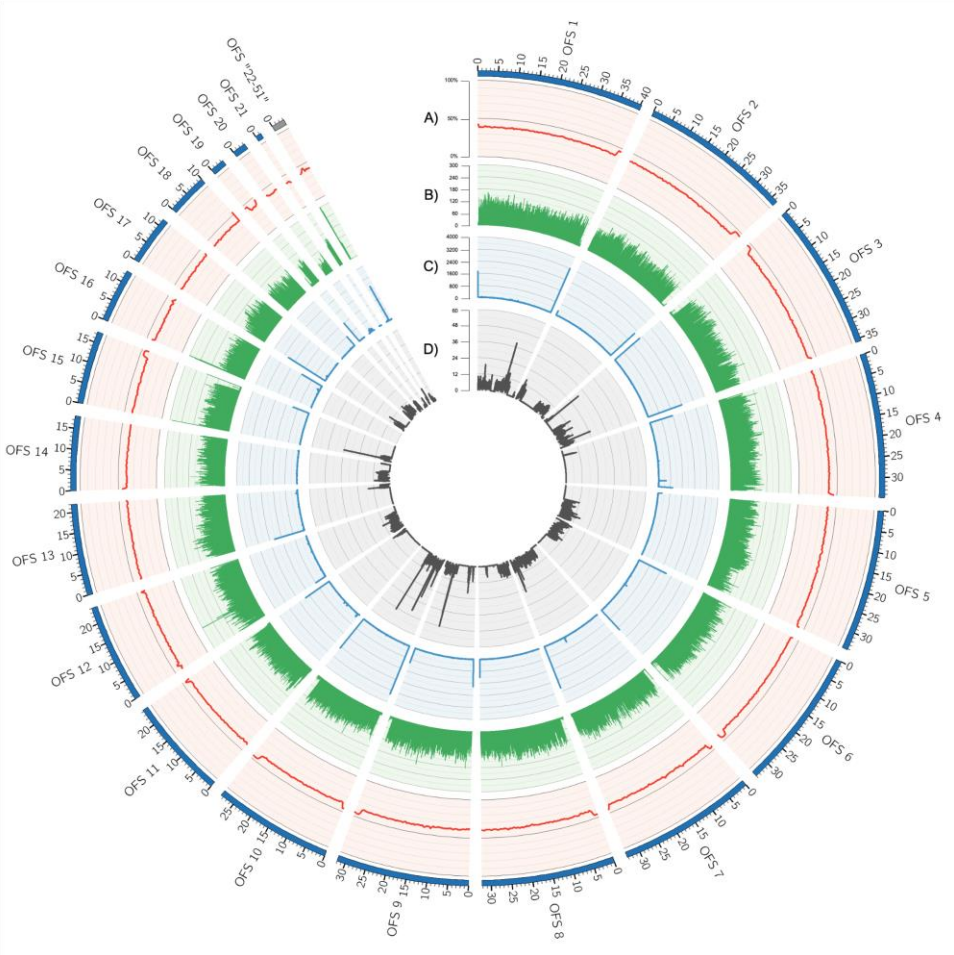
Multi-‘omic analysis of stony coral tissue loss disease resistance in restoration genotypes of *Orbicella faveolata*



FLORIDA'S CORAL REEF

Multi-‘omic analysis of stony coral tissue loss disease resistance in restoration genotypes of *Orbicella faveolata*

Final Report

Prepared By:

Michael Studivan[1,2], Erinn Muller[3], Stephanie Rosales[1,2], Sara Williams[3], Benjamin Young[1,2], Nicholas MacKnight[1,2], Stephanie Sirotzke[3]

[1]Cooperative Institute for Marine & Atmospheric Studies
University of Miami Rosenstiel School of Marine, Atmospheric, & Earth Science

[2]Atlantic Oceanographic & Meteorological Laboratory
National Oceanic & Atmospheric Administration

[3]Mote Marine Laboratory

June 14, 2024

Completed in Fulfillment of C1FB43 for

Florida Department of Environmental Protection
Coral Protection and Restoration Program
8000 N Ocean Dr.
Dania Beach, FL 33004

This report should be cited as follows:
Studivan MS, Muller E, Rosales SM, Williams S, Young BD, Sirotzke S, MacKnight N. 2024. Multi-‘omic analysis of stony coral tissue loss disease resistance in restoration genotypes of *Orbicella faveolata.* Florida Department of Environmental Protection. Dania Beach, FL. 55 pp.

This report was funded through a contract agreement from the Florida Department of Environmental Protection’s (DEP) Coral Protection and Restoration Program. The views, statements, findings, conclusions, and recommendations expressed herein are those of the author(s) and do not necessarily reflect the views of the State of Florida or any of its subagencies.

Acknowledgements

We thank Nikki Traylor-Knowles, Natalia Andrade, and Lys Isma for their productive collaboration over the last four years on the transcriptomic analysis on this project. We thank Mote's Florida Keys Marine Field Operations Program and Coral Reef Restoration Program staff for contributing outplant survival data. In particular, we acknowledge Sarah Hamlyn, Joseph Mandara, Zachary Craig, Erich Bartels, & Cory Walter for their collaboration and data on the targeted outplant disease observation study. We thank Nick Kron and Nicolas Locatelli for guidance in development of genome and transcriptome assembly pipelines. This project would not have been possible without the unwavering support of Sam Cook, Tori Barker, and Britney Swiniuch at FDEP, as well as the CPR program for funding this work.

