Title:

**Authors:** Jason D. Selwyn1\*, INSERT PEOPLE, Steven V. Vollmer1

# Affiliations:

1: Department of Marine and Environmental Sciences, Northeastern University, Nahant, MA, USA.

\*Corresponding author. Email: [j.selwyn@northeastern.edu](mailto:j.selwyn@northeastern.edu)

# Keywords

# Abstract

# Introduction

Caribbean coral reefs are inundated with global and local stressors leading to stark declines in coral cover, abundance, and diversity throughout the region (**CITE**). Historically these reefs were dominated by the sister *Acropora* taxa, *A. cervicornis* and *A. palmata* (**CITE**). However, since the 1980’s these two species have been decimated by the spread of an apparently novel disease causing > 95% mortality (**CITE**). This novel disease is typified by an advancing bleaching margin ahead of a nectrotic tissue margin named white band disease (WBD; **CITE**). Identification of the culprit pathogen(s) has been difficult with multiple taxa being potentially implicated by different research avenues of the years (**CITE**).

* Holobiont makes identification of pathogenic culprit difficult
  + *Vibrio charcharia* like species always present in surface mucopolysaccharide layer (Ritchie and Smith 1998) – type 2
  + *Vibrio harveyi* (Gil-Agudelo et al. 2006) – type 2

While the specific pathogen has yet to be identified to satisfy the Henle–Koch postulates it has been determined to likely be a bacterial infection due to the inhibition of the disease with antibiotics (**CITE**). **CONTINUE HERE TALKING ABOUT SOME OF THE IMPLICATED BACTERIA**

* ML methods can supplement traditional disease press experiments to narrow candidate list

Beyond simple pathogen identification the entire coral holobiont is potentially implicated in the disease spread with a cascading succession of microbial types including commensalists, mutualists, pathogens, and opportunists all likely changing in abundance through the process of dysbiosis undertaken in the transition from healthy to diseased coral (Vega Thurber et al. 2020). Using a suite of traditional and novel analytical techniques we aim to decipher the roles of microbes observed in, and putatively associated with white band disease in *A. cervicornis.*

# Methods

## Field Collections

* Field data collection (surveys and microbial samples)

The prevalence of WBD was modelled using a generalized linear mixed model with a logit link function using sampling time as a fixed effect and a random effect of site with significance assessed using a likelihood ratio test (Bates et al. 2015). Pairwise contrasts were used to identify significant differences between sampling timepoints using the Westfall p-value adjustment to control the familywise error rate (Westfall 1997, Hothorn et al. 2008). In addition to test for significant differences between years and season by combining nested sampling times and using additional pairwise contrasts (Lenth 2019).

## Microbiome Characterization

* 16s sequencing
* ASV Identification

Samples were pruned to keep only those with more than 1,000 reads and to keep ASVs identified as bacteria after the removal of cyanobacteria, mitochondria, and chloroplast sequences. ASVs were further filtered to retain those found in at least 10% of samples and in all four sampling timepoints. Read counts of the remaining ASVs were normalized for variable sequencing depth using the trimmed mean of M-values method with singleton pairing implemented in edgeR (TMMwsp; Robinson et al. 2010, Robinson and Oshlack 2010). Normalized read counts were converted to log2 counts per million reads with all subsequent analyses being performed on these normalized and log-transformed counts per million.

To visualize differences between healthy and diseased microbiomes we estimated the microbial distance between coral colonies using the Raup-Crick distance based on presence/absence of ASVs (Raup and Crick 1979, Chase et al. 2011). These distances were visualized using a non-metric multidimensional scaling (NMDS) plot (Legendre and Legendre 2012, Oksanen et al. 2013). We used a permutational ANOVA with 10,000 permutations to first test for differences between healthy and diseased colony microbiomes and secondarily test for differences based on sampling season, year, and site.

