Platax 15: 1-35, 2018

A statistical platform for evaluating coral health in an era of changing global climate-I: a case study from Fiji's Lau Archipelago

Anderson B. Mayfield[1-2*], Alexandra C. Dempsey[3], Javed Inamdar[4], & Chii-Shiarng Chen[1-2, 5-6]

[1]Taiwan Coral Research Center, Checheng, Pingtung, Taiwan

[2]National Museum of Marine Biology and Aquarium, Checheng, Pingtung, Taiwan

[3]Khaled bin Sultan Living Oceans Foundation, Annapolis, Maryland, USA

[4]SAS Research and Development, Pune, India

[5]Department of Marine Biotechnology, National Dong-Hwa University, Checheng, Pingtung, Taiwan

[6]Department of Marine Biotechnology and Resources, National Sun Yat-Sen University, Kaohsiung, Taiwan

*corresponding author E-mail: andersonblairmayfield@gmail.com

Abstract

Given the significant threats against coral reef ecosystems, there is an urgent need to develop the capacity to make predictions as to which coral reefs are most stress-susceptible, as well as which and are most resilient. However, there is such extensive variation in coral physiology, even in conspecifics reared in the same laboratory tank, that prior works have been characterized by too low statistical power to even explain previously obtained datasets with confidence, let alone predict the behavior of to-be-sampled corals. To obtain a better grasp of the environmental and organismal factors that contribute to variation in coral physiology, a published coral reef dataset from Fiji's remote, understudied Lau Archipelago was re-analyzed herein with a variety of both univariate and multivariate statistical approaches. Of the 12 environmental parameters hypothesized to influence reef coral physiology, only two both significantly drove variation in the multivariate coral physiological response *and* featured readily in the best-fit models produced by stepwise regression and partial least squares analyses: island and host coral

species. That being said, the majority of the models were characterized by low predictive capacity; more data are clearly needed to generate statistical algorithms capable of forecasting coral behavior with confidence in this era of rapidly changing global climate.

Keywords: *biomarkers; coral reefs; dinoflagellates; invertebrate physiology; molecular biology; multivariate statistics; predictive modeling; South Pacific*

Abbreviations. Please see Table 3 for target gene abbreviations.

Akaike's information criterion=AICc	Multivariate ANOVA=MANOVA
Austral Islands+Cooks Islands=Australs-Cooks or A-C	Multivariate statistical approaches=MSA
Average live coral cover=ALCC	Not applicable=NA
Analysis of variance=ANOVA	Not statistically significant=NS
Bayesian information criterion=BIC	Ocean acidification=OA
Canonical correlation analysis=CCA	Partial least squares=PLS
Carbon dioxide partial pressure=pCO_2	Permutational ANOVA=PERMANOVA
Discriminant analysis=DA	Photosynthetically active radiation=PAR
Distance-based linear modeling=DistLM	Predicted residual sum of squares=PRESS
Environmental parameter=EP	Principal components analysis=PCA
Euclidean distance matrix=EDM	Principal components analysis primary axis loading score=PC axis 1 or PC1
Genome copy proportion=GCP	Principal coordinates ordination=PCO
Global climate change=GCC	Response variable=RV
Global Reef Expedition=GRE	Ribulose-1,5-bisphosphate carboxylase/ oxygenase (large subunit)=RuBisCO
Green fluorescent protein=GFP	Standard deviation=std. dev.
Heat map score=HMS	Stepwise regression=SR
Living Oceans Foundation=LOF	Stress-targeted gene=STG
Mahalanobis distance=MD	*Symbiodinium*=Sym
Maximum length=max. length	Temperature=temp.
Mitochondrial open reading frame=mORF	Trans-generational=trans-gen
Molecular-physiological response variable=MPRV	Variability index=VI
Multi-dimensional scaling=MDS	Variable importance parameter=VIP

Introduction

As global temperatures continue to rise, Earth's coral reefs are becoming ever more imperiled (Hoegh-Guldberg et al., 2007; Mayfield & Gates, 2007; Hughes et al., 2018); there is consequently an urgent need to make data-driven predictions regarding coral health. Upon having carried out a plethora of laboratory studies on the responses of two model Indo-Pacific reef coral species (Table 1), *Seriatopora hystrix* (e.g., Mayfield et al., 2011) and *Pocillopora damicornis* (likely to have actually been *P. acuta* in the majority of these studies; e.g., Mayfield et al., 2013a-b, d), we are now beginning to develop an understanding of the responses of reef-building scleractinians, which associate with photosynthetically active dinoflagellates of the genus *Symbiodinium* (Mayfield et al., 2014c), to global climate change (GCC) scenarios.

From Table 1, it is clear that ocean acidification (OA) is not a major threat to stony corals of Taiwan's Hengchun Peninsula. However, prolonged exposure to temperatures above 31℃ can elicit bleaching (Mayfield et al., 2013a), as has been documented in countless other locations across the planet (e.g., Lesser, 1997; Putnam et al., 2017; Sheppard et al., 2017). Some corals, however, *do* resist bleaching and instead acclimate to temperatures that lead to thermal stress in corals elsewhere (Barshis et al., 2013; Krueger et al. 2017). We do not yet possess a clear understanding of the cellular biology underlying the marked capacity of acclimation of some corals. One reason for this knowledge dearth is because the majority of works on the cellular and molecular mechanisms of coral acclimation to environmental change have focused on gene expression (e.g., Kenkel et al., 2011, Kenkel & Matz, 2016); however, as there is no correlation between gene expression and protein concentration in reef-building corals or *Symbiodinium* (Mayfield et al., 2016b-c, 2018a-b), mRNA data cannot be used to model or infer protein behavior in coral cells. Doing just this is, unfortunately, still common practice in marine biology, and particularly in reef coral biology (e.g., Palumbi et al., 2014). This means that the vast majority of the reef coral-*Symbiodinium* mRNA data gathered are of little utility in explaining coral cell biology.

Despite the fact that we, as a field, do not yet have a comprehensive handle on the molecular basis of coral acclimation to, for instance, high temperatures, it is nevertheless possible that we can exploit the molecular- and physiological-scale data acquired from past controlled tank experiments to make *predictions* about

Table 1. Global climate change manipulation studies carried out at Taiwan's National Museum of Marine Biology and Aquarium. Please note that, in most cases, *Pocillopora acuta* was inadvertently classified as *P. damicornis* in the published manuscripts. In certain experiments, more than just the two target model species (*Seriatopora hystrix* and *P. acuta*) listed in the table were used in experiments (e.g., *Acropora nana* and *Porites lutea* in McRae, Mayfield et al., in prep.). Table abbreviations: NA=not applicable. ppm=parts per million (μatm). Temp.=temperature. Trans-gen=trans-generational (adults=>larvae=>recruits).

Target species	**Life history stage**	**Temp. treatment**	**High pCO_2 (ppm)**	**Salinity effects tested?**	**Light effects tested?**	**Nutrient effects tested?**	**Time-scale**	**Acclimation?**	**Reference(s)**
S. hystrix	adult	27 vs. 30℃	NA	no	no	no	hours	yes	Mayfield et al., 2011, 2014b
S. hystrix	adult	26 vs. 23-29℃ over 6-hr	NA	no	no	no	days	yes	Mayfield et al., 2012a, 2013c, 2014b, 2016c, 2018a
P. acuta	adult	NA	NA	no	yes	no	hours	yes	Mayfield et al., 2010, 2012b
P. acuta	adult	NA	NA	yes	no	no	hours	yes	Mayfield et al., 2013d
P. acuta	larvae	26 vs. 29℃	415 vs. 635	no	no	no	days	yes	Putnam et al., 2013
P. acuta	adult	26 vs. 29℃	415 vs. 850	no	no	no	weeks	yes	Putnam, Mayfield et al., in prep.
P. acuta	adult	26.5 vs. 29.7℃	NA	no	no	no	months	yes	Mayfield et al., 2013b, 2014d, 2018b
P. acuta	adult	31.5℃-return to ambient at night	NA	no	no	no	weeks	yes	Mayfield et al., 2013a
P. acuta	adult	31.5℃-sustained	NA	no	no	no	weeks	no-bleached	Mayfield et al., 2013a, 2014a
P. acuta	adult	25℃	400 vs. 800	no	no	no	months	yes	Liu, Mayfield et al., in review.
S. hystrix	adult	25℃	400 vs. 800	no	no	no	months	yes	Liu, Mayfield et al., in prep.
P. acuta	adult	25, 28, or 31℃	400 vs. 800	no	no	yes	months	yes	Liu, Mayfield et al., in prep.
P. acuta	adult	26 vs. 29.5℃	NA	no	no	no	months	yes	McRae, Mayfield et al., in prep.
P. acuta	adult	26 vs. 32℃	NA	no	no	no	hours	yes	McRae, Mayfield et al., in prep.
P. acuta	trans-gen	26 vs. 29.5℃	NA	no	no	no	months	yes	McRae, Mayfield et al., in prep.

coral behavior *in situ*. Indeed, in most cases, aquarium experiments are undertaken with the explicitly expressed purpose of gaining a better understanding of how marine animals will fair in the oceans, yet explanatory data acquired during such laboratory experiments are rarely used in a predictive setting. This may be due to the fact our collective ability to explain previously documented observations or datasets is not necessarily associated with a commensurate ability to predict behavior (Shmueli, 2011). As an example, we have billions of data on the United States stock market and can adeptly *explain* past market phenomena; however, such explanatory data are of little use in *predicting* future market behavior, and technical analysis consistently underperforms the act of selecting stocks at random in building an ideal portfolio. Of course, there are plentiful examples of when congruency *does* exist between explanation and prediction; international websites such as Google, Facebook, and Amazon have acquired massive datasets on consumer behavior and have used these data to successfully predict, at least in the case of Amazon, what consumers may desire to buy in the future.

It is currently unclear whether our ability to explain coral behavior in the laboratory or in the ocean (i.e., from previously acquired field datasets) is associated with a commensurate capacity to predict the phenotypes or physiologies of corals in other environments. If we could use data from laboratory exposures (*sensu* Table 1) and/or published field datasets (e.g., Mayfield et al., 2015) on reef coral physiology to make predictions of how conspecifics would respond to environmental heterogeneity in other locations (or in the same study locations at later dates), then we would likewise possess the capacity to determine which reefs (and/or coral colonies/populations) are most likely to persist in the face of GCC. The statistical package JMP® (version 14; Cary, NC, USA; www.jmp.com) features a plethora of modeling programs for testing the ability of hitherto obtained datasets to predict future organismal responses. Among the approaches at our disposal as statisticians is partial least squares (PLS), which is typically used in cases, such as all coral molecularly-focused datasets, in which many response variables co-vary (Chen et al., 2015).

Not only could PLS be used to develop predictive models for coral behavior, but it could also be used in conjunction with other approaches (e.g., stepwise regression [SR] and principal components analysis [PCA]) to determine the environmental parameters (EP), or

combinations thereof, that best account for spatio-temporal variation in coral physiology. This would address the additional need of uncovering just why coral physiology varies so substantially (Louis et al., 2017; Parkinson et al., 2018), even across colonies of the same source population exposed to very similar environmental conditions (Mayfield, 2016); indeed, this issue has thwarted progress in the field due to the low statistical power of the data from the associated experiments and/or field datasets (Mayfield et al., 2009). Therefore, to both 1) attempt to develop predictive models for gauging future coral performance *in situ* and 2) uncover the EP that most drive variation in coral physiology, a published dataset from Fiji's Lau Archipelago (Mayfield et al., 2017b), which was acquired during the Khaled bin Sultan Living Oceans Foundation's (LOF) "Global Reef Expedition" (GRE; the largest coral reef survey ever undertaken; see Mayfield et al., 2017a for details.), was re-explored with a variety of a univariate and multivariate statistical approaches (MSA). Based on our prior work in the Austral and Cook Islands (Mayfield et al., 2016a), we hypothesized that host species would contribute significantly to physiological variation in the colonies sampled across this remote, understudied, South Pacific frontier province (Fig. 1).