Management Summary

There is a need for innovative research into the resistance and susceptibility of corals to stony coral tissue loss disease (SCTLD), particularly at the molecular level, to support conservation of impacted coral reef populations as well as active restoration of degraded populations. Here, we investigated SCTLD resistance in Mote Marine Laboratory's nursery genotypes of the endangered coral *Orbicella faveolata* through disease transmission experiments, while developing quantitative metrics to assess susceptibility, genetic structure, and microbial bioindicators of disease resistance. By conducting disease challenges across 170 putative genotypes of *O. faveolata* in the largest *ex-situ* coral disease transmission experiment to date, we identified five susceptibility categories that accurately predict disease outcomes in field settings. We also developed the most complete genome/transcriptome assemblies of any Caribbean coral species to date, and incorporated high-resolution genotyping of these experimental genotypes and all additional nursery *O. faveolata*, providing critical information to maximize genetic diversity in propagation and outplanting efforts. While we did not find a genetic basis for SCTLD resistance in this species, we identified microbial taxa in resistant genotypes with a co-evolved relationship to the coral host. These taxa are a priority for development as bioindicators of SCTLD resistance in order to screen additional coral populations, as well as in further investigation in their role during SCTLD pathogenesis. These 'omics datasets have been made available as public resources through the National Center for Biotechnology Information (NCBI) and the Disturbance Advisory Committee (DAC) to improve future 'omics investigations into coral resilience. The outcomes of this project have been incorporated into an ongoing coral disease response effort which seeks to improve understanding about the scale and severity of the coral disease outbreak on Florida's Coral Reef, identify primary and secondary causes, identify management actions to remediate disease impacts, restore affected resources, and ultimately prevent future outbreaks.

Executive Summary

Stony coral tissue loss disease (SCTLD) has devastated Florida's Coral Reef since 2014, affecting many endangered coral species and particularly *Orbicella faveolata*. While there exists anecdotal evidence of disease resistance in *O. faveolata* populations, no study to date has quantitatively assessed the potential for certain genotypes to survive the SCTLD epidemic. Previous efforts have focused on field experiments, which cannot standardize disease exposure potential and often have covarying impacts of environmental variability and non-disease-associated mortality. With previous support from FDEP (CPR C2002; Muller et al., 2023), a collaborative team from Mote Marine Laboratory, University of Miami Rosenstiel School, and NOAA's Atlantic Oceanographic and Meteorological Laboratory (AOML) conducted the largest coral disease transmission study to date (170 putative genotypes, 345 total fragments, 38% with >2 replicates) using *O. faveolata* genotypes from Mote's land-based nursery. This study also prioritized sampling of corals at multiple time points, including initial, pre-exposure, early exposure, initial lesion signs, and >10% tissue mortality to better understand disease responses and progression using multi-'omic analyses. A total of 2,565 'omics samples were collected for population genomics, microbial genomics, transcriptomics, metabolomics, and histological analyses. This comprehensive sampling approach facilitated the greatest possible examination of molecular responses for any coral disease to date. In this current project, we analyzed these multi-'omic datasets, specifically to address the following goals: to 1) screen nursery-reared *O. faveolata* genotypes for SCTLD resistance profiles using updated genome and transcriptome assemblies, 2) evaluate the natural evolutionary adaptation of *O. faveolata* microbial communities to SCTLD resistance, and 3) develop a SCTLD susceptibility hierarchy of restoration genotypes combining transmission and genetic datasets. To this end, we developed the most complete genome and transcriptome assemblies of any Caribbean coral species using PacBio circular consensus (CCS) long-read and ISO-seq sequencing approaches, respectively. We then conducted high-resolution genotyping of 174 putative genotypes from the transmission experiments using whole-genome sequencing (WGS), as well as 173 additional genotypes from Mote's nurseries using restriction-site associated DNA sequencing (2bRAD). By screening over 4.5 million single-nucleotide polymorphisms (SNPs) throughout the *O. faveolata* genome, we did not find evidence of a genomic basis for SCTLD resistance. Through microbial community profiling (16S rRNA) of 1,652 samples collected during the transmission experiments, however, we found one amplicon sequence variant (ASV; *Rhodospirillales*) to exhibit signs of co-evolution with resistant genotypes of *O. faveolata*. This bacterial taxa represents a bioindicator to screen additional genotypes of this species, and perhaps other coral species affected by SCTLD. Finally, through hierarchical clustering, we classified SCTLD resistance of *O. faveolata* into five categories based on results of the lab-based transmission experiments, which translate to field-based observational data. Taken together, this study identified priority *O. faveolata* genotypes for targeted propagation and outplanting of SCTLD-resistant individuals, and provided critical multi-'omic tools, datasets, and analytical pipelines to evaluate disease resistance in additional populations. With these resources, Mote Marine Laboratory and restoration partners can proceed with

large-scale production of *O. faveolata* genotypes for outplanting in the face of ongoing and future SCTLD outbreaks.

