Nutrient enrichment and symbiotic algae are predictive of coral bleaching

Gilia Patterson, Nick Sun, Emily Schmeltzer, Rebecca Mostow

Introduction

Coral reefs are among the most valuable ecosystems on the planet, harboring more than a quarter of marine biodiversity [1] and supporting coastal communities worldwide by providing both ecological and economic benefits [2, 3]. However, the combined effects of multiple anthropogenic impacts alter the coral reef environment and will contribute to the disappearance of these essential assets within our lifetime [4, 5]. Anthropogenic changes in the environment, including increased temperature, pH, and pollution, disrupt coral reefs down to the microbial scale. Thermally stressed corals may bleach by expelling their symbiotic algae, Symbiodiniaceae, which provide essential nutrients and energy resources to the coral animal [6, 7]. While it is possible for surviving corals to recover from bleaching with the uptake of new symbionts from the water column, this recovery process is not well-studied. Furthermore, some species of symbionts have been shown to be more thermotolerant than others [8].

In addition to the stress of bleaching, nutrient enrichment from natural and anthropogenic sources may play a role in how coral reef microbial communities change in response to their environment. It has previously been shown that nitrate (NO_3^-) enrichment often impedes coral growth whereas ammonium (NH_4^+) does not [9]. Nitrate is a common nutrient from anthropogenic sources while urea from fish pee is a common ammonium input. We were afforded the unique opportunity to study the compounding effects of multiple stressors on bacterial and algal symbiont communities during a coral bleaching event.

To assess the interacting effects of bleaching and nutrient enrichment on the algal symbiont communities, we sampled and monitored coral colonies in nutrient enriched plots before, during, and after the 2016 bleaching event on the north shore of the island of Moorea, French Polynesia. This experiment provides important insight not only into temporal shifts in phylotype species over time, but which phylotype species may be contributing to a coral's ability to withstand a bleaching event under increasingly hostile environmental conditions. Additionally, we demonstrate that many different statistical tests and tools may be necessary for the analysis of this type of multivariate dataset in order to attempt to answer multiple meaningful biological questions. Many researchers are beginning to breed corals, purposely selecting for traits that may assist coral survival in the face of climate change [7]; the results of this study may prove a useful tool for future projects of this kind.

Biological Question

How does the composition of symbiotic algae, coral species, and nutrient enrichment affect coral bleaching? Which environmental variables are contributing to bleaching?

Data Structure

The Vega-Thurber lab collected ITS2 data on 67 individual corals over five time points, January 2016, March 2016, May 2016, July 2016, and January 2017. For each coral and time point, the dataset is comprised of the relative abundances of 76 phylotypes of symbiotic algae in the sample, the species of coral, the nutrient enrichment plot that the coral was from (urea, nitrate, or control), and whether the coral had any bleaching (coded as 0 or 1). Many combinations of coral and time point are missing, so in total we have data from 287 samples. Table 1 shows the number of corals that had any bleaching for each of the timepoints. Only 6.6% of the samples were from corals that bleached and almost all of the bleached samples were from a single time point, May 2016.

	Date	Unbleached	Bleached
1	January 2016	36	0
2	July 2016	68	2
3	March 2016	59	0
4	May 2016	52	17
5	January 2017	53	0

Table 1: Number of corals that were bleached at each timepoint in the study.

Sample Preparation

Coral colonies of genera *Acropora*, *Pocillopora*, and *Porites*, some of the most abundant and ecologically important corals [10] in Moorea, were tagged at 10m depth in January 2016 in one site on the fore reef of the island. After the initial sampling of 1cm coral nubbins from each tagged colony, we then established control plots and nutrient enrichment plots using slow-release diffusers filled with nitrate-only or urea-only fertilizer to mimic anthropogenic and natural nitrogen inputs. Diffusers were replaced as needed. 6 replicates per coral species were sampled within each of the 27 total nutrient enrichment plots. The bleaching event was expected to occur in the spring, and coral nubbins were again sampled in the same manner in March, May, and July of 2016, and again in January 2017. Sample sites saw the greatest proportion of bleaching in May 2016. DNA was extracted from the nubbins using the MoBio PowerSoil kit®, and PCR was used to amplify the ITS2 rRNA gene of the algal symbionts present within each coral sample. Amplicon libraries were then sequenced using the Illumina MiSeq platform at Oregon State University, and subsequent sequences underwent quality control, filtering, and taxonomic Symbiodiniaceae phylotype assignment using the SymPortal platform [11].