Materials and Methods

Description of the dataset. Details of the June, 2013 research expedition to Fiji's Lau Archipelago can be found in Mayfield et al. (2017b), and the target coral was the model reef coral *P. damicornis* (Traylor-Knowles et al., 2011). Although comprehensive coral reef assessments had never before been undertaken in the Lau group at the time of surveying, readers interested in the geology and archaeology of this understudied South Pacific region, as well as a treatise on human-ocean interactions, should consult Jones (2009). In total, 70 reef sites were surveyed, and pocilloporid corals were sampled from 47. Of the 153 sampled colonies, 91 were genotyped (Mayfield et al., 2017b), and 90 (from 34 sites) were analyzed for the molecular-physiological response variables (MPRV) discussed below (all but 5 of these 90 were genotyped.). A variety of EP (Table 2) were assessed at each of the reefs surveyed to attempt to uncover which factors are most important in driving physiological differences between coral colonies, and the data for several such continuous EP (namely temperature, salinity, and average live coral cover [ALCC]) have been plotted in Fig. 1. As

outlined in Table 2, only 12 of the 15 EP assessed were generally considered in the MSA discussed below; reasons for the exclusion of the other 3 can be found in Table 2. Likewise, of the 13 response variables (RV) assessed in each of the 153 sampled colonies, only 11 were included in the MSA herein; reasons for exclusion of the remaining 2 can be found in Table 3, and the spatial distribution of the data for the 11 MPRV (as well as maximum [max.] colony length), can be found in Fig. 2.

Since most MSA are highly sensitive to missing data, only those 70 coral samples for which no data were missing were included in the statistical analyses outlined below. Although the target species was *P. damicornis*, *P. acuta* was synonymized with *P. damicornis* at the time of surveying; the two species were not formally distinguished until the publication of Schmidt-Roach et al. (2014). Likewise, although most sampled colonies appeared as *P. damicornis in situ*, such was not always confirmed upon genotyping (Mayfield et al., 2017b), and, of this 70-sample subset, 1, 18, 6, 26, and 19 were *P. brevicornis* (1.5%), *P. verrucosa* (26%), *P. meandrina* (8.5%), *P. acuta* (37%), and *P. damicornis* (27%), respectively. For this reason, and because we hypothesized that physiological differences may exist between these closely related species, "host species/genotype" was included as an EP (Table 2). Since only one *P. brevicornis* was genotyped, it was excluded from the analysis; the resulting final sample size was 69.

Data analysis overview. In addition to providing baseline data for an under-surveyed region of the South Pacific, we were interested in identifying the EP (Table 2) that are most important in driving variation in the physiological response (Table 3) of pocilloporid corals, and a variety of both univariate and multivariate statistical approaches were employed to achieve this goal. For all MSA, RV data were standardized (i.e., converted to *z*-scores) prior to analysis such that all would be on the same scale. Multivariate ANOVA (MANOVA) and ANOVA were initially used to analyze the data (Mayfield et al., 2017b), though they were ultimately found to be inappropriate methods for testing effects of environment on coral physiology with this dataset due to, amongst other issues, the heteroskedastic, non-normally distributed nature of the underlying data.

Table 2. Environmental parameters (EP) assessed at each survey site. In total 12 of the 15 EP were incorporated into the statistical models discussed in the text, and the text colors associated with these 12 EP are used throughout the manuscript's tables and figures (though see exceptions in Figs. 4-5.). Site, sampling date, and *Symbiodinium* assemblage were originally hypothesized to be EP that could affect coral physiology, but there were typically too few corals sampled from any one site, or on the same day, to include the former two, and nearly all sampled colonies hosted *Symbiodinium* of clade C exclusively; these three EP were consequently excluded from statistical analyses. Nearly all sites were virtually devoid of sea cucumbers, which had been harvested to near-local extinction by Chinese fishing vessels (see LOF field report for details: https://www.livingoceansfoundation.org/global-reef-expedition/pacific-ocean/fiji-islands/new-report-on-fijis-reefs/). Table abbreviations: ALCC=average live coral cover. NA=not applicable. PAR=photosynthetically active radiation. STG=stress-targeted gene. Temp.=temperature.

EP	#bins/categories	Hypothesis
1. Island	9 islands: Totoya, Matuku, Moala, Fulaga, Kabara, Tuvuca, Cicia, Mago, Vanua Balavu	Significant effect of island[a]
2. Reef exposure	3 categories: protected, intermediate, or exposed	Significant effect of exposure[a]
3. Reef type	4 types: barrier reef, fringing reef, patch reef, or pinnacles	Significant effect of reef type[a]
4. Reef zone	2 zones: forereef vs. lagoon	Significant effect of reef zone[a]
5. ALCC (%)	5 bins: 10-20, 20-30, 30-40, 40-50, or >50%	Significant effect of ALCC[a]
6. Temp. (℃)	2 bins: 26-27 or 27-28℃	Significant effect of temp.[b]
7. Salinity (unitless)	3 bins: 34.7-35.0, 35.0-35.3, or 35.4-35.6	Significant effect of salinity[a]
8. PAR (µmol m^{-2} s^{-1})	4 bins: <50, 50-100, 100-200, or >200 µmol $m^{-2}s^{-1}$	Significant effect of PAR[b]
9. Sampling time	3 bins: <10:00, 10:00-14:00, or >14:00	Significant effect of time[b]
10. Coral colony depth	7 bins: <5, 5-10, 10-15, 15-20, 20-25, 25-30, or >30 m	Significant effect of depth[a]
11. Coral host genotype[c]	4 species: *P. damicornis*, *P. acuta*, *P. verrucosa*, or *P. meandrina*[d]	Significant effect of host[a]
12. Coral colony color[c]	4 colors: normal, pale, very pale, or bleached	Higher STG expression in bleached corals[b]

[a]confirmed by the majority of statistical approaches (Tables 4-6). [b]not statistically supported by the majority of statistical approaches (Tables 4-6). [c]technically a property of the sampled colony (rather than the environment) but nevertheless hypothesized to influence coral physiology. [d]The lone *P. brevicornis* colony sampled was excluded from the statistical analysis (see main text.).

Table 3. Response variables (RV) assessed in the sampled pocilloporid coral colonies. Of the 13 RV assessed in each of the 153 sampled coral colonies, only 12 were generally included in the suite of univariate and multivariate statistical analyses (MSA) described in this article. Of these 12, only the 11 molecular-physiological response variables (MPRV; i.e., excluding maximum [max.] length) were included in the MSA. All RV data film have been overlaid onto images of the sampled coral colonies on http://coralreefdiagnostics.com/lau-archipelagooverview/. Links to mitochondrial open reading frame sequences (mORF: used for genotyping), which are hosted on the National Center for Biotechnology Information web server, have also been overlaid onto these images. Table abbreviations: GCP=genome copy proportion. GFP=green fluorescent protein. NA=not applicable. RuBisCO=ribulose-1,5-bisphosphate carboxylase/oxygenase (large subunit).

Response variable	**Unit/ abbreviation**	**Proxy**	**Included in analyses?**	**Reason for exclusion**	**Hypothesis**
Size and biological composition response variables (n=4)					
1. Max. length	cm	size	No	Size not hypothesized to be reflective of coral health.	
2. Planar surface area	cm^2	size	No	Collinear with max. length.	
3. RNA/DNA ratio	unitless	total gene expression	Yes	NA	Very high or very low values indicative of aberrant behavior.
4. *Symbiodinium* GCP	unitless	*Symbiodinium* density	Yes	NA	Very low values indicative of aberrant behavior (bleaching).
***Symbiodinium* gene expression (n=5)**		**Function**			
5. heat shock protein 90	*hsp90*	molecular chaperone	Yes	NA	Very high values indicative of aberrant behavior.
6. ascorbate peroxidase	*apx1*	oxidative stress	Yes	NA	Very high values indicative of aberrant behavior.
7. ubiquitin ligase	*ubiq-lig*	cellular stress response/proteolysis[a]	Yes	NA	Very high values indicative of aberrant behavior.
8. RuBisCO	*rbcL*	photosynthesis	Yes	NA	Very high values indicative of aberrant behavior.
9. zinc-induced facilitator-like 1-like	*zifl1l*	metabolism	Yes	NA	Very high values indicative of aberrant behavior.
Host coral gene expression (n=4)		**Function**			
10. GFP-like chromoprotein	*gfp-cp*	light absorption	Yes	NA	Very high values indicative of aberrant behavior.
11. copper-zinc super-oxide dismutase	*cu-zn-sod*	oxidative stress	Yes	NA	Very high values indicative of aberrant behavior.
12. lectin	*lectin*	cell adhesion	Yes	NA	Very high values indicative of aberrant behavior.
13. carbonic anhydrase	*ca*	metabolism	Yes	NA	Very high values indicative of aberrant behavior.

[a]Welchman et al. (2005)

PERMANOVA and PERMDISP. Given the aforementioned issues with using parametric statistical approaches with reef coral datasets, PERMANOVA (permutational ANOVA of [raw] similarity) was carried out with PRIMER 6 featuring the PERMANOVA+ plug-in (Anderson et al., 2008) to test for the effects of each of the target EP (Table 2) on the similarity among samples after first creating a Euclidean distance matrix (EDM) with the standardized data. Additionally, the PERMDISP feature of PRIMER 6 (Anderson, 2006) was utilized with this same EDM to test for homogeneity of multivariate dispersions. The latter was carried out because we hypothesized that the dispersion in the coral physiological response might differ across environmental gradients, particularly ALCC levels. As discussed in detail in Mayfield et al. (in review), the variability in the coral response is likely to be more important than the mean response in terms of identifying aberrantly behaving coral colonies. For both similarity-based approaches, an alpha level of 0.05 was set.

Modeling the coral physiological response with SR and PLS. When dealing with a large number of predictor variables (i.e., the EP of Table 2), hypothesis-driven approaches such as ANOVA may lead to type I errors, even when dramatically adjusting the alpha level as in Mayfield et al. (2016a, 2017a-c; from 0.05 to 0.004 in these published works). Furthermore, hypothesis testing is poorly suited for identifying the optimal model for explaining a dataset, even when a multivariate similarity approach like PERMANOVA is used (Anderson et al., 2000). For these reasons, information theory has been increasingly used in ecology (Anderson & Burnham, 2002) in order to develop the most appropriate, parsimonious, best-fit model for explaining the behavior of a particular RV (Mazerolle, 2006). Information theory is used routinely in coral reef ecology (Jorgensen et al., 2005) and ichthyology (Conover et al., 2006), but, to date, not to any great extent in coral physiology research (where hypothesis testing is far more common; e.g., Mayfield et al., 2013c-d; Putnam et al., 2013; Mayfield et al., 2014c).