Table of Contents

1. BACKGROUND

Florida's Coral Reef is the only barrier reef system in the continental United States, and is the third largest barrier reef system in the world. It is critical to the state's coastal marine ecosystems, supports >70,000 jobs, provides >$8 billion to the state's economy, and serves as the primary coastal defense from major storms. Florida's coral populations have declined substantially over the past four decades, leading to habitat degradation and loss of economic and coastal resiliency benefits (Lane et al., 2013; Micheli et al., 2014; Storlazzi et al., 2021). Multiple stressors are to blame, including thermal-stress associated bleaching (Hoegh-Guldberg et al., 2023), disease outbreaks (Jackson et al., 2014; Walton et al., 2018), coastal development and habitat destruction (Enochs et al., 2023), and ocean acidification (Cornwall et al., 2021; Hoegh-Guldberg et al., 2017; Palacio-Castro et al., 2023). There is now little time for recovery to take place before the next disturbance event(s) (Hoegh-Guldberg et al., 2023). Adaptive management strategies are paramount to ensure the success of coral reef conservation and restoration initiatives, particularly in the face of multiple co-occurring environmental stressors.

There is therefore a dire need for rapid human intervention in the form of large-scale coral restoration efforts to mitigate further losses of critical coral populations. One such stressor substantially impacting the success of coral restoration in Florida is the stony coral tissue loss disease (SCTLD) outbreak, which has continued largely unabated since its first observation off Miami, Florida in 2014. This disease event has likely killed millions of coral colonies across nearly half of reef-building species in Florida alone (Hawthorn et al., 2024; Muller et al., 2020; NOAA, 2018), making its impacts unprecedented relative to other coral diseases. Despite perhaps the world's largest collaboration of coral reef researchers, restoration practitioners, and managers related to a coral disease outbreak response, the pathogen(s) has not yet been identified. Evidence suggests that SCTLD may be transmitted by direct coral-coral contact, through the water column, or potentially through ship's ballast water transfers (Aeby et al., 2019; Dobbelaere et al., 2020; Muller et al., 2020; Studivan, Baptist, et al., 2022; Studivan, Rossin, et al., 2022), which makes it difficult to contain. To date, SCTLD has now spread to many regions throughout the Caribbean (Kramer et al., 2019).

At present, it is largely unknown what impacts co-occurring stressors will have on restoration efforts, particularly in reference to disease-associated mortality. Many restoration groups are actively seeking to identify and propagate resilient genotypes as they pose the highest rate of success following outplanting, however, the process to characterize resilience in corals has posed difficulties. Disease resistance, in light of the SCTLD outbreak in Florida, represents a unique challenge This project culminates four years of highly collaborative and innovative research to significantly advance our ability to restore populations of *Orbicella faveolata* to Florida's Coral Reef using novel research and restoration techniques to increase the resilience of this critical population of major reef-building corals. By combining advanced multi-'omic characterization of coral and algal genetics and bacterial communities with disease susceptibility monitoring, we can identify

high-priority *O. faveolata* genotypes for targeted propagation and outplanting, and compare lab-based transmission assays with field-based monitoring observations.

2. METHODS

2.1. Conduct genotyping and genome-wide association study (GWAS)

This task focused on using next-generation sequencing approaches to provide genotypic data on both corals used in the 2022 SCTLD transmission experiments, as well as additional corals in Mote's land-based nurseries. This accomplished two goals: 1) to confirm identity metadata of all nursery genotypes in Mote's restoration pipeline, therefore aiding in maintenance of the highest possible genetic diversity with propagation and outplanting efforts, and 2) to screen for markers of SCTLD resistance across the species' genome. Previous attempts to accomplish the latter goal in a field-based study were unsuccessful (Klein et al., 2024), and the authors hypothesized that the lack of genetic resolution was due to a reduced-representation approach (2bRAD; (Klein et al., 2024) that sequenced a relatively small portion (10–20%) of the coral's genome to identify single-nucleotide variations called single nucleotide polymorphisms (SNPs) that vary among individuals. Learning from this, we sought to use whole-genome sequencing (WGS) as a means to sequence the entire genome of our coral samples. While WGS provides a complete picture of the genome, it is an order of magnitude more expensive and computing-intensive, so we prioritized experimental samples used in the transmission experiments. For the remaining nursery samples that were not used in the transmission experiments, we opted for the 2bRAD approach to maximize comparability among our dataset and previous studies. 2bRAD is comparatively much cheaper and captures a fraction of the genome, but still allows the identification of SNPs for genotyping and genetic relatedness purposes. This dual-sequencing strategy was used to maximize the cost-benefit analysis of data generation for high-value samples (experimental genotypes), while also providing sequencing data of all of Mote's restoration genotypes.

The WGS data were then further analyzed to identify genomic signatures of disease resistance with comparisons to phenotypic (resistance/susceptibility) data using genome-wide association study (GWAS), which finds SNPs with significant associations to quantitative disease resistance metrics. From this, we aimed to glean a list of SNPs that are indicative of SCTLD resistance (or susceptibility), which could then be used to screen additional coral populations without needing to run SCTLD transmission experiments, such as the remaining Mote nursery genotypes, the genotypes from Klein and colleagues (2024), or wild populations of *O. faveolata*.

