Enabling coral reef triage through molecular biotechnology and artificial intelligence

Anderson B. Mayfield[1-3]*, Chii-Shiarng Chen[1, 4-5]

[1]National Museum of Marine Biology and Aquarium, Checheng, Pingtung 944, Taiwan

[2]Cooperative Institutes for Marine and Atmospheric Studies, University of Miami, Miami, FL 33149, USA

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

[4]Graduate Institute of Marine Biotechnology, National Dong-Hwa University, Checheng, Pingtung 944, Taiwan

[5]Department of Marine Biotechnology and Resources, National Sun Yat-Sen University, Kaohsiung 80424, Taiwan

*corresponding author. Email: anderson.mayfield@noaa.gov

Abstract

Coral reef ecosystems are threatened by an onslaught of anthropogenic stressors, most notably global climate change (GCC); indeed, no regions have been spared from our wide-ranging human impact. Consequently, there have been urgent pushes to 1) model how marine organisms will respond to changes in their environments and 2) make data-driven predictions as to which populations are most stress sensitive. Given our recently elevated understanding of how GCC affects reef corals, we are now in a position in which it may be possible to make projections as to which corals are most susceptible to GCC, as well as which will demonstrate resilience. Herein we explore the potential for artificial intelligence-based approaches to generate models that can accurately predict coral stress susceptibility (CSS). Specifically, we advocate that coral reef-focused partial least squares and neural networking algorithms should be developed, with their prognostic capability then field-tested at sites spanning a gradient of human impact and ecological resilience in the high-biodiversity "Coral Triangle." If the developed actuarial models are characterized by the analytical capacity to forecast CSS, we will possess one means of identifying reefs that should be prioritized for conservation (*i.e.,* coral reef "triage").

Key words: artificial intelligence, coral reefs, global climate change, molecular diagnostics, stress biology

Introduction

Coral reefs harbor immense biodiversity and provide a wealth of benefits to humankind (*e.g.,* as nurseries for numerous commercial fish species). Unfortunately, these ecosystems are threatened by myriad anthropogenic stressors (Mayfield and Gates, 2007), from global-scale impacts like climate change (Hoegh-Guldberg et al., 2007) to local ones like seawater pollution (Huang et al., 2011). The elevated temperatures associated with global climate change (GCC) are especially concerning since most corals live near the upper threshold of their thermotolerance (Brown, 1997). Even increases in temperature of only 1ºC above the summer mean can cause a collapse of the mutualistic relationship between reef-building scleractinians and the photosynthetically active dinoflagellates (family Symbiodinaceae) that inhabit their gastroderms; this phenomenon is known as "bleaching" due to the paling of the coral tissues (Gates, 1990). Since corals rely on the energy obtained from Symbiodinaceae-fixed carbon to not only meet their metabolic needs, but also to accrete the calcium carbonate skeletons that serve as the structural foundation of coral reefs, bleaching can lead to both coral death and reef degradation.

Upon having carried out a plethora of controlled laboratory studies with several Indo-Pacific coral species, including *Seriatopora hystrix* (*e.g.*, Mayfield et al., 2011) and *Pocillopora acuta/damicornis* (Table 1), we have developed an understanding of the responses of scleractinians to GCC scenarios (Mayfield et al. 2013a-b, 2019b). In fact, our knowledge has finally advanced to the point where it would be fruitful to attempt to use the explanatory data acquired during such tank studies to predict coral behavior *in situ*; if we could use data from laboratory exposures (*sensu* Table 1) and/or published field datasets on coral physiology (*e.g.*, Mayfield et al., 2015, 2016a) to make predictions as to how conspecifics would respond to environmental heterogeneity in other locations (or at the same site at later dates), then we would likewise possess the capacity to determine which reefs (and/or coral populations) are most likely to persist in the face of GCC.

Although it is true that many corals bleach when exposed to elevated temperatures over prolonged durations, others have demonstrated a marked capacity for resilience (Barshis et al., 2013; Krueger et al., 2017), and numerous investigators around the globe are currently attempting to elucidate the genetic basis of such thermotolerance. Indeed, we do not yet even have a grasp

of the cellular pathways underlying coral bleaching (Jones et al., 1998). In an ideal world, we would carefully elucidate the molecular underpinnings of bleaching resilience, or lack thereof, in replicated coral populations from diverse locations across the globe. Alongside such studies, we would model the molecular mechanisms underlying coral bleaching at the single-cell level, and such experiments would be repeated with corals sampled from reef sites spanning large-scale latitudinal gradients that are either 1) bleaching-prone or 2) bleaching resistant. Several years may be required to acquire such data, and, unfortunately, time is not on our side given the rate at which seawater temperatures are rising. We therefore propose herein that we not only continue to remedy these deficiencies in our knowledge of coral bleaching and the coral stress response, but also attempt to use the data *already in hand* to make predictions about future coral physiological behavior. We present one such approach for doing so in the following paragraph, with details found further on in the article.

Upon analyzing the "proteomes" (population of all synthesized proteins) of corals that resisted bleaching (Table 1-experiment [exp.]#1), as well as those that instead succumbed to high-temperature stress and bleached

(Table 1-exp.#2), we would possess the capacity to develop both "bottom-up" (molecules=>physiology) and "top-down" (environment=>physiology) predictive models for gauging future coral performance *in situ* using data from not only the aforementioned GCC manipulation studies (Table 1) carried out at the National Museum of Marine Biology and Aquarium's (NMMBA, Taiwan) state-of-the-art coral reef mesocosm facility (Liu et al., 2009), but also from the most wide-ranging coral reef survey ever undertaken: the Khaled bin Sultan Living Oceans Foundation's "Global Reef Expedition" (GRE; see Mayfield et al., 2017a-c for details.). Upon incorporating all environmental and coral molecular-physiological data acquired during such studies into software packages like JMP® Pro (Cary, NC, USA; www.jmp.com), artificial intelligence (AI)-based partial least squares (PLS) and neural networking (NN) models could be developed, with the predictive capacity of the resulting bottom-up and top-down algorithms field tested at well-studied reef sites in the high-biodiversity "Coral Triangle" (not limited to those four we have regularly studied+surveyed in Southern Taiwan that differ dramatically in 1) oceanography and 2) the environmental resilience of the resident coral communities; Fig. 1a-b). If AI can

be used to identify coral colonies, or coral reefs in the case of the top-down model, that are markedly stress sensitive prior to visible, late-stage manifestations of coral health decline, then managers could be alerted such that they could mitigate local-scale stressors (*e.g.*, overfishing) in order to promote coral resilience and potentially thwart bleaching. This would further enable "coral reef triage," or else a system by which reefs could be prioritized for conservation efforts. We now detail all ideas/steps addressed in this paragraph.