## Model Training and ASV Identification

ASVs associated with diseased corals were identified using an ensemble set of five distinct machine learning models (lasso logistic regression (Friedman et al. 2010), random forest (Wright and Ziegler 2017), multilayer perceptron (Collobert et al. 2011, Kuhn and Falbel 2022, Falbel and Luraschi 2023), linear support vector machine (Karatzoglou et al. 2004, 2022), partial least squares (Rohart et al. 2017), and K-nearest neighbors (Schliep and Hechenbichler 2016)). Prior to model fitting the dataset was split into training (75%) and testing (25%) sets to reduce overfitting with all model tuning being done using the training set and only using the test set to calculate the final model metrics (e.g., accuracy). All ASVs were preprocessed by first removing ASVs with near-zero variance. The remaining ASV counts were normalized using the Yeo Johnson transformation and then center and scaled to have a mean of 0 and standard deviation of 1 (Yeo and Johnson 2000). The models were chosen to represent a diversity of machine learning classification models with the thought that they may select different important features for predicting coral disease state (Bolón-Canedo and Alonso-Betanzos 2019). Looking at ASVs identified across multiple models therefore are likely to be candidates for investigation as putative pathogens, opportunists or beneficial bacteria (Vega Thurber et al. 2020).

Model hyperparameters were individually tuned to identify the hyperparameter combination which minimizes the Brier score, a metric designed to penalize misclassifications and reward confident, correct classifications (Brier 1950, Kruppa et al. 2014). All model hyperparameters were tuned by fitting the models on the training dataset cross-validated using 10-fold cross-validation repeated ten times. Each model was fit to these datasets using an initial random grid with 50 random combinations of parameters. This random grid was used to fit a Bayesian gaussian process model to locate hyperparameter space which minimizes the Brier score (Kuhn and Silge 2022). Following prediction of the hyperparameter combination minimizing Brier score the model is refit to the training data and assessed for improvement. This method is then repeated up to 200 iterations or after 25 iterations with no improvement of the Brier score.

Equivalently high-quality models of coral disease were identified by identifying the region of practically equivalent models which are 80% likely to be within 1% overall quality as the best fit model. The overall quality metric used to compare models was a composite metric combining model accuracy, area under the receiver operator curve (ROC AUC), and Brier score (Derringer and Suich 1980). The overall quality metric for the repeated cross validation fitted results were fitted using a hierarchical Bayesian model to identify differences in model quality (Kuhn and Silge 2022). Random effects for repeats and folds within repeats were included to account for the repeated measurements of these data subsets (Kuhn and Silge 2022).

We identified ASVs important to identifying coral disease state for all of the top quality models by calculating Monte Carlo based Shapley values (Shapley 1953, Štrumbelj and Kononenko 2014, Greenwell 2023). Shapley values were calculated independently for each model and ASV using 500 simulations with ASV importance being calculated as the mean of the absolute values of the Shapley values (Molnar 2022). Using an ensemble feature selection approach ASVs which were consistently highly ranked were identified by performing a set of one-sided Wilcoxon rank-sum tests comparing the ASV rankings across models to the average of all other ASVs (Bolón-Canedo and Alonso-Betanzos 2019). These tests were corrected using the sequential Bonferroni adjustment (Holm 1979).

## Field Consistency

The abundance of microbes was also modelled in the field to identify consistency in microbial abundance across sampling years and coral health conditions. We used a linear mixed effects model with a fixed effect of coral health (diseased or healthy) year, and season and all interactions with site as a random effect (Bates et al. 2015). As a follow-up post-hoc analysis we compared healthy and diseased corals within each timepoint to confirm the consistency of significant differences in microbial abundance across sampling times in all identified important ASVs.