In order to have the best chances of screening for disease resistance in this species, however, the genome reference used for sequence alignment and gene identification needed to be improved. Prior to this project, the most recent genome reference for *O. faveolata* was released in 2016 (Prada et al., 2016). Both sequencing technologies and annotation pipelines have greatly improved since then, and therefore the 2016 genome represents an incomplete reference. In particular, PacBio long-read and ISO-seq

sequencing (https://github.com/PacificBiosciences/IsoSeq) approaches have revolutionized the ability to produce highly contiguous genome and transcriptome reference assemblies, both at a relatively low cost (<$10k). Not only do these genome/transcriptome references support this project, but they also provide an excellent resource for all future studies assessing molecular mechanisms of resilience with *O. faveolata*, such as genotyping, genome screening, and transcriptomic analyses.

2.1.1. Genome and transcriptome assemblies

Genome and transcriptome sequences were obtained as described in the final report for CPR C2002 (Muller et al., 2023), and were processed using the bioinformatics scripts available in an associated GitHub repository (https://github.com/benyoung93/orbicella_faveolata_pacbio_genome_transcriptome). The long-read PacBio *O. faveolata* genome was assembled using HifiASM (Cheng et al., 2021) and scaffolded using Longstitch (Coombe et al., 2021). Genome completeness was assessed using BUSCO (Manni et al., 2021) and the metazoa_odb10 database, and Quast (Mikheenko et al., 2016, 2018) was used to evaluate genome metrics such as N50, GC content, longest contig, and L50. Gene prediction, annotation, and reference transcriptome generation of the scaffolded assembly was done using Funnannotate (Palmer & Stajich, 2020) and utilized high-quality transcripts generated from ISO-seq sequencing and previously available short-read RNA-seq reads from *O. faveolata* (MacKnight et al., 2022). To assess completion of the reference transcriptome and predicted genes, BUSCO was again used (database metazoa_odb10), as well as an OrthoFinder (Emms & Kelly, 2019) analysis to identify single-copy genes and orthogroups between other stony coral species with available long-read genome resources.

2.1.2. Whole genome sequencing (WGS)

WGS sequences were obtained as described in the final report for CPR C2002 (Muller et al., 2023), and were processed using the bioinformatics scripts available in an associated GitHub repository (https://github.com/mstudiva/SCTLD-resistance-genomics). Raw sequences were evaluated for read quality using FastQC (Andrews, 2010), and low-quality reads were removed using Trim-galore (Krueger, 2016/2024), a wrapper of cutadapt (Martin, 2011). Trimmed reads were then mapped to the *O. faveolata* reference genome (see section 2.1.1.) using bowtie2 (Langmead & Salzberg, 2012), which splits the files into aligned reads for genotyping, and unaligned reads for alignment to symbiont genomes. The latter reads were aligned to a concatenated Symbiodiniaceae reference (https://github.com/RyanEckert/Stephanocoenia_FKNMS_PopGen), where aligned reads were used for downstream symbiont typing (see section 2.1.4.), while unaligned reads were discarded.

Coral host-aligned reads were then processed using the Genome Analysis ToolKit (GATK) v4.0 (Auwera & O'Connor, 2020) following their best practices guide. The full bioinformatics script can be found in the associated GitHub repository (https://github.com/mstudiva/SCTLD-resistance-genomics), but in brief: 1) individual

variant (SNP) calls were first conducted using the function HaplotypeCaller, 2) resulting individual variant call format (vcf) files were combined into a genomics database using the function GenomicsDBImport, 3) then joint genotyping (comparisons among individual variant calls) was conducted using the function GenotypeGVCFs. Raw global SNPs were then hard-filtered based on GATK best practices, as 'truth' and 'training' datasets were not available for this species to use Variant Quality Score Recalibration (VQSR). True SNPs (excluding insertions/deletions [indels]) were isolated for all samples using the function SelectVariants, then the following filtering steps were applied using the function VariantFiltration: QUAL<30.0, QD<2.0, FS>60.0, SOR>3.0, MQ<40.0, MQRankSum<-12.5, ReadPosRankSum<-8.0. Quality scores (QUAL) measure the quality of each variant call. QualByDepth (QD) is a metric of the confidence that a particular variant locus is high-quality (i.e., real, and not a false positive); two peaks at QD values of ~17 and ~30 represent homozygous and heterozygous variants, respectively. FisherStrand (FS) is the Phred-scaled probability that there is strand bias at the site, and StrandOddsRatio (SOR) is another way to estimate strand bias. RMSMappingQuality (MQ) represents the mapping quality of the locus to the genome. MappingQualityRankSumTest (MQRankSum) is a z-score approximation from the Rank Sum Test to assess mapping qualities, and ReadPosRankSumTest (ReadPosRankSum) is a z-score approximation from the Rank Sum Test for site position within reads. The latter identifies whether seeing an allele only near the ends of reads is indicative of error, since sequencing platforms tend to make more errors at the end of reads. SNPs passing/failing these quality thresholds were visualized using the R statistical environment (R Core Team, 2020). SNPs passing these quality filters were then exported for downstream analyses using the function SelectVariants.

High-quality SNP calls from WGS samples were analyzed using custom R scripts (https://github.com/mstudiva/SCTLD-resistance-genomics) to visualize genetic relatedness as a genetic distance matrix and dendrograms with the packages poppr (Kamvar et al., 2014) and ggdendro (Vries, 2011/2024), respectively. Pairwise genetic distance values were compared for sequenced duplicates to determine an appropriate threshold for recognizing clonal genotypes. The custom clonal threshold was used to identify multi-locus genotype identities ('true' distinct genotypes) for all samples. Multi-locus genotype assignments were then compared to Mote's provenance metadata for validation and QAQC purposes (see section 2.3. below).