Fig. 1. Map of Taiwanese field sites. Taiwan's location at the northern-most region of the "Coral Triangle" (a) has been expanded in (b) to show the study area in more detail. Nanwan Bay, where the "Outlet" and Houbihu are located, is characterized by summer upwelling of cold, deep-ocean seawater (Mayfield et al., 2012a); corals there consequently have special adaptations for accommodating not only elevated (see description for Outlet elsewhere.), but also highly variable, temperature regimes (22-29ºC over a day; see temperature profile insets.). In contrast, the two Taiwan Strait sites are characterized by a more typical coral reef temperature profile (22-29ºC over a calendar year), nor are these corals routinely exposed to dramatically elevated seawater temperatures; they may, then, be more susceptible to environmental change.
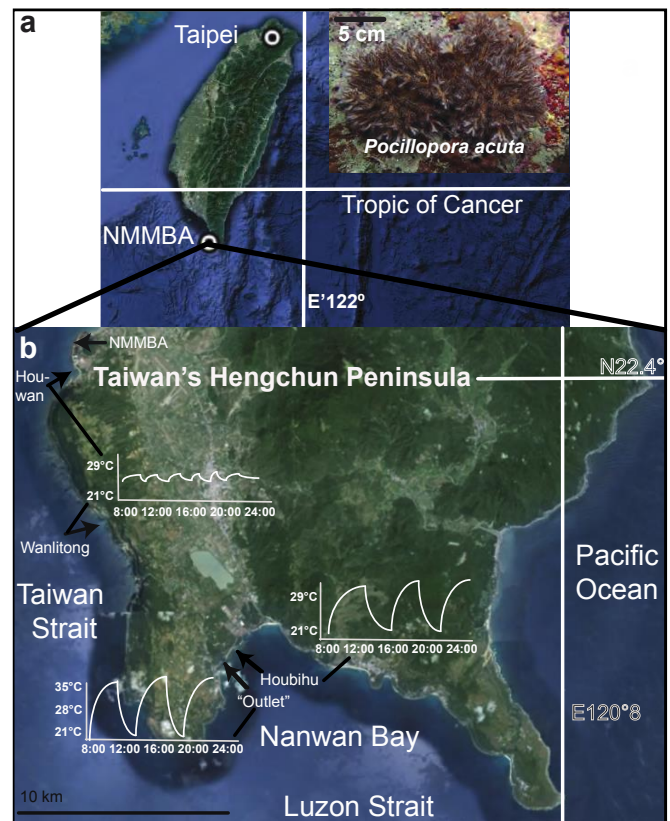
Goals and objectives

To gauge the coral response to environmental change using AI, two different models will be built using a NN/machine learning approach in conjunction with PLS. In the first model (Aim I-referred to as "bottom-up" since it will utilize molecular information ["bottom"] to make physiological ["up"] conjectures), the predictors will be biomarker proteins uncovered from proteomic analysis of experimentally stressed aquarium samples (Table 1-exps.#1-2; described in detail below). The coral response will be binomial: stressed (*i.e.*, bleaching-prone) vs. unstressed (bleaching-resistant). PLS and NN analyses are particularly well suited for these sorts of "omics" analyses, in which the concentrations of thousands of proteins are quantified simultaneously, and both analytical techniques permit response variable reduction to where only those proteins whose concentrations are tightly linked with coral health decline will be incorporated into the final training models. PLS, in particular, is also adept at handling those datasets featuring a high degree of collinearity, such as all coral gene expression datasets characterized to date (*e.g.,* Mayfield et al., 2014d, 2016c). Other attributes of these modeling techniques are highlighted below.

We will also build PLS+NN models in an "ecosystem ('top') to physiology ('down')" direction (Aim II); instead of using molecular data to make inferences about future declines in coral performance (*sensu* Aim I), we will exploit reef coral and coral reef datasets to verify whether we can use aquarium simulation and environmental data, respectively, to predict which corals and reefs, respectively, will be most stress prone. The first data source will be our extensive series of GCC simulation studies carried out with *P. acuta* at NMMBA (Table 1). Secondly, data from the GRE (Fig. 2) will be incorporated; as *P. acuta* was sampled at hundreds of reefs across the Indo-Pacific as part of the GRE, we have the opportunity to identify the environmental parameters (EP) that are most influential in driving variation in coral physiology when combining these GRE data with those of the controlled GCC simulations carried out in Taiwan. These EP (*i.e.*, the models' predictors) include, but are not limited to, biotic parameters like coral cover, invertebrate diversity, and algal abundance, and abiotic parameters like temperature and salinity (see Fig. 3 and Mayfield et al., 2019a for details.). The population densities of the settlements nearest the GRE field sites will also be included in the models to determine whether reefs nearby human habitations are more stress-sensitive than

remote ones; preliminarily, the opposite appears to be true (healthy corals are more likely to be found in marginalized areas.). Although a variety of both physiological and molecular response variables were assessed in corals of these experiments and field surveys (Fig. 4), there will be a particular focus on those PLS+NN models that best predict, more simply, coral survival.

Global Reef Expedition- data summary
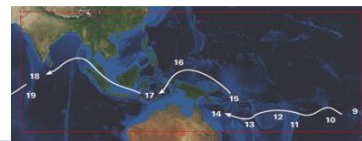
Country	Notes	Environmental data	Coral physiological data	Coral molecular data	Deliverables (to date)
Taiwan	Efforts ongoing	✓	✓	✓	Manuscript*s*[a-b]
French Polynesia		✓	✓	✓	Manuscript*s*[a]
Fiji		✓	✓	✓	Manuscript[a]
Tonga		✓	✓	✓	Manuscript[a]
New Caledonia		✓	✓	✓	Manuscript[a-b]
Australia (GBR)			No samples taken		Manuscript[a]
Solomon Islands		✓	✓	✓	Data not yet analyzed
Palau		✓	✓	✓	Data not yet analyzed
Chagos/BIOT		✓	✓	in prep.	Data not yet acquired
Maldives		✓	✓	in prep.	Field report only
Philippines	Grant under review				Preliminary data
Indonesia	Grant under review				Survey data

[a]Data analyzed (somewhat) sophomorically.   [b]Only a portion of the dataset was analyzed.