Statistical Methods

We fit logistic regression models with the lasso penalty to determine which phylotypes, treatments, and coral genera were predictive of bleaching. We fit two models, one with the data from all five timepoints and one with just the data from May 2016. The model with all timepoints uses random effects to handle repeated measures on the same coral. We chose to fit the second model on May 2016 because 17 of the 19 bleached samples were from that timepoint and focusing on just one timepoint allowed us to fit a simpler model without random effects. The response for both models is the binary bleaching variable and the potential predictors are the relative abundances of the phylotypes, coral species, and treatment groups. We did not include any interaction terms. In our analysis we ignored the fact that the nutrient enrichment treatments were applied to plots of 18 corals rather than individual corals. Future analyses should take into account the correlation between corals in the same plot by including a random plot effect.

Model 1: All timepoints

We followed the procedure outlined by [12] to fit a Lasso-penalized generalized linear mixed effects model using the R package `glmmLasso` [13]. To account for correlation between measurements on the same coral, we included individual coral as a random effect. In order to test the prediction capabilities of this model, we first partitioned the data into a training set and a test set. We selected the training and test sets so that the proportion of bleaching events was about the same in each.

To reduce the dimensionality of our dataset, we removed phylotypes that were not significantly correlated with the response using a correlation test with Spearman's rank based correlation statistic. We did not adjust for multiple comparisons in our tests since we did additional variable selection later in our model. We also dropped phylotypes that were present in less than two samples.

To choose the penalty parameter, λ, that minimized the deviance of the model, we used five-fold cross validation on the training set. Cross-validation is a statistical technique in which the data is split into equal folds and is trained and tested five times, using a different fold as the test set each time.

We fit the model on the training set using the chosen λ and then tested the model on the test set. Given our original small data set, we did not have enough final test data to get a large confusion matrix, but we should still be able to get some intuition for how our final model predicts new observations.

Model 2: May 2016 only

We used the R package `glmnet` [14] to fit a lasso-penalized logistic regression model on just the data from May 2016. Before fitting, we removed the 24 phylotypes that were present in less than two samples, leaving 52 phylotypes.

The logistic regression model is:

$$y_i = \begin{cases} 1 & \text{coral } i \text{ is bleached} \\ 0 & \text{otherwise} \end{cases}$$
$$y_i \sim \text{Bernoulli}(p_i)$$
$$\log \frac{p_i}{1-p_i} = \beta_0 + \boldsymbol{\beta}^T \mathbf{x_i}$$

Lasso shrinks many of the coefficients to 0 by estimating $(\beta_0, \boldsymbol{\beta})$ that minimize the negative penalized log-likelihood of $y_1, ..., y_n$,

$$-\left(\frac{1}{n}\sum_{i=1}^{n} y_i \left(\beta_0 + \boldsymbol{\beta}^T \mathbf{x_i}\right) - \log\left(1 + e^{\beta_0 + \boldsymbol{\beta}^T \mathbf{x_i}}\right)\right) + \lambda \|\boldsymbol{\beta}\|_1$$

The R package `glmnet` uses cyclic coordinate descent to estimate $(\beta_0, \boldsymbol{\beta})$ for many values of λ. For a given value of λ, the algorithm starts with initial values for the coefficients and calculates a quadratic approximation of the negative penalized log-likelihood by taking the 2^{nd} order Taylor expansion around the initial values. The quadratic approximation is then minimized by minimizing for one parameter at a time until the parameters do not change. To estimate $(\beta_0, \boldsymbol{\beta})$ for a range of values, `glmnet` starts with the smallest value of λ for which all the coefficients $(\beta_0, \boldsymbol{\beta})$ are estimated to be 0. At each step, the algorithm decreases λ and minimizes the negative penalized log-likelihood using the estimated values of $(\beta_0, \boldsymbol{\beta})$ from the previous step as the initial values.

We used ten-fold cross validation to choose the value of λ that minimized the deviance of the model. We then fit a regular logistic regression model in R [15] using just the explanatory variables with non-zero coefficients.

Results

Model 1: All time points

After filtering the 76 phylotypes using Spearman's Rho correlation with our bleaching response, we were left with 12 significant phylotypes. Through 5 fold cross validation on our training data set, we found that the optimal penalty parameter was 1.3. We finally refit the model using our entire training set with this penalty parameter. The estimated coefficients for the phylotypes, timepoints, and treatments are in Table 2. One phylotype, C120b, was shrunk to zero.

The factor variable for the time point May 2016 has a positive coefficient, indicating that it increases the probability of bleaching. In addition, both the nitrate and urea treatments have positive relationships to bleaching. Some of the phylotypes such as C3au have large correlations with bleaching.