Herein we used an information theory-based SR program in JMP to select the best-fit model for each of 12 RV: max. colony length, *Symbiodinium* genome copy proportion (GCP), the RNA/DNA ratio, and expression of five *Symbiodinium* stress-targeted genes (STGs) and four host coral genes (Table 3). Backwards model selection was used such that all EP were initially considered,

and a stopping rule was enacted based on minimizing the "Bayesian information criterion" (BIC). Finally, a "combine" rule was used such that, when necessary to minimize the BIC, certain EP were partitioned heuristically into sub-categories/bins. As a comparison to SR, PLS was used after converting the EP data into integers. For instance, the islands were scored as follows: Totoya=1, Matuku=2, Moala=3, etc. The other EP were converted to continuous numerical terms in a similar fashion. PLS not only attempts to find a best-fit model for the response data (y), but also the EP (x; predictor) data. A NIPALS method was used with "leave-one-out" validation, with 12 factors initially considered. The model featuring the minimum root mean predicted residual sum of squares (PRESS) was selected, except for when 0 latent factors were deduced by JMP, in which case a model featuring a single latent factor was instead chosen. A stringent variable importance parameter (VIP) threshold of 1 was chosen (as opposed to the JMP-recommended VIP threshold of 0.8), and both positive and negative coefficients were chosen (i.e., EP that were positively or negative correlated, respectively, with the target RV). To simulate the predictive power of the PLS models, random sub-samples of the dataset were used for training, and training models were validated with another random subset of the dataset.

The aforementioned analyses sought to identify the EP (or combinations thereof) that best accounted for variation in each of 12 RV (the 11 MPRV+max. length). However, we were also interested in the suite of EP that best modeled variability in the multivariate coral response. Therefore, as a distance-based analog to the information theory+linear modeling analyses (SR & PLS), PRIMER's "DistLM" (distance-based linear modeling; Clarke et al., 2014) program was used with the "best" selection procedure and a minimum AICc selection criterion. In other words, the EP(s) best accounting for between-sample differences in the underlying EDM was/were identified. As a comparison, models were also built based on minimizing the adjusted r^2 values. A multivariate PLS was also considered in the initial analyses, but the resulting power of the models developed was generally too low to be worthy of inclusion herein.

Outlier analysis. A detailed treatise on identification of outliers in the Lau Archipelago dataset can be found in Mayfield et al. (2017b). Briefly, the Mahalanobis distance (MD)

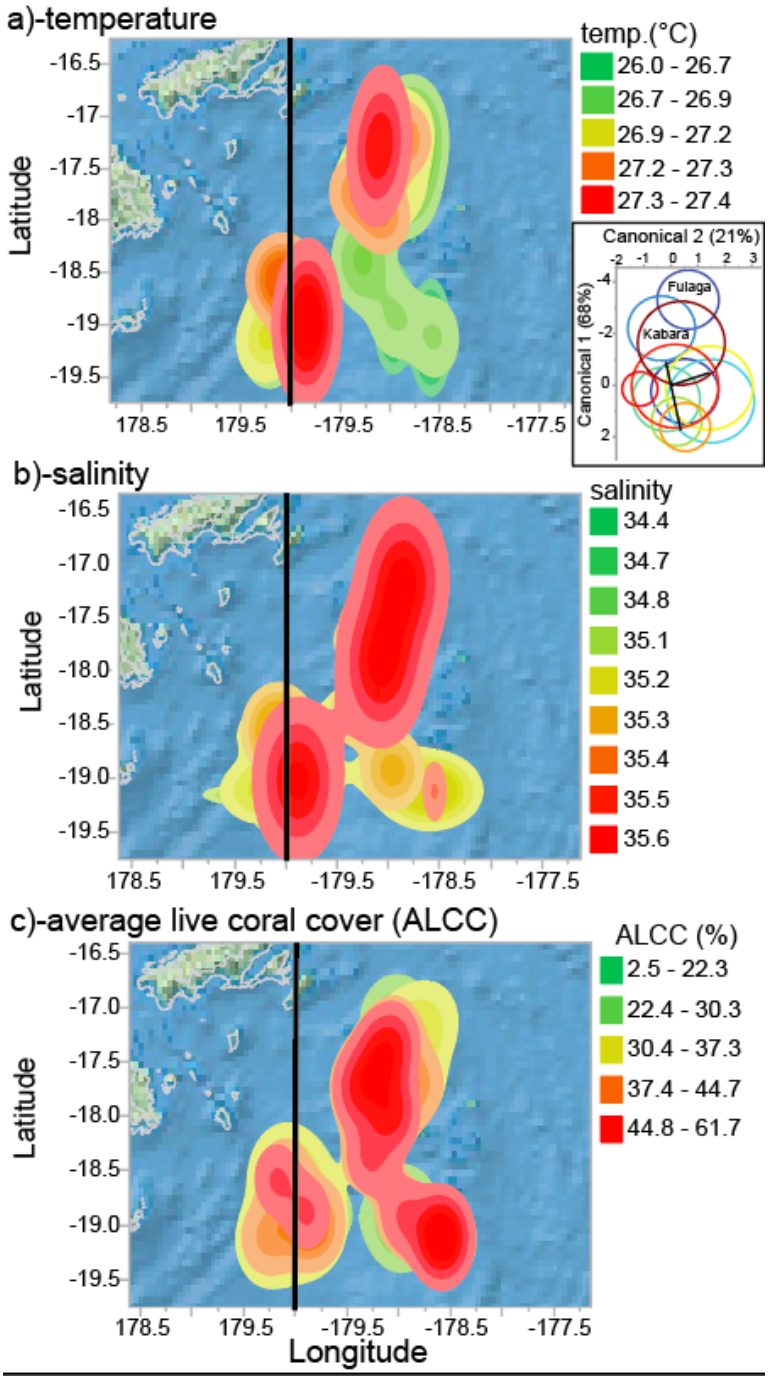

Fig. 1. Contour plots of environmental data in Fiji's Lau Archipelago. The sizes of the data clouds are proportional to the data variability, and, given generally high variability, all islands are masked. An inset adjacent to (a) shows a canonical correlation analysis (CCA) plot of the continuous environmental data (temperature [temp.], salinity, and average live coral cover [ALCC]) vs. island, and there was a statistically significant effect of island (Wilks' lambda, $p<0.0001$). Centroids represent 95% confidence, and the black EP biplot rays at ~11:00, ~2:00, and ~5:00 represent salinity, ALCC, and temp., respectively. Due to spatial constraints, the plot positions of only two islands (Fulaga and Kabara) have been denoted; others are available upon request (or by reconstructing one's own CCA plot using the raw data provided in the online data supplement). Please note that, although the effect of island was statistically significant, 52% of the sites were misclassified by the associated discriminant analysis. For a data-free map of the Lau Archipelago, please see Mayfield et al. (2017b). The international date line has been highlighted in bold.

was calculated across the 11 MPRV to serve as the primary means of identifying outliers, and those samples whose MD values were >5 and whose heat map scores (HMS) were ≥1 were considered to be outliers. Briefly, a sample featuring a MPRV with a *z*-score >2 would be given an HMS of 1. To corroborate this outlier assignment, the principal coordinate from the primary axis of a principal coordinates ordination (PCO) analysis carried out by PRIMER 6 was calculated ("PC1") and regressed against the MD. It should be noted that, because this PCO was carried out with an EDM, it is comparable to the more commonly employed PCA. The MD and PC1 score are collectively referred to as "multivariate variability terms" throughout the manuscript. We also calculated a second (after the HMS) "univariate variability term" known as the "variability index" (VI), which was first described by Mayfield et al. (2017a-b). This term is simply the standard deviation of the *z*-score-transformed (i.e., standardized) data across all RV for an

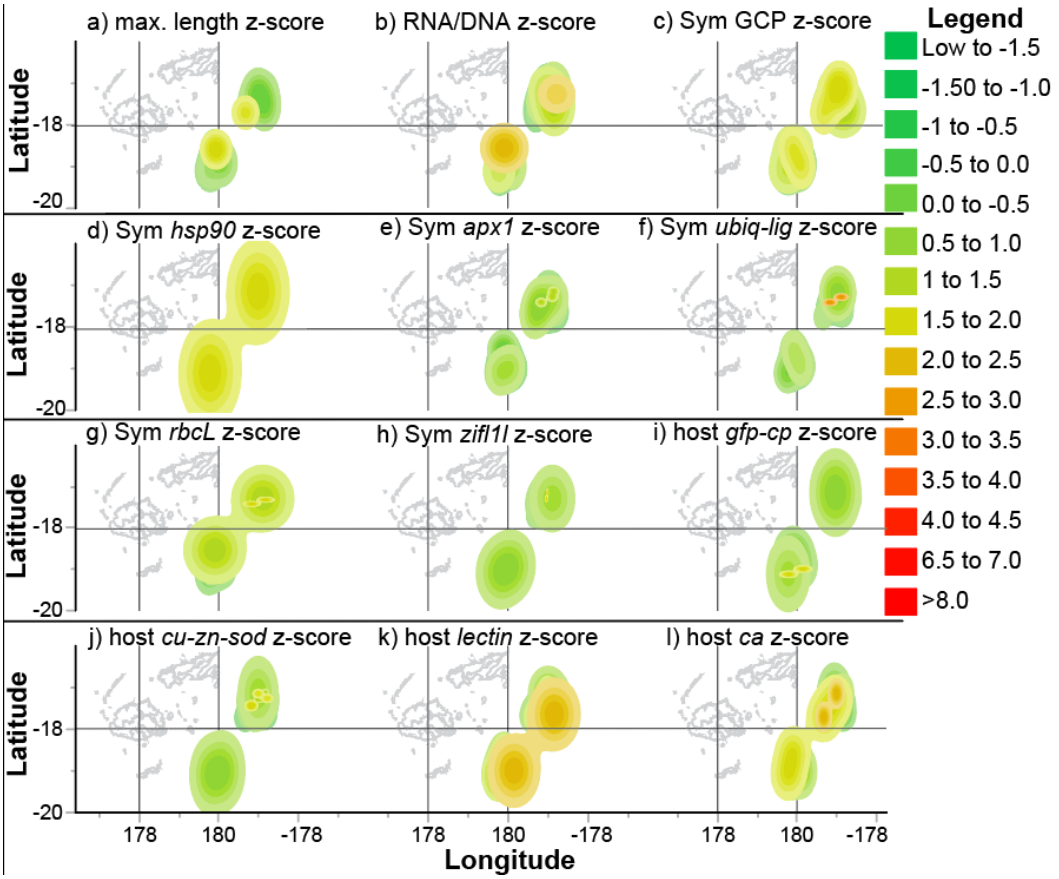

Fig. 2. Contour plots of standardized pocilloporid coral data in Fiji's Lau Archipelago. Please see Table 3 for full gene names.

individual coral biopsy. For instance, if the *z*-scores for the 11 MPRV measured herein were 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, and 11 for a particular coral sample, its VI would be 3.3.

We hypothesized that samples with high VI (within-sample variation) would also tend to be characterized by relatively high MD (between-sample variation; i.e., a greater distance from the global centroid), and linear regression analysis was used to determine the significance of this relationship across the 69-sample dataset. JMP's "predictor screening" program was used to calculate the relative effects of the MPRV on the MD. These findings can be found in Mayfield et al. (2017b). Finally, a discriminant analysis (DA) was used with outlier status (yes vs. no) as the x (categorical) variable in order to create a best-fit model in which environmental data could be used to predict whether or not a sample would be considered an outlier. A similar DA was performed with the 11 MPRV in order to uncover which are most important in determining whether or not a sampled colony was displaying aberrant behavior. A more simplistic, X^2-based outlier frequency analysis (as a function of EP) can be found in Mayfield et al. (2017b), in which it was found that outlier distribution was effectively random across environments.

