國立海洋生物博物館
NATIONAL MUSEUM OF MARINE BIOLOGY & AQUARIUM
Platax 17: 27-52, 2020
doi: 10.29926/platax.202012_17.0003

Exploiting the power of multivariate statistics for probing the cellular biology of environmentally challenged reef corals

Anderson B. Mayfield[1,2]

[1]Cooperative Institute for Marine and Atmospheric Studies, University of Miami, Miami, FL 33149, USA

[2]Atlantic Oceanographic and Meteorological Laboratory, National Oceanic and Atmospheric Administration, Miami, FL 33149, USA

*Corresponding author. Email: abm64@miami.edu or anderson.mayfield@noaa.gov

Abstract

The coral reef research field has grown markedly in terms of both human power and technological capacity in recent years, a fortuitous occurrence given the rapidly diminishing nature of Earth's reefs on account of climate change and other anthropogenic stressors. Unfortunately, most coral biologists lack the statistical background to realize the full analytical potential of "big" datasets emerging from (non-exhaustively) 1) expanding reef survey efforts, 2) satellite and in-water (e.g., photomosaic) coral reef imaging projects, and 3) "next-generation" molecular approaches (i.e., 'OMICs); statistical training has not advanced commensurately with dataset size, a significant short-coming when considering the utility of these data in informing coral reef ecosystem management and conservation. One notably pervasive issue in 'OMICs research in particular is the general omission of multivariate statistical approaches (MSA), which universally outperform the more commonly employed, less statistically conservative univariate alternatives when attempting to A) model experimental results and B) make predictions about reef coral health and fate. Herein I have attempted to make a case for coral biologists to strongly re-evaluate the merit of MSA, as well as explain why relying on univariate approaches alone may actually lead to spurious findings that do not advance our knowledge of corals and coral reefs.

Key words: Climate change, Coral reefs, Dinoflagellate, Molecular biology, Predictive modeling, Upwelling Statistics

Introduction

Over the past 20 years, the collective research effort of the coral reef field has grown immeasurably as more and more young scientists in particular seek to study and/or save these incredibly beautiful (Fig. 1), yet rapidly vanishing, ecosystems. Serendipitously, technological capacity has also increased dramatically as of recent, and many large-scale questions on the fundamental nature and persistence of coral-dinoflagellate endosymbioses can now be addressed in ways, and with levels of rigor and depth, that would have been unheard of even a decade ago (Mayfield et al., 2019b). It could be argued that advances in molecular biology (Mayfield et al., 2011) and high-level data analytics (Mayfield & Chen, 2019) have benefited the coral reef field (& certainly many others) to greater degrees than any and all other emerging discipline(s). As an example, a Ph.D. dissertation on scleractinian coral gene expression that required 5-6 years of concerted field and laboratory bench work (~60-70 hr/week; Mayfield, 2009; Mayfield et al., 2010) can now be completed in several days given the existence of optimized protocols (Mayfield et al., 2009) and well curated and annotated DNA/mRNA sequence databases (e.g., http://symbiont.iis.sinica.edu.tw/coral_pdltte/static/html/index.html#home for *Pocillopora acuta* & http://symbiont.iis.sinica.edu.tw/s_hystrix/static/html/ for *Seriatopora hystrix*).

In many such cases, though, "big" datasets are generated (e.g., sequenced genomes or transcriptomes) and explored using "tried and true" (bio)informatics pipelines (e.g., Xiang et al., 2015), yet key facets of the underlying biology are nevertheless overlooked due to analytical errors or oversights stemming from lack of training in statistics. This arises not only from poor data stewardship, but also from the "castles in the sky" phenomenon, in which a faulty script or statistical test is published in a respected, peer-reviewed journal, regularly cited, and then ingratiated within the field's *Zeitgeist* to where it is no longer questioned. For instance, failure to rigorously assess the validity of the ITS2 marker in the initial Symbiodiniaceae genotyping protocols developed in the 1990s culminated in the publication of hundreds of articles on the importance of endosymbiont diversity in coral health (e.g., LaJeunesse, 2001) that were later refuted through a series of simple, cloning-based laboratory experiments undertaken by a highly motivated graduate student who uncovered extensive sequence heterogeneity of this gene *within* the Symbiodiniaceae genome (Apprill &

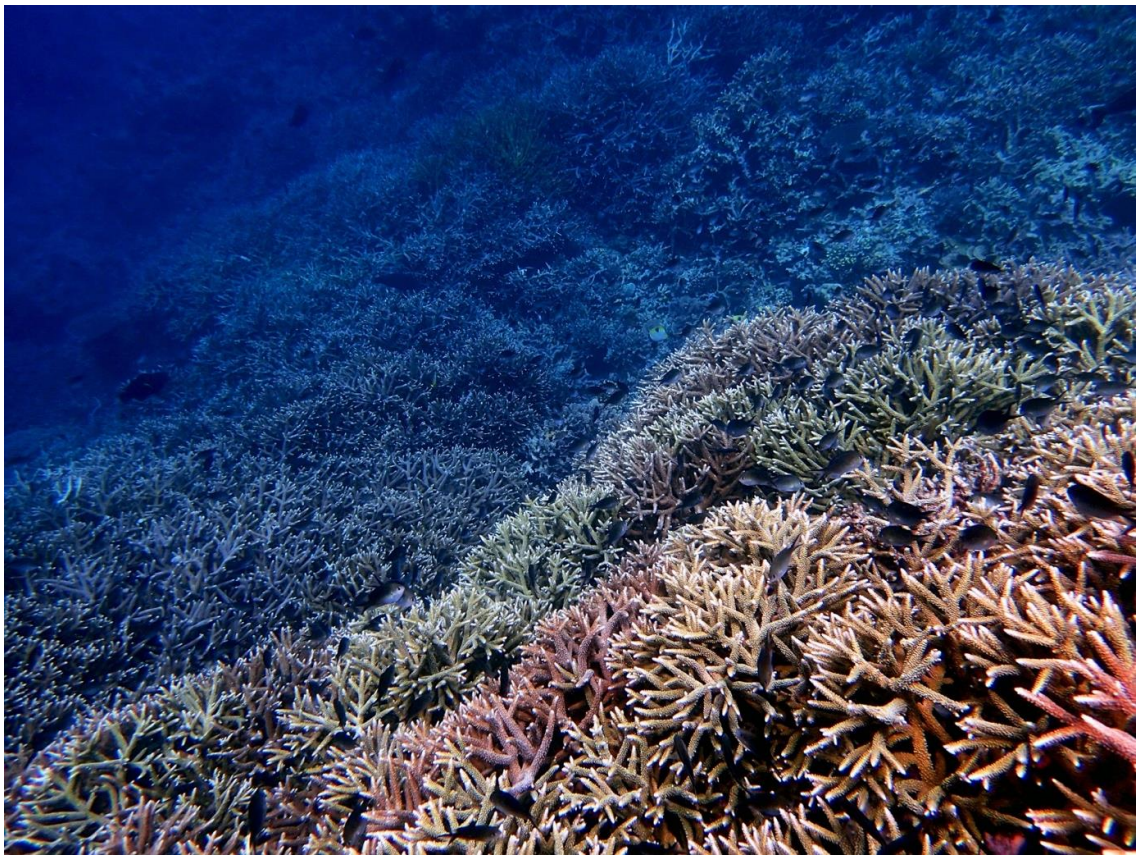

Fig. 1. A recent lava flow overtaken by several species of the reef-building coral genus *Acropora*. "Lava Flow," off Gunung Api (volcano), Banda Islands, Maluku, Indonesia. Photo credit=Anderson Mayfield (a.k.a. 珊瑚醫生).

Gates, 2007); every ITS2 "type" can be found within a single cell, thereby invalidating all published research on the use of this marker for assessing dinoflagellate diversity (Putnam et al., 2017). Rather than go "back to the drawing board" and admit defeat, the field has continued to build upon the ITS2-driven dinoflagellate diversity "castle" (e.g., LaJeunesse et al., 2009), which sits atop a decidedly flimsy foundation ("the air"). Entire careers and "paradigm-shifting" stories have been built off of a mistake that was pointed out from data collected across only several days of benchwork. Let us not allow sophomoric statistical errors currently rife within the coral 'OMICs literature (e.g., DeSalvo et al., 2008; Meron et al., 2019) invalidate the potentially immense value of the myriad coral-dinoflagellate 'OMICs datasets now being generated by researchers across the

globe. Before addressing this central theme of this article, allow me to make one additional conceptual recommendation for those, like myself, interested in anthozoan health and physiology (Mayfield et al., 2014a,c).