To check for collinearity in our model, we calculated the correlation between relative abundances of the phylotypes in the model (Figure 1). Unsurprisingly, there are some phylotypes that are correlated such as all of the C42 phylotypes. However, there is nothing overly concerning with this correlation matrix. Upon

Variable	Estimated Coefficient
(Intercept)	-5.86
C3	-1.88
C1	-0.09
C15	-1.65
C42.2	-0.03
C120b	0.00
C15ae	-8.30
C1h	-100.31
C42i	46.34
C42j	-2.53
C42l	41.87
C3au	1030.53
C15i	-27.96
as.factor(Timepoint)16-Jul	0.85
as.factor(Timepoint)16-Mar	0.28
as.factor(Timepoint)16-May	4.19
as.factor(Timepoint)17-Jan	-0.28
as.factor(Coral)poc	-0.36
as.factor(Coral)por	-0.64
as.factor(Treatment)N	1.06
as.factor(Treatment)U	0.78

Table 2: Estimated coefficients for the model including all time points.

closer inspection of our training data, C3au was found to be a relatively rare phylotype and only present in 5 samples, 2 of which bleached. This proportion is the probable cause of this large coefficient.

The test set we partitioned before creating our model contains 53 non-bleaching events and 3 bleaching events. Using our final `glmmLasso` model, we predicted whether or not each coral in the test set bleached using a cutoff of predicted probability greater than 0.5 (Table 3). We were able to correctly predict 2 of 3 bleaching events and 52 of 53 non-bleaching events, giving us an approximate accuracy of 96%.

	Observed 0	Observed 1
Predicted 0	52	1
Predicted 1	1	2

Table 3: Confusion Matrix for our final model on withheld data

Model 2: May 2016

Table 4 shows the results of the model fit on samples from May 2016. Six phylotypes and treatment had non-zero coefficients. We found that both urea and nitrate treatment increased the probability that a coral bleached compared to the control. We found that increasing the relative abundance of the phylotypes C42.2, C1d, C3f, C3bj, C42i, or C3au increased the probability that a coral bleached. One phylotype, C3au, had a very large coefficient. C3au was in 3 bleached and 2 unbleached corals. The very large coefficient is probably because the relative abundance of C3au was 0.001 in all five corals, so a very low relative abundance was associated with increased probability of bleaching.

To evaluate the fit of the model, we plotted the change in deviance if each observation was removed (ΔD_i) vs fitted values (Figure 2). The observations seem to fit the model well except for one coral sample with very high ΔD_i. The sample was from a bleached coral given the nitrogen treatment with relative abundances of 0 for C42.2, C1d, C3f, C3bj, C42i, and C3au.

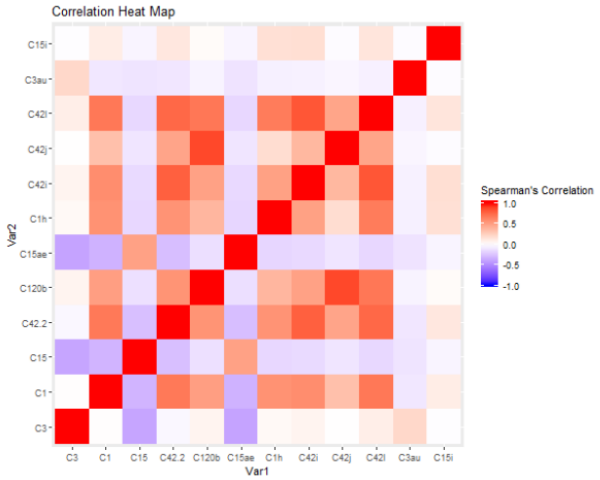

Figure 1: Correlation between relative abundances of phylotypes with non-zero coefficients in the model with all time points.

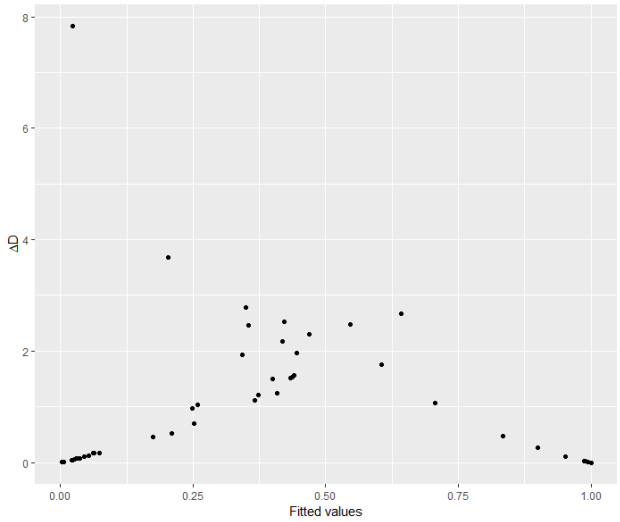

Figure 2: Plot of change in deviance vs fitted probabilities for the May 2016 model.

	Variable	Estimated Coefficient
1	C42.2	5.11
2	C1d	4.15
3	C3f	17.87
4	C3bj	27.52
5	C42i	153.30
6	C3au	2035.45
7	TreatmentN	2.25
8	TreatmentU	2.22

Table 4: Estimated coefficients for the May 2016 model.