Results and discussion

Multivariate findings. A variety of both univariate and multivariate statistical methods were taken to attempt to understand the EP (or combinations of EP) that best accounted for spatio-temporal, or otherwise environmental, variation in coral physiology. Although data tended to be non-normally distributed and heteroskedastically variable (even upon log, rank, or other transformations), MANOVA was nevertheless carried out by Mayfield et al. (2017b) with the 11 MPRV (Table 4), and only island and host genotype had a significant effect on the multivariate mean. When using a more conservative distance-based approach known as PERMANOVA to model environmental effects on the multivariate coral response (Table 4), only host genotype affected mean sample similarity; there was also a marginal effect of reef zone. For a graphical view of this host effect on the coral dataset, please see Fig. 3a, in which some clustering by species is evident. Additionally, a DA-based CCA of host can be found in Fig. 5c. In contrast, forereef and lagoonal samples appeared inter-mixed in the MDS plot (not shown).

Table 4. Permutational ANOVA (PERMANOVA), PERMDISP, discriminant analysis (DA), and multivariate ANOVA (MANOVA). The former two approaches were carried out with a Euclidean distance matrix of standardized data (the 11 molecular-physiological response variables [MPRV] of Table 3), and the 12 environmental parameters (EP) of Table 2 (excluding site, date, and *Symbiodinium* assemblage) were considered as predictors. For PERMANOVA raw data were permutated in an unrestricted manner (type III, partial), and, in certain cases, the *df* is higher than would be expected from Table 2 due to the inclusion of "missing data" as a category in the analysis. JMP®'s partial least squares-based DA algorithm was used to build best-fit models of categorical environmental data with the 11 MPRV (as continuous y-variables), and the percent of samples misclassified with respect to each EP was calculated. MANOVA was also carried out upon excluding the 11 outliers uncovered (discussed below and in Mayfield et al., 2017b; values in cells represent *p*-values.). Table abbreviations: NS=not significant, PA=*Pocillopora acuta*, PB=*P. brevicornis*, PD=*P. damicornis*, PM=*P. meandrina*, PV=*P. verrucosa*, and temp.=temperature.

	PERMANOVA				PERMDISP (centroid)		DA	MANOVA excluding 11 outliers & PB
EP	***df***	**Pseudo-*F***	***P* (perm)**	***post-hoc* test**	***F***	***p*(perm)**	**% mis- classified**	
island	8, 61	1.14	0.13		1.26	0.65	51*	NS
exposure	2, 67	1.22	0.28		1.42	0.36	40	<0.05
reef type	3,66	1.35	0.20		0.39	0.93	31	NS
reef zone	1, 68	2.14	0.05	forereef≠lagoon	2.51	0.17	21	NS
ALCC	4, 65	0.80	0.69		1.57	0.36	50	NS
temp.	1, 68	0.96	0.44		0.77	0.46	40	NS
salinity	7, 62	1.08	0.33		0.93	0.79	51	NS
PAR	4, 65	0.79	0.79		1.57	0.39	49	NS
time	2, 67	0.86	0.59		0.19	0.88	46	<0.01 w/ transformed data
depth	7, 62	0.89	0.54		1.60	0.76	57	NS
host	4, 65	1.98	0.04	PA≠PV≠PD	2.87	0.16	31*	<0.01 w/ & w/o data transformation
color	4, 65	0.94	0.52		1.64	0.32	47	NS
					mean % misclassified		43	

*Wilks' lambda, $p<0.05$ (see Fig. 5c.).

Fig. 3. Multi-dimensional scaling (MDS) plot of the 69 Fijian samples analyzed in full and summary of environmental effects. In the MDS plot (a), the 11 outliers have been labeled, and the species icon legend is found in the bottom left corner. Images of two outliers have been included, and the yellow and white colony tags (the scaling objects) are 6.1 (from the metallic protrusion on the left side to the right end) and 4.7 (width) cm, respectively. For labeled biplot ray axes (red arrows; some of which overlap) in the principal components loading plot, please see Mayfield et al. (2017b). In (b), the icons are proportional to the number of analytical approaches that yielded a statistically significant finding (p<0.05); three univariate (one-way ANOVA, stepwise regression, and partial least squares) and four multivariate (last row only) approaches (MANOVA, DistLM, PERMANOVA, and PERMDISP) were utilized, meaning that the maximum frequencies were three and four, respectively. As an example of how to interpret this panel, one of the three univariate statistical approaches documented a significant effect of reef exposure on maximum (max.) colony length (bubble size=1), whereas there was no effect of reef exposure on the multivariate molecular-physiological response (no bubble present).

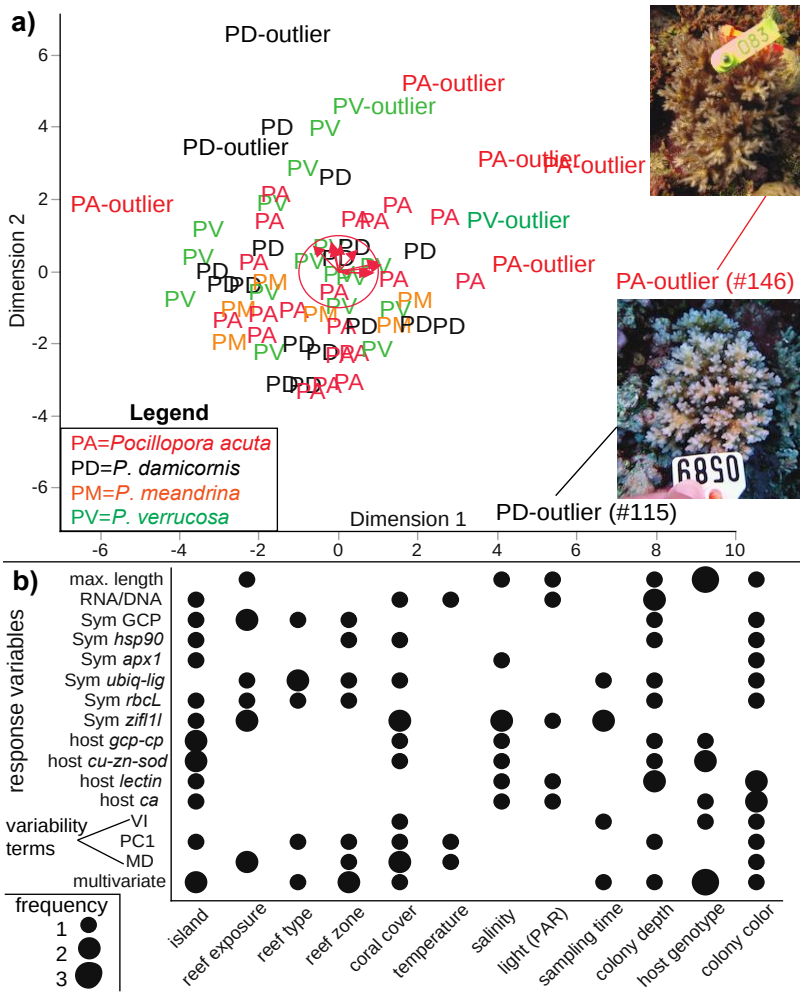

Fig. 4. Graphical breakdown of the environmental parameters (EP) included in the best-fit stepwise regression (SR) and partial least squares (PLS) models. The inner and outer pie graphs in (a-o) depict the SR and PLS model EPs, respectively, and the EPs characterized by the highest *F*-test *p*-values have been listed at the bottom left and top right of each panel, respectively. In (p), the proportional breakdown of the EPs in all 15 response variable (RV) models (maximum [max.] colony length, the 11 molecular-physiological response variables, the variability index [VI], the principal components analysis primary axis loading score [PC axis 1 or PC1]), and the Mahalanobis distance [MD]) has been included for SR and PLS, and EPs that were represented in a significantly higher proportion of models generated by one approach over the other (X^2 test, $p<0.05$) have been underlined. The r^2 values (x 100; adjusted r^2 for SR and raw r^2 for PLS) have been plotted across RV in (q), and the global mean SR and PLS r^2 values have been plotted as red and blue horizonal lines, respectively. Please note that, in certain cases, the EP colors do not sync with those in the manuscript's other figures and tables (e.g., colony color).

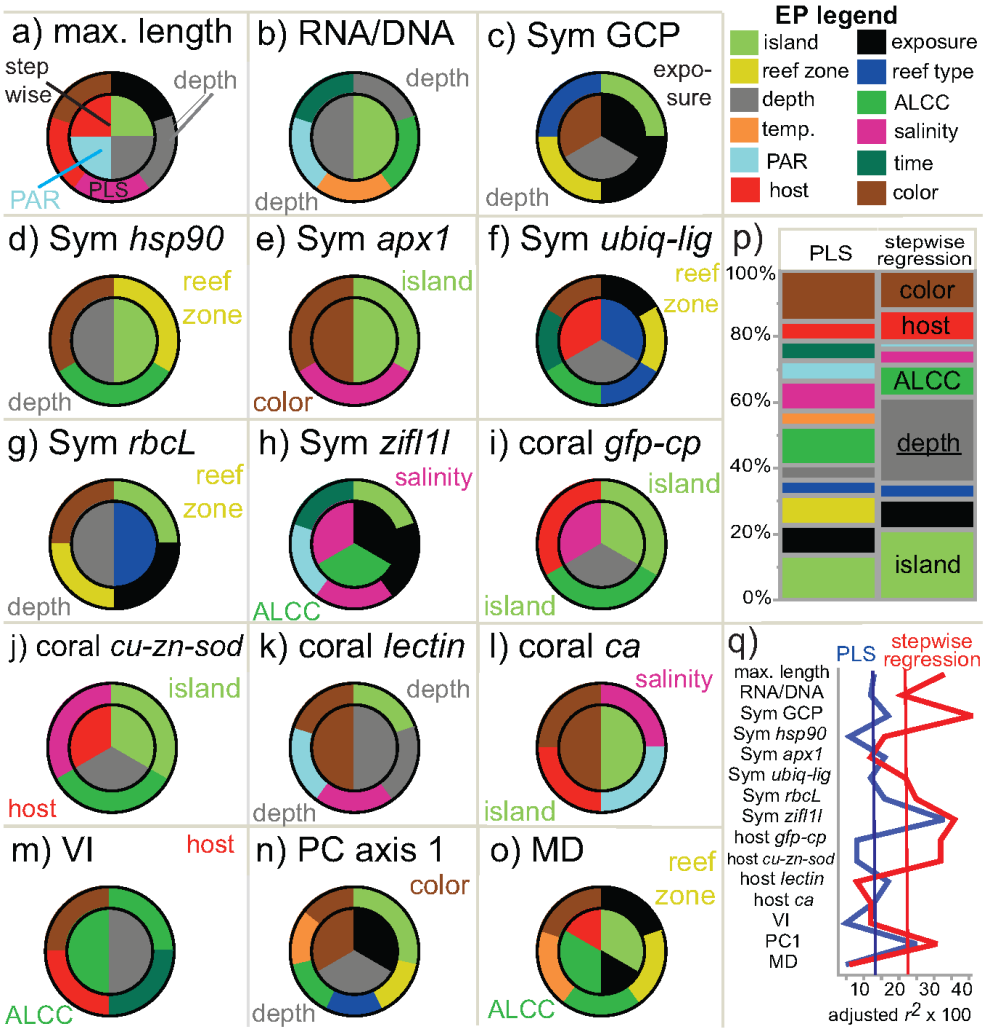

Fig. 5. Proportional breakdown of the environmental parameters (EP) across all stepwise regression (SR) and partial least squares (PLS) models and a canonical correlation analysis (CCA) of host coral species. As shown in Table 5, 28 and 64 statistically significant EP were featured in the 15 SR (a) and 15 PLS models (b), respectively. There was a relatively higher proportion of the EP "depth" in the SR models (32%) than in the PLS ones (5%; see Fig. 4p for details.). Please note that, in certain cases, the EP colors do not sync with those in the manuscript's other figures and tables (e.g., reef zone, which is yellow in most other tables and figures.). In the CCA (c), the centroids represent 95% confidence, and their colors match the font color of the four sampled pocilloporid coral species. Only one positive biplot ray has been labeled; the (negative) ray pointing towards 8:30 (clock position; towards the *Pocillopora verrucosa* centroid) is the *Symbiodinium* genome copy proportion (GCP), which was marginally higher in *P. verrucosa* samples when compared to the other three species (NS). CCA plot icons: L=outlier, plus sign (+)=<5 m depth, circle (o)=5-10 m depth, square=10-15 m depth, upside-down v=15-20 m depth, triangle=20-25 m depth, diamond=25-30 m depth, and z=greater than 30 m depth.

