e.g., latitude) and categorical (e.g., island) predictors. For multiple correspondence analysis (MCA), the five continuous ENV were recoded as categorical data; for all other analyses, they were analyzed as in **Table 1**. Because the ENV dataset (266 site x depth bins across the 59 sites surveyed spanning three island regions: Ha'apai, Va'vau, and Niuatoputapu [see maps in Mayfield et al., 2017a].) included both data types (continuous and categorical), the dataset complexity reduction approaches used for the CPD and ECO were not employed with the ENV. The ECO data were either analyzed as raw percentages of the 82 benthic categories (**Table 1**; see Purkis et al., 2017 for survey methods.) or as 37 factors derived from a factor analysis carried out with JMP® Pro 16 (which was used for all remaining statistical analyses) with the intention of reducing dataset complexity. With the ECO data as the model Y's (82 or 37 columns), the 14 ENV parameters were included as model X terms (i.e., putative predictors) in either single, squared, or cubed factorial designs. JMP Pro's "response surface" predictor design was used as a fourth means of building models. These input data were analyzed via partial least squares (PLS) featuring a non-linear iterative PLS (NIPALS) algorithm (kfold validation of 7).

In a separate analysis, we built predictive models for coral cover (%; Y) using single, squared, or cubed factorial combinations of the 14 ENV as predictors. For first-order factorials, a JMP Pro "model screen" was used to test each of the following models in parallel: logistic regression, PLS, generalized multivariate regression (multiple

algorithms: lasso, elastic net, and others), stepwise regression, Naïve Bayes, XGBoost, bootstrap forest, decision tree, partition, support vector machines, k-nearest neighbors, and neural network (NN; **Figure 2B**). When the latter was found to best fit the validation data (highest  $R^2$ ), a JMP Pro NN model-tuning add-in from Diedrich Schmidt (GUI ver. 3.0) was used to identify the optimal suite of model tuning parameters: number of hidden layers (1 or 2), type(s) of activation (sigmoidal [TanH], linear, or radial [Gaussian]), number of nodes of each type of activation (0, 1, 2, 3, or 4), number of boosts (only for single-layer models; 0, 1, 2, or 20), learning rate (only with boosted models; 0.1, 0.2, or 0.5), number of tours (1, 20, 50, or 100), covariate transformation (yes or no), and robust fit (yes or no). For additional details on this topic, please see Mayfield ([under review](#)).

### ***Analytical approaches-coral eco-physiology***

The second major goal of this work was to model variation in coral physiology across various within-country ecological and environmental gradients. First, we sought to look at ENV- (**Figure 2A**) and ECO-driven (**Figure 2C**) variation in 17 CPD, as analyzed in a multivariate (i.e., multiple Y's) framework; for these analyses, PLS (NIPALS), rather than the model screen, was used. As a descriptive counterpart, 15 continuous CPD were analyzed via MDS, and the coordinates for the first four dimensions were used as the model Y terms; this is known as non-parametric MANOVA (NP-MANOVA), and it was undertaken mainly to corroborate findings from permutational MANOVA found in our prior work with these samples (Mayfield et al., 2017a). In these analyses, only the 14 ENV were tested as model X's, and an alpha of 0.01 was set for this and all other multivariate analyses.



Secondly, we looked at the endogenous properties of the coral with respect to their driving of differences in the CHI (the models' singular Y term). Since factor analysis determined that a four-factor model was optimal for reducing the complexity of the 17-CPD+70 sample dataset (see the online supplemental data file [OSDF]; 47 *P. acuta*+23 *P. damicornis* samples for which no data were missing), the models were tested two ways: 1) with 13 CPD (the eight genes, RNA/DNA ratio, Symbiodiniaceae GCP, maximum colony length, polyp extension, and *Symbiodiniaceae assemblage*); or 2) with the loading scores from the four factors; the qualitative (diver-based) color score, Mahalanobis distance, heat map score, and variability index were excluded since they were used to define the CHI and would consequently covary with it.

### ***Coral health index predictions***

Most importantly, we aimed to uncover the ENV (**Figure 3A**) and ECO (**Figure 3B-C**) drivers of variation in the CHI (**Table 2**). For the latter, the 26 factor loading scores from an analysis undertaken with only the ECO data associated with the transects from which corals were sampled were also tested as predictors (**Table 2**). For first-order factorials, JMP Pro's model screening platform was used, and the same NN GUI described above was used to tune NN models when 1) they possessed the highest validation  $R^2$  of all modeling types tested and 2) the validation  $R^2$  of the default model screen NN (typically "NTanH(3)-Boost(20)") was  $<0.80$ .

## Results and Discussion

### *Environmental effects on the benthos*

A factor analysis reduced the complexity of the 82-category benthic dataset to 37 factors (53% of the variation; **Table 2**), whose loading scores were then used in a NP-MANOVA to assess the influence of each of the 14 ENV on the benthic composition; all 14 significantly affected the structure (**Table S1**). Please note, though, that some ENV inherently covary. For instance, the date effect is driven by the fact that different sites were surveyed on different days. Site and GPS coordinates are also inherently linked, meaning that certain predictors are redundant (e.g., site and latitude). In a more sophisticated multivariate analysis, we looked at ENV effects on the benthic composition using differing factorial combinations of predictors with either the raw, 82-category percentage data or the aforementioned 37 factor loading scores (**Table 2**). Despite the statistically significant NP-MANOVA effects, the ENV did not explain a high percentage of the variation in the benthic structure when undertaking PLS (**Table 2**); a 2-factor model featuring 2,744 predictors ( $14 \text{ ENV}^3$ ) still explained less than 15% of the variation in the benthos (approximated from the loading scores from the 37 factors). This suggests that other, unmeasured factors are contributing significantly to variation in the benthic composition of Tonga's coral reefs.

### *Predicting coral cover from environmental data*

In contrast to the multivariate analyses described above, predictive models for coral cover were more robust (**Table 2**). A model screen of coral cover versus the 14 ENV revealed a NN model with an  $R^2$  of 0.50 (**Table 2**); the NN GUI was then used to uncover a NN with a superior  $R^2$ , and the model shown in **Figure 6A** (and described in

**Table 2)** was characterized by an  $R^2$  of 0.95. A predictor screen of the 14 ENV vs. coral cover (**Figure 4A**) revealed that reef site explained 67% of the variation in cover. In the NN model itself, date (total effect=0.82) and site (0.78) were the two most influential model terms based on an independent resampled inputs analysis. When looking at this NN in detail (**Figure 6A**), it is clear that it is markedly complex; the four Gaussian activation nodes were boosted 20 times, with a learning rate of 0.5 for each boost. Boosting is an ensemble learning approach in which weak learners are combined into a stronger learner that reduces training error, and 20 was the maximum set *a priori* in our NN GUI model-tuning design. This explains why 80, rather than 4, nodes are depicted in **Figure 6A**. Using this model in a desirability analysis of coral cover (*sensu* Chen et al., 2019), in which maximum values were “desired,” a coral cover of 85% could hypothetically be reached in Tonga under the following conditions: 12-18 m depth on a protected, emergent back reef in Niuatoputapu at 24.5°C (**OSDF**). Since these conditions did not exist, the highest coral cover (73%) was instead documented at site TOVA52 (a shallow, lagoonal patch reef in Va’vau).

### ***Coral physiology-univariate***

A treatise on the majority of the CPD can be found in Mayfield et al. (2017a). The strongest univariate statistical effects (highest false discovery rate [FDR]-log[worth]) were documented for maximum colony length vs. the spatial ENV factors (site, latitude, and longitude); colony size varied markedly across both small and large spatial gradients. In terms of the molecular response variables, the Symbiodiniaceae zinc-induced facilitator-like 1-like gene (*zifl1l*) varied significantly over time, with morning

levels over 5-fold higher than afternoon ones (and 2-fold higher than midday levels). This finding is discussed in detail in Mayfield et al. (2017a).

### ***Coral physiology-multivariate***

**Table 3** shows the results of the NP-MANOVA for both species analyzed together as well as individually upon having dimensionalized (via MDS) the data. Unlike the aforementioned analysis of ENV effects on ECO, few ENV affected the multivariate coral phenotype; when looking at all 70 samples (i.e., both species), only lagoon (inside vs. outside) significantly affected the coral phenotype, and this appears to be entirely biased by the larger sample size for *P. acuta* (no lagoon effect was documented for the 23 *P. damicornis* samples; **Table 3**). In fact, not a single ENV affected the *P. damicornis* phenotype at the alpha of 0.01, though this is likely a testament to the small sample size and high degree of variation. With the larger *P. acuta* sample set (n=47), half of the 14 ENV significantly affected the phenotype (**Table 3**). Of note, *P. acuta* physiology differed significantly between intermediately protected and fully protected habitats, as well as between the fore reef and lagoonal zones. The phenotype of this coral also differed between fringing and patch reefs (**Table 3**), as well as over time; the latter observation is likely driven by the aforementioned temporal change in *zifl1* expression and is discussed in a prior work (Mayfield et al., 2017a).

### ***Endogenous contributors to the coral health index***

A predictor screen of the CHI vs. the other coral response variables (the 13 *not* used to calculate it) found that *Symbiodiniaceae zifl1* was the CPD that contributed most to variation in the CHI (28%), with the host green fluorescent protein-like chromoprotein

(*gfp-cp*) the second highest contributor (17%); only host coral *lectin* contributed >15% of the variation in the CHI of the remaining 11 CPD. When looking instead at raw correlations between these three CPD and the CHI,  $R^2$  of 0.25, 0.15, and 0.13, respectively, were calculated ( $p < 0.001$  for all); all were negatively associated with the CHI (i.e., high expression levels of these genes being indicative of healthier corals). This finding is unsurprising since, of the eight target genes, all but these three, host carbonic anhydrase (*ca*), and Symbiodiniaceae ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO [*rbcL*]) are stress genes (in which high expression levels are reflective of stress). High expression levels of *lectin* have been previously hypothesized to be diagnostic of intact, optimally functioning coral-dinoflagellate endosymbioses (Mayfield and Dempsey, [under review](#)), whereas the role of *gfp-cp* in coral health is less clear (Mayfield et al., 2014). However, given its putative role in shading the dinoflagellate endosymbionts during high-light periods (Smith et al., 2013), an elevated concentration could be a desirable attribute (Mayfield et al., 2016).