## Tank Experiments

To the identify individual roles in coral dysbiosis of the most important ASVs associated with discriminating between healthy and diseased corals we performed a tank exposure experiment (Vega Thurber et al. 2020). In January 2017 six coral genotypes were collected from \_\_\_\_. Coral genotypes were then fragmented into replicate fragments spread across three disease and three healthy exposure 18-liter recirculating tanks at ambient seawater temperatures. Prior to placing in the tanks each fragment was experimentally lesioned with a waterpik to facilitate transmission (Gignoux-Wolfsohn et al. 2012). The disease exposure tanks were exposed to 50ml of disease slurry produced from 10 WBD infected coral fragments while healthy exposed tanks were exposed to 50ml of healthy slurry from 10 healthy fragments. Slurries were produced by waterpiking disease or healthy coral tissue off the sampled corals in filtered seawater (FSW) and normalizing the slurry doses to a standard ocular density of 0.6 at 600nm. Fragment microbiomes were sampled at three timepoints, after fragmentation and placement in tanks (day 0), immediately following tank dosage with either diseased or healthy slurries (day 2), and one-week post-exposure (day 8). Samples were sequenced and bioinformatically processed in the same way as field sample collections (see above).

The abundance of ASVs consistently identified as discriminatory between healthy and diseased corals in the field were analyzed in the tank exposure experiment using a before-after control-impact design (**CITE**). To account for the repeated measurements we included a random effect for coral fragment nested within random effects for both genotype and tank. The fixed effect treatments analyzed were the exposure to the disease and whether the genotype got diseased when exposed to it (i.e., susceptible) or not (i.e., resistant).

## ASV Correlation Network

* Identify ASVs which across all field & tank samples correlate strongly with implicated ASVs from previous steps

## ASVs typing

* How do we group together similar ASVs and how can they be classified into the schema described by vega Thurber
  + Pathogen
    - ML – important
    - Field – D > H @ all timepoints
    - Tank
      * ↑ susceptible post disease exposure
      * Constant other treatments
  + Opportunist –
    - ML
    - Field – D > H but doesn’t have to always be true
    - Tank
      * ↑ susceptible & resistant post disease exposure
      * Constant healthy exposure
  + Mutualist –
    - ML
    - Field
    - Tank
      * ↓ post disease exposure (more-so susceptible?)
  + Beneficial –
  + Commensalist -
    - Unchanging across tank/field & not ML important

ASVs with significant increases in abundance in the susceptible fragments exposed to the disease (i.e., those which got the disease) without a comparable increase in the other treatments were classified as putative pathogens. ASVs significantly more abundant in resistant genotypes when exposed to disease were classified as potential

# Results

The prevalence of white band disease in Bocas del Toro, Panama was found to significantly vary through time (χ2(4) = 99.98, *p* < 0.0001, Figure 1). In particular we found that there is a somewhat circular oscillation between a relative high of 56.9% ( ± 3.4% SE) which we observed in January 2016 and a relative low of 25.5% ( ± 2.3% SE) occurring during the July 2015 and January 2017 sampling. This oscillatory period does not appear to reflect a seasonal cycle (January: 38.7% ± 2.9% SE, July: 38.9% ± 2.5%, *p* = 0.949) but may be related to annual differences with 2015 (28.0% ± 2.9% SE) and 2017 (34.4% ± 2.7% SE) having significantly lower disease prevalence than 2016 (49.5% ± 2.9% SE, *p* < 0.001) and only marginally different disease prevalence between them (*p* = 0.044). We found strongly significant differentiation in the microbial communities of healthy and diseased corals (r2 = 0.52, *p* < 0.0001, Figure 2A). The differences in microbial community between healthy and diseased colonies was of the same magnitude as between sampling sites (r2 = 0.53, *p* < 0.0001, Figure 2B) with communities significantly differing by year (r2 = 0.11, *p* < 0.0001, Figure 2C) and season (r2 = 0.17, *p* < 0.0001, Figure 2D) to a lesser extent.

Of the six tested predictive models four were found to be of equivalent high quality in classifying the training data and similarly high quality with the test dataset (Table 1, Figure 3). Specifically, we found after tuning the best model using a metric combining accuracy, ROC/AUC and Brier score to be a multilayer perceptron with five hidden units, a dropout rate of 0.01 and a learning rate of 0.17 trained for 34 epochs using the elu activation function. The other equivalently good models were the support vector machine (cost = 0.00098), lasso regression (penalty = 1 x 10-10), and random forest (mtry = 55, trees = 438, min\_n = 2) which all had test set accuracies greater than 97%. While the k-nearest neighbor (test accuracy = 96.2%) and partial least squares (test accuracy = 99.0%) were both substantially better than the null model (test accuracy = 65.4%) neither was equivalent to the multilayer perceptron model across all quality metrics with lower accuracy and ROC/AUC in the k-nearest neighbor model and less confident classifications (Brier score) in the partial least squares model (Table 1).