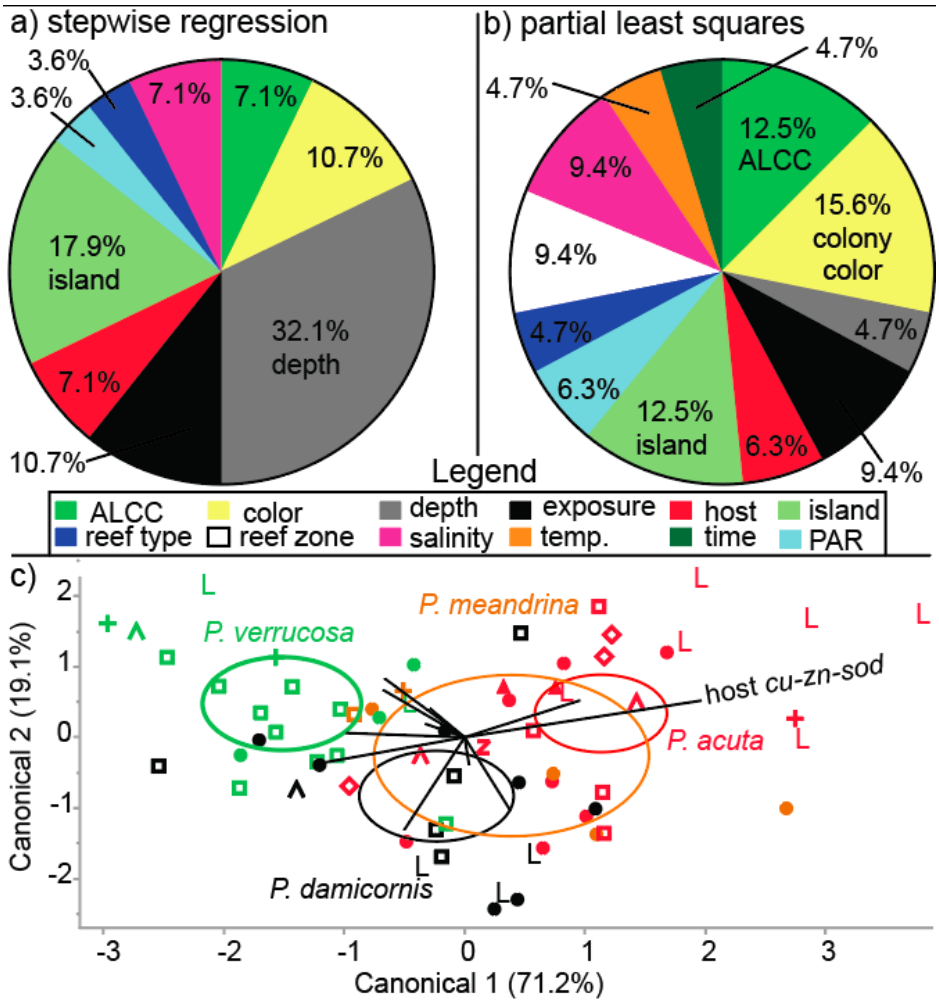

Table 5. Summary of stepwise regression (SR) and partial least squares (PLS) results. Details of SR can be found in Supplemental table 2. Details of the PLS analysis, which was not undertaken with the Australs-Cooks (A-C) dataset, can be found in Supplemental table 3. Statistically significant *F*-test ($p<0.01$) Fiji SR model terms have been underlined; for A-C, only statistically significant terms have been included. When the mean adjusted (adj.) r^2 values for PLS (Fiji only) and SR (Fiji+A-C) were both under 0.19, the models were deemed "poor-fit" in the "Important drivers of variation" column; for mean adj. r^2 >0.19 (i.e., >19% cumulative response variable variation explained), the EP that featured in the models derived from multiple approaches (or in multiple regions) have instead been listed in this column. The Fiji data found in this table have been plotted in Fig. 4. PC1=principal components analysis primary axis loading score. MD=Mahalanobis distance. VI=variability index. VIP=variable importance parameter.

Response variable	SR-Fiji (model terms)	SR adj. r^2	PLS-Fiji (VIP>1)	PLS-Fiji r^2	A-C dataset SR (model terms)	SR-A-C adj. r^2	Important drivers of variation (mean r^2)
Max. length	PAR>host>island>depth	0.37	depth>host>exposure>color>salinity	0.13	host	0.12	depth & host (0.21)
RNA/DNA	depth>island	0.21	depth>PAR>temp.>ALCC	0.12	ALCC	0.44	depth & ALCC (0.26)
Sym GCP	depth>color>exposure	0.31	exposure>island>reef zone>reef type	0.17	host>depth>ALCC	0.47	depth & exposure (0.32)
Sym *hsp90*	depth>island	0.16	reef zone>ALCC>color	0.06	island	0.21	poor-fit models
Sym *apx1*	color>island	0.12	island>salinity>color	0.16	salinity	0.37	color, salinity, & island (0.22)
Sym *ubiq-lig*	depth>reef type>host	0.22	reef zone>reef type>time>ALCC>exposure>color	0.12	host	0.28	host & reef type (0.21)
Sym *rbcL*	depth>reef type	0.16	reef zone>island>exposure>color	0.16	not analyzed		poor-fit models
Sym *zifl1l*	ALCC>exposure>salinity	0.37	salinity>island>exposure>PAR>time	0.33	not analyzed		salinity & exposure (0.35)
host *gfp-cp*	island>depth>salinity	0.32	island>ALCC>host	0.08	not analyzed		island (0.20)
host *cu-zn-sod*	host>depth>island	0.31	island>salinity>ALCC	0.08	not analyzed		island (0.20)
host *lectin*	depth>color	0.08	depth>PAR>salinity>island>color	0.17	not analyzed		poor-fit models
host *ca*	island>color	0.12	salinity>host>color>PAR	0.13	not analyzed		poor-fit models
PC1	depth>color>exposure	0.25	color>reef type>reef zone>island>temp.>ALCC	0.25	temp.>host>ALCC	0.37	color, temp., & ALCC (0.29)
VI	ALCC>depth	0.12	host>color>ALCC>time	0.05	host	0.19	poor-fit models
MD	ALCC>exposure>island	0.18	reef zone>exposure>ALCC>temp.>color	0.05	island>host (0.40)	0.40	ALCC, island, & exposure (0.21)
	Mean SR-Fiji adj. r^2	0.22	Mean PLS-Fiji r^2	0.14	Mean SR-A-C r^2	0.32	Mean overall r^2 =0.23

Table 6. Summary of the effects of 12 environmental parameters (EP) on the univariate and multivariate coral physiological response (11 molecular-physiological response variables [MPRV] only). In the univariate ANOVA cells, significantly affected response variables have been listed, whereas the mean (±std. dev.) adjusted (adj.) r^2 values across the 1) maximum colony length, 2) 11 MPRV, 3) variability index, 4) Mahalanobis distance, and 5) principal components analysis primary axis loading scores (PC1) in which the respective EP was included have been inserted in the "SR-adj. r^2" and "PLS-VIP>1" columns (the frequency [freq.] of inclusion of the EP in all 15 RV models has been included in parentheses.). For the multivariate statistical approaches (MSA), multivariate *p*-values have instead been included in the cells, except for the case of distance-based linear modeling (DistLM), in which those EP "included" in the model characterized by the minimum Akaike information criterion value have instead been inserted. As no significant effects of PERMDISP were uncovered, this approach has been excluded from the table. When outliers were excluded, "time" was found to significantly affect the multivariate coral response (see Table 4.). Table abbreviations: ALCC=average live coral cover. NS=no response variables were affected by the EP (univariate statistical approaches) or not statistically significant (MSA). PLS=partial least squares. Sym=*Symbiodinium*. Temp.=temperature. VIP=variable importance parameter.

	Univariate analyses			**Multivariate statistical analyses**				
		Information theory			**Similarity analysis**			
EP	**ANOVA[a]**	**SR-adj. r^2 (freq.)[b]**	**PLS-VIP>1 (freq.)[b]**	**MANOVA[a]**	**PERMA-NOVA[c]**	**DistLM[d]**	**Significant driver of coral physiological variation?**	**rank**
Host	host *gfp-cp* & *cu-zn-sod*	0.34±0.03 (2)	0.10±0.02 (4)	$p<0.05$	$p<0.05$	included	Yes	1
Island	NS	0.22±0.09 (5)	0.18±0.08 (8)	$p<0.05$	NS	included	Yes	2
Depth	NS	0.22±0.08 (9)	0.14±0.03 (3)	NS	NS	included	Yes	3
Reef exposure	NS	0.29±0.10 (3)	0.16±0.09 (6)	NS	NS	NS	Yes	4
ALCC	Sym *zifl1l*	0.28±0.13 (2)	0.10±0.07(8)	NS	NS	included	Somewhat	5
Salinity	NS	0.35±0.04 (2)	0.17±0.09 (6)	NS	NS	NS	Somewhat	6
Color	NS	0.17±0.12 (3)	0.13±0.06 (3)	NS	NS	included	Somewhat	7
Reef zone	NS	NS (0)	0.14±0.08 (6)	NS	$p=0.05$	included	Somewhat	8
Reef type	NS	0.16 (1)	0.18±0.07 (3)	NS	NS	included	Somewhat	9
PAR	NS	0.37 (1)	0.19±0.10 (4)	NS	NS	NS	No	10
Time	Sym *zifl1l*	NS (0)	0.17±0.15 (3)	NS	NS	included	No	11
Temp.	NS	NS (0)	0.14±0.10 (3)	NS	NS	NS	No	12

[a]First presented in Mayfield et al. (2017b). [b]analyzed as part of this work (see Table 5 and Supplemental tables 2-3 for details.). [c]analyzed as part of this work (see Table 4 for details.). [d]analyzed as part of this work (see Supplemental table 1 for details.).

In contrast to MANOVA and PERMANOVA, multivariate dispersion (as assessed by PERMDISP) was similar across all 12 EP (Table 4); this suggests that PERMANOVA was a suitable means for analyzing the data. It also indicates that coral physiology was not more variable under certain environmental conditions when compared to others. As further evidence of this, the MD was not affected by any EP; distribution of multivariate outliers was, then, markedly random, and best-fit models (SR and PLS) for modeling the MD were characterized by weak r^2 values (0.18 and 0.05, respectively). This means that it is not currently possible to predict with confidence where, or under which environmental conditions, corals displaying aberrant behavior will be located; however, upon analysis of a greater portion of the GRE dataset, statistical models characterized by higher predictive capacity will surely be developed. It is worth noting there that DistLM, which relies on information theory-based approaches to generate the most parsimonious best-fit models, yielded models with very low r^2 values (<0.10; Supplemental table 1); for this reason, these DistLM data are not discussed at any great length herein.