### ***Environmental drivers of the coral health index***

Predictor screens of CHI vs. the 14 ENV were undertaken for both coral hosts analyzed together (**Figure 4B**), *P. acuta* alone (**Figure 4C**), and *P. damicornis* alone (**Figure 4D**). For the former, temperature was the biggest contributor to variation, and there was a weak, positive, linear association between temperature and CHI ( $R^2 = 0.19$ ,  $p < 0.01$ ). Since temperatures were generally low at the time of surveys (September; cruise mean = 25.5°C), this could be driven by the higher metabolic rates characteristic of tissues exposed to higher temperatures. Had temperatures instead been warmer and nearing the bleaching threshold of these corals (~30.5°C), we likely would have instead

seen an *inverse* relationship between temperature and coral stress loads (lower CHI at higher temperatures). Temperature was also the biggest driver of the CHI for the *P. acuta* dataset (**Figure 4C**) and the third highest contributor for *P. damicornis* (**Figure 4D**). Depth was the third-most and most important contributor to CHI variation for *P. acuta* and *P. damicornis*, respectively, though  $R^2$  values were only 0.03 and 0.05, respectively (both  $p>0.05$ ).

In addition to temperature, only two other ENV significantly affected CHI in an FDR-controlled “response screen:” reef exposure and reef type. Regarding the former, the CHI of corals of intermediately exposed reefs were over 50% higher than congenetics of protected environments; this means that corals from protected reefs might be exposed to higher levels of stress (perhaps due to the higher sediment loads and consequently higher potential for smothering and/or incidental shading). Because fringing reefs tended to be intermediately exposed, rather than protected, there was a similar, 50% higher CHI in corals from fringing reefs vs. those of patch reefs (which tended to be sheltered); the lone barrier reef sampled from excluded from this analysis.

### ***Environmental and ecological effects on the coral health index***

When looking at the relationship between the CHI and the 82 ECO for both coral species (**Figure 5A**), *P. acuta* alone (**Figure 5C**), and *P. damicornis* alone (**Figure 5E**), the top 15 parameters explained 69, 67, and 83% of the variation, respectively.

*Acanthastrea* spp. cover was the most influential predictor for the 70-sample dataset, and there was a weak, statistically insignificant ( $p=0.08$ ), negative relationship between the CHI and *Acanthastrea* spp. cover. This could indicate that pocilloporid corals compete with corals of this genus. *Acropora* spp. cover was also a top-three predictor

for each species in isolation, though acroporid cover was positively associated with the CHI for *P. acuta* ( $R^2=0.02$ ) and negatively associated with it for *P. damicornis* ( $R^2=0.19$ ); the latter correlation narrowly missed the statistically significant difference cutoff ( $p=0.04$ ).

Predictor screens of CHI were also undertaken with 14 ENV+82 ECO (i.e., 96 parameters as predictors), and the top 15 parameters explained 73, 83, and 85% of the variation in the CHI for both coral hosts analyzed together (**Figure 5B**), *P. acuta* only (**Figure 5D**), and *P. damicornis* only (**Figure 5F**), respectively. Site was the best predictor of the CHI for the 70-sample dataset (**Figure 5B**), accounting for over 18% of its variation. The temperature effect documented in the CHI vs. ENV analysis (**Figure 4**) was corroborated with the two-species (**Figure 5B**) and *P. acuta* (**Figure 5D**) datasets; given that temperature was *not* among the top 15 predictors for *P. damicornis*, the two-species finding is clearly being driven by the response of *P. acuta* alone. For *P. damicornis* (**Figure 5F**), *Favites* spp. cover was the top predictor, accounting for just under 15% of the variation in the CHI (vs. only 1% for *P. acuta*). There was a statistically significant, negative association between the CHI and *Favites* spp. cover ( $R^2=0.26$ ,  $p=0.01$ ); more stressed corals (lower CHI) were documented in areas with higher *Favites* cover, suggesting that corals of this genus may compete with pocilloporids.

### ***Coral health index predictive modeling***

In the model screen, the machine learning-based NN consistently generated models with the highest validation sample  $R^2$  (**Table 4**). In all cases, though, the base NN model generated by JMP, “NTanH(3)Boost(20),” did not yield a high enough  $R^2$  for *in situ* coral diagnostics (0.6-0.8). When using the NN model-tuning GUI, however,  $R^2$

values for both species, *P. acuta*, and *P. damicornis* of 0.92 (**Figure 6B**), 0.51, and 1.00 were obtained with the 14 ENV as predictors (0.86-0.93 [**Figure 6C**], 0.99, and 1.00, respectively, with the 82 ECO as predictors, and 0.96 [**Figure 6D**], 0.67, and 1.00, respectively, for all 96 parameters [14 ENV+82 ECO]). Although these findings imply that coral health can be predicted with confidence by simply measuring several environmental factors, the difference in the predictive power of the *P. acuta* vs. *P. damicornis* models is worth exploring; of note, the *P. damicornis* NN models appear to be perfectly fit, even with validation samples. This is likely to be an artifact associated with the small sample size, and it is doubtful that 100% of the variation in the *P. damicornis* CHI can be explained by 14 ENV alone. Instead, the  $R^2$  of the two-species models (0.86-0.96) are likely to be more realistic. The dominant (i.e., highly weighted) predictors of the ENV, ECO, and ENV+ECO models were depth, *Acropora* spp. cover, and date, respectively, when using independent resampled inputs, and date, *Pachyseris* spp. cover, and *Acanthastrea* spp. cover, respectively, with dependent resampled inputs.

Survey date and time will not be useful in future analyses despite being important predictors herein, and island and site would be inappropriate were these models used beyond Tonga. Therefore, the two-species models were re-run with only those 10 ENV that would characterize reefs not yet surveyed/sampled: latitude, longitude, depth, temperature, salinity, reef exposure, reef zone, reef type, reef emergence, and lagoon. This 10-predictor model only suffered a slight decrease in validation  $R^2$  (0.91 vs. 0.92 for the 14-ENV model; **Table 4**), and the most important predictor was depth. When these 10 ENV were added to the 82 ECO (92 parameters in total), the validation  $R^2$  was 0.92 (vs. 0.96 for the 96-factor model), and the most important predictors were reef type (total



effect=0.39) and longitude (total effect=0.37). As mentioned above, the CHI was ~50% higher in corals of fringing reefs vs. patch reefs. In contrast, the association between longitude and the CHI was weak:  $R^2=0.06$  (positive association;  $p>0.01$ ).

When using the NN model derived from the 92 ENV+ECO parameters in a desirability analysis aimed at maximizing the CHI, the following environment was ideal (CHI of 6, which is not technically possible): intermediately exposed, emergent fringing reef in the lagoon at 11 m depth and 26.7°C (salinity=35.2). In actuality, the highest CHI was measured in sample T56.2, a *P. acuta* colony from a fringing fore reef in the Va'vau lagoon (30 m depth). The high pigmentation and color scores for this sample may be driven by the deep depth (amongst the deepest colonies sampled). However, it is clear from **Figure 7** that corals of Va'vau tended to be characterized by higher CHI than congeners of Ha'apai, and of the seven colonies with CHI >4.5, six were from the more remote Va'vau region. It is important to note, though, that the CHI is strictly a hypothetical measure of stress; the fates of the sampled colonies were not tracked in the intervening time since the cruise to determine whether corals with high CHI are more likely to resist high-temperature-induced bleaching events. It will be imperative to undertake such field validation prior to the widespread adoption of this novel benchmark.

Alternatively, since 10 ENV could predict the CHI with high confidence, it may not be necessary to make the underlying measurements required to calculate the CHI, some of which are expensive and require extracting RNAs and DNAs from biopsies. If the environmental characteristics alone can explain a large portion of the resilience demonstrated by corals, then the machine learning models developed herein could be used to identify refugia, as well as where less thermotolerant corals may be found; the latter might be good candidates for *ex situ* “rescue” and culture (Lin et al., 2019; Chang

et al., 2020; Huang et al., 2020). As with corals of other nations assessed during the LOF-GRE, all corals sampled in Tonga were characterized by high stress marker levels, regardless of their CHI; whether or not this “alternative stable state” (i.e., “new normal”) represents a sustainable survival strategy will inevitably be unveiled in the near future.

### **Acknowledgments**

We are deeply indebted to Diedrich Schmidt for designing the GUI that enabled us to test thousands of machine learning models using JMP Pro. This work was funded by the National Oceanic and Atmospheric Administration’s ‘Omics Initiative (NRDD18978), with the research cruise and subsequent laboratory analyses funded by the Khaled bin Sultan LOF. We also thank the Fulbright and MacArthur Foundations for supporting ABM’s time in Taiwan (where laboratory analyses were undertaken).