The high-quality models consistently identified 12 ASVs as of above average importance in classifying healthy and diseased corals (Table 2, Figure 4, Figure 5). Of these 12 important ASVs three were found to have inconsistent differential abundances across sampling timepoints, a Rhodobacteraceae (ASV15) which did not differ between healthy and diseased samples in January 2016 (FC = 1.0 ± 0.54, *p* = 0.066), and an Algicola (ASV49) and an Endozoicomonas (ASV700) which did not differ between healthy and diseased samples in July 2016 (FC = -0.05 ± 0.41, *p* = 0.91, FC = 0.37 ± 0.27, *p* = 0.17, respectively). One ASV, a Sphingomonadaceae (ASV40), was found to be consistently more abundant in healthy field samples (FC = -3.31 ± 0.23, *p* < 0.0001) and in tank samples was significantly more abundant in susceptible samples prior to disease exposure (FC = 2.74 ± 0.98, *p* = 0.011). The remaining eight ASVs which consistently distinguish healthy and diseased corals were consistently more abundant in diseased samples (Table 2). Of these one, an Endozoicomonas (ASV322), was significantly more abundant in susceptible corals prior to treatment exposure (FC = 0.86 ± 0.37, *p* = 0.026). An additional ASV (ASV8), an Aliivibrio, was significantly elevated in disease susceptible corals both prior to treatment (FC = 3.26 ± 1.43, *p* = 0.028) and remained more abundant after treatment with the disease dose (FC = 2.35 ± 1.07, *p* = 0.040). Finally, three ASVs all only differed in the post disease exposure samples with significantly more in susceptible corals. These ASVs include a Francisellacea in the [Caedibacter] taeniospiralis group (ASV25, FC = 3.27 ± 0.81, *p* = 0.0006), a Thalassolituus (ASV26, FC = 1.49 ± 0.64, *p* = 0.032), and an Endozoicomonas (ASV361, FC = 0.40 ± 0.13, *p* = 0.005).

# Discussion

* Panama 9 dominant ASVs distinguish healthy and diseased coral microbiomes representing X microbial families
* Classification of each of the 9 ASVs into functional categories
* Dive into likely pathogens and what we know about them
* Caveats – may not be pathogen because could be extremely opportunist
* Unlike previous Florida based studies we found nearly 0 Ricketsias but a vast amount of Endozoicomonas
* However – bring in other studies and mention overlaps of high order taxonomic levels as candidates
* Suggest culture and introduction of one of two to finally close Henle–Koch postulates for WBD
  + Alternatives to Koch’s postulates

# Literature Cited

Bates, D., M. Mächler, B. Bolker, and S. Walker. 2015. Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software 067.

Bolón-Canedo, V., and A. Alonso-Betanzos. 2019. Ensembles for feature selection: A review and future trends. Information Fusion 52:1–12.

Brier, G. W. 1950. Verification of forecasts expressed in terms of probability. Monthly weather review 78:1–3.

Chase, J. M., N. J. B. Kraft, K. G. Smith, M. Vellend, and B. D. Inouye. 2011. Using null models to disentangle variation in community dissimilarity from variation in α-diversity. Ecosphere 2:art24.

Collobert, R., K. Kavukcuoglu, and C. Farabet. 2011. Torch7: A Matlab-like Environment for Machine Learning. Page BigLearn, NIPS Workshop.

Derringer, G., and R. Suich. 1980. Simultaneous Optimization of Several Response Variables. Journal of Quality Technology 12:214–219.

Falbel, D., and J. Luraschi. 2023. torch: Tensors and Neural Networks with “GPU” Acceleration.