Fig. 2. A summary of data acquired during the Khaled bin Sultan Living Ocean Foundation's "Global Reef Expedition" (GRE), alongside an inset featuring a dissecting microscope image of the model reef-building coral *Pocillopora acuta.* Only the Indo-Pacific leg of the GRE has been depicted in the inset map. "Deliverables" (right-most column) correspond to coral molecular-physiological data only; for most countries/regions, ecological data and general field reports have already been published on www.livingoceansfoundation.org. Fulbright (USA) funds have been awarded to ABM to continue his coral diagnostics research in Southern Taiwan in 2020. BIOT=British Indian Ocean Territory. GBR=Great Barrier Reef.
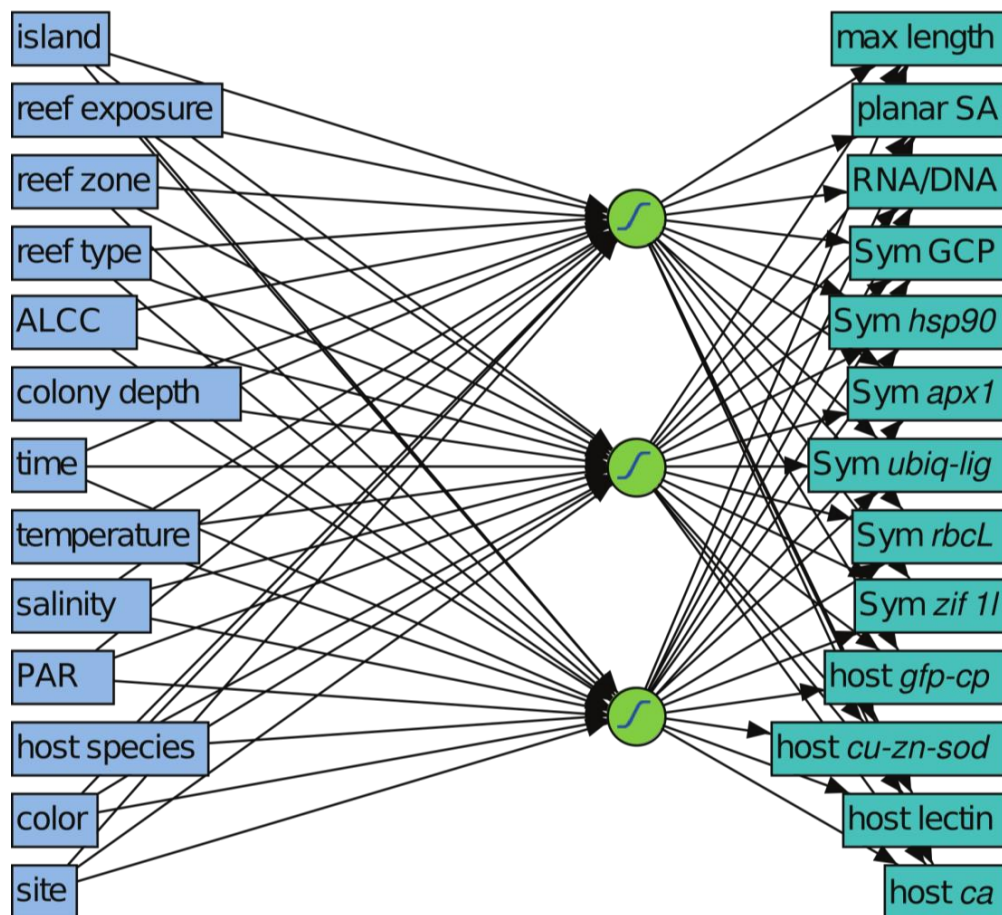
Fig. 3. An example of a neural network (NN) featuring coral reef environmental data (left side) and reef coral response variables (right side). In the case of this 3-node NN example, the predictive capacity of the model was <30% due to having trained it with a dataset featuring only 70 samples. "z-scores" refer to standardized data: (value-mean)/standard deviation. ALCC=average live coral cover. GCP=genome copy proportion. PAR=photosynthetically active radiation. Sym=Symbiodiniaceae dinoflagellates. Abbreviations for the gene mRNAs (*hsp90*, *apx1*, *rbcL*, *ubiq-lig*, *zifl1l*, *gfp-cp*, *cu-n-sod*, and *ca*) can be found in references cited throughout the manuscript.

Table 1. Global climate change manipulation studies carried out at Taiwan's National Museum of Marine Biology and Aquarium. In most cases, *Pocillopora acuta* was inadvertently classified as *Pocillopora damicornis* in the published manuscripts. It is worth noting that ocean acidification (*i.e.*, elevated carbon dioxide partial pressures [pCO_2]) did *not* adversely affect corals in any experiment (exp.). Only corals exposed to 31.5°C for several weeks (exp.#2) bleached. In the coming year, we propose to undertake proteomic examinations of samples highlighted in yellow (exps.#1-2 [n=12 & 18, respectively] and exps.#3-8 [n=10 for each exp; 60 samples in total]). Table abbreviations: NA=not applicable. *S. hystrix*=*Seriatopora hystrix*. Trans-gen=trans-generational (adults=>larvae).