Univariate findings. A one-way ANOVA matrix of the 12 EP against the 12 RV (max. length+11 MPRV; n=144 tests) can be found in Mayfield et al. (2017b); briefly, very few univariate effects of environment were detected at the highly stringent, Bonferroni-adjusted alpha level of 0.004, and all such differences are listed in the summary table (Table 6). There was a statistically significant effect of ALCC and sampling time on *Symbiodinium zifl1l* mRNA expression, as well as a significant effect of host coral genotype on host *gfp-cp* and *cu-zn-sod* mRNA expression.

Outlier analysis. A detailed treatise on the 11 outliers uncovered can be found in Mayfield et al. (2017b). A DA (categorical x-variable=outlier status: yes vs. no) was carried out with the 11 MPRV (as y [RV] covariates) used to calculate the MD, and only one sample was misclassified (4%) by a PLS-like model featuring a mix of host and *Symbiodinium* gene mRNAs as the dominant, standardized loading coefficients (detailed information not shown; Wilks' lambda, p<0.001). However, when using a DA with the EP converted to integers, 33% of the samples were misclassified, and Wilks' lambda was not statistically

significant. This is further evidence for the observation, noted above, that it is not currently possible to predict where, or under what environmental conditions, corals displaying aberrant behavior are most likely to be found. It is interesting to note, though, that those samples characterized by high VI (within-sample variation) were also those with the highest MD (inter-sample variation; linear regression *t*-test, r^2=0.83, p<0.001 [data not plotted]), as was documented previously in another pocilloporid coral dataset (Mayfield et al., in review). In other words, corals displaying aberrant behavior were more likely to demonstrate high variability across RV. Although it is premature to speculate on whether these 11 outliers were significantly more stressed than non-outliers, it is worth noting that high transcriptional variability is a hallmark of many cancers (Han et al., 2016; Sharma et al., 2018) and points to a loss of homeostasis. Additionally, and as was pointed out by Mayfield et al. (2017b), these outliers were also characterized by lower *Symbiodinium* densities and higher stress gene expression, also pointing to their being stressed. Future works should, then, attempt to uncover whether these outliers are, for instance, more likely to bleach than those found to be displaying statistically normal behavior at the time of sampling. Indeed, we hypothesize that, more generally, variability in the coral response is more important than the mean response (*sensu* Cleophas et al., 2006) when attempting to gauge the environmental sensitivity of a sampled colony; this idea could be directly tested in a controlled aquarium setting.

SR and PLS. Both SR and PLS (summarized in Table 5, with details found in Supplemental tables 2 and 3, respectively) were used to determine the EP (or combinations thereof) that best modeled variation in max. colony length, the 11 MPRV, the VI, PC1, and the MD (n=15 models built), and the data have been depicted graphically in Fig. 3b and Fig. 4a-o. Please note that in the former figure, the frequencies (as bubbles) could feature not only SR and PLS, but also one-way ANOVA (i.e., RV for which SR, PLS, and one-way ANOVA all determined the EP in question to be significant would be characterized by a frequency of 3.). Although SR and PLS tended to feature the same EP in the models for each of the RV (Table 5), depth was featured in a significantly higher proportion of the SR models (Fig. 4p). Furthermore, SR models

tended to yield higher r^2 values (0.22±0.09) than PLS models (0.14±0.08; Fig. 4q). However, as both the predictor variables (e.g., temperature and salinity) and response variables (e.g., *Symbiodinium* gene expression) tended to co-vary (Mayfield et al., 2017b), it is likely that PLS is nevertheless a superior, more robust approach for developing predictive models with environmental datasets featuring sampled reef coral colonies (Cox & Gaudard, 2013).

As mentioned above, 11 outliers were deliberately left in the dataset during the aforementioned model generation exercises, as we feel these aberrantly behaving colonies may be those of most interest to those seeking to develop means of identifying corals and coral reefs that are potentially stress sensitive (or, alternatively, of a marked capacity for resilience to GCC or other anthropogenic stressors). That being said, future predictive modeling efforts with reef coral datasets may be better served by excluding such outliers; although information will be lost on potentially the most stress-sensitive or, alternatively, bleaching-resistant samples, the ensuing, outlier-free models will almost surely be characterized by improved predictive capacity. As an example of this, MANOVA was undertaken herein upon the removal of the outliers (Table 4), and two additional EP *not* found by other multivariate methods to be significant drivers of variation in coral physiology, sampling time and reef exposure, yielded statistically significant results. It should be noted, though, that the SR and PLS models generated herein did not improve markedly upon exclusion of the 11 outliers (data not shown), though this could simply due to the small size of the resulting dataset (n=58 without the outliers).

When looking at the shear frequency at which an EP was included in a model (Tables 5-6 and Fig. 4p), island ranked highest; this EP was found in 5 and 8 of the 15 SR and 15 PLS best-fit models, respectively (n=13/30 models; 43% of all models). This equated to 18% (Fig. 5a; n=28 total model terms for SR) and 13% (Fig. 5b; n=64 total model terms for PLS) of the best-fit model terms (14% of the 92 best-fit model terms). When compared to the PLS models, depth was featured in a relatively higher proportion of the SR models (Fig. 4p); depth comprised nearly 1/3 of all 28 SR model terms (Fig. 5a) and was featured in 9 of the 15 SR best-fit models for the individual RV (Table 5 and Supplemental table 2). In contrast, it

featured in only 3 of the 15 PLS best-fit models (Table 5 and Supplemental table 3), equating to less than 5% of all PLS model terms (Fig. 5b). Across both modeling methods, then, 12 of the 30 RV models constructed featured depth (40%), and depth comprised 12 of the 92 best-fit model terms (13%). For this reason, the CCA plot of host coral described below (Fig. 5c) features icons labeled by depth.

Although host coral genotype was featured as a significant model term in only 2 and 4 SR and PLS RV models, respectively (Table 5; 6/30 models built; 20%), the global average r^2 for models featuring "host" as a significant term was relatively high (0.22), and, furthermore, host genotype was found to significantly affect the multivariate coral phenotype (MANOVA and PERMANOVA, $p<0.05$; Table 4). In contrast, although island and depth were featured in more models and were characterized by statistically similar global r^2 values (0.20 and 0.18, respectively; Table 6), neither was found to significantly affect coral physiology by multiple MSA; for this reason, we consider host coral genotype to be the biggest driver of variation in coral physiology in the Lau Archipelago dataset (see ranks in Table 6.).

When looking at the Fiji dataset analyzed alongside the Australs-Cooks one (Table 5 and Supplemental tables 2-3), host and island were also the two most important EP with respect to driving variation in pocilloporid coral physiology. It is perplexing as to why island itself featured in best-fit models to a greater extent than the underlying characteristics defining the islands' coral reef ecosystems (e.g., reef type, reef zone, reef exposure, temperature, etc.). Were it simply driven by latitudinal differences between the islands (see Figs. 1-2), we would expect temperature to also feature heavily in the models; in fact, this was the least important term (Table 6). That being said, there are myriad properties and characteristics of the individual islands whose reefs were surveyed that could have driven this island effect. Of note, population density varies greatly across the Lau Archipelago (data not shown); different land-use practices (e.g., agriculture) could have contributed to differences in nutrient levels in the seawater surrounding the coral reefs. Unfortunately, nutrient levels were not assessed herein, and we, in fact, highly recommend that those undertaking similar eco-physiological assessments of coral reefs (and reef corals) take, at a minimum, measurements of seawater nitrogen levels

in their future survey projects.

Another reason why island may have featured in so many best-fit models is due to the co-variation between island and host genotype frequency (Mayfield et al., 2017b); the assemblage of pocilloporid corals varied across islands, and, as is evidenced by this manuscript's data (as well as those of other works; e.g., Mayfield et al., 2016a, 2017c), there are clear differences in the molecular physiology of closely-related (con-generic) pocilloporid corals. In the Lau Archipelago dataset, *P. acuta* and *P. verrucosa* were particularly well partitioned in a CCA (Fig. 5c; Wilks' lambda, $p<0.001$); this was driven, in part, by higher host coral *cu-zn-sod* mRNA expression in the former species (3.5-fold) relative to the latter. *P. verrucosa* colonies tended to have higher *Symbiodinium* densities than the other three species, though this difference was not statistically significant. Despite the significant MANOVA/CCA of host and the visible separation of *P. acuta* and *P. verrucosa* in the CCA plot, nearly 31% of the samples were misclassified in the associated DA; in other words, the underlying predictive model could only correctly "call" the correct species 2/3 of the time. Of the four coral species, only *P. verrucosa* was characterized by a stable enough phenotype to be correctly classified in a high percentage of cases by PLS-based DA (16/18 samples were properly classified; 89%).

Conclusions

With the exception of the statistical classification of *P. verrucosa*, nearly all predictive models were characterized by low r^2 values (Tables 5-6; 22 and 14% for SR and PLS, respectively) and high misclassification rates (Table 4; mean=nearly 50%). Although to an ecologist attempting to make explanations of the natural world, these values might be acceptable, a manager in need of predicting coral behavior with confidence based on local environmental conditions would do no better than flipping a coin. However, upon analyzing the far larger dataset produced during the LOF's GRE, only a small portion of which has been assessed to date, we feel confident that we will ultimately possess the capacity to develop eco-physiological predictive models for reef corals such that we can, for instance, insert abiotic data into an equation and predict, with some confidence, how corals at the site/reef in question will behave. Such predictive models will become increasingly critical

given the aforementioned need to forecast how coral reefs will respond to the increases in temperature, in particular, that will characterize their habitats in the coming years (Cinner et al., 2016).

Acknowledgements

We would like to thank the crew of the *M.Y. Golden Shadow* for their support and hospitality during the June 2013 cruise to Fiji's Lau Archipelago. This work was funded through an international postdoctoral research fellowship from LOF to ABM. Finally, we would like to thank Mr. John Powell for aid in the use of interactive HTML to display the manuscript's data online on http://coralreefdiagnostics.com.

References

Anderson, D.R., K.P. Burnham & W.L. Thompson. 2000. Null hypothesis testing: problems, prevalence, and an alternative. The Journal of Wildlife Management, 64:912-923.

Anderson, D.R. & K.P. Burnham. 2002. Avoiding pitfalls when using information-theoretic methods. The Journal of Wildlife Management, 66:912-918.

Anderson, M.J. 2006. Distance-based tests for homogeneity of multivariate dispersions. Biometrics, 62:245-253.

Anderson, M.J., R.N. Gorley & K.R. Clarke. 2008. PERMANOVA+ for PRIMER: Guide to Software and Statistical Methods. PRIMER-E, Plymouth, UK.

Barshis D.J., J.T. Ladner, T.A. Oliver, F.O. Seneca, N. Traylor-Knowles & S.R. Palumbi. 2013. Genomic basis for coral resilience to climate change. Proceedings of the National Academy of Sciences of the United States of America, 110:1387-1392.

Chen, H.K., S.N. Song, L.H. Wang, A.B. Mayfield, Y.J. Chen, W.N.U. Chen, et al. 2015. A compartmental comparison of major lipid species in a coral-*Symbiodinium* endosymbiosis: evidence that the coral host regulates lipogenesis of its cytosolic lipid bodies. PLoS ONE, e0132519.

Cinner, J.E., C. Huchery, M.A. MacNeil, N.A. Graham, T.R. McClanahan, J. Maina, et al. 2016. Bright spots among the world's coral reefs. Nature, 535:416-419.