Friedman, J., T. Hastie, and R. Tibshirani. 2010. Regularization Paths for Generalized Linear Models via Coordinate Descent. Journal of Statistical Software 33:1–22.

Gignoux-Wolfsohn, S. A., C. J. Marks, and S. V. Vollmer. 2012. White Band Disease transmission in the threatened coral, *Acropora cervicornis*. Scientific Reports 2:804.

Gil-Agudelo, D. L., G. W. Smith, and E. Weil. 2006. The white band disease type II pathogen in Puerto Rico. Revista de Biología Tropical 54:59–67.

Greenwell, B. 2023. fastshap: Fast Approximate Shapley Values.

Holm, S. 1979. A Simple Sequentially Rejective Multiple Test Procedure. Scandinavian Journal of Statistics 6:65–70.

Hothorn, T., F. Bretz, and P. Westfall. 2008. Simultaneous Inference in General Parametric Models. Biometrical Journal 50:346–363.

Karatzoglou, A., A. Smola, and K. Hornik. 2022. kernlab: Kernel-Based Machine Learning Lab.

Karatzoglou, A., A. Smola, K. Hornik, and A. Zeileis. 2004. kernlab – An S4 Package for Kernel Methods in R. Journal of Statistical Software 11:1–20.

Kruppa, J., Y. Liu, H.-C. Diener, T. Holste, C. Weimar, I. R. König, and A. Ziegler. 2014. Probability estimation with machine learning methods for dichotomous and multicategory outcome: Applications. Biometrical Journal 56:564–583.

Kuhn, M., and D. Falbel. 2022. brulee: High-Level Modeling Functions with “torch.”

Kuhn, M., and J. Silge. 2022. Tidy Modeling with R: A Framework for Modeling in the Tidyverse. 1st edition. O’Reilly Media, Sebastopol, CA.

Legendre, P., and L. F. J. Legendre. 2012. Numerical Ecology, Volume 24, Third Edition. 3 edition. Elsevier, Amsterdam.

Lenth, R. 2019. emmeans: Estimated Marginal Means, aka Least-Squares Means.

Molnar, C. 2022. Interpretable Machine Learning: A Guide For Making Black Box Models Explainable. Independently published, Munich, Germany.

Oksanen, J., F. G. Blanchet, R. Kindt, P. Legendre, P. R. Minchin, R. B. O’Hara, G. L. Simpson, P. Solymos, M. H. Stevens, and H. Wagner. 2013. vegan: Community Ecology Package. R package version 2.0-8.

Raup, D. M., and R. E. Crick. 1979. Measurement of Faunal Similarity in Paleontology. Journal of Paleontology 53:1213–1227.

Ritchie, K. B., and G. W. Smith. 1998. Type II white-band disease. Revista De Biologia Tropical 46:199–203.

Robinson, M. D., D. J. McCarthy, and G. K. Smyth. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26:139–140.

Robinson, M. D., and A. Oshlack. 2010. A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biology 11:R25.

Rohart, F., B. Gautier, A. Singh, and K.-A. L. Cao. 2017. mixOmics: An R package for ‘omics feature selection and multiple data integration. PLOS Computational Biology 13:e1005752.

Schliep, K., and K. Hechenbichler. 2016. kknn: Weighted k-Nearest Neighbors.

Shapley, L. S. 1953. A value for n-person games. Page 409 Contributions to the Theory of Games (AM-28). Princeton University Press.

Štrumbelj, E., and I. Kononenko. 2014. Explaining prediction models and individual predictions with feature contributions. Knowledge and Information Systems 41:647–665.

Vega Thurber, R., L. D. Mydlarz, M. Brandt, D. Harvell, E. Weil, L. Raymundo, B. L. Willis, S. Langevin, A. M. Tracy, R. Littman, K. M. Kemp, P. Dawkins, K. C. Prager, M. Garren, and J. Lamb. 2020. Deciphering Coral Disease Dynamics: Integrating Host, Microbiome, and the Changing Environment. Frontiers in Ecology and Evolution 8.