Target species	Life history stage	Temperature treatment (°C)	High pCO_2 (ppm)	Salinity effects tested?	Light effects tested?	Nutrient effects tested?	Time-scale	Acclima-tion?	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, 2016b-c, 2018a, 2019a-b
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- exp. 1	adult	26.5 vs. 29.7	NA	no	no	no	months	yes	Mayfield et al., 2013b, 2014d, 2018b
P. acuta- exp. 2	adult	31.5-sustained	NA	no	no	no	weeks	no-bleached	Mayfield et al., 2013a, 2014a
P. acuta exp. 3	adult	26 vs. 29	415 vs. 850	no	no	no	weeks	yes	Putnam, Mayfield, et al., in prep.
P. acuta	adult	31.5-return to ambient at night	NA	no	no	no	weeks	yes	Mayfield et al., 2013a
P. acuta exp. 4	adult	25	400 vs. 1,000	no	no	no	months	yes	Liu, Mayfield, et al., in prep.
S. hystrix	adult	25	400 vs. 1,000	no	no	no	months	yes	Liu, Mayfield, et al., in prep.
P. acuta exp. 5	adult	25, 28, or 31	400 vs. 800	no	no	yes	months	yes	Liu et al., in press
P. acuta exp. 6	adult	26 vs. 29.5	NA	no	no	no	months	yes	McRae, Mayfield, et al., in prep.(a)
P. acuta exp. 7	adult	26 vs. 32	NA	no	no	no	hours	yes	McRae, Mayfield, et al., in prep.(b)
P. acuta exp. 8	trans-gen	26 vs. 30	NA	no	no	no	months	yes	McRae, Mayfield, et al., in prep.(c)

To determine the predictive capacity of the bottom-up and top-down models (Aim III), we will field-test them at four study sites in Southern Taiwan (Fig. 1b): Wanlitong and Houwan in the Taiwan Strait and Houbihu and "Outlet" in Nanwan Bay. The latter site is named such due to its location at the outlet where the thermal effluent from a nearby nuclear power plant enters the ocean; seawater temperatures there can reach 35ºC, yet the coral communities are thriving (see the caption for Fig. 1 for additional details.). We have been 1) sampling tagged coral colonies and 2) collecting environmental data from these sites for several years, and we will use protein biomarker concentration data and field environmental data, respectively, acquired over the course of 2020 to feed the bottom-up and top-down models, respectively. If the former and latter models can predict which coral colonies and reef sites, respectively, are most bleaching prone prior to future, high-temperature, field seasons, then we will have validated the capacity to use such computational approaches to prioritize reefs for targeted, proactive, and "triaged" management. Details of this AI+molecular biotechnological approach have been outlined below.

Methodological details

Aim I-bottom-up model training. SCIEX's "isobaric tags for relative and absolute quantification" (iTRAQ) reagents will be used to sequence and profile the proteomes of the following 30 *P. acuta* samples with a Q Exactive™ (Thermo-Fisher) mass spectrometer (MS): 1) 3 samples maintained at control temperature+3 sub-lethally bleached samples at each of three sampling times (1, 7, & 14 days; exp.#2 of Table 1; n=18) and 2) 3 samples maintained at control temperature+3 samples that resisted high-temperature bleaching (30°C) at each of two sampling times (2 and 36 weeks [9 months]; exp.#1 of Table 1; n=12). The protein concentration data from the sub-lethally stressed samples of exp.#2 will be compared to experimental controls, and the differentially concentrated proteins uncovered will be used to train the bottom-up PLS+NN model. Corals of exp.#1 instead acclimated to high temperatures, and so the resulting data will be used to better understand the molecular pathways underlying such long-term, high-temperature acclimation.

Aim II-top-down model training. Although the majority of samples from the GCC simulation studies of Table 1, which will be used, in part, to train the top-down PLS+NN model, have been analyzed for an array of physiological-

and molecular-scale response variables (Fig. 4), we will further profile the proteomes of *P. acuta* samples from several other experiments listed in Table 1 (see table caption for details; n=60 in total). Furthermore, we have processed the majority of the coral samples from the GRE (Fig. 2), whose data will be used alongside the GCC simulation data for the top-down model training; of note, we have not yet extracted macromolecules those coral samples from the remote Chagos Archipelago (also known as the "British Indian Ocean Territory"). As corals of this uninhabited region of the Indian Ocean are affected only by GCC (and no other anthropogenic stressors), these samples represent some of the most important coral samples ever collected and will be critical in the development of models seeking to elucidate environmental effects on corals. We will profile the proteomes of 12 such samples in 2020.

Aim III-field testing of the bottom-up and top-down predictive models. Upon generating the bottom-up (protein biomarker) and top-down (GCC+GRE eco-physiological datasets) PLS+NN models, we will test their predictive capacity *in situ*.

We have been sampling corals from two of our four study sites (Wanlitong [hypothetically bleaching-susceptible] and Outlet [hypothetically bleaching-tolerant]; Fig. 1b) since 2017 and will continue to do so at quarterly intervals over the course of 2020 and beyond. To validate the predictive capacity of the bottom-up model, protein biomarker concentrations will be measured in corals (n=48; detailed below) sampled from these two sites (with corals of Houwan and Houbihu analyzed, as well, if funding permits), and biomarker signatures will be fed into the bottom-up model to predict which colonies will bleach as temperatures rise over the course of the 2020 and 2021 summers. As a more simplistic, similarity-based approach, we will compare the protein profiles of field-sampled *P. acuta* colonies to those of the GCC experiments of Table 1 using permutational analysis of variance (PERMANOVA; PRIMER ver. 7). This multivariate statistical approach will allow us to visualize, using multi-dimensional scaling, and quantify (with PERMANOVA) whether the overall protein profiles of the field colonies are more similar to those of bleaching-tolerant (exp.#1) or bleaching-prone (exp.#2) corals.

Simultaneously, the predictive capacity of the top-down approach will be verified by monitoring the EP determined by the PLS+NN model to be most important in gauging the coral response at

all four field sites (Fig. 1b); upon inputting the field environmental data into the models, we will rank the four sites in terms of bleaching susceptibility and monitor the condition of the corals on the reefs to determine if those sites predicted to be most bleaching-prone indeed bleach prior to (or more severely than) those reef sites predicted to be more resilient. In the event that the predictive capacity of both model types is verified, the cheaper, top-down approach will be promoted given that, unlike the bottom-up approach, it requires neither expensive instrumentation (*e.g.,* MS) nor highly trained personnel.
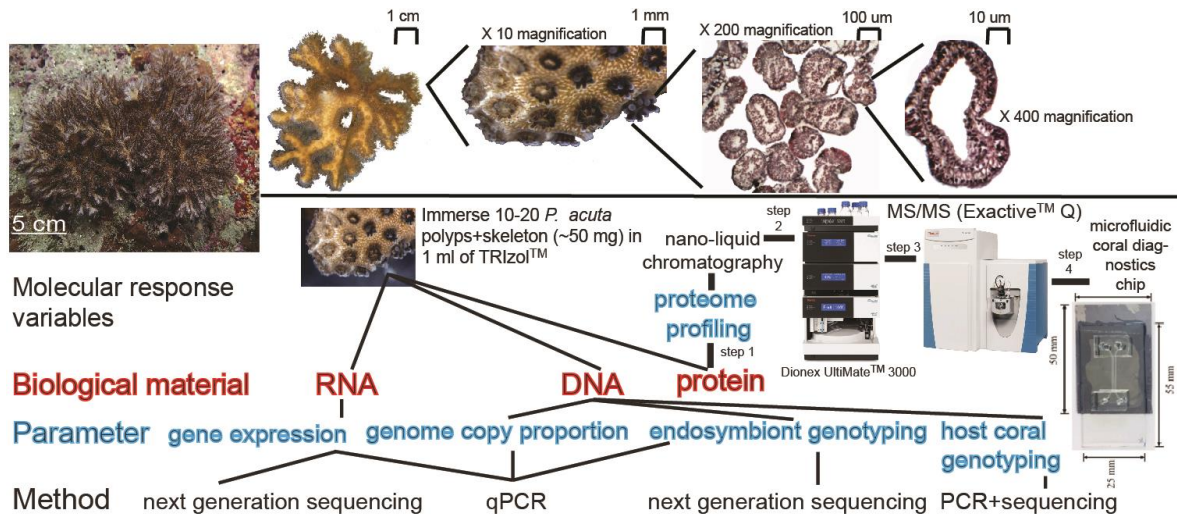