Most studies of coral biology focus on only one member of the mutualism: the coral host (e.g., Palumbi et al., 2014) or the dinoflagellate endosymbionts. To be fair, science is inherently reductionistic, particularly molecular biology; if one is only interested in, for instance, coral gene expression, why need one concern his/herself with the dinoflagellates' transcriptomes? This may indeed be a valid question for those seeking to elucidate coral thermo-biology, but for researchers (such as myself) interested in questions pertaining to coral health, resilience, and persistence, it is decidedly myopic and overly simplistic to consider the biology only one of the plethora of "compartments" within the holobiont (Cheng et al., 2016); only by probing the biology of the anthozoan host (~60%), its gastrodermal symbionts (30%), *and* the diverse microbial consortium (10%) collectively comprising the holobiont (Rosales et al., 2019) can we hope to paint a holistic picture of the health of a coral polyp or colony (Mayfield & Gates, 2007). In contrast, by characterizing host physiology alone, which is the current norm in the field (but see Mayfield, 2016.), a researcher will be unable to demarcate a level of health or make confident predictions regarding coral fate in this era of rapidly changing global climate (Mayfield & Chen, this issue).

With respect to 'OMICs technologies specifically, in almost all cases, the microbial (including Symbiodiniaceae) data are collected alongside those of the host at no extra expense and with no additional need to undertake further laboratory benchwork; the only extraneous effort required to incorporate these data would be in the time required to analyze them (granted, this is not a trivial endeavor for "big" datasets.). As such, all suggestions outlined below apply to data from both the anthozoan and microbial fractions. Whether the coral and microbial analytes are pooled or separated (e.g., Chen et al., 2017a) prior to statistical analyses, however, is project-dependent, and health-informing, diagnostic findings could be garnered from either approach.

As did quality control fall to the wayside in early (& some more recent) works on Symbiodiniaceae genetics and multi-compartmental coral molecular eco-physiology, some common statistical mistakes that would not be accepted by reviewers of traditional articles featuring "small" datasets (e.g., Enochs et al., 2020) are often overlooked because, in many

cases, those having undertaken "big data" analyses (e.g., transcriptomics) surely spent a considerable amount of money and so could not readily redo the experiment(s) (e.g., Mayfield et al., 2018b); this could potentially assuage reviewers to be more sympathetic. As an example, in the pursuit of attempting to understand the implications of global climate change (GCC) on coral-dinoflagellate biology (sensu Mayfield et al., 2013a-d, 2014a; Putnam et al., 2013), there has been a focus on uncovering differentially concentrated analytes (DCAs; e.g., genes, proteins, lipids, etc.) in a "*p*-value-mining" fashion; although these response screening analyses (RSA) are typically false discovery rate (FDR)-governed, if researchers fail to first uncover more general multivariate patterns in the dataset, one could make a compelling argument that RSA are actually unsupported or inappropriate. After all, an article presenting post-hoc differences (e.g., Tukey's tests) among means in the absence of an overall ANOVA treatment effect would be flagged by a competent reviewer. In 'OMICs research, however, this fundamental statistical flaw is oft overlooked (i.e., multiple univariate comparisons in the absence of a treatment effect). Whether this is more due to reviewer ignorance of statistics or the aforementioned willingness to overlook issues in datasets that have impressed in their size, scope, and/or cost of acquisition, remains to be determined.

However, I would actually argue *against* the latter excuse: were one to invest heavily (in terms of funding, effort, & time) in a study in which a massive, potentially rich dataset was compiled, it would be my hope that both authors *and* reviewers would aim to ensure that it be exploited to the fullest extent possible. This idea is, in fact, another goal of this article: to highlight the notion that insightful multivariate data exploration algorithms, techniques, and presentation styles (beyond the few commonly adopted by coral biologists at present) exist and should be more seriously incorporated, particularly in the analysis of "big" datasets. I envision this as a call to "up the (statistics) ante" and urge researchers to take advantage of both "time-tested" and more contemporary data analysis approaches for making sense of the evermore complex datasets being generated nowadays. There will be a particular focus on 'OMICs research as the associated datasets are amongst the most analytically daunting generated to date in the coral biology field, especially when these data are mapped onto habitats (e.g., Chen et al., 2017b, 2019) to attempt to understand the environmental drivers of variation in coral health and resilience

(Mayfield et al., 2015, 2016a, 2017a-c, 2018c, 2019a, in prep.). My over-arching hypothesis is that most 'OMICs researchers (not limited to coral biologists) would do well to familiarize themselves with the multivariate statistical approaches (MSA) beyond the lone method commonly found in the current literature: principal components analysis (PCA). Not only are these additional MSA more robust, but, as mentioned in the previous paragraph, it is actually statistically invalid to progress to the far more commonly reported univariate findings *unless* the MSA provide strong support for doing so. This extra degree of statistical conservation should not be seen as a barrier to publication, but instead as a means of tapping into the power of MSA (& preferably MSA combined with information theory) while simultaneously limiting the possibility of type I statistical errors near-ubiquitous within *p*-value-mining studies (Anderson et al., 2000).

Materials & methods

As mentioned above, a plethora of studies have attempted to utilize cutting-edge molecular approaches like RNA-Sequencing (RNA-Seq) and mass spectrometry (MS)-based proteomics to elucidate the implications of GCC and other anthropogenic stressors on the cellular biology of marine organisms (Monteiro et al., 2020); the coral reef field has been an especially strong advocate of such 'OMICs approaches (Wang et al., 2013; Mayfield et al., 2014d; McRae et al., in review). In typical eco-physiological experiments, corals or cultured Symbiodiniaceae are exposed to experimental treatment(s) in the laboratory for a pre-set period of time prior to sampling biopsies for later macromolecular extraction (e.g., Mayfield et al., 2011). Upon sequencing the resulting DNAs, mRNAs, proteins, lipids, and/or metabolites via genomic (amplicon sequencing or meta-genomics), transcriptomic (RNA-Seq), proteomic (MS), lipidomic (MS), and metabolomic (MS) technologies, respectively, the researcher must typically attempt to discern patterns among tens of thousands, tens of thousands, thousands, hundreds, and hundreds of molecules, respectively. In very few cases are the number of analyzed samples greater than the number of analytes; even when profiling lipids, for which only 100-200 can typically be quantified in a single study, it is unlikely that the researcher's budget would lend itself to analyzing 100-200 samples at current rates of $200USD/sample (similar to late-2020 prices for the other 'OMICs technologies). Although a dataset with 10-20 samples in total is publishable in even the most highly respected journals (assuming a robust experimental design &

preferably 3-5 replicates per treatment), what this signifies is that traditional MSA, namely multivariate ANOVA (MANOVA), *cannot* be performed with such datasets because MANOVA generally necessitates that the number of analytes be *less than* the number of samples. In other words, MANOVA cannot be used in >99% of 'OMICs studies.

It is for this reason, perhaps, that MSA have fallen out of favor among 'OMICs researchers. Additionally, many are only interested in uncovering DCAs (e.g., differentially expressed genes [DEGs] & differentially concentrated proteins [DCPs] in their transcriptomic & proteomic projects, respectively), since, even when using a FDR-controlled *p*-value, it is likely that at least several analytes will differ significantly in concentration across treatments (i.e., results around which a manuscript-selling story could be generated). However, as mentioned above, this is analogous to proceeding directly to inter-mean post-hoc tests without having first carried out the requisite ANOVA; is it appropriate to search for univariate effects (i.e., on an analyte-by-analyte basis) *without* having previously documented an overall multivariate treatment effect? The general consensus seems to be that no such multivariate assessment is needed if one simply wants to search for DCAs across treatments, though this could simply be due to the aforementioned inability to use parametric statistical approaches (PSA) to uncover multivariate treatment effects. In fact, there *are* a number of ways to document multivariate treatment effects with 'OMICs datasets, and I 1) describe them herein using a sample dataset and 2) advocate for their more widely adopted use in future 'OMICs analyses (not limited to corals). Only when the approaches outlined below identify a treatment effect should the more rudimentary, commonly employed RSA be undertaken, preferably using an FDR-adjusted alpha of no higher than 0.01.