Clonal genotypes were then removed using the bioinformatics package vcftools (Danecek et al., 2011) so that only one representative from each multi-locus genotype was present in the dataset, and the dendrogram was re-plotted with the susceptibility hierarchy metrics (see section 2.3. below) as a color overlay to identify any potential relationships between genetic identity and genotype-specific susceptibility. Additional dendrograms with color overlays corresponding to sampling location, sample type (e.g., sexual recruits versus corals of opportunity), and dominant algal symbiont genus were also produced to explore potential associations with provenance metadata.

Prior to screening for genome-susceptibility associations, SNPs were then further filtered using the bioinformatics tool PLINK (Purcell et al., 2007) to produce a set of loci for use

in GWAS with the R package LEA (Frichot & François, 2015). The additional filtration steps were necessary to 1) remove SNPs with missingness >0.2, 2) remove SNPs with a minor allele frequency (MAF) <0.05, 3) remove SNPs which were not in Hardy Weinberg Equilibrium (HWE), 4) and remove heterozygosity rate outliers; otherwise, the GWAS would be negatively skewed. Similarly, structure analysis and principal component analysis (PCA) were performed in LEA to identify sample groupings due to ancestral populations, as ancestry can influence genotype-phenotype associations. Three iterations of GWAS were performed, incorporating genetic structure information, to determine whether regions of the *O. faveolata* genome (SNP loci) corresponded with the quantitative metrics: proportion of healthy replicates per genotype (resistance), proportion of diseased replicates per genotype (susceptibility), and proportion of *Durusdinium* abundance (*Durusdinium*).

2.1.3. Restriction-site associated DNA sequencing (2bRAD)

2bRAD samples were processed according to the bioinformatics pipeline described in the associated GitHub repository (https://github.com/mstudiva/SCTLD-resistance-genomics) using custom scripts found in another GitHub repository (https://github.com/RyanEckert/Stephanocoenia_FKNMS_PopGen). Raw sequence files were first deduplicated based on dual-index, Illumina-standard barcode primers that were incorporated during library preparation (CPR C2002; Muller et al., 2023). Following deduplication of pooled libraries into individual sample files, reads were trimmed using cutadapt. All remaining bioinformatics steps, including genotyping with GATK, were conducted in the same manner as described for the WGS samples in section 2.1.2.

High-quality SNP calls from 2bRAD samples were analyzed as described previously to visualize genetic relatedness with a genetic distance matrix and dendrogram. Pairwise genetic distance values for sequenced duplicates once again determined the threshold for clonal genotype detection. Multi-locus genotype assignments were compared to Mote's provenance metadata for validation and QAQC purposes (see section 2.3. below). Finally, following removal of clonal genotypes, the dendrogram was re-plotted to visualize genetic relatedness in Mote's nursery genotypes.

2.1.4. Algal symbiont communities

WGS and 2bRAD sequencing data from coral samples naturally included gene sequences from algal symbionts (Symbiodiniaceae). During the alignment of reads to the coral host genome, unaligned reads were separated and then aligned to a reference genome assembly containing sequences from the four main Caribbean coral-associated Symbiodiniaceae genera: *Symbiodinium*, *Breviolum*, *Cladocopium*, and *Durusdinium* (see section 2.1.2. above). The bioinformatics package samtools (Danecek et al., 2021) was used to count the number of sequence alignments to each of the symbiont references in the concatenated genome, which was then used to calculate the proportion of the four symbiont genera in each sample across both WGS and 2bRAD datasets. Symbiont assemblages were

visualized using the R package ggplot2 (Wickham, 2016) in a custom R script (https://github.com/mstudiva/SCTLD-resistance-genomics).

2.2. Identify microbial signatures of SCTLD resistance across a susceptibility gradient

2.2.1. Differential abundance and microbial diversity

For this task, we aimed to understand how the SCTLD microbiome changed through time and how the microbiome played a role in SCTLD disease resistance from samples obtained during the 2022 transmission experiments (Muller et al., 2023). A total of 1,652 16S Samples were processed through a Qiime2 pipeline (Bolyen et al., 2019), which can be found in detail (https://github.com/nmacknight/Ofav.16s.SCTLD). Samples with less than the lower quartile read depth (less than 5,141 reads) were removed (442 samples) to retain the maximum number of samples while observing the highest number of unique amplicon sequence variants (ASVs). Post-quality control, microbial diversity analysis was performed to identify microbial community differences among coral health states. First, the changes in SCTLD at four different time points (pre-exposure, early exposure, initial disease, and final) were characterized. Pre-exposure included corals that had not yet been exposed to a coral with SCTLD lesions. Early exposure contained corals that had been exposed to a coral with SCTLD lesions for 2 days. Initial disease time points described corals that developed a disease phenotype. The final time point was when the coral was pulled from the experiment due to >10% tissue loss.