Westfall, P. H. 1997. Multiple Testing of General Contrasts Using Logical Constraints and Correlations. Journal of the American Statistical Association 92:299–306.

Wright, M. N., and A. Ziegler. 2017. ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. Journal of Statistical Software 77:1–17.

Yeo, I.-K., and R. A. Johnson. 2000. A new family of power transformations to improve normality or symmetry. Biometrika 87:954–959.

# Tables

Table 1: Model quality metrics showing the overall metric used to identify equivilent models and the components of the overall metric, accuracy, area under the receiver operator curve (ROC/AUC), and Brier score. Finally the probability that each model is practically equivilent (within 1%) of the best overall model. Numbers indicate the mean (± SE) evaluated on the training data. Numbers in parentheses indicate the value when assessed on the test set.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Algorithm | Overall | Accuracy | ROC/AUC | Brier Score | Practical Equivalence |
| MLP | 97.9% ± 0.3% (98.9%) | 98.0% ± 0.3% (99.0%) | 0.995 ± 0.001 (1) | 0.017 ± 0.002 (0.01) | 100.0% |
| SVM | 97.6% ± 0.2% (98.8%) | 96.7% ± 0.3% (98.1%) | 0.996 ± 0.001 (1) | 0.015 ± 0.002 (0.007) | 99.2% |
| LASSO | 97.4% ± 0.3% (97.4%) | 97.4% ± 0.3% (97.1%) | 0.997 ± 0.001 (0.998) | 0.022 ± 0.002 (0.021) | 97.7% |
| Random Forest | 97.2% ± 0.2% (97.8%) | 97.7% ± 0.3% (98.1%) | 0.994 ± 0.001 (0.999) | 0.026 ± 0.002 (0.022) | 82.5% |
| KNN | 94.3% ± 0.4% (95.3%) | 95.4% ± 0.4% (96.2%) | 0.972 ± 0.004 (0.98) | 0.044 ± 0.003 (0.038) | 0.0% |
| PLS | 92.7% ± 0.2% (93.0%) | 98.0% ± 0.2% (99.0%) | 0.997 ± 0.001 (1) | 0.094 ± 0.001 (0.096) | 0.0% |
| Null | 53.0% ± 0.1% (53.0%) | 65.4% ± 0.1% (65.4%) | 0.5 ± 0 (0.5) | 0.226 ± 0 (0.226) | 0.0% |