PERMANOVA. Permutational MANOVA (PERMANOVA; Anderson et al., 2008) is likely the preferred means of assessing multivariate differences within 'OMICs datasets because it can be carried out when there are more analytes than samples and is robust even when analyte concentrations are collinear (a hallmark of 'OMICs datasets; discussed in more detail below). Furthermore, since it is based on assessment of inter-sample similarity, it is also robust to datasets featuring numerous 0s (i.e., analytes that went unsequenced in certain samples; Clarke et al., 2014). PERMANOVA can be undertaken with PRIMER (UK) or with open-access scripts (e.g., R). In most cases, rows and columns represent samples and analytes, respectively, and transformations are

critical since there may be large variation in mean analyte concentrations; some analytes might be at far higher concentrations than others (& so would be over-weighted in the analysis). The preferred means of accommodating inter-analyte variation is standardization (Mayfield, in review), whereby the data for each analyte are transformed such that the mean is 0 and the standard deviation is 1. This then puts each analyte on an "even playing field."

Given that such standardized data will inherently feature numerous negative values, the Euclidean distance matrix (EDM) is recommended over the Bray-Curtis similarity matrix (BCSM), which cannot be computed with negative values. If a BCSM is preferred, then a "dummy" variable can be added to the standardized data to where all values are positive. Alternatively, a fourth-root transformation of the raw data might down-weigh high-concentration analytes sufficiently to justify its use over standardization, in which case a BCSM could be constructed. Ultimately, the transformation and distance matrix used may be justified by determining which best limits heterogeneity in multivariate dispersion across treatments (Anderson, 2006); although likely less critical than when undertaking parametric ANOVA, in which homogeneity of (univariate) variance is critical (Levene's test [or comparable], $p>0.05$ to justify use assuming a normal distribution), multivariate dispersion should ideally not vary significantly across experimental treatments (i.e., PERMDISP [PRIMER], $p>0.05$).

As an example, please consider the "*Seriatopora hystrix* variable temperature study" (SHVTS) of Mayfield et al. (2012a), in which we exposed corals from an upwelling site known as Houbihu (see Lee et al., 1997; Ribas-Deulofeu et al., 2016; & Liu et al., 2012 for oceanographic, ecological, & human-impact data, respectively.) or a non-upwelling site (Houwan) to either upwelling-simulating (23-29°C over 6 hr) or stable-temperature (26°C) conditions for a week and profiled the molecular physiology of corals "transplanted" (ex situ) to foreign temperature environments (i.e., Houbihu➔stable-temperatures & Houwan➔variable-temperatures) vs. non-transplanted controls (e.g., Houbihu corals exposed to upwelling-simulating conditions). In this experiment, then, there were two fixed factors: site of origin (SO; n=2) and temperature treatment (TT; stable vs. variable). We were particularly interested in their interaction, since we hypothesized that corals would demonstrate superior physiological performance when exposed to their "native" temperature conditions. Although this was

indeed documented to a certain extent (Mayfield et al., 2012a), the Houwan corals actually appeared unstressed at the highly variable temperature regime (one they never encounter in situ). It was later found that all corals were of the same genotype (Mayfield et al., 2014b), signifying that all phenotypic plasticity can be attributed to environmental, rather than genetic, differences. We later characterized the transcriptomes (Mayfield et al., 2016c) and proteomes (Mayfield et al., 2016b, 2018a; Mayfield, 2020) of a subset of 12 samples to make conjectures about how these corals acclimated to variable temperatures in the laboratory, with the ultimate goal of shedding light on the molecular basis of thermo-acclimatization and phenotypic plasticity *in situ*. In an effort to promote data dissemination and transparency (discussed in more detail below), the physiological and target gene expression data from Mayfield et al. (2012a, 2014b) from this same subset have been posted on JMP® Public: https://public.jmp.com/packages/Response-of-corals-from-two-sites-to-sta/js-p/4blb_hY7xkwQ4srBMp55N.

The transcriptome features over 125,000 contigs of both host coral and Symbiodiniaceae dinoflagellate (*Cladocopium* spp.) origin and can be downloaded from the *S. hystrix*-Symbiodiniaceae SHVTS transcriptome server referenced above, as well as NCBI. To demonstrate the utility of PERMANOVA, I first standardized the data from a subset of ~93,000 contigs that passed all quality control. It is important to note that, unlike virtually all RNA-Seq pipelines referenced in the literature (e.g., Barshis et al., 2013), there was no requirement that a contig must be expressed by all samples; even those for which only 1 of the 12 samples expressed a particular gene were maintained in the analysis. The omission of 0s from 'OMICs datasets likely stems from the early days of next-generation technologies, in which few (if any) replicates were analyzed. When replication is limited, it is difficult to discern whether a concentration of 0 actually reflects an absence of that analyte in a sample or whether it was present but simply went unsequenced. Assuming proper replication (n>2), if there is a gene that is expressed by all replicates of one treatment and in none of the remainders, then this mRNA is likely to actually be the *most* informative with respect to developing mechanistic hypotheses describing the treatment effect (Mayfield et al., in review). If a second DEG is measured at 2-fold higher expression levels in one SO in the SHVTS, for instance, it is unclear whether this difference would have significant

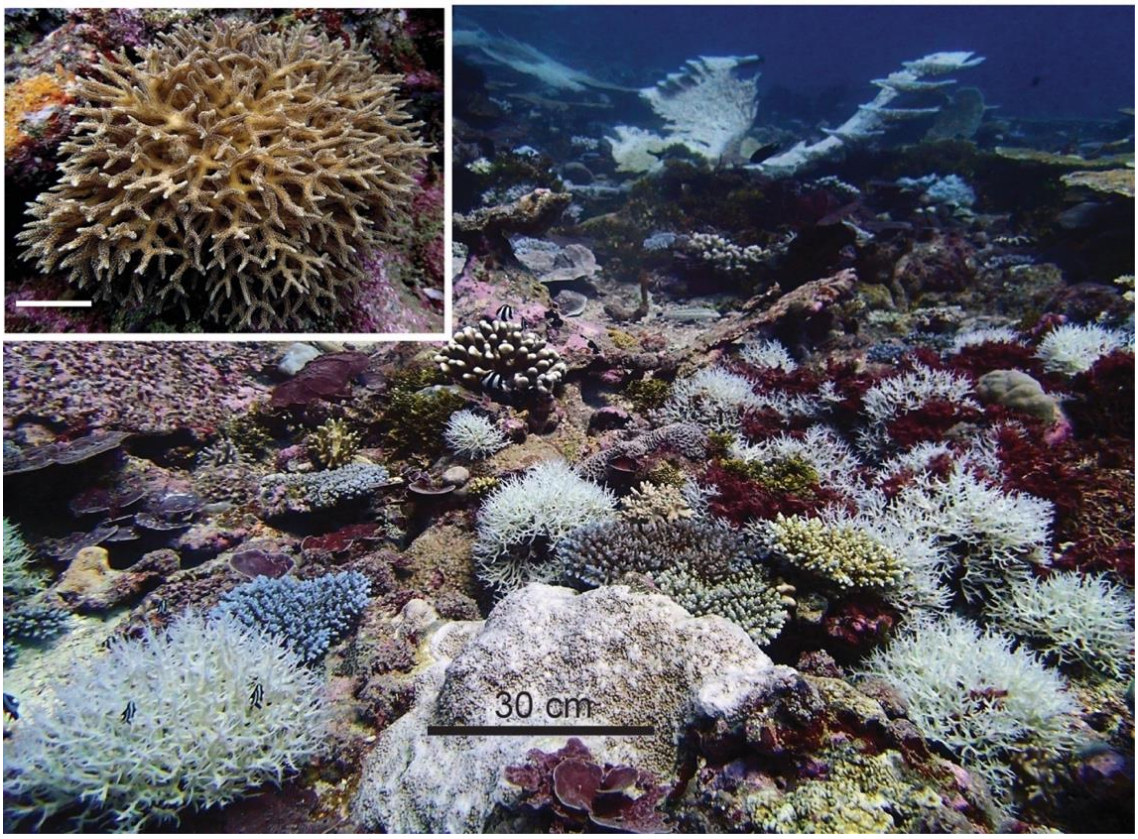

Fig. 2. A healthy *Seriatopora hystrix* colony (inset) and an actively bleaching, *Seriatopora* spp.-dominated coral reef in the British Indian Ocean Territory (i.e., Chagos; May 2016). The white scale bar in the inset is approximately 5 cm in length. Photo credits=珊

implications for cellular biology, especially given the well-established absence of congruency between gene expression and concentration of the respective protein in reef corals or Symbiodiniaceae (Mayfield et al., 2016c). In contrast, if a gene is absent from replicates of the stable TT, yet documented at high levels in those of the variable TT, it is far more likely to play a role in acclimation to variable temperatures than the 2-fold-DEG.