Fig. 4. *Pocillopora acuta* over a variety of biological scales and the molecular protocol routinely used to gauge its health. Protein biomarkers uncovered from analysis of samples of Aim I will be incorporated into a miniaturized microfluidics chip such that coral health can be ascertained while on the diving vessel, not days (or weeks) later in the laboratory. MS=mass spectrometry. qPCR= real-time PCR.

Molecular approaches. There is no correlation between gene expression and protein concentration in corals or their endosymbiotic dinoflagellates (Mayfield et al., 2018a-b). As such, widespread transcriptome profiling efforts by coral biologists (our own included) have not greatly improved our understanding of the coral response to GCC. Are we to develop proactive, pre-bleaching coral health assessment biomarkers and models, we instead require quantitative coral protein concentration data. By using iTRAQ protein labeling in conjunction with MS and customized bioinformatic scripts (known as "MS-SCAN") implemented on our *P. acuta* transcriptome+proteome server (http://symbiont.iis.sinica.edu.tw/coral_pdltte/static/html/index.html#home), we could acquire such data and determine which proteins are involved in coral bleaching (exp.#2 of Table 1); those proteins only synthesized by corals that ultimately bleach will be used to train the bottom-up predictive models described elsewhere.

Molecular approaches-detailed. For all aims, we will use a series of molecular protocols developed over the past decade for work with reef-building coral-Symbiodinaceae dinoflagellate endosymbioses (*e.g.*, Mayfield et al., 2010, 2012b). RNAs, DNAs, and proteins will be extracted with TRIzol™ from 50-mg biopsies from each sampled *P. acuta* (Figs. 1 and 4), and the RNAs will be assayed for expression of several target genes with real-time quantitative PCR (qPCR) as described by Mayfield et al. (2013d, 2014b-c). Furthermore, qPCR assays for genes expressed *only* by temperature-stressed samples are currently being designed. Next generation sequencing (NGS)-based transcriptome profiling (*i.e.*, RNA-Seq) will also be undertaken with a subset of samples from Aims I-II using an Illumina platform.

Although, as mentioned above, gene expression *cannot* be used to make physiological inferences about coral-Symbiodinaceae dinoflagellate "holobionts" (host+endosymbionts), mRNA-level data *can* be used to identify corals displaying aberrant behavior (Fig. 5); specifically, those coral colonies identified as outliers based on a suite of multivariate statistical approaches (Mayfield, 2016), most importantly the Mahalanobis distance (a multivariate equivalent of the standard deviation) will be targeted for proteome profiling as described below; given the prohibitive expense of profiling the proteomes (~$150/sample) of all ~150 coral biopsies proposed to be generated across all three aims, we intend to first employ this previously developed "aberrancy

detection" approach to screen the collective sample set for those coral biopsies that are most likely to be stressed (see Fig. 5 for a representative output of this analysis.). This mRNA focus is due in part to the ease of carrying out qPCR, as well as its low expense, relative to proteome profiling (though the latter is rapidly dropping in price). Furthermore, we have found that the variability across response variables, rather than the mean values of each, is actually a better testament to the degree of stress within coral cells (Mayfield et al., 2019b).

The DNAs co-extracted from the same coral biopsies from which RNAs were isolated will be used for four different purposes (Fig. 4). First, an RNA/DNA ratio will be calculated to estimate levels of total gene transcription. The DNAs will also be used to 1) estimate Symbiodinaceae dinoflagellate density by calculating the "genome copy proportion" (*sensu* Putnam et al., 2013) and 2) genotype the host corals (Mayfield et al., 2018c), and 3) their *in hospite* dinoflagellate populations (*sensu* Correa et al., 2008). Next, proteins will be 1) extracted from the same coral biopsies from which RNAs and DNAs were isolated (*sensu* Mayfield et al., 2014c), 2) purified (*sensu* Mayfield et al., 2016b), and 3) labeled with iTRAQ reagents such that up to eight coral protein samples can

be analyzed simultaneously by nano-liquid chromatography (nano-LC) followed by MS. Instead of identifying differentially concentrated protein spots from 2-dimensional (2D) protein gels (the original proteomics approach), which is subjective and results in semi-quantitative data, this iTRAQ-nano-LC-MS/MS approach sequences all proteins synthesized by each sample in a manner analogous to RNA-Seq and provides spectral data that can be used to extract quantitative information on their cellular concentrations.

Bioinformatics approaches will then be used to uncover those proteins uniquely translated by samples of one treatment and not the other. For Aim I there will be a particular focus on proteins synthesized *only* by samples experiencing temperature stress (i.e., "core bleaching proteins" [CBPs]). A coral-customized proteomic script ("MS-SCAN") will be used to characterize the sequenced proteins upon uploading the MS data files (.MGF) onto our coral transcriptome+proteome server (http://symbiont.iis.sinica.edu.tw/coral_pd ltte/static/html/index.html#home); this interactive coral bioinformatics resource/repository has been described in detail by Mayfield et al. (2014d). We can identify the CBPs the same day we receive the data files from the proteomics

facility, allowing for their rapid characterization. In addition to these RNA-, DNA-, and protein-level response variables (Figs. 4 and 6), several physiological indices of performance will also be assessed in experimental coral specimens from all aims. Please see our prior publications for examples, but these

include (non-exhaustively) growth, Symbiodinaceae photosynthetic efficiency (*sensu* Mayfield et al., 2013c), and chlorophyll content (*sensu* Mayfield et al., 2012a), gastroderm/epiderm ratio (*sensu* Mayfield et al., 2013b), and reproductive output.