## References

- Chang, T.C., A.B. Mayfield & T.Y. Fan. 2020. Culture systems influence the physiological performance of the soft coral *Sarcophyton glaucum*. *Scientific Reports* 10: 20200
- Chen, T.Y., G.W. Hwang, H.J. Lin, A.B. Mayfield & C.P. Chen. 2019. The development of a habitat suitability model for sub-tropical tidal flat fiddler crabs. *Ocean and Coastal Management* 182: 104931
- Cruz-García, R., A.P. Rodríguez-Troncoso, F.A. Rodríguez-Zaragoza, A.L. Cupul-Magaña & A.B. Mayfield. 2020. Ephemeral effects of El Niño southern oscillation events on an eastern tropical Pacific coral community. *Marine and Freshwater Research* 71(10): 1259-1268
- Demerlis, A., A. Kirkland, M. Kaufman, A.B. Mayfield, N. Formel, D.P. Manzello, D. Lirman, N. Traylor-Knowles & I.C. Enochs. **under review**. Variable temperature treatments alter the response of *Acropora cervicornis* to acute thermal stress. *Coral Reefs*
- Enochs, I.C., N. Formel, D.P. Manzello, J. Morris, A.B. Mayfield, A. Boyd, G. Kolodziej G. Adams. 2020. Coral persistence despite extreme periodic pH fluctuations at a volcanically acidified Caribbean reef. *Coral Reefs* 39: 523-528
- Grottoli, A., A.B. Mayfield & RCN Bleaching Working Group. 2021. Increasing comparability among coral bleaching experiments. *Ecological Applications* 31(4): e02262
- Huang, Y.L., A.B. Mayfield & T.Y. Fan. 2020. Effects of feeding on the physiological performance of the stony coral *Pocillopora acuta*. *Scientific Reports* 10: 19988
- Lin, C., S. Tsai & A.B. Mayfield. 2019. Physiological differences between cultured and wild coral eggs. *Biopreservation and Biobanking* 17: 370-371
- Mayfield, A.B. 2016. Uncovering spatio-temporal and treatment-derived differences in the molecular physiology of a model coral-dinoflagellate mutualism with multivariate statistical approaches. *Journal of Marine Science & Engineering* 4: 63
- Mayfield, A.B. 2020a. Exploiting the power of multivariate statistics for probing the cellular biology of thermally challenged reef corals. *Platax* 17: 27-52
- Mayfield, A.B. 2020b. Proteomic signature of corals from thermodynamic reefs. *Microorganisms* 8(8): 1171
- Mayfield, A.B. **under review**. Differential proteomics of the massive Caribbean coral *Orbicella faveolata*. *Diversity*

- Mayfield, A.B., C. Aguilar, I.C. Enochs, G. Kolodziej & D.P. Manzello. 2021. Shotgun proteomics of thermally challenged Caribbean reef corals. *Frontiers in Marine Science* 8: 660153
- Mayfield, A.B., P.H. Chan, H.M. Putnam, C.S. Chen & T.Y. Fan. 2012. The effects of a variable temperature regime on the physiology of the reef-building coral *Seriatopora hystrix*: results from a laboratory-based reciprocal transplant. *The Journal of Experimental Biology* 215: 4183-4195
- Mayfield, A.B. & C.S. Chen. 2019a. Enabling coral reef triage via molecular biotechnology and artificial intelligence. *Platax* 16: 23-47
- Mayfield, A.B. & C.S. Chen. 2020. A coral transcriptome in the Anthropocene as an “alternative stable state.” *Platax* 17: 1-26
- Mayfield, A.B., C.S. Chen & A.C. Dempsey. 2017a. Biomarker profiling in reef corals of Tonga’s Ha’apai and Vava’u Archipelagos. *PLoS ONE* e0185857
- Mayfield, A.B., C.S. Chen & A.C. Dempsey. 2017b. Identifying corals displaying aberrant behavior in Fiji’s Lau Archipelago. *PLoS ONE* e0177267
- Mayfield, A.B., C.S. Chen, A.C. Dempsey & A.W. Bruckner. 2016. The molecular ecophysiology of closely related pocilloporids from the South Pacific: a case study from the Austral and Cook Islands. *Platax* 13: 1-25
- Mayfield, A.B. & A.C. Dempsey. **under review**. Environmental variation in New Caledonian reef corals. *Oceans*
- Mayfield, A.B., T.Y. Fan & C.S. Chen. 2013a. Physiological acclimation to elevated temperature in a reef-building coral from an upwelling environment. *Coral Reefs* 32: 909-921
- Mayfield, A.B., T.Y. Fan & C.S. Chen. 2013b. Real-time PCR-based gene expression analysis in the model reef-building coral *Pocillopora damicornis*: insight from a salinity stress study. *Platax* 10: 1-29
- Mayfield, A.B. & R.D. Gates. 2007. Osmoregulation in anthozoan-dinoflagellate symbiosis. *Comparative Biochemistry and Physiology A: Molecular and Integrative Physiology* 147: 1-10
- Mayfield, A.B., M.B. Hirst & R.D. Gates. 2009. Gene expression normalization in a dual-compartment system: a real-time PCR protocol for symbiotic anthozoans. *Molecular Ecology Resources* 9: 462-470
- Mayfield, A.B., A.W. Bruckner, C.H. Chen & C.S. Chen. 2015. A survey of pocilloporids and their endosymbiotic dinoflagellate communities in the Austral and Cook Islands of the South Pacific. *Platax* 12: 1-17

- Mayfield, A.B., C.S. Chen & P.J. Liu. 2014. Decreased green fluorescent protein-like chromoprotein gene expression in specimens of the reef-building coral *Pocillopora damicornis* undergoing high temperature-induced bleaching. *Platax* 11: 1-23
- Mayfield, A.B., Y.J. Chen, C.Y. Lu & C.S. Chen. 2018a. The proteomic response of the reef coral *Pocillopora acuta* to experimentally elevated temperature. *PLoS ONE* e0192001
- Mayfield, A.B., A.C. Dempsey, J. Inamdar & C.S. Chen. 2018b. A statistical platform for assessing coral health in an era of changing global climate-I: a case study from Fiji's Lau Archipelago. *Platax* 15: 1-35
- Mayfield, A.B., S. Tsai & C. Lin. 2019b. The Coral Hospital. *Biopreservation and Biobanking* 17: 355-369
- Mayfield, A.B., L.H. Wang, P.C. Tang, Y.Y. Hsiao, T.Y. Fan, C.L. Tsai & C.S. Chen. 2011. Assessing the impacts of experimentally elevated temperature on the biological composition and molecular chaperone gene expression of a reef coral. *PLoS ONE* e26529
- McRae, C., A.B. Mayfield, T.Y. Fan, W.B. Huang & I. Cote. 2021. Differing proteomic responses to high-temperature exposure between adult and larval reef corals. *Frontiers in Marine Science* 8: 716124
- Peng, S.E., A. Moret, C. Chang, A.B. Mayfield, Y.T. Ren, W.N.U. Chen & C.S. Chen. 2020. A shift away from mutualism under food-deprived conditions in an anemone-dinoflagellate association. *PeerJ* 8: e9745
- Peng, S.E., W.N.U. Chen, H.K. Chen, C.Y. Lu, A.B. Mayfield, L.S. Fang & C.S. Chen. 2011. Lipid bodies in coral-dinoflagellate endosymbiosis: ultrastructural and proteomic analyses. *Proteomics* 17: 3540-3455
- Purkis, S., A. Dempsey, R. Carlton, B. Samaniego, K. Lubarsky & P.G. Renaud. 2017. Global Reef Expedition: Kingdom of Tonga. Final Report. Khaled Bin Sultan Living Oceans Foundation, Annapolis, MD. Vol. 8
- Putnam, H.M., A.B. Mayfield, T.Y. Fan, C.S. Chen & R.D. Gates. 2013. The physiological and molecular responses of larvae from the reef-building coral *Pocillopora damicornis* exposed to near-future increases in temperature and  $p\text{CO}_2$ . *Marine Biology* 160: 2157-2173
- Rodríguez-Troncoso, A.P., F.A. Rodríguez-Zaragoza, A.B. Mayfield & A.L. Cupul-Magaña. 2019. Temporal variation in invertebrate recruitment on an Eastern Pacific coral reef. *Journal of Sea Research* 145: 8-15

- Rubin, E., I. Enochs, C. Foord, G. Kolodziej, I. Basden, D.P. Manzello & A.B. Mayfield. 2021. Molecular mechanisms of coral persistence within highly urbanized locations in the Port of Miami, Florida. *Frontiers in Marine Science* 8: 695236
- Smith, E.G., C. D'Angelo, A. Salih & J.J.C.R. Wiedenmann. 2013. Screening by coral green fluorescent protein (GFP)-like chromoproteins supports a role in photoprotection of zooxanthellae. *Coral Reefs* 32: 463-474

**Table 1. Data types.** Please note that the total number of bins in parentheses corresponds to the subset from which corals were sampled; these cells are left blank for continuous predictors (latitude, longitude, depth [m], temperature [°C], and salinity [unitless]). In certain instances, polyp extension and Symbiodiniaceae (Sym) assemblage (*Cladocopium* only or mixed assemblage of *Symbiodinium* and *Cladocopium*) were excluded from models that could not accommodate categorical predictors. See the online supplemental data file (**OSDF**) for the exact algal (n=6 taxa) and coral (n=74 genera) benthic categories. GCP=genome copy proportion. GFP=green fluorescent protein. RuBisCO=ribulose-1,5-bisphosphate carboxylase/oxygenase. SOD=superoxide dismutase

Coral physiological data (CPD) (n=19)		Environmental data (ENV) (n=14)		Ecological data (ECO) (n=82; all in %)	
Parameter	Unit/abbreviation/details	Parameter	# bins	Parameter	# bins
Color (qualitative)	1, 2, 3.5, or 5	Island <sup>a</sup>	3 (2)	Barren substrate	1
Mean color score	See main text.	Site <sup>a</sup>	59 (27)	Invertebrate cover	1
Colony size	cm	Exposure <sup>a</sup>	3 (2)	Algal cover	6
Polyp extension <sup>a</sup>	yes or no	Reef zone <sup>a</sup>	3 (3)	Coral cover	74
Sym assemblage <sup>a</sup>		Reef type <sup>a</sup>	4 (3)		
RNA/DNA ratio	unitless	Lagoon <sup>a</sup>	2 (2)		
Sym GCP	unitless	Emergence <sup>a</sup>	2 (2)		
Mahalanobis distance		Latitude			
Heat map score	See Mayfield et al. (2017a).	Longitude			
Variability index	See Mayfield et al. (2017a).	Depth			
Coral health index	0-5; see main text.	Temperature			
Gene expression		Salinity			
Sym <i>ubiq-lig</i>	ubiquitin ligase	Sampling time <sup>a</sup>	3 (3)		
Sym <i>rbcL</i>	RuBisCO (large subunit)	Sampling date <sup>a</sup>	20 (16)		
Sym <i>zif11</i>	zinc-induced facilitator 1				
Sym <i>hsp90</i>	heat shock protein 90				
Host <i>gfp-cp</i>	GFP-like chromoprotein				
Host <i>ca</i>	carbonic anhydrase				
Host <i>lectin</i>					
Host <i>cu-zn-sod</i>	copper-zinc SOD				

<sup>a</sup>Categorical variable.