To do this, in R we used the diversity function from the vegan package (Oksanen et al., 2013) to calculate Shannon diversity by using raw read counts, while Pielou's evenness, Simpson diversity, and beta diversity was calculated using proportionally normalized count data. Shannon diversity was calculated with raw microbial read counts to reflect unfiltered species richness and evenness. In contrast, Pielou's evenness, Simpson diversity, and beta diversity was calculated with normalized counts to ensure accurate proportion-based metrics and fair comparisons across samples, minimizing the impact of varying sequencing depths. Beta diversity was calculated using the package betapart (Baselga & Orme, 2012), which partitions beta diversity into turnover and nestedness components. Alpha and beta diversity data were not normally distributed, so comparisons among disease outcomes were tested using non-parametric Kruskal-Wallis tests. To visualize ASV and order changes from the four different time points, taxa above 3% relative abundance were visualized using 100% stacked bar graphs. To identify changes of SCTLD through time, only disease-exposed samples were analyzed. These samples were also ordinated using a supervised linear discriminant analysis (LDA) using the package MASS (Venables & Ripley, 2013) for time points in SCTLD-exposed samples and overlaid with vectors of bacteria that had the highest correlations with axes 1 and 2. ANCOM-BC2 (Lin & Peddada, 2024) was used to test microbial differential abundance between pre-exposure and the final time point in SCTLD-exposed samples at two taxonomic levels (ASV and order), and also to compare genotype susceptibility classes (see section 2.3.1.) in the healthy control state. Input data included raw read counts, and multiple comparisons were corrected with a Bonferroni test.

2.2.2. Environmentally introduced and co-evolved microbial signatures of resilience

Next, we examined if microbial taxa were associated with disease resistance. To do this, the expression variance and evolution model (EVE; Rohlfs & Nielsen, 2015) was used to identify taxa that have evolved with resilient and susceptible genotypes. This novel application in coral disease research has only recently been applied (Avila-Magaña et al., 2021; MacKnight et al., 2022). The microbial abundance and genotype tree (see section section 2.1.2.) were input into the EVE model using the packages evemodel (Grønvold, 2021) and ape (Paradis & Schliep, 2019). The EVE model links the microbial abundance data to the coral host's phylogenetic position in the tree. The beta shared test (i.e., phylogenetic ANOVA) can detect bacteria with increased or decreased ratios of abundance divergence to diversity, represented as the beta parameter. EVE can be used for identifying bacteria with high abundance divergence between genotypes as candidates for abundance-level adaptation, and bacteria with high abundance diversity within genotypes as candidates for abundance-level plasticity. This works by using an Ornstein-Ulhenbeck process of optimization to identify an ancestrally optimal abundance value for each bacterium where variance from this optimum is represented by beta. The log-likelihood ratio between the individual and shared beta fit indicates whether the individual beta was a better fit (i.e., the bacteria has an increased or decreased ratio of abundance divergence to diversity). Significant deviations of beta from the optimal abundance value were determined through the log likelihood ratio test statistic which follows a chi-squared distribution with one degree of freedom. Multiple comparisons were corrected using a false discovery rate (FDR).

2.3. Determine susceptibility hierarchy and transmission risk among restoration *O. faveolata* genotypes

2.3.1. Transmission experiment susceptibility

During the four transmission studies conducted in 2022 (CPR C2002; Muller et al., 2023), 170 putative genotypes of *O. faveolata* were sampled and tested for resistance to SCTLD. We used a hierarchical Bayesian framework to quantify the relative risk of SCTLD analysis for the four transmission experiments, as described in an associated GitHub repository (https://github.com/saradwms/DEP_OFAV_2023). This analysis closely followed Muller et al. (2018) that used a binomial likelihood distribution and a uniform-Beta prior distribution (additional relative risk model details described below). Samples from additional Mote Restoration *O. faveolata* genotypes were also collected as part of the CPR C2002 project. Background information on Mote's Restoration *O. faveolata* genotypes (including provenance information), the whole-genome sequencing (WGS) multi-locus genotype assignments, and the results of the transmission experiments were quality-checked and combined to inform the SCTLD susceptibility analyses described below. The resulting metadata contained 188 genotypes, 180 of which had WGS-assigned genotypes and 8 did not have WGS data due to library preparation failures. 154 of those genotypes were exposed within the transmission experiments conducted in 2022, of which 53 had 3 or more replicates exposed to SCTLD.

Three disease metrics, each on a 0 (highly susceptible) to 1 (resistant) scale, were quantified for the 154 WGS-genotypes exposed: resistance, progression, and transmission. Resistance was quantified as the fraction of replicates exposed that remained healthy. Progression was the average number of days fragments had disease signs before being removed (>10% tissue loss) standardized by the experiment length in days. Transmission was the average number of days until disease signs occurred on fragments (initial disease date) standardized by the experiment length in days. To categorize the genotypes into susceptibility groups, we used a hierarchical clustering analysis of the disease metrics for all 154 genotypes with the package mclust (Scrucca et al., 2016). Ward's hierarchical clustering with a Gower dissimilarity matrix was performed to group the genotypes into five susceptibility categories: highly susceptible, susceptible, intermediate, resistant, and 1/1 resistant (any genotype that only had 1 exposed replicate and it remained healthy). The hierarchical clustering was repeated for the 53 genotypes that had 3 or more replicates exposed, but only 4 clusters were used: highly susceptible, susceptible, intermediate, & resistant.