Clarke, K.R., R.N. Gorley, P.J. Somerfield & R.M. Warwick. 2014. Change in marine communities: an approach to statistical analysis and interpretation, 3rd edition. PRIMER-E, Plymouth, UK.

Cleophas T.J., A.H. Zwinderman & T.F. Cleophas. 2006. Clinical data where variability is more important than averages. In: Statistics Applied to Clinical Trials. Springer, Dordrecht.

Conover, D.O., L.M. Clarke, S.B. Munch &

G.N. Wagner. 2006. Spatial and temporal scales of adaptive divergence in marine fishes and the implications for conservation. Journal of Fish Biology, 69:21-47.

Cox, I. & M. Gaudard. 2013. Discovering partial least squares with JMP®. SAS Institute, Inc., Cary, NC, USA.

Han, R., G. Huang, Y. Wang, Y. Xu, Y. Hu, W. Jiang, et al. 2016. Increased gene expression noise in human cancers is correlated with low p53 and immune activities as well as late stage cancer. Oncotarget, 7(44): 72011-72020.

Hoegh-Guldberg, O., P.J. Mumby, A.J. Hooten, R.S. Steneck, P. Greenfield, E. Gomez, et al. 2007. Coral reefs under rapid climate change and ocean acidification. Science, 318:1737-1742.

Hughes, T.P., J.T. Kerry, A.H. Baird, S.R. Connolly, A. Dietzel, C.M. Eakin, et al. 2018. Global warming transforms coral reef assemblages. Nature, 21:492-496.

Jones, S. 2009. A long-term perspective on biodiversity and marine resource exploitation in Fiji's Lau Group. Pacific Science, 63(4):617-648.

Jørgensen, S.E., F.L. Xu, F. Salas & J.C. Marques. 2005. Application of indicators for the assessment of ecosystem health, in: Jørgensen, S.E., F.L. Xu, R. Costanza (Eds). Handbook of Ecological Indicators for Assessment of Ecosystem Health, CRC Press, Boca Raton, pp. 5-66.

Kenkel, C.D., G. Aglyamova, A. Alamaru, R. Bhagooli, R. Capper, R. Cunning, et al. 2011. Development of gene expression markers of acute heat-light stress in reef-building corals of the genus *Porites*. PLoS ONE, e26914.

Kenkel, C.D. & M.V. Matz. 2016. Gene expression plasticity as a mechanism of coral adaptation to a variable environment. Nature Ecology and Evolution, 1:0014.

Krueger, T., N. Horwitz, J. Bodin, M.E. Giovani, S. Escrig, A. Meibom, et al. 2017. Common reef-building coral in the Northern Red Sea resistant to elevated temperature and acidification. Royal Society Open Science, 4:170038.

Lesser, M.P. 1997. Oxidative stress causes coral bleaching during exposure to elevated temperatures. Coral Reefs, 16:187-192.

Louis, Y.D., R. Bhagooli, C.D. Kenkel, A.J. Baker & S.D. Dyall. 2017. Gene expression biomarkers of heat stress in scleractinian corals: promises and limitations. Comparative Biochemistry and Physiology C-Toxicology and Pharmacology, 191:63-77.

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 and Engineering, 4:63.

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., P.S. Chan, H.M. Putnam, C.S. Chen & T.Y. Fan. 2012a. 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 & A.C. Dempsey. 2017a. Biomarker profiling in 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. 2017c. The molecular ecophysiology of closely related pocilloporid corals of New Caledonia. Platax, 14:1-45.

Mayfield, A.B., C.S. Chen & A.C. Dempsey. in review. A statistical platform for assessing coral health in an era of changing global climate-II: the Austral and Cook Islands. Journal of Sea Research.

Mayfield, A.B., C.S. Chen, A.C. Dempsey & A.W. Bruckner. 2016a. 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., C.S. Chen & P.J. Liu. 2014a. 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., M. Chen, P.J. Meng, H.J. Lin, C.S. Chen & P.J. Liu. 2013a. The physiological response of the reef coral *Pocillopora damicornis* to elevated temperature: results from coral reef mesocosm experiments in Southern Taiwan. Marine Environmental Research, 86:1-11.

Mayfield, A.B., Y.H. Chen, C.F. Dai & C.S. Chen. 2014b. The effects of temperature on gene expression in the Indo-Pacific reef-building coral *Seriatopora hystrix*: insight from aquarium studies in Southern Taiwan. International Journal of Marine Science, 4(50):1-23.

Mayfield, A.B., Y.J. Chen, C.Y. Lu & C.S Chen. 2016b. Proteins responsive to variable temperature exposure in the reef-building coral *Seriatopora hystrix*, in: Ortiz, S. (Ed.) Coral Reefs: Ecosystems, Environmental Impact and Current Threats. NOVA

Publishing, New York, pp. 1-60.

Mayfield, A.B., Y.J. Chen, C.Y. Lu & C.S. Chen. 2018a. Exploring the environmental physiology of the Indo-Pacific reef coral *Seriatopora hystrix* using differential proteomics. The Open Journal of Marine Science, 8:223-252.

Mayfield, A.B., Y.J. Chen, C.Y. Lu & C.S. Chen. 2018b. The proteomic response of the reef coral *Pocillopora acuta* to experimentally elevated temperature. PLoS ONE, e0192001.

Mayfield, A.B., T.Y. Fan & C.S. Chen. 2013b. 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. 2013c. The physiological impact of *ex situ* transplantation on the Taiwanese reef-building coral *Seriatopora hystrix*. Journal of Marine Biology, Article ID 569369.

Mayfield, A.B., T.Y. Fan & C.S. Chen. 2013d. 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., Y.Y. Hsiao, H.K. Chen & C.S. Chen. 2014c. Rubisco expression in the dinoflagellate *Symbiodinium* sp. is influenced by both photoperiod and endosymbiotic lifestyle. Marine Biotechnology, 16:371-384.

Mayfield, A.B., Y.Y. Hsiao, T.Y. Fan, C.S. Chen. 2012b. Temporal variation in RNA/DNA and protein/DNA ratios in four anthozoan-dinoflagellate endosymbioses of the Indo-Pacific: implications for molecular diagnostics. Platax, 9:1-24.

Mayfield, A.B., Y.Y. Hsiao, T.Y. Fan, C.S. Chen & R.D. Gates. 2010. Evaluating the temporal stability of stress-activated protein kinase and cytoskeleton gene expression in the Pacific corals *Pocillopora damicornis* and *Seriatopora hystrix*. Journal of Experimental Marine Biology and Ecology, 395:215-222.

Mayfield, A.B., L.H. Wang, P.C. Tang, Y.Y. Hsiao, T.Y. Fan, C.L. Tsai, et al. 2011. Assessing the impacts of experimentally elevated temperature on the biological composition and molecular chaperone gene expression of a reef coral. PLoS ONE,

e26529.

Mayfield, A.B., Y.B. Wang, C.S. Chen, S.H. Chen & C.Y. Lin. 2014d. Compartment-specific transcriptomics in a reef-building coral exposed to elevated temperatures. Molecular Ecology, 23:5816-5830.

Mayfield, A.B., Y.B. Wang, C.S. Chen, S.H. Chen & C.Y. Lin. 2016c. Dual-compartmental transcriptomic +proteomic analysis of a marine endosymbiosis exposed to environmental change. Molecular Ecology, 25:5944-5958.

Mazerolle, M., 2006. Improving data analysis in herpetology: using Akaike's Information Criterion (AIC) to assess the strength of biological hypotheses. Amphibia-Reptilia, 27:169-180.

Palumbi, S.R., D.J. Barshis, N. Traylor-Knowles & R.A. Bay. 2014. Mechanisms of reef coral resistance to future climate change. Science, 344:895-898.

Parkinson, J.E., E. Bartels, M.K. Devlin-Durante, C. Lustic, K. Nedimyer, S. Schopmeyer, et al. 2018. Extensive transcriptional variation poses a challenge to thermal stress biomarker development for endangered coral. Molecular Ecology, 27:1103-1119.

Putnam, H.M., K. Barott, T.D. Ainsworth & R.D. Gates. 2017. The vulnerability and resilience of reef-building corals. Current Biology, 27:R528-R540.

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 pCO_2. Marine Biology, 160:2157-2173.

Schmidt-Roach, S., K.J. Miller, P. Lundgren, & N. Andreakis. 2014. With eyes wide open: a revision of species within and closely related to the *Pocillopora damicornis* species complex (Scleractinia; Pocilloporidae) using morphology and genetics. Zoological Journal of the Linnaean Society, 170:1-33.

Sharma, A., C. Jiang & S. De. 2018. Dissecting the sources of gene expression variation in a pan cancer analysis identifies novel regulatory mutations. Nucleic Acids Research, 46:4370-4381.

Sheppard, C., A. Sheppard, A. Mogg, D. Bayley, A.C. Dempsey, R. Roche et al., 2017. Coral bleaching and mortality in the Chagos Archipelago. Atoll Research Bulletin, 613.

Shmueli, G. 2011. To explain or to predict? Statistical Science, 25(3):289-310.

Traylor-Knowles, N., B. Granger, T. Lubinski, J. Parikh, S. Garamszegi, Y. Xia, et al. 2011. Production of a reference transcriptome and

a transcriptomic database (PocilloporaBase) for the cauliflower coral, *Pocillopora damicornis*. BMC Genomics, 12(1):585.

Welchman, R.L., C. Gordon & R.J. Mayer. 2005. Ubiquitin and ubiquitin-like proteins as multifunctional signals. Nature Reviews Molecular Cell Biology, 6:599-609.

Supplemental tables

Supplemental table 1. PRIMER's (ver. 6) distance-based linear modeling (DistLM) function using a Euclidean distance similarity matrix of 70 Lau Archipelago (Fiji) samples for which no data were missing. Categorical environmental data were converted into integers (e.g., Totoya Island=1, Matuku Island=2, etc.). Environmental parameters (EP) that were significantly correlated with the multivariate response in the marginal test (p<0.05) have been underlined in the "Exact model terms" column. Table abbreviations: Adj.==adjusted. AICc=Akaike's information criterion. ALCC=average live coral cover. BIC=Bayesian information criterion.

Selection criterion	Selection procedure	r^2	Adj. r^2	AICc/ BIC	# terms	Exact model terms
adjusted r^2	all specified	0.22	0.06	NA	12	all
adjusted r^2	forward	0.19	0.09	NA	8	island, reef zone, reef type, depth, ALCC, time, host, color
adjusted r^2	backward	0.19	0.09	NA	8	island, reef zone, reef type, depth, ALCC, time, host, color
adjusted r^2	stepwise	0.19	0.09	NA	8	island, reef zone, reef type, depth, ALCC, time, host, color
adjusted r^2	best	0.19	0.09	NA	8	island, reef zone, reef type, depth, ALCC, time, host, color
When attempting to create a model with the minimum adj. r^2, one featuring island, reef zone, reef type, colony depth, ALCC, time, host species, and colony color was optimal.						
minimum AICc	all specified	0.22	NA	180	12	all (only island and reef zone were significant)
minimum AICc	forward	0.04	NA	166	1	island
minimum AICc	backward	0.09	NA	167	3	island, reef zone, reef type
minimum AICc	stepwise	0.04	NA	166	1	island
minimum AICc	best	0.04	NA	166	1	island
minimum BIC	all specified	0.22	NA	203	12	all (only island and reef zone were significant)
minimum BIC	forward	0.04	NA	171	1	island
minimum BIC	backward	unable to calculate				
minimum BIC	stepwise	unable to calculate				
minimum BIC	best	0.04	NA	171	1	island
When attempting to create a model with the minimum AICc or BIC, island was generally the most important EP in explaining the multivariate response (with reef zone being #2).						