Of the ~93,000 contigs, I imported a random subset of 65,000 into PRIMER (ver. 6) and created a EDM to assess inter-sample similarity. Sub-selection was carried out because PRIMER currently limits analyses to 65,000 analytes (i.e., columns). After generating a multi-dimensional scaling (MDS) plot (Fig. 3A) and a PCA biplot (Fig. 3B), I then ran a 2-way PERMANOVA of SO, TT, and their

interaction. No random effects were considered. It is clear from Tab. 1 that all three simple random samples (SRS) yielded the same overall finding: strong SO effect on the transcriptome (both coral host & Symbiodiniaceae gene mRNAs were pooled in all analyses discussed herein.) yet no TT or interaction (SO x TT) effects (p>0.01). Given these findings, one might question our treatise on TT-sensitive DCPs (Mayfield et al., 2016c); although several genes were expressed only by samples of one TT and in no replicates of the other, the multivariate findings point to

Tab. 1. Permutational multivariate analysis of variance (PERMANOVA) of a partial *Seriatopora hystrix*-Symbiodiniaceae transcriptome (Mayfield et al., 2016c). A type III sum of squares (SS) model (partial) with unrestricted permutation of raw data was run thrice (simple random samples [SRS]1-3). Significant multivariate site of origin (SO; upwelling site [Houbihu] vs. non-upwelling site [Houwan]) effects were documented in all three iterations, and the mean SO effect p-value was 0.001. NA=not applicable. TT=temperature treatment (stable vs. variable).

Source	**df**	**SS**	**Mean square**	**Pseudo-*F***	***p***	**#Permu-tations**
SRS1						
SO	1	1779	1779	1.930	**0.001**	992
TT	1	917.4	917.4	0.9951	0.444	989
SO x TT	1	853.9	853.9	0.9263	0.668	985
Residual	8	7375	921.9			
Total	11	10930				
SRS2						
SO	1	1865	1865	2.013	**0.001**	991
TT	1	928.3	928.3	1.002	0.397	988
SO x TT	1	860.4	860.4	0.9286	0.640	985
Residual	8	7412	926.5			
Total	11	11070				
SRS3						
SO	1	682.4	682.4	2.778	**0.001**	985
TT	1	206.9	206.9	0.8420	0.587	985
SO x TT	1	210.65	210.6	0.8570	0.594	982
Residual	8	1965	245.7			
Total	11	3065				
Average of three SRS						
SO	1	1442	1442	2.240	**0.001**	NA
TT	1	684	684	0.9457	0.476	NA
SO x TT	1	642	642	0.9040	0.634	NA
Residual	8	5584	698			
Total	11	8355				

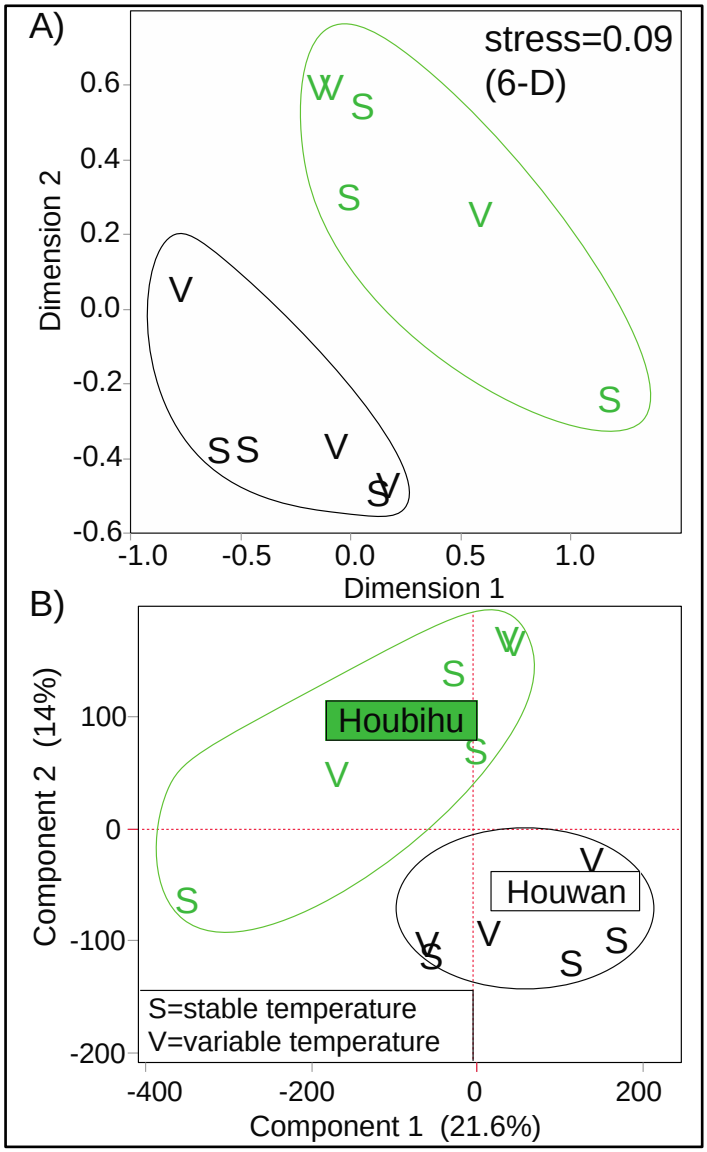

Fig. 3. MSA. Multi-dimensional scaling plot depicting similarity (Euclidean distance) among the 12 samples of the *Seriatopora hystrix* variable temperature study (A) and a principal components analysis scoring plot (2-D) showing the same data (B). In both cases, data were standardized prior to ordination. Biplot rays have not been shown in B since they would obscure the 12 data points.

a more compelling SO effect, in which case we should have prioritized the SO findings in the manuscript. We were consequently guilty of data (*p*-value)-mining in a statistically unsupported fashion.

Non-parametric MANOVA (NPMANOVA). When one is confident that the distance matrix robustly represents similarity among samples (i.e., low MDS "stress" in the predicted vs. actual similarity [Shepard] analysis), permutation may be superfluous; the mean distances within and among treatments in a EDM or BCSM may be sufficient for testing treatment effects. Although only two or three dimensions are normally portrayed in figs. (e.g., two dimensions in Fig. 3), increasing dimensionality leads to a better fit between predicted and actual similarity (higher R^2 in the Shepard diagram, or lower stress; Tab. 2). Instead, the number of dimensions will likely be dictated by the final NPMANOVA model. In the example herein, there were 12 coral samples (3 in each of 4 interaction groups). However, since SO and TT have only 1 df each, the model's residual term is 8. This means that the maximum number of dimensions that can be included as Y terms in the NPMANOVA is 8. For this reason, I have computed the stress, the MANOVA Wilks' lambda for the entire model, and the MANOVA SO effect *p*-values for dimensions ranging from 2 (the minimum) to 8 (the maximum) in Tab. 2. It can be seen that, although the predicted vs. actual similarity R^2 (i.e., Shepard's R^2) increased with increasing dimensionality (as expected since more complex models inherently fit a higher percentage of variation within the dataset), the ability to detect an SO effect was inversely correlated; statistically significant SO effects were more likely to be documented with only two dimensions. In this example, the optimal dimensionality was six since this ordination is associated with an acceptable stress (<0.2) and a highly significant SO *p*-value. Please note that, although the MDS plot in Fig. 3A was computed across six dimensions, only the first two have been presented.