Conclusions and Discussion

Both the full and May 2016 models found that nutrient enrichment will positively impact the probability of a coral bleaching during increased temperature anomalies. Both the nitrate and the urea nutrient treatments were found to be predictive of higher bleaching probabilities, an interesting result given that only the nitrate treatment has been previously shown to negatively affect coral growth over time [9].

The reduced, single time-point model identified higher relative abundances of C42.2, C1d, C3f, C3bj, C42i, and C3au phylotypes as predictive of bleaching probability. There are two possible explanations for the association of these symbiodinia with coral bleaching. Either a) they could be indicators of a thermal stress response in their coral host or b) they may be filling niche space previously occupied by other phylotypes prior to bleaching. The C1 and C3 phylotypes have both previously identified as thermosensitive [16]. Additionally the C42.2 phylotype is usually present in samples with low variants of C1 and C3.

A few statistical caveats limit our interpretation of these models. It is generally difficult to test the predictive power of logistic regression models because the model predicts probabilities but our observed data is binary and consists only of 0s and 1s. These two metrics are not directly comparable and therefore further tests are needed to validate logistic regression models. Coral bleaching events are relatively rare in our data set, which additionally compounds the difficulty of getting a good estimate of prediction error. A larger dataset with a higher proportion of bleached to unbleached corals would greatly improve the prediction power of this model. The usual method of validating a model of this type by dividing it into test and training datasets is not ideal in this situation because the test and training sets are not independent. Ideally, we would be able to test the model on a completely new data set, made up of the same variables but conducted by different researchers.

Future Analyses

A model capable of predicting which corals will bleach during a temperature anomaly would be of profound scientific and management utility. A predictive model of that kind could be used to set management priorities, define scientific study areas, and better prepare local communities for the effect of global climate change. To further refine this model, we suggest using more extensive, long term ecological datasets. It would potentially be more advantageous to build different models for each coral species. More testing and validation would be needed to understand the prediction error rates of any model.

References

1. Plaisance L, Caley MJ, Brainard RE, Knowlton N. The diversity of coral reefs: What are we missing? PLoS One. 2011;6:e25026.

2. Brander LM, Van Beukering P, Cesar HS. The recreational value of coral reefs: A meta-analysis. Ecological Economics. 2007;63:209–18.

3. Moberg F, Folke C. Ecological goods and services of coral reef ecosystems. Ecological economics.

1999;29:215–33.

4. Hoegh-Guldberg O. Coral reef ecosystems and anthropogenic climate change. Regional Environmental Change. 2011;11:215–27.

5. Vega Thurber RL, Burkepile DE, Fuchs C, Shantz AA, McMinds R, Zaneveld JR. Chronic nutrient enrichment increases prevalence and severity of coral disease and bleaching. Global change biology. 2014;20:544–54.

6. Bourne DG, Morrow KM, Webster NS. Insights into the coral microbiome: Underpinning the health and resilience of reef ecosystems. Annual Review of Microbiology. 2016;70:317–40.

7. Oppen MJ van, Oliver JK, Putnam HM, Gates RD. Building coral reef resilience through assisted evolution. Proceedings of the National Academy of Sciences. 2015;112:2307–13.

8. Stat M, Gates RD. Clade d symbiodinium in scleractinian corals: A "nugget" of hope, a selfish opportunist, an ominous sign, or all of the above? Journal of Marine Biology. 2011;2011.

9. Shantz AA, Burkepile DE. Context-dependent effects of nutrient loading on the coral–algal mutualism. Ecology. 2014;95:1995–2005.

10. Lenihan HS, Edmunds PJ. Response of pocillopora verrucosa to corallivory varies with environmental conditions. Marine Ecology Progress Series. 2010;409:51–63.

11. Hume B, Smith E, Ziegler M, Warrington H, Burt J, LaJeunesse T, et al. SymPortal: A novel analytical framework and platform for coral algal symbiont next-generation sequencing its 2 profiling. Molecular ecology resources. 2019.

12. Tipton L, Cuenco KT, Huang L, Greenblatt RM, Kleerup E, Sciurba F, et al. Measuring associations between the microbiota and repeated measures of continuous clinical variables using a lasso-penalized generalized linear mixed model. BioData Mining. 2018;11.

13. Groll A. GlmmLasso: Variable selection for generalized linear mixed models by l1-penalized estimation. R package version. 2014;1.

14. Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. Journal of statistical software. 2010;33:1–22.

15. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2018. https://www.R-project.org/.

16. Swain TD, Chandler J, Backman V, Marcelino L. Consensus thermotolerance ranking for 110 symbiodinium phylotypes: An exemplar utilization of a novel iterative partial-rank aggregation tool with broad application potential. Functional Ecology. 2017;31:172–83.