**Table 2. Analytical approaches.** The data types include coral physiological data (CPD; either 15 for continuous data analyses or 17 for categorical+continuous analyses), environmental (ENV) data (14 parameters), and ecological (ECO; i.e., benthic) data (82 categories). Only analyses that resulted in a reduced-complexity model with >5% explanation of variation in the associated Y parameter(s) (and whose partial least squares [PLS] PRESS was minimized with >0 factors) have been included; analyses that did not pass quality control can instead be found in **Table S2**. The coral-specific analyses were carried out with the datasheet “Coral sample data” while the ENV vs. ECO analyses were undertaken with the 266-row “Benthic data” datasheet in the online supplemental data file (**OSDF**). For the singular Y responses, the coral health index (CHI) and percent coral cover, JMP® Pro 16’s “model screen” platform was used to test a large number of predictive models in parallel (see main text.). Please note that the ECO factor analysis reduced the data columns from 82 to 37 dimensions in the benthic dataset; this value was only 28 in the 70-row data table since corals were not sampled at all 59 sites surveyed. Third-order factorials were not possible for the ECO or ENV+ECO datasets since the number of model terms was too high (>500,000). Model screening and PLS used kfold validations of 5 and 7, respectively, unless denoted by asterisks (\*; validation column instead used). NA=not applicable. NP-MANOVA=non-parametric multivariate analysis of variance. PCA=principal components analysis.

To be uncovered	Model/analysis type	Model Y (#)	Model X (#)	Conclusion/ data location
Relationship among corals	PCA and MDS	CPD (15)	NA	Mayfield et al. (2017a)
Relationship among corals	Factor analysis	CPD (15)	NA	4 factors (72%)
Relationship among reef sites	MCA	ENV (14)	NA	2 factors (11.5%)
Relationship among reef sites	PCA and MDS	ECO (82)	NA	<b>OSDF</b>
Relationship among reef sites	Factor analysis	ECO (82)	NA	37 factors (52.7%)
Endogenous drivers of coral health	Model screen	CHI (1)	CPD <sup>a</sup> (13)	<b>Table 4</b>
Endogenous drivers of coral health	Predictor screen	CHI (1)	CPD <sup>a</sup> (13)	Data not shown
Environmental effects on coral health	NP-MANOVA	CPD (4 dimensions)	ENV (14)	<b>Table 4</b>
Environmental effects on coral health	PLS (NIPALS)	CPD (17)	ENV (14)*	1 factor (11.4%)
Environmental effects on coral health	PLS (NIPALS)	CPD (17)	ENV (14 <sup>2</sup> )*	1 factor (15.4%)
Environmental effects on coral health	PLS (NIPALS)	CPD (17)	ENV (14 <sup>3</sup> )*	1 factor (17.2%)
Environmental effects on coral health	PLS (NIPALS)	CPD (17)	ENV-response surface*	1 factor (15.3%)
Environmental effects on coral health	PLS (NIPALS)	CPD (4 factors)	ENV (14)*	2 factors (15.1%)
Environmental effects on coral health	PLS (NIPALS)	CPD (4 factors)	ENV (14 <sup>2</sup> )*	1 factor (12.3%)
Environmental effects on coral health	PLS (NIPALS)	CPD (4 factors)	ENV-response surface*	1 factor (12.2%)
Environmental effects on coral health	PLS (NIPALS)	CPD (4 factors)	ECO (28 factors <sup>3</sup> )	5 factors (40.6%)
Environmental effects on coral health	PLS (NIPALS)	CPD (17)	ENV (14)+ECO (82)*	1 factor (13.4%)
Environmental effects on coral health	PLS (NIPALS)	CPD (17)	(ENV (14)+ECO (82)) <sup>2*</sup>	1 factor (18.8%)
Environmental effects on coral health	PLS (NIPALS)	CPD (17)	ENV+ECO-response surface (96)*	1 factor (18.8%)
Environmental effects on coral health	PLS (NIPALS)	CPD (4 factors)	ENV (14)+ECO (82)*	1 factor (8.8%)
Environmental effects on coral health	Predictor screen	CHI	ENV (14)	<b>Figure 4B-D</b>
Environmental effects on coral health	Predictor screen	CHI	ECO (82)	<b>Figure 5A, 5C, 5E</b>
Environmental effects on coral health	Predictor screen	CHI	ENV (14)+ECO (82)	<b>Figure 5B, 5D, 5F</b>
Environmental effects on coral health	Neural-GUI	CHI	ENV (10)	<b>Table 4</b>



**Table 2. Analytical approaches.** The data types include coral physiological data (CPD; either 15 for continuous data analyses or 17 for categorical+continuous analyses), environmental (ENV) data (14 parameters), and ecological (ECO; i.e., benthic) data (82 categories). Only analyses that resulted in a reduced-complexity model with >5% explanation of variation in the associated Y parameter(s) (and whose partial least squares [PLS] PRESS was minimized with >0 factors) have been included; analyses that did not pass quality control can instead be found in **Table S2**. The coral-specific analyses were carried out with the datasheet “Coral sample data” while the ENV vs. ECO analyses were undertaken with the 266-row “Benthic data” datasheet in the online supplemental data file (**OSDF**). For the singular Y responses, the coral health index (CHI) and percent coral cover, JMP® Pro 16’s “model screen” platform was used to test a large number of predictive models in parallel (see main text.). Please note that the ECO factor analysis reduced the data columns from 82 to 37 dimensions in the benthic dataset; this value was only 28 in the 70-row data table since corals were not sampled at all 59 sites surveyed. Third-order factorials were not possible for the ECO or ENV+ECO datasets since the number of model terms was too high (>500,000). Model screening and PLS used kfold validations of 5 and 7, respectively, unless denoted by asterisks (\*; validation column instead used). NA=not applicable. NP-MANOVA=non-parametric multivariate analysis of variance. PCA=principal components analysis.

To be uncovered	Model/analysis type	Model Y (#)	Model X (#)	Conclusion/ data location
Environmental effects on coral health	Model screen	CHI	ENV (14), ENV (14 <sup>2</sup> )	<b>Table 4</b>
Environmental effects on coral health	Model screen	CHI	ECO (82), ECO (82 <sup>2</sup> )	<b>Table 4</b>
Environmental effects on coral health	Model screen	CHI	ECO (28 factors), 28 <sup>2</sup>	<b>Table 4</b>
Environmental effects on coral health	Model screen	CHI	ENV+ECO (96), 96 <sup>2</sup>	<b>Table 4</b>
Environmental effects on coral health	Model screen	CHI	ENV (10)+ECO (92)	<b>Table 4</b>
Environmental effects on coral health	Model screen	CHI	ENV+ECO (42), 42 <sup>2</sup>	<b>Table 4</b>
Environmental effects on reef ecology	PLS (NIPALS)	ECO (82)	ENV (14 <sup>2</sup> )	1 factor (5.5%)
Environmental effects on reef ecology	PLS (NIPALS)	ECO (82)	ENV (14 <sup>3</sup> )	1 factor (6.0%)
Environmental effects on reef ecology	PLS (NIPALS)	ECO (82)	ENV-response surface	1 factor (5.5%)
Environmental effects on reef ecology	PLS (NIPALS)	ECO (37 factors)	ENV (14)	1 factor (5.3%)
Environmental effects on reef ecology	PLS (NIPALS)	ECO (37 factors)	ENV (14 <sup>2</sup> )	1 factor (6.1%)
Environmental effects on reef ecology	PLS (NIPALS)	ECO (37 factors)	ENV (14 <sup>3</sup> )	2 factors (14.7%)
Environmental effects on reef ecology	PLS (NIPALS)	ECO (37 factors)	ENV-response surface	1 factors (6.1%)
Environmental effects on reef ecology	Model screen	% coral cover	ENV (14)	R <sup>2</sup> =0.50
Environmental effects on reef ecology	Neural-GUI	% coral cover	ENV (14)	R <sup>2</sup> =0.95 ( <b>Figure 6A</b> )
Environmental effects on reef ecology	PLS (NIPALS)	% coral cover	ENV (14)	1 factor (6.0%)
Environmental effects on reef ecology	PLS (NIPALS)	% coral cover	ENV (14 <sup>2</sup> )	1 factor (7.0%)
Environmental effects on reef ecology	PLS (NIPALS)	% coral cover	ENV (14 <sup>3</sup> )	1 factor (7.8%)
Environmental effects on reef ecology	PLS (NIPALS)	% coral cover	ENV-response surface	1 factor (7.0%)

<sup>a</sup>Excluding the five terms from which the CHI were derived: qualitative color score, mean color score, Mahalanobis distance, heat map score, and variability index.

**Table 3.** Non-parametric MANOVA of the first three multi-dimensional scaling dimensions (stress=0.28 and 0.30 for *Pocillopora damicornis* and *Pocillopora acuta*, respectively) derived from Euclidean distances among coral samples (assessed from multivariate assessment of 12 coral response variables: maximum colony length, mean color score, Symbiodiniaceae GCP, RNA/DNA ratio, and expression levels of eight genes [Table 1]). For environmental factors (ENV) featuring only two groupings (e.g., island), the exact *F* statistics have been shown; all others are approximations ("approx. *F*"). Statistically significant findings ( $p < 0.01$ ) have been highlighted in **bold**.