A PCA of the same dataset will give a near-identical solution, and the ordination will actually be 100% identical if the data are standardized prior to MDS (as was conducted herein; Fig. 3B). Although PCA yields eigenvectors and eigenvalues that can be used to estimate treatment effects and the relative influence of various analytes, it cannot, in and of itself, statistically demarcate treatment effects (despite their being obvious in the case of the SO effect shown in Fig. 3B). Therefore, I recommend that 'OMICs researchers strongly consider MDS in place of (or, at least, in addition to) the far more commonly utilized PCA in presenting multivariate 'OMICs data simply because of the direct link between the coordinates in the MDS plot and the corresponding PERMANOVA and/or NPMANOVA. In contrast, PCA may better demonstrate the

aforementioned high collinearity among analytes (not evident in Fig. 3B since biplot rays were omitted), which would actually further invalidate the use of PSA (in addition to the #analytes>#samples issue raised above). Principal coordinates ordination (PCoA) could represent a compromise since it depicts inter-sample similarity *and* features biplot rays showing relationships among response variables and experimental predictors (alongside the percent variation explained by each axis). On the topic of inter-analyte collinearity, it is worth noting here that, for those interested in searching for DCAs in 'OMICs datasets, generalized regression (gen-reg) is superior to (& more robust than) FDR-governed RSA since it penalizes against collinear model terms; although the total number of DCAs may be less than generated by RSA, the collective DCA suite will be more informing (& less redundant) in model-building. Analyte clustering could also be considered.

Partial least squares (PLS). PERMANOVA and NPMANOVA are not the only means of depicting and analyzing multivariate, highly collinear, "wide" datasets in which there are many more analytes than samples (>92,000 vs. 12 herein); PLS is another such approach that is rarely taken advantage of by coral

Tab. 2. Non-parametric multivariate analysis of variance (NPMANOVA) of a standardized coral transcriptome. A Euclidean distance matrix was used in the multi-dimensional scaling analysis with 2 to 8 dimensions. Please note that NPMANOVA cannot be undertaken with >8 dimensions given the model's degrees of freedom (df)- site of origin (SO; df=1) x temperature treatment (TT; df=1)- and replication (n=3/SO x TT=12). Statistically significant *p*-values (<0.01) have been highlighted in **bold**, as have typically publishable stress levels (<0.1). No TT or interaction effects were documented for any model. Based on these results, a similarity analysis with six dimensions represents the best compromise in terms of limiting ordination stress while still enabling sufficient statistical power to detect a significant SO effect (see Fig. 3.).

#Dimensions	Shepard's R^2	Stress	Wilks' lambda (whole model) x 10^3	*p*-value (SO)
8	0.7964	**0.0605**	0.2	0.1646
7	0.7357	**0.0746**	0.4	0.0218
6	0.6701	**0.0904**	2.7	**0.0040**
5	0.6374	0.1141	3.2	**0.0004**
4	0.6133	0.1452	4.4	**<0.0001**
3	0.5831	0.1973	15.2	**<0.0001**
2	0.5847	0.2913	56.9	**<0.0001**

biologists. In PLS, which bears similarity to PCA in its data reduction approach, latent variables ("projections to latent structures" being the original name) are established to simultaneously model covariance between predictor (X) and response (Y) matrices (Cox & Gaudard, 2013). In addition to the usual model output terms (e.g., root mean square error), "variable importance" plots (VIP) are generated such that users can best determine which suites of analytes are most influential in delineating treatment effects, as well as which treatment factors drive the most variation across response variables. PLS is widely used in marketing and has been more recently adopted in metabolomic research, likely due to its accommodation of highly correlated response (& even predictor) variables. It also shines in response variable reduction and so may be an alternative to gen-reg in robust DCA identification.

Herein several PLS models were built. In the first, the six dimensions from the MDS analysis (Fig. 3) were the Y terms (response variables), and the two treatment factors (SO & TT) were the predictors (X). The resulting non-iterative PLS (NIPALS) algorithm (KFold cross-validation of seven) determined that the root mean predicted residual sum of squares (PRESS) of 0.42 was minimized with a single latent factor. Although the model explained only 21% of the variation in Y, it did clearly corroborate the aforementioned finding that the transcriptome was more strongly impacted by SO than TT (Tab. 3). It should be noted that the full PLS featuring 92,960 genes (standardized data) as Y and the SO, TT, and SO x TT terms as X cannot be analyzed on a computer with 8 GB of RAM and a 1.7-Ghz Intel Core i7 processor, though it is likely to yield a similar finding: the SO effect on the transcriptome >TT effect. When using the SO, TT, and their interaction (three separate models) as Y and the 92,960 genes as predictors, a single latent factor in each accommodated 47, 46, and 26% of the variation, respectively, and 43,903 (PRESS=0.64), 41,203 (PRESS=0.76), and 62,763 (PRESS=0.49) genes were characterized by VIP>0.8. Since nearly (or over in the case of SO x TT) 50% of the genes were included in the final models, this suggests that PLS may not be the preferred means of reducing dataset complexity (or identifying the most informative DCAs) in this particular example.

When using the 6D coordinates as Y and SO alone as X, a PLS-based discriminant analysis (DA) model could correctly predict the SO for 100% of samples. When using all 92,960 genes as the Y in the PLS-DA, the misclassification rate was far higher: 5/12 (42%) called

incorrectly with respect to site. For TT effects, 6D and 92,960-gene PLS-DA model misclassification rates were 2/12 (17%) and 6/12 (50%), respectively. In other words, the models had a greater chance of correctly guessing the corals' SO (mean success rate of prediction=79% correct) than their experimental TT (67% correct); this is unsurprising given the multivariate SO>TT effect trend yielded/suggested by PCA, MDS, PERMANOVA, & NPMANOVA. Finally, the PLS-DA classification rates for the SO x TT interaction group using the 6D and 92,960-gene data as Ys were the same as for SO: 100 and 58%, respectively (mean=79% correctly predicted).

Data dissemination

Good data stewardship transcends simply identifying the best statistical approach for addressing the research question at hand; it also hinges upon rapid and transparent data dissemination. Given the ever-constant demand on most scientists to "publish or perish," data sharing typically occurs well after article publication, if at all. Although this indeed limits data theft, in many cases such a significant amount of time passes between data collection and data dissemination to the public that the coral(s) from which such data were derived may no longer exist. For instance, a coral may be sampled in May, used in an experiment in June, with

Tab. 3. Partial least squares variable importance plot (VIP) data generated upon using either the multi-dimensional scaling coordinates from six dimensions ("6D") or the concentrations of the 637 unique differentially expressed genes ("DEG") as the Y and the two treatment effects (& their interaction) as X. Model terms whose VIP were >0.8 (the commonly accepted threshold for model inclusion; Cox & Gaudard, 2013) have been denoted by asterisks (*). Of the 762 DEGs whose FDRlogworths were >2, 558, 6, and 198 were significantly affected by site of origin (73.2%), temperature treatment (0.8%), and their interaction (26.0%), respectively (with 125 having been affected by more than one experiment parameter, i.e., 637 unique DEGs).

X	VIP-6D	VIP-DEG
site of origin[Houbihu]	1.4132*	1.4680*
site of origin[Houwan]	1.4132*	1.4680*
site of origin[Houbihu] x temperature treatment[stable]	0.7416	0.9074*
site of origin[Houbihu] x temperature treatment[variable]	0.6715	0.9734*
site of origin[Houwan] x temperature treatment[stable]	0.7636	0.9295*
site of origin[Houwan] x temperature treatment[variable]	0.6495	0.8602*

biopsies analyzed with 'OMICs technologies in December. Then, the data may be processed the next May, with the article written the next December. After 6-12 months of peer review and another 6-12 months of revisions, the data may finally reach the public 2-3 years later. Although the article certainly benefits the authors' careers, it does beg the question as to whether the data achieved what may, in many cases, have been the primary goal of the study: to aid in the ultimate conservation of that coral (& the reef framework in which it is embedded). Yes, the data may *later* be used to better manage the same reef, even if the study corals have died, but the point is that the data dissemination process is nearly always too slow to benefit the resident corals in a timely fashion (Mayfield & Chen, 2019), and proactive endeavors are certainly doomed to fail.