Modeling approaches. Although

Fig. 5. The results of an "aberrancy detection system" developed for corals of Fiji's ("F") Lau Archipelago (Mayfield et al., 2017b). Briefly, the four outer wedges correspond to coral+symbiont physiological response axes, with the inner circle reflecting their average. The degree of aberrancy is indicated by color, with red being reserved for corals demonstrating the most atypical behavior. The HMS corresponds to the number of response variables characterized by values >2 standard deviations above the mean (*i.e.*, *z*-score>2). For instance, if gene expression level *z*-scores of 1, 1.5, and **3** were obtained for a particular coral sample, its HMS would be 1.

Gene expression (real-time PCR & next generation sequencing)

Cell morphology & ultrastructure (histology, SEM, TEM)

Metabolite profiling (TLC, HPLC, NMR)

Proteome profiling (nano-liquid chromatography-mass spectrometry)
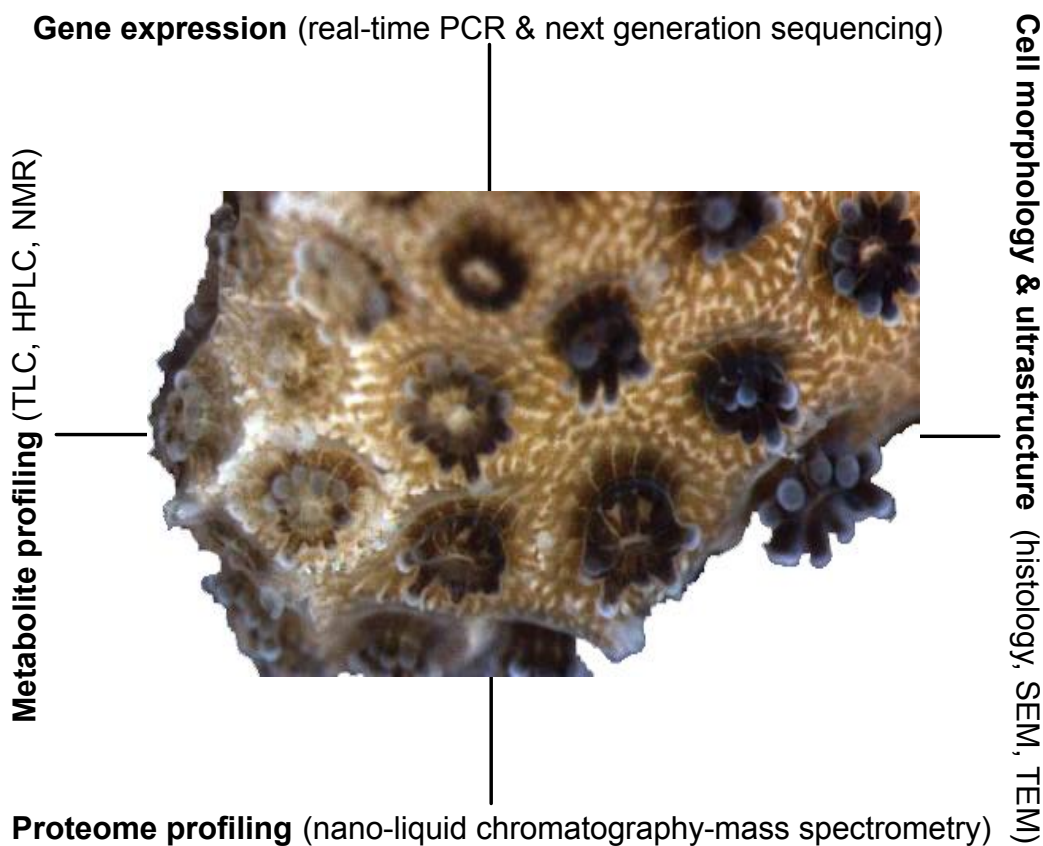
Fig. 6. A dissecting microscope image of the model coral *Pocillopora acuta* and four approaches for generating health-indicative biomarkers. Each polyp is ~1 mm in diameter. HPLC=high-performance (or pressure) liquid chromatography, NMR=nuclear magnetic resonance imaging, SEM=scanning electron microscopy, TEM= transmission electron microscopy, and TLC=thin-layer chromatography. Proteome profiling via nano-liquid chromatography and mass spectrometry is typically preceded by a labeling approach, such as the iTRAQ method outlined herein ("tandem mass tags" [TMT] being the competing technology).

multiple regression and its various incarnations *(e.g.,* stepwise regression) can theoretically handle large, complex environmental datasets (such as those proposed to be analyzed herein), the high collinearity among response variables (a hallmark of all reef-building coral molecular datasets; see above.) leads to models that lack robustness. For this reason, statisticians have been making a strong case for molecular physiologists to instead employ PLS (Cox and Gaudard, 2013), a predictive modeling approach that can be used when there are more response variables than samples (an issue with all "omic" analyses, regardless of target species), as well as when there is a high degree of collinearity between response variables. Given these attributes, PLS appears to lend itself well to the assessment of coral reef data; not only does this algorithm attempt to model the response variable data (y) in such datasets, but it also designs parsimonious, best-fit models for the predictor variables (the experimental treatment in the case of the GCC simulation studies [Table 1] or the EP in the case of the GRE dataset [Fig. 2]).

Issues that plague all modeling approaches (namely tradeoffs between over- and under-fitting) nevertheless persist with PLS, even with large (and growing) datasets such as the two featured herein. For these reasons, AI is quickly being exploited by biologists because of its capacity to "learn from mistakes" made from past simulations when building predictive models, and we propose to use NN to build adaptable models capable of forecasting the likelihood of coral bleaching. Although a description of how AI works is beyond the scope of this article, NN-based model generation is entrenched in the computer program learning from non-optimal projections in which data were over- or under-fit such that ensuing iterations strike a better balance in fitting the training and validation datasets; this balancing process is termed "boosting" by JMP. By integrating data from controlled tank studies (Table 1) and the field (Fig. 1a-b and Fig. 2), we hypothesize that these AI-optimized models (e.g., Fig. 3) will possess an elevated capacity to identify reefs of compromised resilience within a timespan in which management intervention could seek to mitigate local-scale stressors and therefore thwart coral bleaching. In contrast, the bleaching model developed by NOAA ("Coral Watch"), which is based on "degree-heating weeks," suffers from predicting bleaching likelihood only days before it is likely to begin, at which point it may be too late to enact any legislative changes.