ENV	df	<i>F</i>	<i>p</i>	Trend
<b>Island:</b> Ha'apai vs. Va'vau (no pocilloporids sampled at Niuatoputapu)				
Both species	1	1.21	0.31	
<i>P. acuta</i> only	1	2.27	0.09	
<i>P. damicornis</i> only	1	0.16	0.92	
<b>Site:</b> 23 reef sites (a subset of all surveyed)				
Both species	22	1.24	0.12	
<i>P. acuta</i> only	22	1.51	0.02	
<i>P. damicornis</i> only	22	0.58	0.95	
<b>Latitude</b>				
Both species	24	0.86	0.46	
<i>P. acuta</i> only	24	2.35	0.08	
<i>P. damicornis</i> only	24	1.20	0.33	
<b>Longitude</b>				
Both species	24	1.22	0.31	
<i>P. acuta</i> only	24	2.48	0.07	
<i>P. damicornis</i> only	24	0.90	0.45	
<b>Reef exposure:</b> intermediate vs. protected (no exposed reefs sampled)				
Both species	1	2.58	0.06	
<i>P. acuta</i> only	1	4.00	<b>0.01</b>	intermediate≠protected
<i>P. damicornis</i> only	1	0.37	0.78	
<b>Reef zone:</b> back reef, fore reef, or lagoon				
Both species	2	1.56	0.16	
<i>P. acuta</i> only	2	2.88	<b>0.01</b>	fore reef≠lagoon
<i>P. damicornis</i> only	2	0.52	0.79	
<b>Reef type:</b> barrier, fringing, or patch				
Both species	2	1.96	0.07	
<i>P. acuta</i> only	2	3.51	<b>&lt;0.01</b>	fringing≠patch
<i>P. damicornis</i> only	2	0.65	0.59	
<b>Lagoon:</b> inside vs. outside				
Both species	1	4.20	<b>&lt;0.01</b>	inside≠outside
<i>P. acuta</i> only	1	5.65	<b>&lt;0.01</b>	inside≠outside
<i>P. damicornis</i> only	1	0.18	0.91	
<b>Reef emergence:</b> emergent vs. submergent				
Both species	1	0.15	0.93	
<i>P. acuta</i> only	1	0.71	0.55	
<i>P. damicornis</i> only	1	1.02	0.40	
<b>Coral cover (%)</b> : scaled continuously except for multiple comparisons (categorical bins)				
Both species	16	2.51	0.06	
<i>P. acuta</i> only	16	6.10	<b>&lt;0.01</b>	20-30%(a)=30-40%(a)=40-
<i>P. damicornis</i> only	16	1.70	0.19	50%(ab)= >50%(b)
<b>Sampling date:</b> 16 days (subset of all 20 survey days)				
Both species	15	1.44	0.04	

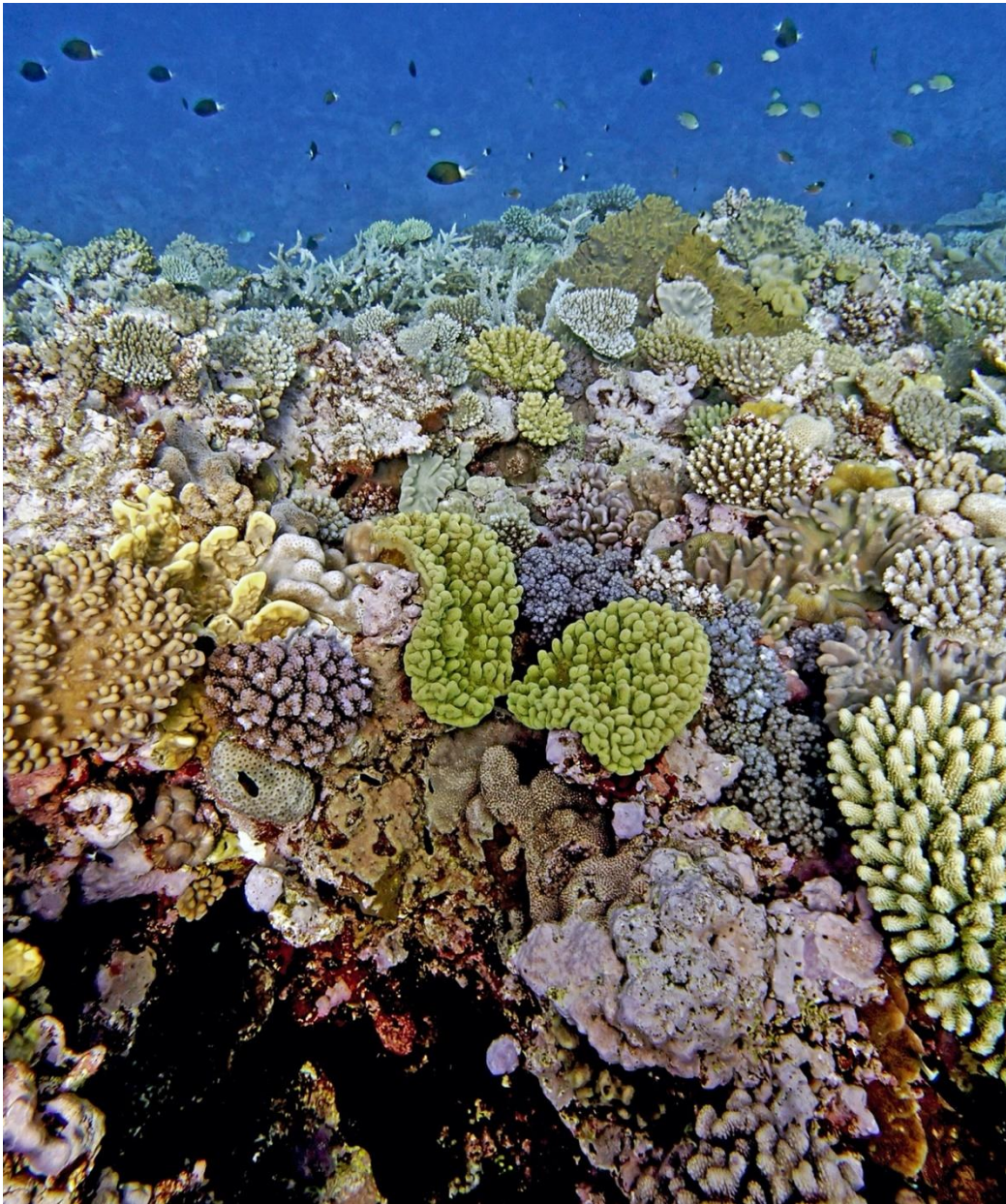
**Table 3.** Non-parametric MANOVA of the first three multi-dimensional scaling dimensions (stress=0.28 and 0.30 for *Pocillopora damicornis* and *Pocillopora acuta*, respectively) derived from Euclidean distances among coral samples (assessed from multivariate assessment of 12 coral response variables: maximum colony length, mean color score, Symbiodiniaceae GCP, RNA/DNA ratio, and expression levels of eight genes [Table 1]). For environmental factors (ENV) featuring only two groupings (e.g., island), the exact *F* statistics have been shown; all others are approximations (“approx. *F*”). Statistically significant findings ( $p < 0.01$ ) have been highlighted in **bold**.

ENV	df	F	p	Trend
<i>P. acuta</i> only	15	1.69	<b>0.01</b>	
<i>P. damicornis</i> only	15	0.62	0.92	
<b>Sampling time:</b> morning, midday, or afternoon				
Both species	2	2.56	0.02	
<i>P. acuta</i> only	2	3.34	<b>&lt;0.01</b>	morning(a)=midday(ab)= afternoon(b)
<i>P. damicornis</i> only	2	0.63	0.70	
<b>Depth</b> (m): scaled continuously				
Both species	79	0.87	0.46	
<i>P. acuta</i> only	79	1.06	0.37	
<i>P. damicornis</i> only	79	0.04	0.99	
<b>Temperature</b> (°C): scaled continuously				
Both species	16	0.74	0.53	
<i>P. acuta</i> only	16	3.14	0.03	
<i>P. damicornis</i> only	16	3.53	0.03	
<b>Salinity</b> (unitless): scaled continuously (only four unique values were documented: 35.2, 35.3, 35.4, and 35.6)				
Both species	3	3.10	0.03	
<i>P. acuta</i> only	3	2.38	0.08	
<i>P. damicornis</i> only	3	1.08	0.38	
<b>Symbiodiniaceae assemblage:</b> <i>Cladocopium</i> only or mixed <i>Cladocopium</i> + <i>Symbiodinium</i>				
Both species	1	1.73	0.17	
<i>P. acuta</i> only	1	0.98	0.41	
<i>P. damicornis</i> only	1	1.80	0.18	
<b>Polyp extension:</b> yes vs. no				
Both species	1	0.42	0.74	
<i>P. acuta</i> only	1	0.57	0.64	
<i>P. damicornis</i> only	1	0.42	0.74	

**Table 4. Predictors of the coral health index (CHI).** All continuous data were standardized prior to analysis, and the mean color score, qualitative color score (“color” in **Table 1** and **Figure 2**), Mahalanobis distance, heat map score, and variability index were excluded as CHI predictors since they were used to calculate this value; this resulted in 13 coral physiological data (CPD) parameters assessed in total. A kfold validation of five was used when the CPD were the predictors; validation columns were used for the ENV and ECO comparisons (training validation of 17/6 for *P. damicornis* [n=23 total] and 35/12 for *P. acuta* [n=47 total]) unless denoted by “-kfold” (kfold of 5). To determine the “Most important predictor,” a “independent resampled inputs” algorithm was used; when the CPD were the predictors, JMP Pro’s “predictor screen” was instead employed. Environmental data (ENV; n=14 parameters unless noted otherwise [e.g., 10 parameters “ENV-10”]). Ecological (i.e., benthic) data (ECO; n=82 parameters [**Table 1**]). When a neural network’s (NN) validation  $R^2$  was <0.80, a NN model-tuning GUI was run to generate an additional 100 models (minimum) in an effort to enhance the predictive power. Sym=Symbiodiniaceae.