Table 2: The twelve ASVs consistently identified as important in distinguishing healthy and diseased *Acropora cervicornis* showing the median importance ranking of each ASV across equivilant models and the FDR corrected p-value of the test that the rank is better than the average ranking of all considered ASVs. The fold-change difference in diseased and healthy corals (± SE) for each field sampling timepoint and each of the three tank exposure timepoints, prior to exposure, after disease slurry exposure, and after healthy slurry exposure. Bold fold-change values indicate they are significantly different from 0.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Family | Genus | ID | Median Rank | FDR | 2016 | | 2017 | | Before | After | |
| Jan | July | Jan | July | Disease | Healthy |
| Arcobacteraceae | Malaciobacter | ASV7 | 4.5 | 0.022 | **4.6 ± 0.66** | **7.9 ± 0.68** | **6.9 ± 0.69** | **4.7 ± 0.63** | 0.1 ± 0.46 | 0 ± 0.37 | -0.5 ± 0.37 |
| Rhodobacteraceae |  | ASV15 | 16.5 | 0.028 | 1 ± 0.54 | **4.6 ± 0.56** | **5.2 ± 0.56** | **1.1 ± 0.51** | 0.1 ± 0.62 | -0.2 ± 0.47 | 0.2 ± 0.48 |
| Sphingomonadaceae |  | ASV40 | 11 | 0.022 | **-4.3 ± 0.45** | **-3.1 ± 0.47** | **-1.6 ± 0.47** | **-4.3 ± 0.43** | **2.7 ± 0.98** | 0.5 ± 0.8 | -0.9 ± 0.81 |
| Colwelliaceae | Thalassotalea | ASV30 | 6 | 0.022 | **3.9 ± 0.54** | **4.6 ± 0.56** | **3.7 ± 0.56** | **4.6 ± 0.51** | -0.4 ± 1.02 | 1.2 ± 0.76 | -0.1 ± 0.78 |
| Pseudoalteromonadaceae | Algicola | ASV49 | 13.5 | 0.028 | **2.4 ± 0.4** | 0 ± 0.41 | **1.5 ± 0.41** | **1.3 ± 0.38** | 0 ± 0.49 | 0.5 ± 0.36 | 0.3 ± 0.37 |
| Francisellaceae | [Caedibacter] taeniospiralis group | ASV25 | 1 | 0.022 | **6.7 ± 0.51** | **3.9 ± 0.52** | **5.7 ± 0.52** | **4.7 ± 0.48** | -0.4 ± 1.08 | **3.3 ± 0.81** | 0.3 ± 0.83 |
| Endozoicomonadaceae | Endozoicomonas | ASV108 | 9 | 0.022 | **1.8 ± 0.39** | **3.4 ± 0.4** | **2.8 ± 0.41** | **3 ± 0.37** | 0.5 ± 0.31 | 0.2 ± 0.24 | 0.1 ± 0.24 |
| ASV700 | 10 | 0.022 | **2.1 ± 0.26** | 0.4 ± 0.27 | **1.3 ± 0.27** | **0.9 ± 0.25** | 0.2 ± 0.2 | 0.2 ± 0.15 | 0 ± 0.15 |
| ASV322 | 15.5 | 0.028 | **2.5 ± 0.31** | **2 ± 0.32** | **1.4 ± 0.32** | **2 ± 0.29** | **0.9 ± 0.37** | 0.2 ± 0.29 | -0.2 ± 0.29 |
| ASV361 | 18 | 0.028 | **1.7 ± 0.32** | **2.7 ± 0.33** | **1.5 ± 0.34** | **2.3 ± 0.31** | 0 ± 0.17 | **0.4 ± 0.13** | 0 ± 0.13 |
| Saccharospirillaceae | Thalassolituus | ASV26 | 4.5 | 0.022 | **4 ± 0.51** | **5.8 ± 0.53** | **3 ± 0.53** | **5.2 ± 0.48** | -0.2 ± 0.86 | **1.5 ± 0.64** | -0.3 ± 0.66 |
| Vibrionaceae | Aliivibrio | ASV8 | 3 | 0.022 | **4.9 ± 0.6** | **9.6 ± 0.62** | **5.7 ± 0.62** | **7.2 ± 0.57** | **3.3 ± 1.43** | **2.4 ± 1.07** | 0.4 ± 1.1 |

**Figure Legends**

Figure 1: Mean white band disease prevalence in *Acropora cervicornis* colonies across sampling timepoints. Letters indicate significant groupings. Error bars mark the 95% confidence intervals.

Figure 2: NMDS ordination of microbial communities of *Acropora cervicornis* indicating differences between (A) healthy and diseased colonies, (B) colonies from spatially distinct locations, (C) colonies sampled in different years, and (D) colonies sampled in different seasons.

Figure 3: Overall model quality (A) evaluated using the training (triangle) and testing (circle) datasets. Probability of model practical equvilance (within 1% overall quality of best model) showing cutoff of models considered equivalent in this study (B).

Figure 4: SHAP importance of each of the ASVs consistently in the top ranks distingishing healthy and diseased *Acropora cervicornis* colonies across all equivilent model types.

Figure 5: Each panel shows the fold-change difference between diseased and healthy coral from the field sampling timepoints and each of the three tank exposure timepoints, prior to exposure, after disease slurry exposure, and after healthy slurry exposure. Positive values show greater abundance in diseased samples while negative values show greater abundance in healthy samples. Error bars indicate the 95% confidence interval.