Slow data dissemination can be combatted on at least two fronts: rapid publishing of data online in open-access format *before* manuscript submission (Fig. 4) and expediting data acquisition itself. Regarding the latter, a typical 'OMICs pipeline was used as the example in the preceding paragraph; although extractions can now be done in bulk, there is nearly always a bioinformatics bottleneck. Even if a well-established pipeline is employed, making sense of the data to where they are both publishable and of conservation benefit may take months or even years with a team of dedicated researchers. For these reasons, there has been a push to develop *in situ* diagnostic systems for coral health (Mayfield & Chen, 2019). To use an example from work currently being conducted by scientists at NOAA's Atlantic Oceanographic and Meteorological Laboratory (AOML), if a proteomic analysis of environmentally challenged corals reveals that certain proteins show spikes in concentration in visibly healthy corals that later proceed to bleach, their concentrations could be measured in "test" field samples to determine which colonies and reefs are at most risk of bleaching as seawater temperatures rise (univariate biomarker approach). Antibodies raised against such protein biomarkers could be incorporated into microfluidic chips to where the data are obtained while still at sea, not months later like in the example above.

If a candidate biomarker-based diagnostic approach is considered overly simplistic or lacks predictive power with respect to health, the more time-consuming 'OMICs-scale alternative could be undertaken; Fig. 5 presents one example of how this could be achieved. As molecular biology and data analysis technologies continue to improve, it is no longer inconceivable that a coral could be

sampled on Monday, with the 'OMICs-scale data obtained by the end of the week. Given the concerted effort by the coral reef research field to elucidate the molecular biology of anthozoan-dinoflagellate endosymbioses (sensu Peng et al., 2011), we will soon know exactly which molecules are most informative with respect to resilience by having analyzed their concentrations in experiments in which common physiological benchmarks like growth and endosymbiont density (Fig.

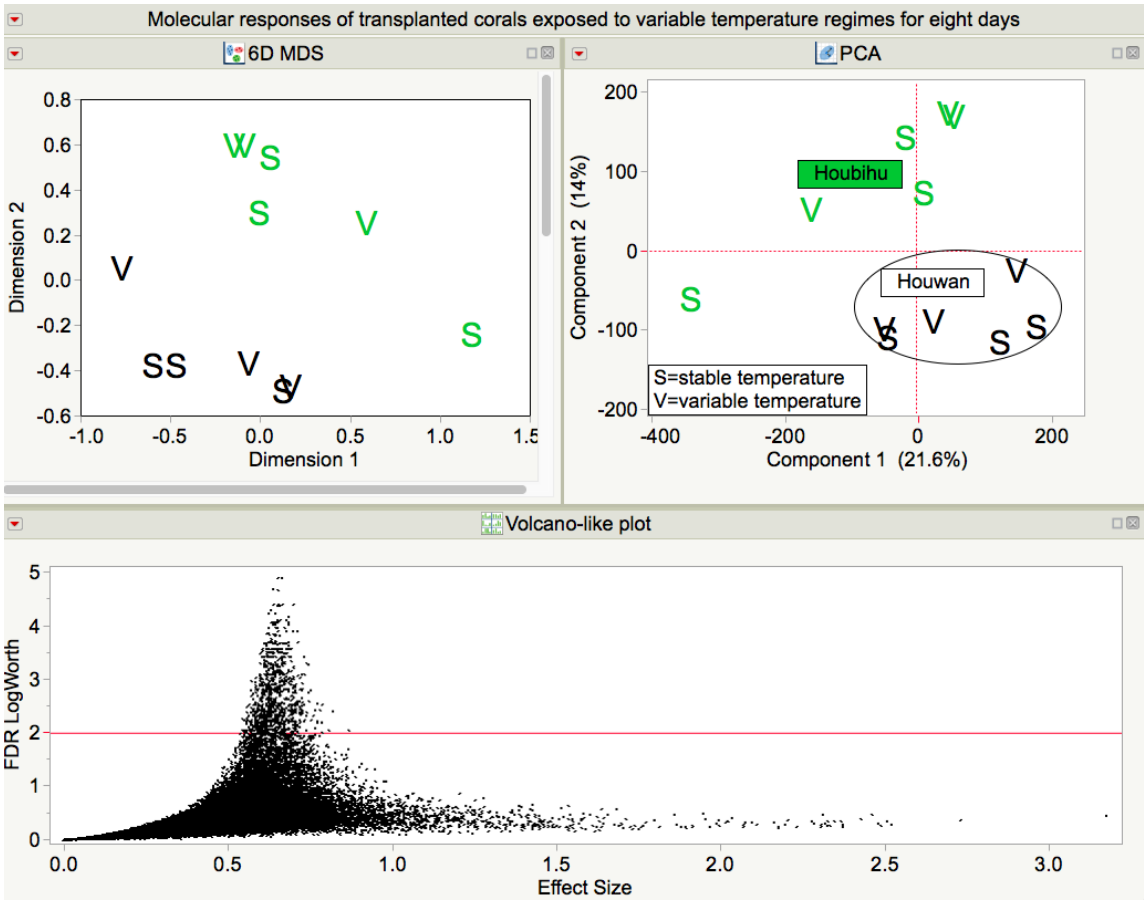

Fig. 4. JMP® Public. In addition to simply uploading data tables alongside manuscripts to promote data transparency, or posting data tables on public repositories (e.g., dryad, Figshare, & Zenodo), data can be shared interactively by websites such as JMP® Public. In this example, the MDS and PCA plots of Fig. 3 are shown above an effect size vs. FDRLogworth (inverse $\log_{10}$ of p) volcano-like plot (logworth of 2=p of 0.01). Readers can download the data table and plots to their devices, or, alternatively, directly manipulate them on the website without the need for third party software. With this approach, interested readers can re-analyze data and attempt to reproduce findings from published manuscripts. Green icons=Houbihu. Black icons=Houwan.

6) are also measured (sensu Mayfield et al., in review). Although the health of the coral would not be known while still on the dive boat with this approach, the diagnostic speed might still be amenable to some sort of proactive conservation measure, even if it simply amounts to "triage," whereby colonies (or, more likely, reefs) are ranked on a scale of 1 (highly likely to bleach) to 10 (the most resilient corals studies to date).

Efforts are currently being made to compare these "bottom-up" (molecular biology-derived) coral health predictions against "top-down" (global-scale) analyses of satellite temperature data (van Hooidonk et al., 2020), which have projected most reefs to begin bleaching annually by 2040. These analyses will allow us to identify "refugia," reefs that, either due to environmental forcing or the unique physiologies of the resident corals (Brown et al., 2000), resist widespread bleaching. We had previously prioritized upwelling reefs such as those discussed herein since variable environmental conditions may "stress-harden" corals (Safaie et al., 2018; Storlazzi et al., 2020, but see Chollett et al., 2010.). Unfortunately, Houbihu experienced catastrophic bleaching in 2020, leading some to wonder whether *ex situ* husbandry (Lin et al., 2019; Mayfield et al., 2019b; Chang et al., 2020; Huang et al., 2020) or even cryopreservation (Lin & Tsai, this issue) represents the only means by which corals and their genetic material will persist beyond 2100. These severe threats also signify that poor data and statistical stewardship not only thwart progress in the field, but they could actually detrimentally affect the very coral reef ecosystems the researchers intend to conserve; statistical mistakes resulting in false positive results could, for instance, lead to management intervention at reefs that should not have been prioritized (i.e., resilient ones) at the expense of more stress-prone reefs (or vice versa, i.e., inadvertently prioritizing reefs with so little ability to withstand climate change and other anthropogenic stressors as to be beyond saving). By employing more statistically rigorous MSA prior to the more commonly utilized univariate approaches featured in most coral biology studies, not only will the dataset have been realized to its full potential, but greater insight into the health of the experimental or field corals will have been gained. Furthermore, MSA will allow us to make predictions as to which corals may have the physiological capacity to acclimatize to GCC and other environmental challenges, and which may only avoid extinction via cryopreservation and other human-assisted preservation efforts (sensu Lin & Tsai, this issue).