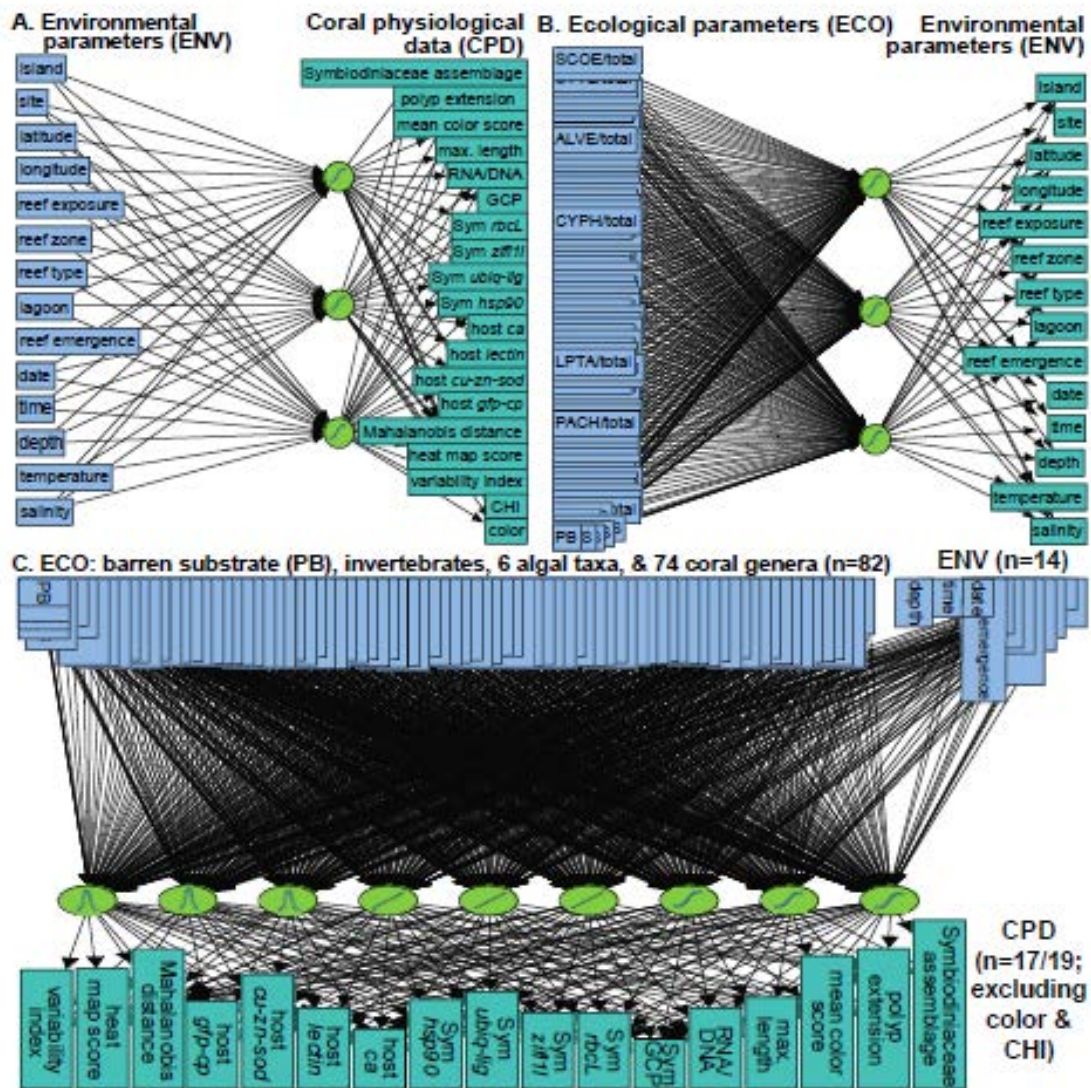
Coral host(s)	X	Model type	Validation $R^2$	Model details	Most important predictor
Both species	CPD	Neural	0.86	NTanH(3)NBoost(20)	Sym <i>zifl11</i>
<i>P. acuta</i>	CPD	Neural	0.83	NTanH(3)NBoost(20)	Sym <i>zifl11</i>
<i>P. damicornis</i>	CPD	Neural	0.97	NTanH(3)NBoost(20)	Host <i>gfp-cp</i>
Both species	ENV	Neural	0.75	NGaussian(3)NBoost(20)	Date
Both species	ENV	Neural-GUI	0.92	NTanH(3)NGaussian(1)NBoost(15)	Salinity ( <b>Figure 6B</b> )
Both species-kfold	ENV-10	Neural-GUI	0.91	NTanH(4)NGaussian(4)	Depth
<i>P. acuta</i>	ENV	Neural	0.45	NTanH(3)NBoost(20)	Site
<i>P. acuta</i>	ENV	Neural-GUI	0.51	NTanH(1)NTanH2(1)NLinear2(2)NGaussian2(2)	Date
<i>P. damicornis</i>	ENV	Neural	0.84	NTanH(3)NBoost(20)	Depth
<i>P. damicornis</i>	ENV	Neural-GUI	1.00	NGaussian(3)NBoost(20)	Site
Both species	ECO	Neural	0.61	NTanH(3)NBoost(20)	<i>Montipora</i> cover
Both species	ECO	Neural-GUI	0.93	NTanH(1)NLinear(3)NGaussian(4)NBoost(20)	<i>Montipora</i> cover ( <b>Figure 6C</b> )
<i>P. acuta</i>	ECO	Neural	0.36	NTanH(3)NBoost(20)	<i>Acanthastrea</i> cover
<i>P. acuta</i>	ECO	Neural-GUI	0.63	NTanH(3)NLinear(4)NBoost(3)	<i>Acanthastrea</i> cover
<i>P. acuta</i> -kfold	ECO	Neural-GUI	0.99	NTanH(1)NLinear(4)NGaussian(3)NTanH2(4)-NGaussian2(4)	<i>Caulastrea</i> cover
<i>P. damicornis</i>	ECO	Neural	0.87	NTanH(3)NBoost(20)	<i>Favites</i> cover
<i>P. damicornis</i>	ECO	Neural-GUI	1.00	NTanH(2)NBoost(7)	<i>Coelastrea</i> cover
Both species	ENV+ECO	Neural	0.71	NTanH(3)NBoost(20)	Depth
Both-kfold	ENV+ECO	Neural	0.65	NTanH(3)NBoost(20)	Site
Both species	ENV+ECO	Neural-GUI	0.96	NTanH(1)NLinear(3)NGaussian(4)NBoost(6-8)	Site ( <b>Figure 6D</b> )
Both species	ENV-10+ECO	Neural-GUI	0.92	NTanH(1)NGaussian(1)NBoost(3)	Reef type
<i>P. acuta</i>	ENV+ECO	Neural	0.50	NTanH(3)NBoost(20)	<i>Acanthastrea</i> cover
<i>P. acuta</i> -kfold	ENV+ECO	Neural	0.85	NTanH(3)NBoost(20)	Site
<i>P. acuta</i>	ENV+ECO	Neural-GUI	0.67	NGaussian(3)NBoost(20)	Site
<i>P. acuta</i> -kfold	ENV+ECO	Neural-GUI	0.93	NTanH(4)NGaussian(4)	Date
<i>P. damicornis</i>	ENV+ECO	Neural	0.86	NTanH(3)NBoost(20)	Date
<i>P. damicornis</i> -kfold	ENV+ECO	Neural	0.98	NTanH(3)NBoost(20)	Date
<i>P. damicornis</i>	ENV+ECO	Neural-GUI	0.97	NTanH(2)NBoost(20)	Date
<i>P. damicornis</i> -kfold	ENV+ECO	Neural-GUI	1.00 <sup>a</sup>	NTanH(1)NLinear(1)NGaussian(3)NBoost(2)	Site

<sup>a</sup>numerous models were characterized by  $R^2$  values of 1.00; the most parsimonious has been presented here.