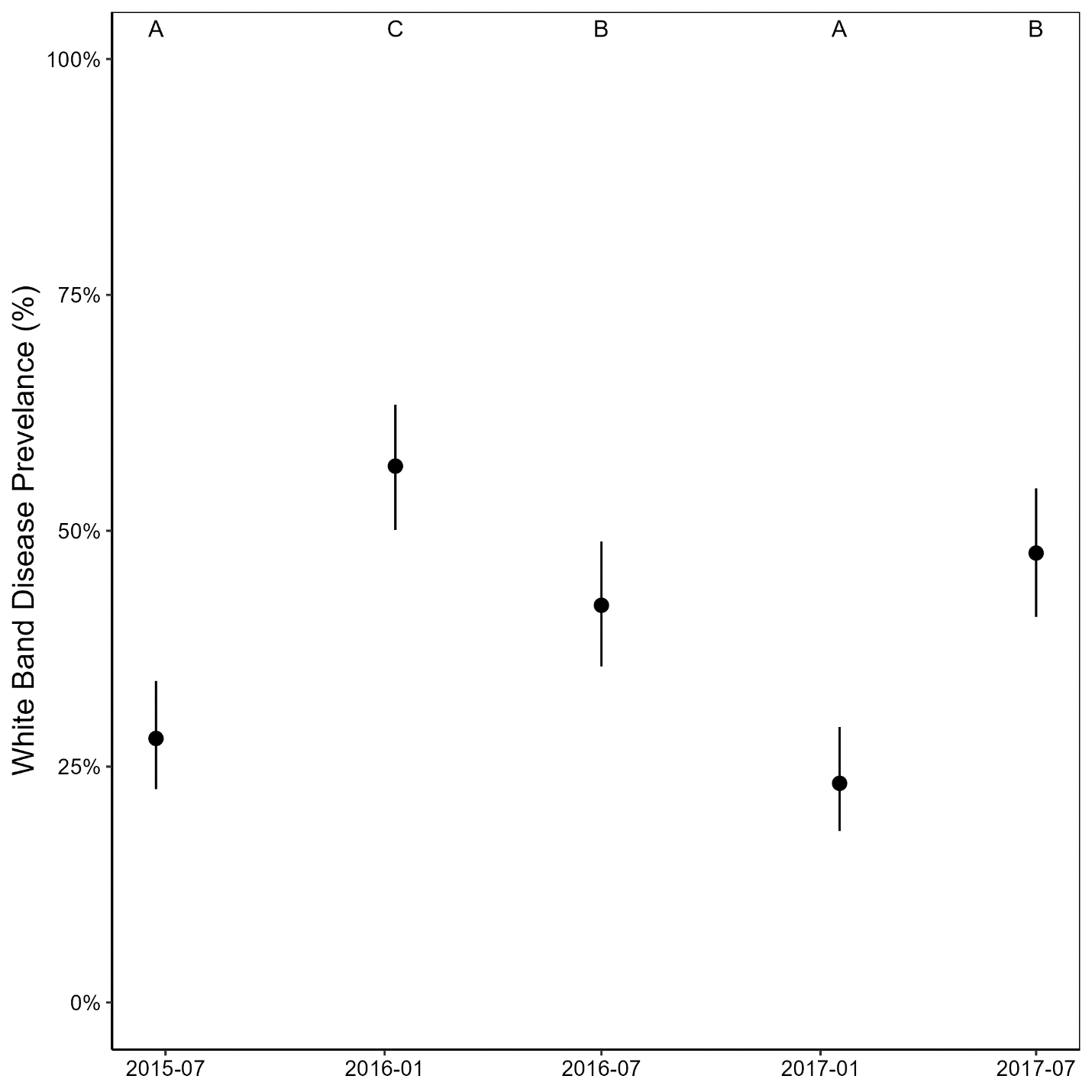


Figure 1:

A diagram of a brain

Description automatically generated with medium confidence

Figure 2:

A graph of a line

Description automatically generated with medium confidence

Figure 3:

A graph of different colored lines

Description automatically generated

Figure 4:

A screenshot of a graph

Description automatically generated

Figure :