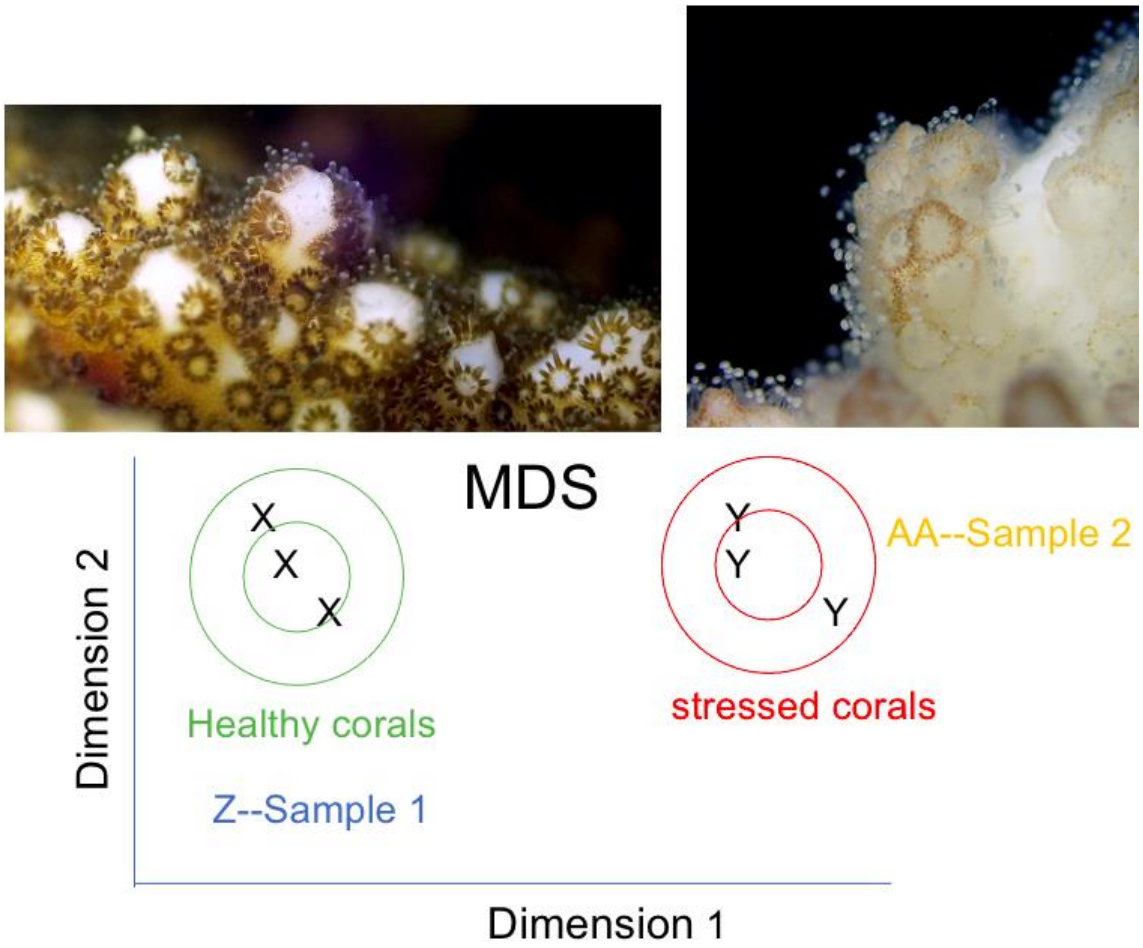

Fig. 5. Similarity analysis of coral health. In this example, molecular signatures are defined with multi-dimensional scaling (MDS) in at least two dimensions for healthy (i.e., resilient) corals (left image & "X's" in MDS plot) and stress-susceptible (i.e., bleaching-prone) corals (right image & "Y's" in MDS plot). "Test" samples "Z-Sample 1" (blue) and "AA-Sample 2" (orange) are then collected from the field, analyzed, and, based on their positions on the MDS plot (i.e., proximity to experimental coral data points), hypothesized to represent healthy and stressed corals, respectively. Each polyp is approximately 1 mm in diameter.

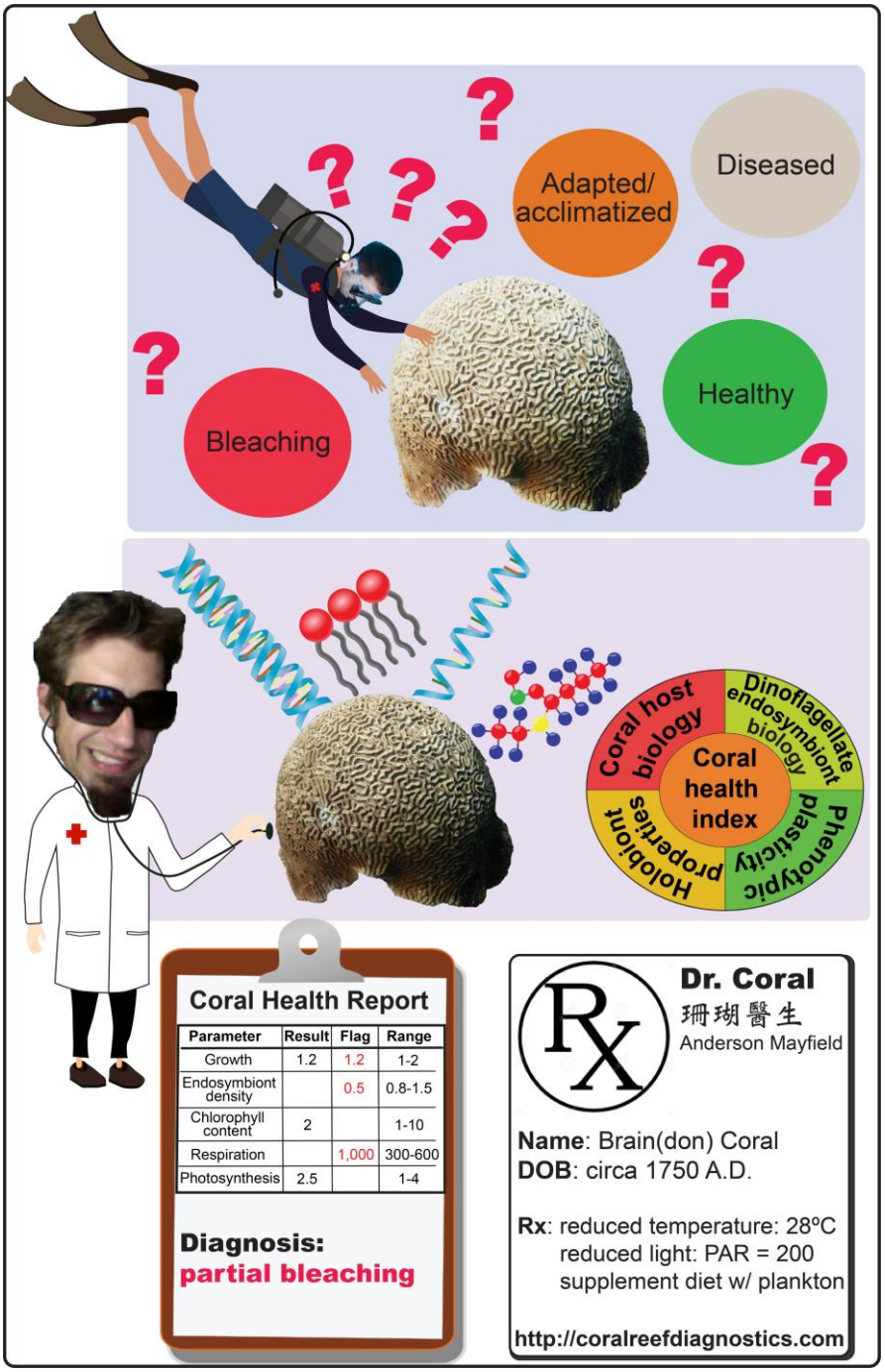

Fig. 6. Coral health diagnostics. Molecular techniques alongside traditional physiological benchmarks (e.g., growth) can be used to make conjectures about current coral health as well as future resilience to environmental change. Please see Mayfield & Chen (same issue) for details on undertaking predictive modeling with 'OMICs data. In this example, "Braindon" has demonstrated aberrant respiration with anomalously low endosymbiont densities; given these ailments, Dr. Coral (珊瑚醫生) recommends that he be cooled off and consume large quantities of plankton under photosynthetically active radiation levels near 200 µmol m^{-2} s^{-1}.