2.3.2. Outplant susceptibility and survivorship

The *O. faveolata* outplant survival data from 2018-2023 was acquired and reviewed to determine what would be appropriate to include within a comparative analysis. Initial quality checks and filtering were conducted on a subset of the dataset to include only the experiment genotypes. As part of the filtering process, we identified a particular dataset that included monthly monitoring specifically focused on documenting SCTLD on outplanted *O. faveolata*. This study included monthly disease observations from September 2018 through November 2019 of 995 *O. faveolata* outplanted at nearshore, mid channel, and offshore sites near Big Pine/Summerland Key and Key West. 9 of Mote's *O. faveolata* genotypes utilized for this outplant study were among the putative genotypes exposed in the 2022 SCTLD experiments (Muller et al., 2023). These 9 genotypes clustered as 8 WGS-assigned genotypes. This outplant monitoring data set included the time when SCTLD was progressing through the Lower Florida Keys, where the outplants occurred, and so captured the susceptibility of genotypes during the invasion and epidemic periods of the outbreak. Because of the specificity associated with this outplant data and the overlap in genotypes between outplanted corals and our experiments conducted within the present study, we focused on this dataset for our targeted comparison purposes. We adapted the relative risk analysis conducted by van Woesik and Randall (2017), which used a standardized expected ratio with a negative binomial distribution for relative risk and a gamma distribution to account for variance (Poisson-Gamma Model; full analysis pipeline on GitHub: https://github.com/saradwms/DEP_OFAV_2023). Models were run using 3000 Markov chain Monte Carlo simulations in OpenBUGS implemented in R using the R2OpenBugs package (Sturtz et al., 2005) to obtain the posterior probability distributions. We then calculated 95% credible intervals for relative risk estimates. Credible intervals that did not overlap 1 were considered significant, and those that were higher than one indicated a higher disease risk. Relative risks to SCTLD in the outplant study were compared with the susceptibility group determined by the hierarchical clustering. This comparison was used to explore and potentially ground-truth levels of risk captured within the lab-based

transmission studies with field-based, 'real world' outplant scenarios. Finally, the overall survival of outplanted *O. faveolata* genotypes, taken at one-year post outplanting, was compared with the SCTLD susceptibility groupings to determine if this trait influenced overall likelihood of survival, which could translate into restoration success.

2.4. Data archival

All metadata generated from this project are available on the SCTLD DataOne portal under identifier urn:uuid:f1c6f769-e7aa-464f-a046-504782f402cd. Additionally, 'omics sequencing data and associated metadata have been uploaded to the National Center for Biotechnology Information, and analysis pipelines are publicly available through GitHub:

Raw reads from the *Orbicella faveolata* genome and ISO-seq transcriptomes are available at the NCBI under project number (PRJNA970355). Completed genome and transcriptome assemblies have also been submitted to the same NCBI project number. The mitochondrial genome identified is available at GenBank (accession number OR906199). Assembled genome and transcriptome assemblies are also openly accessible in a Zenodo repository (http://doi.org/10.5281/zenodo.10151798). The bioinformatics pipeline is accessible via a GitHub repository (https://github.com/benyoung93/orbicella_faveolata_pacbio_genome_transcriptome).

WGS and 2bRAD raw sequences and associated metadata are available at the NCBI under project number (PRJNA1123826), and bioinformatics pipelines and analysis scripts are available on GitHub (https://github.com/mstudiva/SCTLD-resistance-genomics).

16S raw sequences and associated metadata are available at the NCBI under project number (PRJNA955222), and bioinformatics pipelines and analysis scripts are available on GitHub (https://github.com/nmacknight/Ofav.16s.SCTLD).

Disease resistance and susceptibility analysis scripts are available on GitHub (https://github.com/saradwms/DEP_OFAV_2023).

3. RESULTS

3.1. Conduct genotyping and genome-wide association study (GWAS)

3.1.1. Genome and transcriptome assemblies

The *Orbicella faveolata* genome and transcriptome assemblies were published in the open-access journal *BMC Genomics* (Young et al., 2024). The newly assembled genome greatly improved on contiguity (51 scaffolds versus 1,933 scaffolds) and completion (93.6% versus 85.3%) compared to the previously assembled short-read *Orbicella faveolata* genome (Prada et al., 2016). The largest contig was 40,246,328 base pairs (bp), N50 (the sequence length of the shortest contig at 50% of the total assembly length) of 33,295,526 bp, L50 (count of smallest number of contigs whose length sum makes up 50% of genome

size) of 7, and GC content (the proportion of the genome that is guanine or cytosine) of 39.49% (Figure 1). Repeat masking of the scaffolded assembly identified 50.20% (247,928,041 bp) as repetitive regions (Figure 2). Telomeric analysis identified telomeric repeats at either one (telocentric, 12 of 19 scaffolds) or both (7 of 19 scaffolds) ends of scaffolded contigs (Figure 2). BUSCO analysis of the 19 scaffolds with telomeric repeats identified a 90.2% completion. Using ISO-seq and previously available short read RNA sequencing (RNA-seq) reads, we identified 32,172 protein-coding genes and 5,762 transfer-RNAs (tRNAs), with an average read length of 5,977.66 bp. BUSCO analysis of the protein coding genes identified complete orthologs of 95.1%. To assess whether our annotation was comparable to other long read coral genomes, an orthofinder analysis was undertaken. We identified that 29,917 (93%) were within orthogroups, 2,255 (7%) were not assigned to orthogroups, 18,199 (55.7%) genes were shared between coral species, and 1,903 (5.9%) of genes were only present within single coral species (Figure 3).