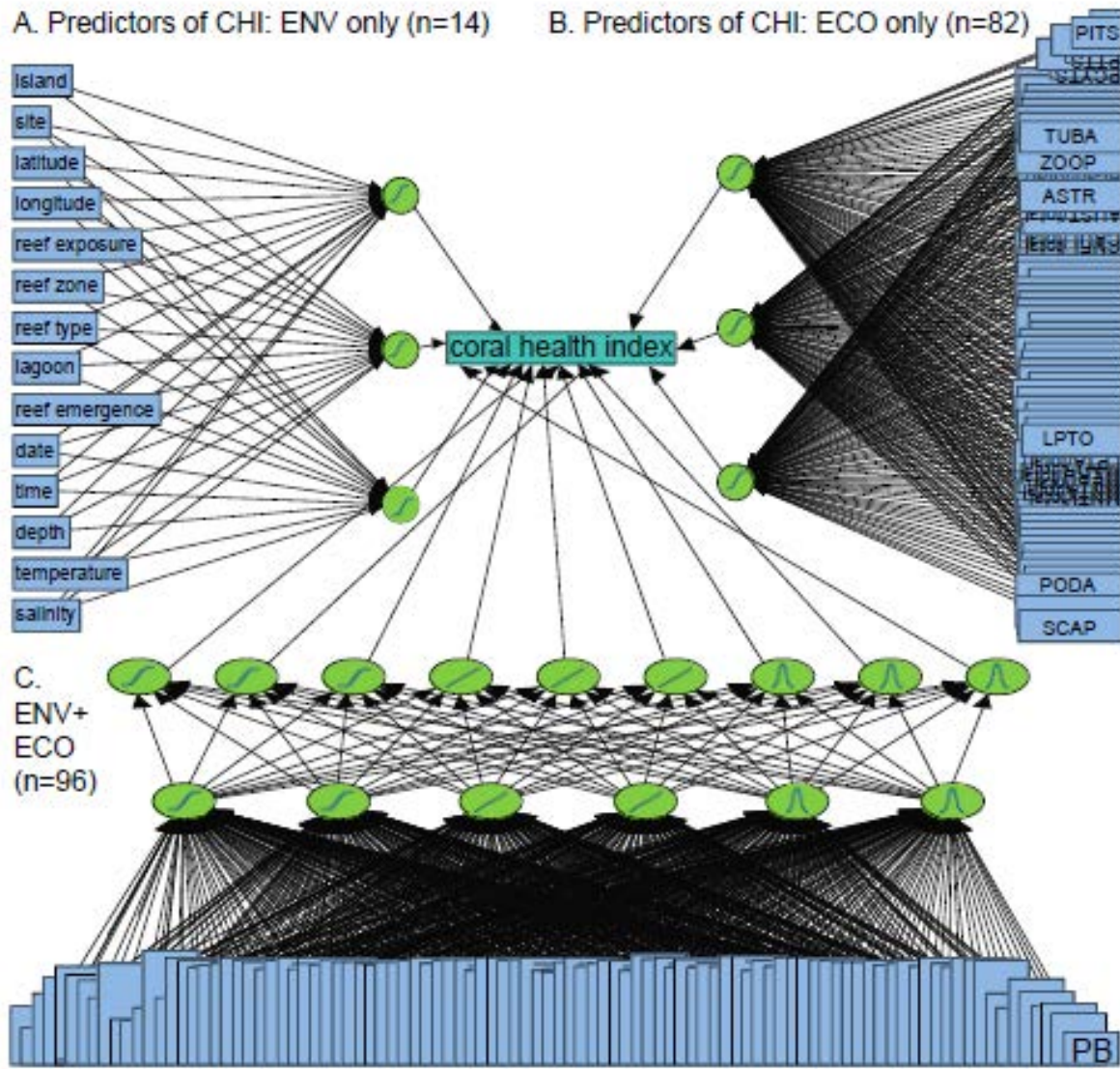


**Figure 1.** A high coral cover reef in the Kingdom of Tonga. Photo credit=ABM.



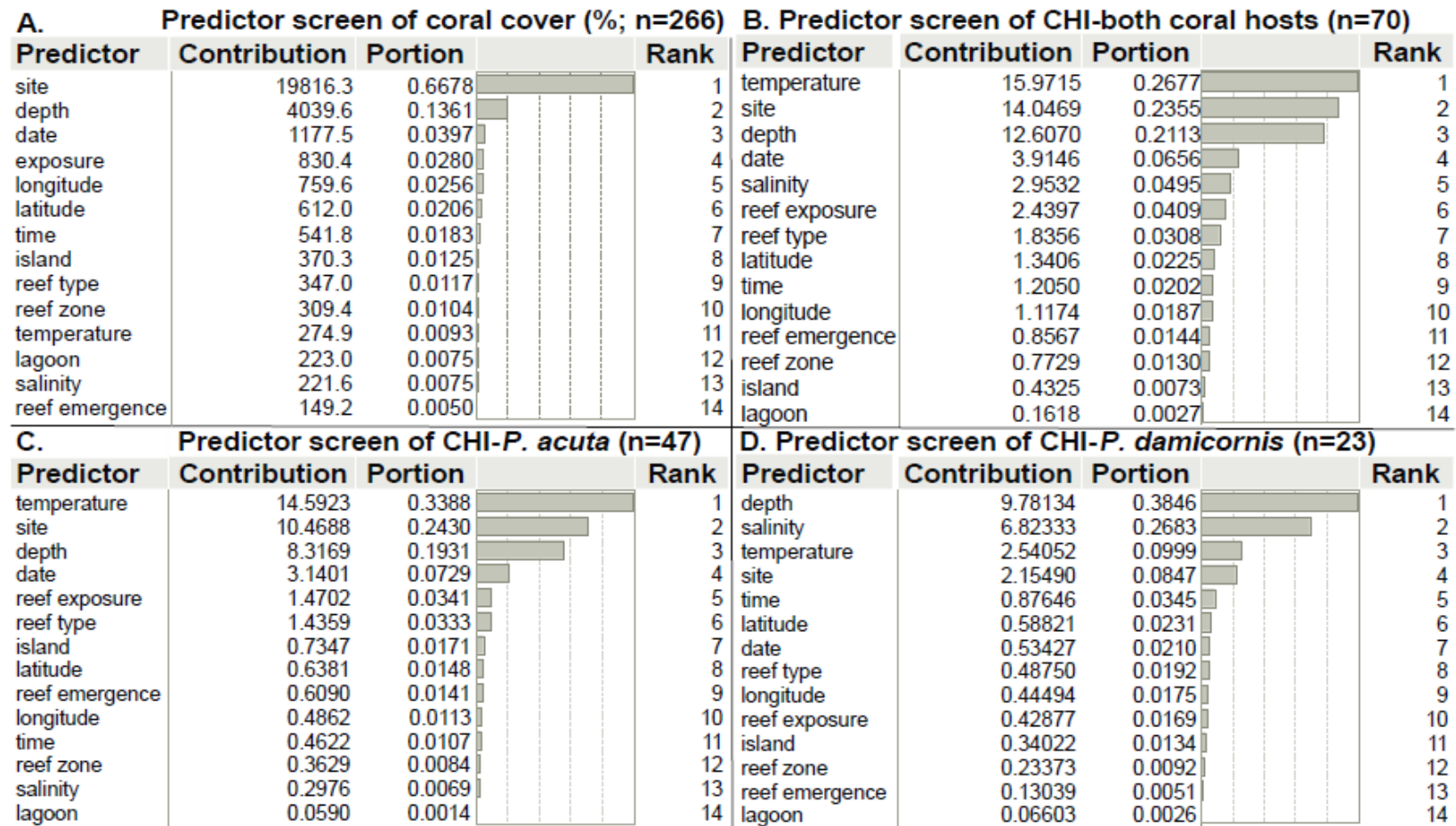


**Figure 2. Neural networks (NN) showing relationships among coral physiological data (CPD), environmental (ENV), and ecological (ECO) parameters.** A. A hypothetical NN with a single hidden layer and three sigmoidal activation nodes for linking ENV (n=14) and CPD (n=19). B. A hypothetical NN with a single hidden layer and three sigmoidal activation nodes for linking ENV with ecological (ECO; i.e., benthic) data (n=82 categories: barren substrate, invertebrate cover, six algal taxa, and 74 coral genera [all in % of total benthos]). C. A hypothetical NN featuring a single hidden layer with three radial, three linear, and three sigmoidal activation nodes for modeling CPD data (n=17/19 parameters [excluding the qualitative color score and the coral health index (CHI)]) with both ENV and ECO data (n=96 parameters in total). Sym=Symbiodiniaceae.



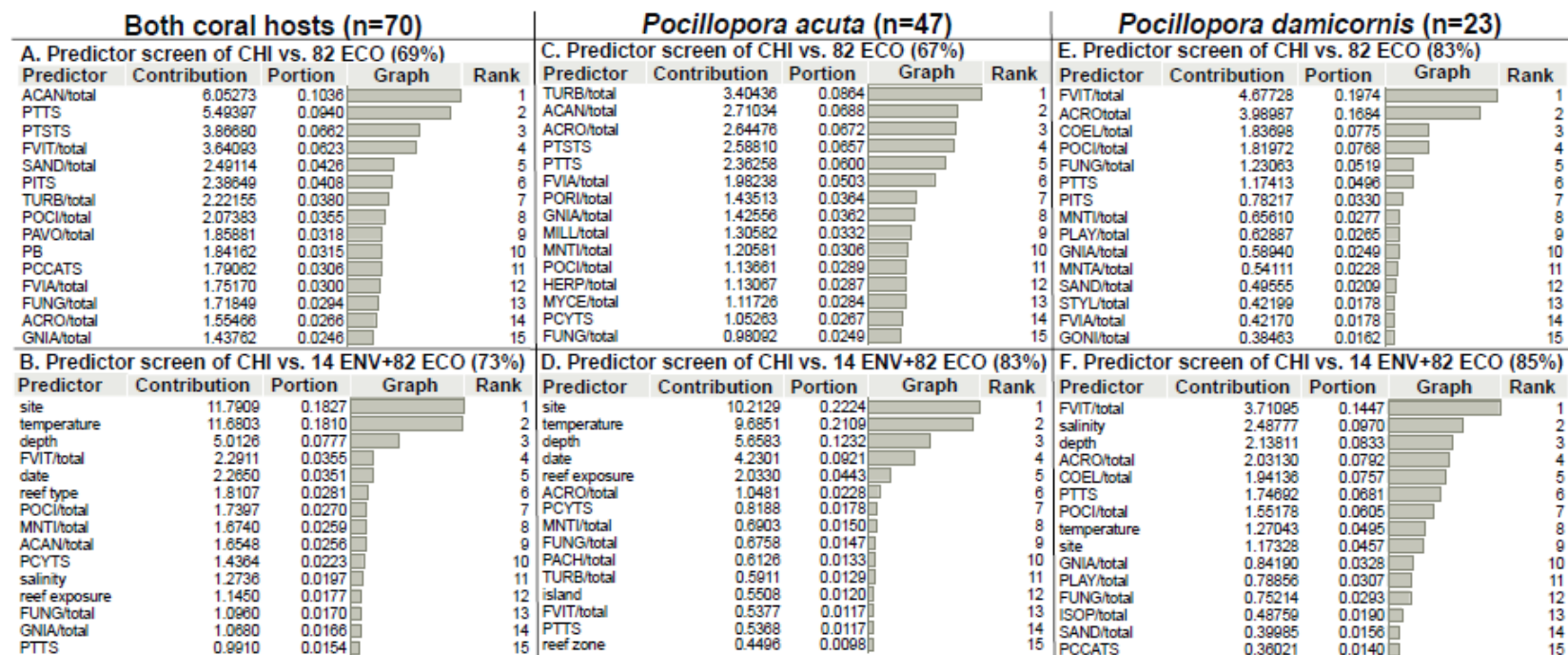
**Figure 3. Neural networks (NN) for predicting the coral health index (CHI) from environmental (ENV) and/or ecological (ECO) data.** A. A hypothetical NN with one hidden layer and three sigmoidal activation nodes for using ENV (n=14) to predict the CHI. B. A hypothetical NN with a single hidden layer and three sigmoidal activation nodes for predicting the CHI from ECO (n=82). C. A hypothetical NN with two hidden layers for predicting the CHI from a combination of 14 ENV+82 ECO (96 parameters). The first hidden layer features two sigmoidal, linear, and radial (i.e., Gaussian) activation nodes, while the second includes three each of these activation types (15 total nodes). In the case of ECO (B-C), only select parameters have been shown due to spatial constraints. PB=percent barren substrate and PITS=percent invertebrate cover. Please consult the online supplemental data file for additional benthic category abbreviations.