Acknowledgements

Thanks are given to the Fulbright and MacArthur Foundations for funding this work. I would also like to thank Mr. John Powell (JMP) for assistance with interactive HTML. All opinions are my own and do not necessarily represent those of the United States government.

References

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

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

Apprill, A.M. & R.D. Gates. 2007. Recognizing diversity in coral symbiotic dinoflagellate communities. Molecular Ecology, 16:1127-1134.

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.

Brown, B.E., R.P. Dunne, M.S. Goodson & A.E. Douglas. 2000. Bleaching patterns in reef corals. Nature, 404:142.

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, H.K., A.B. Mayfield, L.H. Wang & C.S. Chen. 2017a. Coral lipid bodies as the relay center interconnecting diel-dependent lipidomic changes in different cellular compartments. Scientific Reports, 7:3244.

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.

Chen, T.Y., G.W. Hwang, A.B. Mayfield, C.P. Chen & H.J. Lin. 2017b. The relationship between intertidal soil composition and fiddler crab burrow depth. Environmental Engineering, 100:256-260.

Cheng, Y.R., P.J. Meng, A.B. Mayfield & C.F. Dai. 2016. Copepods associated with scleractinian corals: a worldwide checklist and a case study of their impact on the reef-building coral *Pocillopora damicornis* (Linnaeus, 1758) (Pocilloporidae). Zootaxa, 4174(1):291-345.

Chollett, I., P.J. Mumby & J. Cortés. 2010. Upwelling areas do not guarantee refuge for coral reefs in a warming ocean. Marine Ecology Progress Series, 416:47-56.

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.

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

Desalvo, M.K, C.R. Voolstra, S. Sunagawa, J.A. Schwarz, J.H. Stillman, M.A. Coffroth, A.M. Szmant & M. Medina. 2008. Differential gene expression during thermal stress and bleaching in the Caribbean coral *Montastraea* faveolata. Molecular Ecology, 17 (17):3952-3971.

Enochs, I.C., N. Formel, D.P. Manzello, J. Morris, A.B. Mayfield, A. Boyd, et al. 2020. Coral persistence despite extreme periodic pH fluctuations at a volcanically acidified Caribbean reef. Coral Reefs, 39:523-528.

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.

LaJeunesse, T.C. 2001. Investigating the biodiversity, ecology, and phylogeny of endosymbiotic dinoflagellates in the genus *Symbiodinium* using the ITS region: In search of a 'species' level marker. Journal of Phycology, 37:866-880.

LaJeunesse, T.C., R.T. Smith, J. Finney & H. Oxenford. 2009. Outbreak and persistence of opportunistic symbiotic dinoflagellates during the 2005 Caribbean mass coral 'bleaching' event. Proceedings of the Royal Society B-Biological Sciences, 276: 4139-4148.

Lee, H.J., S.Y. Chao, K.L. Fan, Y.H. Wang & N.K. Liang. 1997. Tidally induced upwelling in a semi-enclosed basin: Nan Wan Bay. Journal of Oceanography, 53:467-480.

Lin, C. & S. Tsai. 2020. Fifteen years of coral cryopreservation. Platax, this issue.

Lin, C., S. Tsai & A.B. Mayfield. 2019. Physiological differences between cultured and wild coral eggs. Biopreservation and Biobanking, 17:370-371.

Liu, P.J., K.N. Chung, L.L. Liu, W.H. Twan, J.T. Wang & M.Y. Leu. 2012. Impacts of human activities on the coral reef ecosystems of southern Taiwan: a long-term study. Marine Pollution Bulletin, 64:1129-1135.

Mayfield, A.B. 2009. Osmoregulation in anthozoan-dinoflagellate endosymbiosis. Ph.D. dissertation, University of Hawaii, Manoa.

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. 2020. Proteomic signature of corals from thermodynamic reefs. Microorganisms. 8(8), 1171.

Mayfield, A.B. in review. Environmentally-mediated variation in New Caledonian reef corals. Atoll Research Bulletin.

Mayfield, A.B., C. Aguilar, I. Enochs & D.P. Manzello. in review. Shotgun proteomics of thermally challenged Caribbean reef corals. Coral Reefs.

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. 2019. Enabling coral reef triage via molecular biotechnology and artificial intelligence. Platax, 16:23-47.

Mayfield, A.B. & C.S. Chen. this issue. A coral transcriptome in the Anthropocene as an "alternative stable state." Platax

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. 2019a. Modeling environmentally-mediated variation in reef coral physiology. Journal of Sea Research, 145:44-54.

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 &

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., A.C. Dempsey & C.S. Chen. in prep. Biomarker profiling in pocilloporid corals Of the Solomon Islands.

Mayfield, A.B., A.C. Dempsey, J. Inamdar & C.S. Chen. 2018c. 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., 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., 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. 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. 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. 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., 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., S. Tsai & C. Lin C. 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.

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.

McRae, C., A.B. Mayfield, T.Y. Fan & I. Cote. in review. Differing proteomic responses to high-temperature exposure between adult and larval corals. Marine Ecology Progress Series.

Meron, D., K. Maor-Landaw, E. Weizman, H. Waldman Ben-Asher, G. Eyal, E. Banin, Y. Loya & O. Levy, O. 2019. The algal symbiont modifies the transcriptome of the scleractinian coral *Euphyllia paradivisa* during heat stress. Microorganisms, 7:256.

Monteiro, H.J.A., C. Brahmi, B. Lapeyre, A.B. Mayfield & J. Le Luyer. 2020. Molecular mechanisms of acclimation to long-term elevated temperature exposure in marine symbioses. Global Change Biology, 26:1271-1284.

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.

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.

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.

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.

Randall, C.J., L.T. Toth, J.J. Leichter, J.L. Maté & R.B. Aronson. 2020. Upwelling buffers climate change impacts on coral reefs of the eastern tropical Pacific. Ecology, 101(2):e02918.

Ribas-Deulofeu, L., V. Denis, S. De Palmas, C.Y. Kuo, H.J. Hsieh & C.A. Chen. 2016. Structure of benthic communities along the Taiwan latitudinal gradient. PLoS ONE, e0160601.

Rosales, S.M., C. Sinigalliano, M. Gidley, P.R. Jones & L.J. Gramer. 2019. Oceanographic habitat and the coral microbiomes of urban-impacted reefs. PeerJ, 7:e7552.

Safaie, A., N.J. Silbiger, T.R. McClanahan, G. Pawlak, D.J. Barshis, J.L. Hench, J.S. Rogers, G.J. Williams & K.A. Davis. 2018. High frequency temperature variability reduces the risk of coral bleaching. Nature Communications, 9:1-12.

Storlazzi, C.D., O.M. Cheriton, R. van Hooidonk, R. Z. Zhao, R. Brainerd. 2020. Internal tides can provide thermal refugia that will buffer some coral reefs from future global warming. Scientific Reports, 10:13435.

van Hooidonk, R., J. Maynard, G. Grimsditch, G. Williams, J. Tamelander, J. Gove, H.

Koldewey, G. Ahmadia, D. Tracey, K. Hum, E. Conklin & M. Berumen. 2020. Projections of future coral bleaching conditions using IPCC CMIP6 models: climate policy implications, management applications, and Regional Seas summaries. United Nations Environment Programme, Nairobi, Kenya.

Wang, L.H., H.H. Lee, L.S. Fang, A.B. Mayfield & C.S. Chen. 2013. Normal fatty acid and phospholipid synthesis are prerequisites for the cell cycle of *Symbiodinium* and their endosymbiosis with sea anemones. PLoS ONE, e72486.

Xiang, T, W. Nelson, J. Rodriguez, D. Tolleter & A.R. Grossman. *Symbiodinium* transcriptome and global responses of cells to immediate changes in light intensity when grown under autotrophic or mixotrophic conditions. The Plant Journal, 82:67-80.