Supplemental table 2. Stepwise regression (SR) details. Models were built in stepwise, backwards fashion to minimize the Bayesian information criterion (BIC). Statistically significant (*F*-test, $p<0.01$) environmental parameter (EP) model terms have been underlined, and statistically significant models (<0.05) are denoted by asterisks (*). Of the 90 samples analyzed, the lone *Pocillopora brevicornis* colony sampled was excluded, as were 20 samples featuring missing data (resulting in a final sample size of 69). Please note that "max. length" was included in this analysis but was excluded from the multivariate analyses (e.g., PERMANOVA) discussed in the main text. The Austral and Cook Islands ("Australs-Cooks") dataset was first presented in Mayfield et al. (2015), with a more thorough multivariate statistics analysis undertaken more recently (Mayfield et al., in review), and, unlike for the Fiji data, only the significant model terms have been shown for the Australs-Cooks dataset. Table abbreviations: Adj.=adjusted. ALCC=average live coral cover. NA=not applicable. PAR=photosynthetically active radiation. RV=response variable. temp.=temperature. Please see the "Abbreviations" table in the main text for additional abbreviations (e.g., the variability terms).

Response variable	BIC	Adj. r^2	Fiji dataset SR model terms	Australs-Cooks dataset SR model terms (adj. r^2)	Conclusions
Max. length[a]	559*	0.37	PAR>host>island>depth	host (0.12)	strong effect of host
RNA/DNA[b]	12.1*	0.21	depth>island	ALCC (0.44)*	inconsistent results across regions
Sym GCP	539*	0.31	depth>color>reef exposure	host>depth>ALCC (0.47)*	strong effect of depth
***Symbiodinium* gene expression**					
hsp90[c]	213*	0.16	depth>island	island (0.21)*	poor-fit models
apx1[b]	320*	0.07	color	salinity (0.37)	inconsistent results across regions
ubiq-lig[d]	178*	0.22	depth>reef type>host	host (0.28)*	strong effect of host
rbcL[d]	352*	0.16	depth>reef type	RV not assessed	poor-fit models
zifl1l[d]	329*	0.37	ALCC>exposure>salinity	RV not assessed	NA
Host coral gene expression					
gfp-cp[c]	250*	0.31	island>depth>salinity	RV not assessed	NA
cu-zn-sod[d]	208*	0.31	host>depth>island	RV not assessed	NA
lectin[a]	603*	0.16	depth>color	RV not assessed	poor-fit models
ca[c]	221*	0.12	island>color	RV not assessed	poor-fit models
Variability terms					
PC1[a]	601*	0.25	depth>color>exposure	temp.>host>ALCC (0.37)*	inconsistent results across regions
MD[a]	607*	0.18	ALCC>exposure>island	island>host (0.40)*	strong effect of island
VI[c]	108*	0.12	ALCC>depth	host (0.28)	poor-fit models
Mean±std. dev.		0.22±0.09		0.33±0.11	

[a]rank-transformed data. [b]square root-transformed data. [c]log-transformed data. [d]4th root-transformed data.

Supplemental table 3. Partial least squares (PLS) analysis when treating environmental parameter (EP) data (n=12) as continuous terms (not categorical) in the statistical model input. As an example, the categorical EP "island" was treated as follows: Totoya=1, Matuku=2, Moala=3, etc. Positive coefficients have been underlined for the PLS rows, whereas significant (*F*-test, $p<0.01$) individual model terms have been underlined in the stepwise regression (SR) rows. EP with variable importance parameter (VIP) values >1 ("VIP EP;" right-most column) have been ordered from highest coefficient value (with positive associations underlined) to lowest (typically negative correlations). As a comparison, the SR data from the Austral and Cook Islands ("A-C") dataset (Mayfield et al., in review) have also been presented (PLS was not undertaken with the A-C dataset.). The SR and PLS cumulative ("Cum") y (response variable) variation explained (as percentage data) have also been plotted as an inset in Fig. 4. All error terms represent standard deviation. Table abbreviations: ALCC=average live coral cover. #factors (PLS)= the number of latent PLS factors. NA=not applicable. MD=Mahalanobis distance. PAR=photosynthetically active radiation. PC1=principal components analysis primary axis loading score. VI=variability index.

Response variable	#factors (PLS)	#VIP>1.0	Cum X variation explained	Cum Y varia-tion explained	VIP EP
Max. length-Fiji (PLS)[a]	1	5	16%	13%	depth>host>exposure>color>salinity
Max. length-Fiji (SR)[a]	NA	NA	NA	37%	PAR>host>island>depth
Max. length-A-C (SR)[c]	NA	NA	NA	12%	host
Max. length summary (Fig. 4a): host & depth in Fiji; host in A-C; host in both regions					
global mean max. length variation explained				**21±14%**	
RNA/DNA-Fiji (PLS)[b]	1	3	12%	13%	depth>PAR>temp.>time>ALCC
RNA/DNA-Fiji (SR)[b]	NA	NA	NA	21%	depth>island
RNA/DNA-A-C (SR)[a]	NA	NA	NA	44%	ALCC>host>depth
RNA/DNA summary (Fig. 4b): depth & ALCC in both regions					
global mean RNA/DNA variation explained				**26±16%**	
Sym GCP-Fiji (PLS)	1	4	24%	17%	exposure>island>reef zone>reef type
Sym GCP-Fiji (SR)	NA	NA	NA	31%	depth>color>exposure
Sym GCP-A-C (SR)	NA	NA	NA	47%	host>depth>ALCC>color
Sym GCP summary (Fig. 4c): 1) exposure & reef type in Fiji; host, depth, & ALCC in A-C; color & ALCC in both regions					
global mean Sym GCP variation explained				**32±15%**	
Sym *hsp90*-Fiji (PLS)[c]	1	3	22%	5.6%	reef zone>ALCC>color
Sym *hsp90*-Fiji (SR)[c]	NA	NA	NA	16%	depth>island
Sym *hsp90*-A-C (SR)[c]	NA	NA	NA	21%	island
Sym *hsp90* summary (Fig. 4d): generally inconsistent findings across methods; models characterized by low predictive power					
global mean Sym *hsp90* variation explained				**14±7.9%**	
Sym *apx1*-Fiji (PLS)[b]	1	3	12%	16%	island>salinity>color
Sym *apx1*-Fiji (SR)[b]	NA	NA	NA	12%	color>island
Sym *apx1*-A-C (SR)[c]	NA	NA	NA	37%	salinity>color>depth> temp.>host>island>ALCC
Sym *apx1* summary (Fig. 4e): color, island, & salinity in both regions					
global mean Sym *apx1* variation explained				**22±13%**	
Sym *ubiq-lig*-Fiji (PLS)[d]	1	6	21%	12%	reef zone>reef type>time>ALCC>exposure>color
Sym *ubiq-lig*-Fiji (SR)[d]	NA	NA	NA	22%	depth>reef type>host
Sym *ubiq-lig*-A-C (SR)[a]	NA	NA	NA	28%	host>time
Sym *ubiq-lig* summary (Fig. 4f): reef type, time, & host in both regions					
global mean Sym *ubiq-lig* variation explained				**21±8.0%**	
Sym *rbcL*-Fiji (PLS)[d]	1	4	21%	16%	reef zone>island>exposure>color

Sym *rbcL*-Fiji (SR)[d]	NA	NA	NA	16%	depth>reef type
Sym *rbcL* summary (Fig. 4g): generally inconsistent findings across methods; models characterized by low predictive power					
global mean Sym *rbcL* variation explained				**16%**	
Sym *zifl1l*-Fiji (PLS)[d]	1	5	17%	33%	salinity>island>exposure>PAR>time
Sym *zifl1l*-Fiji (SR)[d]	NA	NA	NA	36%	ALCC>exposure>salinity
Sym *zifl1l* summary (Fig. 4h): salinity in both regions					
global mean Sym *zifl1l* variation explained				**35±2.1%**	
Host coral *gfp-cp*-Fiji (PLS)[c]	1	3	14%	8%	island>ALCC>host
Host coral *gfp-cp*-Fiji (SR)[c]	NA	NA	NA	32%	island>depth>salinity
Host coral *gfp-cp* summary (Fig. 4i): island in both regions					
global mean host *gfp-cp* variation explained				**20±17%**	
Host coral *cu-zn-sod*-Fiji (PLS)[d]	1	3	19%	8%	island>salinity>ALCC
Host coral *cu-zn-sod*-Fiji (SR)[d]	NA	NA	NA	31%	host>depth>island
Host coral *cu-zn-sod* summary (Fig. 4j): island in both regions					
global mean host *cu-zn-sod* variation explained				**20±16%**	
Host coral *lectin*-Fiji (PLS)[a]	2	5	27%	17%	depth>PAR>salinity>island>color
Host coral *lectin*-Fiji (SR)[a]	NA	NA	NA	8%	depth>color
Host coral *lectin* summary (Fig. 4k): island in both regions, though models characterized by low predictive power					
global mean host *lectin* variation explained				**13±6.4%**	
Host coral *ca*-Fiji (PLS)[c]	1	4	12%	13%	salinity>host>color>PAR
Host coral *ca*-Fiji (SR)[c]	NA	NA	NA	12%	island>color
Host coral *ca* summary (Fig. 4l): color in both regions, though models characterized by low predictive power					
global mean host *ca* variation explained				**13±0.7%**	
VI-Fiji (PLS)[c]	1	4	10%	5%	host>color>ALCC>time
VI-Fiji (SR)[c]	NA	NA	NA	12%	ALCC>depth
VI-A-C (SR)[c]	NA	NA	NA	28%	host>ALCC>island
VI summary (Fig. 4m): ALCC in both regions, though models characterized by low predictive power					
global mean VI variation explained				**15±12%**	
PC1-Fiji (PLS)[c]	1	6	31%	25%	color>reef type>reef zone>island>temp.>ALCC
PC1-Fiji (SR)[a]	NA	NA	NA	25%	depth>color>exposure
PC1-A-C (SR)[a]	NA	NA	NA	37%	temp.>host>ALCC>salinity
Principal components (PC1) summary (Fig. 4n): color & ALCC in both regions					
global mean PC1 variation explained				**29±6.9%**	
MD-Fiji (PLS)[a]	1	5	23%	5%	reef zone>exposure>ALCC>temp.>color
MD-Fiji (SR)[a]	NA	NA	NA	18%	ALCC>exposure>island
MD-A-C (SR)[a]	NA	NA	NA	40%	island>host>ALCC
MD summary (Fig. 4o): ALCC & exposure in Fiji; island & host in A-C; ALCC in both regions					
global mean MD variation explained				**21±18%**	

Mean PLS r^2 (x 100)-Fiij (y)	14±7.6[a]	color was most common term in PLS best-fit models (10/15)
Mean SR r^2 (x 100)-Fiji	22±9.6[b]	depth was most common term in Fiji SR best-fit models (9/15)
Mean SR r^2 (x 100)-A-C	34±8.6%[c]	host was most common term in A-C SR best-fit models (5/9)
Mean r^2 (x 100) of both approaches (PLS & SR)-Fiji	19±9.4%[b]	island (13/30) & color (13/30) were most common terms in Fiji best-fit models
Mean r^2 (x 100) of both approaches (PLS & SR)-Fiji+A-C	22±12%[b]	island was most common term in all best-fit models (15/39)

[a]rank-transformed. [b]square root-transformed. [c]log-transformed. [d]4th root-transformed.