Expected outcomes

If the molecular biomarker- and environmental dataset-based machine learning AI models are able to predict which coral colonies and reef sites, respectively, are most susceptible to environmental change, then the developed analytical system will represent the first proactive means of assessing sub-lethal levels of stress in corals and will consequently aid us in determining which reefs are most stress-prone prior to more visible, late-stage manifestations of severe health decline (*e.g.*, bleaching). This novel technology seeks to replace the retroactive, vision-based manner in which we currently assess coral reef health (*i.e.*, by instead documenting death; Liu et al., 2012). In the event that the former, bottom-up PLS-NN model is characterized by such predictive capacity, the underlying protein biomarkers will be integrated into a microfluidic chip; this device (discussed in detail in the next section) would allow for us to make predictions about coral stress susceptibility (CSS) within minutes of sampling (*i.e.*, while still at sea). Using the (current) approach outlined herein, at least several days are required to generate coral health-indicative proteomic data.

If the bottom-up and environmental data-based top-down PLS-NN models are able to accurately forecast coral bleaching events with confidence, the algorithms will be published on open-access websites (*e.g.*, coralreefdiagnostics.com) prior to publication in freely accessible journals (*e.g.*, *PLoS ONE*). Although publication in the peer-reviewed literature will be critical for our future job security, it is the models themselves that are most important for conservation, and we aim to interact with Kenting National Park (KNP) officials and other Taiwanese and Coral Triangle citizens working in the ecosystem management, tourism, and conservation sectors (including those working on cryopreservation; Lin et al., 2019). It will therefore be feasible to ensure that our data reach those capable of doing the most good with them over the course of this project. We aim to further alert managers and concerned members of the general public alike, of our growing capacity to begin to monitor Earth's coral reef ecosystems in a proactive (*i.e.*, pre-death), data-driven manner.

Although corals and the reefs they construct will be the direct beneficiaries of this project, those millions of seafood-obsessed Taiwanese, Filipino, and Indonesia nationals whose livelihoods depend on these Coral Triangle reefs will undoubtedly benefit from a proactive management plan for the targeted protection of their local reefs. In addition to the societal and conservation impacts of

this project, it would provide ample opportunities for interaction among local scientists and students in the Coral Triangle such that we may further explore ways we can work together to understand how corals will respond to the changes in their abiotic milieu that will come to pass over the coming decades. Will even such science-drive, local-scale efforts be enough, though? Global over-population, and the increasing carbon footprint of those of rapidly developing countries (including all nations within the Coral Triangle with the exception of Taiwan, whose population will actually begin decreasing in the coming decades [albeit with an increased carbon footprint per person]), are worrisome indeed.

Data dissemination

We will publish all results in open access journals and on ABM's personal website (coralreefdiagnostics.com) such that any interested individual has access to all data generated (*e.g.*, gene expression levels, protein profiling results, images of the sampled coral colonies and reef sites, etc.), as well as both the bottom-up and top-down models. In fact, we will alert local marine managers at KNP (described above) of such findings well before the respective manuscripts are published. The reason for doing so is because it can take months or even years to publish a

scientific manuscript; during that time, the bleaching-prone coral colonies likely to be identified herein will have already died. Data will be shared with coral reef managers through email or Microsoft's OneDrive data cloud, which we use regularly to share coral imagery data, presentations, and other large files. We will also archive data (at cost) on the open-access website "dryad.org," which we have used previously as a repository for coral health and imagery data from the South Pacific. Also, we are advocates of data transparency and are currently working with developers at JMP to make interactive data plots both on ABM's website and in manuscripts so that those interested can recreate the same figures that we made (and therefore gain greater confidence in the analysis and results).

It should be noted that, although software developed by a for-profit software company, JMP (a subsidiary of the much larger SAS), will be used to build the PLS and NN-based predictive models, the model codes themselves are likely to be quite simple, especially for PLS; several lines of universal scripting code is all that will likely need to be exported to a website, manuscript, email, etc., and JMP allows for the exporting of all relevant code in a format known as "JMP scripting language" (JSL), which is convertible to more common languages

like Python or Java. In other words, one does not need to purchase a JMP license to be able to interpret the code. In the case of the inherently more complex NN algorithms, it is likely that a JSL-based NN code could nevertheless be written such that an individual could use freeware such as R to read, interpret, and execute the program using their own protein biomarker (bottom-up) or environmental (top-down) data.

Regardless of the approach, or the number of model terms, both PLS and NN take only seconds to run on a standard personal laptop, despite the elegant nature of the latter, in particular. As such, although AI has the potential to revolutionize coral health diagnostics, the actual amount of computing power needed is minimal. We therefore anticipate that a plethora of scientists, even in the least developed nations, will have the potential to utilize these technologies; anyone with internet connection and the capability of executing R (3 MB and free of charge) will potentially benefit from the developed models in the event that the cheaper, top-down alternative is found to have predictive capacity.

If, on the other hand, only the more expensive, biotechnology-driven, bottom-up AI/PLS+NN model is found to have high predictive capacity with respect

to coral health, we will work with a Taiwanese microchip laboratory (that of Dr. Gwo-Bin Lee at National Tsing-Hua University) to manufacture small (5-6 cm) microfluidic chips (Fig. 4) featuring custom probes or antibodies that target the proteins found to be indicative of bleaching sensitivity (the "bleaching susceptibility chip") after validating the efficacy of the associated bottom-up CSS model. Although we do not intend to profit from this endeavor, which will enable us to assess coral health while still in the field, we may need to sell the chips at-cost since the microchip lab is non-profit.