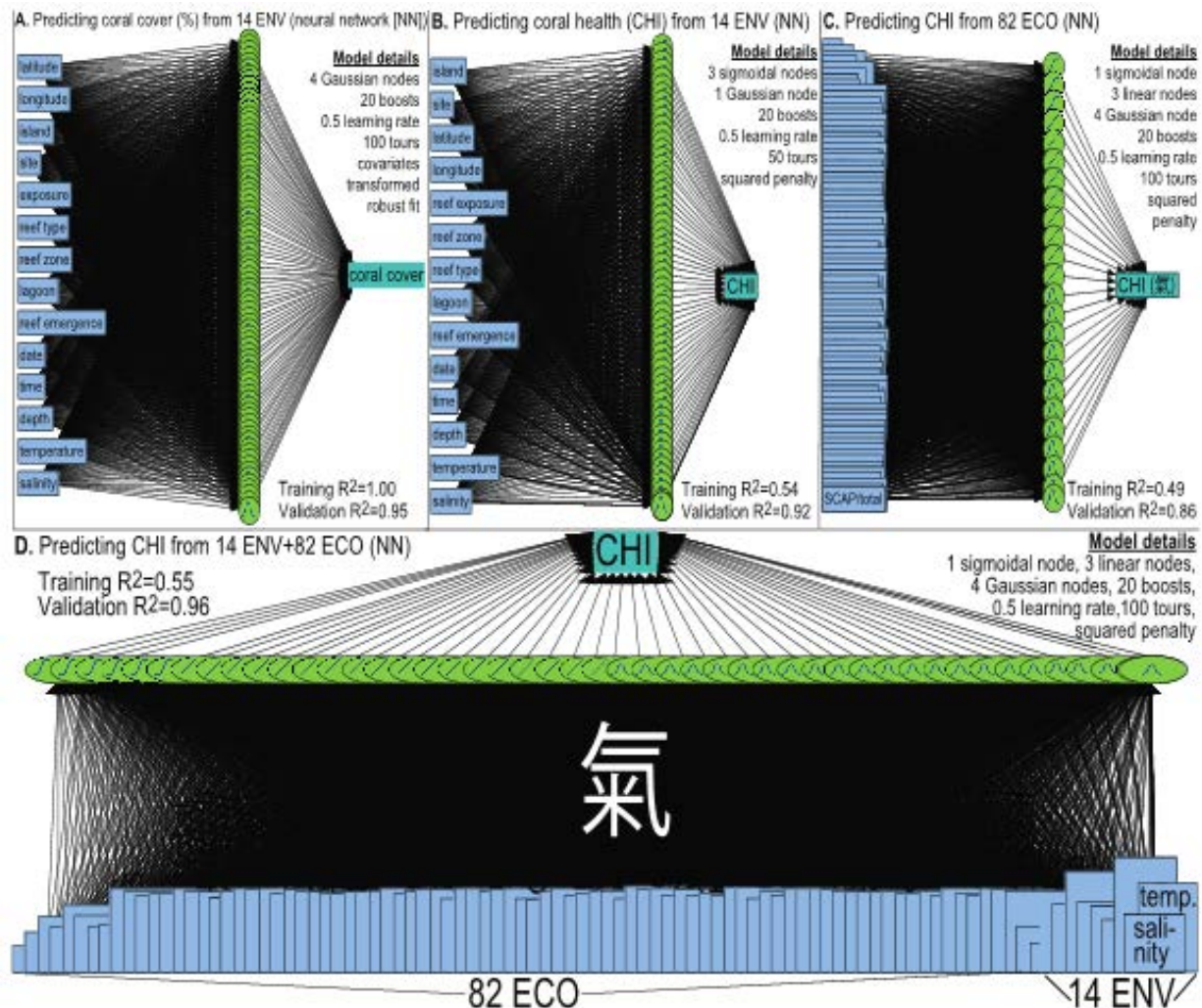


**Figure 4. Predictor screening analysis-I: accommodating variation in percent coral cover and the coral health index (CHI) from environmental data.** Separate screens were carried out for percent coral cover (A) and the CHI (B), the latter also shown individually for the two target species: *Pocillopora acuta* (C) and *Pocillopora damicornis* (D).

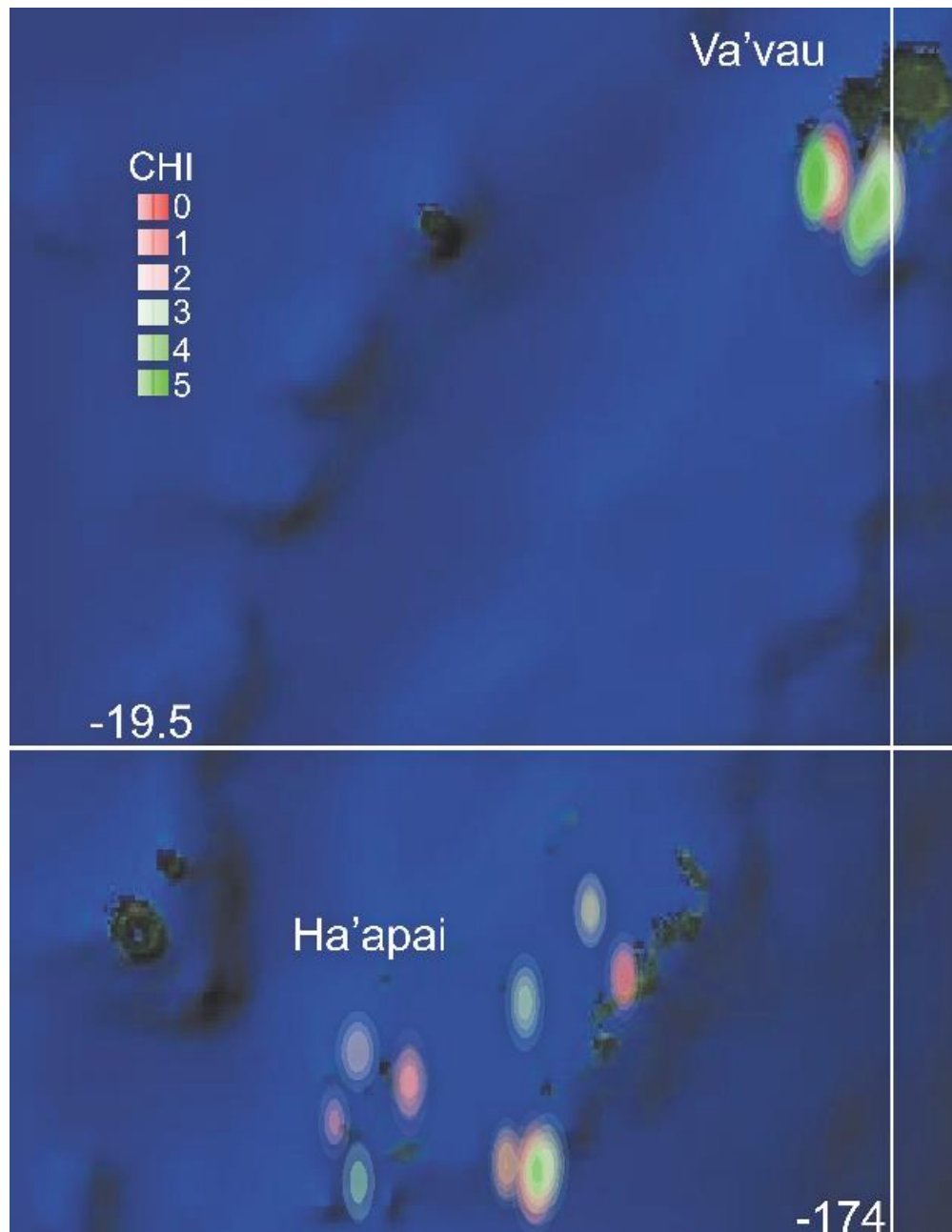




**Figure 5. Predictor screening analysis-II: accommodating variation in the coral health index (CHI) from environmental (ENV) and ecological (ECO) data.** Separate screens were carried out for the 82 ECO parameters alone (A, C, and E) and the 14 ENV+82 ECO (96 total) parameters (B, D, and F) for both coral hosts (n=70; A and B, respectively), *Pocillopora acuta* only (n=47; C and D, respectively), and *Pocillopora damicornis* only (n=23; E and F, respectively), and only the top 15 parameters have been shown; the total variation explained by these 15 parameters is included in parentheses and ranged from 67 to 85%. Please see the online supplemental data file for benthic category abbreviations.



**Figure 6. Machine-learning models for predicting percent coral cover (A) and the coral health index (CHI; 氣; B-C).** In A-B, the 14 environmental (ENV) parameters were the predictors, whereas in C-D, both the 14 ENV and the 82 ecological (ECO) parameters were the models' X's; the latter have generally been masked due to spatial constraints, though a select ECO has been shown in C (SCAP=*Scapophyllia*). Unless otherwise mentioned, covariates were not transformed, nor were robust fit methods employed. The neural network (NN) model terms necessary to reproduce the data have been shown, though please note that, because multiple tours were used in all models, re-run analyses will be characterized by slightly different  $R^2$  values. temp.=temperature.



**Figure 7. Map of Ha'apai and Va'vau with coral health index (CHI) values overlaid as averaged contours.** There was a marginally statistically significant effect of island on the CHI (student's  $t$ -test,  $p=0.02$ ), with mean values of 2.0 and 2.9 for Ha'apai and Va'vau, respectively (CHI=0 and 5 signify highly stressed and healthy corals, respectively).