Evaluation of results

An entire third of this project (Aim III) will be dedicated to evaluating the experimental models. Four to five *P. acuta* colonies will be tagged along each of two transects at each of two depths at each of four sites in Southern Taiwan (each of which being characterized by differing degrees of human impact and coral bleaching susceptibility; Fig. 1b) at each of four sampling times: January, March, May, and July of 2020. Then, iTRAQ protein labeling+nano-LC+MS/MS (iTRAQ-LC-MS/MS) biomarker profiling will be carried out as described above with a subset of two colonies from Wanlitong (Taiwan Strait) and Outlet

(Nanwan Bay) at each transect x depth and sampling time (n=48 proteomes). When combined with the 30 GCC simulation samples to be analyzed from Aim I and those of other GCC simulations (n=60 across five experiments) and GRE field sites (n=12 from Chagos/BIOT) proposed to be analyzed from Aim II, the proteomes of 150 corals will be profiled in 2020-2021.

The resulting proteomic data for the field-sampled Southern Taiwanese corals will be input into the bottom-up PLS+NN model of Aim I; those found to be over-expressing proteins known to be associated with bleaching will be given high "bleaching susceptibility index" (BSI) scores and will be expected to bleach at high-temperature sampling times (*e.g.,* September 2020). If, instead, the PLS-NN model yields low BSI scores (*i.e.,* proteomes characterized by high concentrations of proteins associated with bleaching resistance upon comparison to those identified in exp.#1 of Aim I) in samples collected during the cooler sampling times (January, March, and May) will instead be given low BSI scores and will not be predicted to bleach during periods of elevated temperatures.

Similarly, environmental data will be gathered from the same four field sites bimonthly (January, March, and May, and July of each year). Some such data, such

as temperature and light can be measured by deployed loggers, which will be left at the sites to log data at 10-min intervals. Other such data, such as coral cover, can only be acquired by dedicated surveyors. Upon inputting the environmental data into the PLS+NN top-down models developed from assessment of the GCC simulation and GRE datasets (Aim II), we will make predictions as to which reefs are most likely to bleach as temperatures rise over the course of the following summer. Then, we will return to the same field sites when temperatures typically peak to determine the diagnostic capacity of our models.

If these models (either or both) are found to have predictive capacity with respective to coral health and bleaching, we can proceed to not only publish the associated manuscripts, but, more importantly, alert government officials of these findings. For instance, if we discover that a high percentage of corals on a particular reef are bleaching-prone based on their biomarker signatures (bottom-up model), yet they have not yet bleached, managers could be alerted to attempt to promote coral resilience by, for instance, closing down the reef to fishing; doing so would diminish the likelihood of algal overgrowth of corals (since herbivorous fish would become more abundant). Active shading or cooling of

threatened corals could also be explored (Mayfield et al., 2019b). Upon validation of the BSI, we will work with the microchip lab discussed above to miniaturize and expedite the coral diagnostics process to where it can be employed *in situ*, with data derived only moments later. Ideally, though, the cheaper, environmental data-based top-down model will likewise be characterized by a high prognostic capability, as this simpler approach better lends itself to developing nations lacking in the funds or infrastructure for proteomic analyses (or the purchasing of diagnostic chips that may cost several hundred USD).

External capacity development

We routinely collaborate with researchers in the countries in possession of the world's most beautiful (arguably) and high-biodiversity coral reefs, the Philippines and Indonesia (Fig. 7), and will seek share findings from this project to our Coral Triangle collaborators over the course of 2020 and into the future. As mentioned elsewhere, the simpler, cheaper, top-down model featuring environmental data-based predictions of coral reef health will ideally prove to have high predictive capacity; in this case, the computer scripts for the resulting algorithms could simply be emailed to our collaborators. If, on the other hand, only the more expensive and complex bottom-up model proves effective, then we will produce and distribute CSS diagnostic chips throughout Taiwan, the Philippines, Indonesia, and elsewhere in the Coral Triangle in 2020-2021.

Finally, we will collaborate with scientists at LOF (where ABM was once a research fellow), the institute that funded and undertook the world's largest ever coral reef survey (the GRE discussed above), as these scientists routinely analyze data that are used to develop educational tools for classroom training of students about coral reef ecosystems. As such, findings to emerge from this proposed work will not only result in the development of computer scripts that may only be interpretable by a few, highly trained individuals, but the more general biological findings of interest (*e.g.*, HOW corals bleach) will be distilled into more basic, fundamental principles that will be integrated into the open-source marine biology teaching materials under constant development by LOF. See www.livingoceansfoundation.org/educationaltools for details and examples.

Fig. 7. A lush, high-biodiversity coral reef ("Siaba Kecil") in Komodo National Park, Indonesia (taken in early 2019). Photograph by ABM.

Acknowledgements

References

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. 1997. Coral bleaching: causes and consequences. Coral Reefs, 16: s129-s138.

Correa, A.M.S., M.D. McDonald & A.C. Baker. 2009. Development of clade-specific *Symbiodinium* primers for quantitative PCR (qPCR) and their application to detecting clade D symbionts in Caribbean corals. Marine Biology, 156: 2403-2411.

Cox, I. & M. Gaudard. 2013. Discovering partial least squares with JMP®. SAS Institute, Inc., Cary, NC, USA. Gates, R.D. 1990. Seawater temperature and sublethal coral bleaching in Jamaica. *Coral Reefs*, 8: 193-197.

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.

Huang, Y.C.A., H.J. Hsieh, S.C. Huang, P.J. Meng, Y.S. Chen, S. Keshavmurthy, et al., 2011. Nutrient enrichment caused by marine cage culture and its influence on subtropical coral communities in turbid waters. Marine Ecology Progress *Series*, 423: 83-93.

Jones R.J., O. Hoegh-Guldberg, A.W.D. Larkum & U. Schreiber. 1998. Temperature-induced bleaching of corals begins with impairment of the CO_2 fixation metabolism in zooxanthellae. Plant Cell and Environment, 21: 1219-1230.

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.

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., H.F. Chang, A.B. Mayfield & H.J. Lin. in press. Influence of the seagrass *Thalassia hemprichii* on coral reef mesocosms exposed to ocean acidification and experimentally elevated temperatures. Science of the Total Environment

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

Liu, P.J., S.M. Lin, T.Y. Fan, P.J. Meng, K.T. Shao & H.J. Lin. 2009. Rates of overgrowth by macroalgae and attack by sea anemones are greater for live coral than dead coral under conditions of nutrient enrichment. Limnology and Oceanography, 54: 1167-1175.

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. 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, 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, 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.

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.