3.1.2. Whole genome sequencing (WGS)

A total of 180 WGS samples (174 putative genotypes, including 6 sequenced duplicates) produced a total of 4.6 billion sequenced reads, with a mean read count of 26 million per sample. Following quality trimming and alignment to the *O. faveolata* genome (see section 3.1.1. above), an average of 16.3 million coral host-aligned reads remained, which was an average alignment rate of 62.5% (see associated DataOne project for full metadata; section 2.4.). WGS samples processed through the GATK pipeline resulted in 24,097,804 raw SNP loci. Quality filtering resulted in 10,366,466 high-quality SNPs that were used for genotype assignments of experimental samples (Figure 4). Based on 3 out of the 6 sequenced duplicates that had sufficient read counts and alignment rates, a clonal detection threshold of 0.015 (1.5%) was used on the genetic distance matrix, where any samples with a genetic dissimilarity below that value was considered a true clonal genotype. Genotype assignment resulted in 157 multi-locus genotypes (out of 174 putative genotypes sequenced; Figure 5), which corresponded to a clone rate of 12.6%. None of the multi-locus genotype assignments clashed with Mote's provenance metadata (e.g., such as if a sexual recruit matched with a wild-collected coral of opportunity as a clone-mate), and therefore analyses could proceed with the combined genotype and provenance metadata.

Following removal of clonal genotypes with the lowest genome alignment rates and highest proportion of missing SNP coverage (Figure 6), visualization of the 157 unique multi-locus genotypes did not reveal any consistent patterns between genetic relatedness and SCTLD susceptibility (see section 3.3.1.; Figure 7), sampling location (Figure 8), sample type (Figure 9), or dominant algal symbiont genus (Figure 10). In particular, SCTLD-resistant genotypes were found across all genetic clusters of samples, where there was no apparent grouping of resistant genotypes in a similar genetic cluster.

Further filtering of WGS SNPs resulted in 4,192,708 out of 10,366,466 high-quality SNPs for GWAS. Interestingly, structure analysis of the remaining SNPs predicted 5 ancestral lineages (K=5) present in the data (Figure 11), which correlated with observed genetic similarity among samples, but did not correspond to original sampling location (Figure 12).

The three GWAS models that accounted for genetic structure did not identify any significant associations between susceptibility metrics or *Durusdinium* abundance and SNPs in the *O. faveolata* genome (Figures 13-15).

3.1.3. Restriction-site associated DNA sequencing (2bRAD)

A total of 185 2bRAD samples (173 putative genotypes, with 12 sequenced duplicates/triplicates) produced a total of 311.5 million sequenced reads, with a mean read count of 1.7 million per sample. Following quality trimming and alignment to the *O. faveolata* genome (see section 3.1.1. above), an average of 1.4 million coral host-aligned reads remained per sample, corresponding to an average alignment rate of 80.7% (see associated DataOne project for full metadata; section 2.4.). 2bRAD samples processed through the GATK pipeline produced 61,812 raw SNP loci, which were then quality-filtered, resulting in 29,767 high-quality SNPs for use in genotype assignments (Figure 16). 3 out of 7 sequenced duplicates/triplicates had sufficient read counts and alignment rates, resulting in a clonal detection threshold of 0.027 (2.7%) to distinguish true clonal genotypes (Figure 17). Genotype assignment produced 151 multi-locus genotypes (out of 173 putative genotypes sequenced; Figure 18), a clone rate of 12.7%. No mismatches between multi-locus genotype assignments and Mote's provenance data were observed, however, one set of clone-mates in Mote's metadata (2bRAD-99 / OF12, 2bRAD-102 / OF27, and 2bRAD-73 / F12) were not identified as genetic clones due to low sequencing success of these samples (Figure 17).

3.1.4. Algal symbiont communities

Samples sequenced using both WGS and 2bRAD approaches were found to be largely dominated by *Durusdinium*, with background levels of *Cladocopium*, *Symbiodinium*, and *Breviolum*, in that order (Table 1; see associated DataOne project for full metadata; section 2.4.). Of the 180 WGS samples, only 3 had dominance by a genus other than *Durusdinium* (*Breviolum*: OF100, UK 36, and OF96). Of the 185 2bRAD samples, only 2 were dominated by *Cladocopium* (M22 and a sample with missing metadata), while 1 was dominated by *Breviolum* (OF629). Symbiont abundance was generally in congruence among sequenced duplicates/triplicates and between sequencing approaches (WGS versus 2bRAD), although WGS typically demonstrated higher abundance of background symbionts due to increased sequencing depth (Figure 19).

3.2. Identify microbial signatures of SCTLD resistance across a susceptibility gradient

3.2.1. Differential abundance and microbial diversity

After evaluation of the data for QC, the remaining samples were normalized to a read depth of 5,141 reads per sample with the exception of calculating Shannon diversity and ANCOM-BC2 input data which utilized raw count data (see associated DataOne project for full metadata; section 2.4.). Diversity indices initially showed an even abundance of

unique bacteria taxa that began to gradually deviate from this into a convergent microbiome in the final time point (Figure 20). Diversity metrics of the bacterial community varied significantly among treatment, disease outcome, and between time points of disease-exposed samples (Table 2). This trend aligns with relative abundance plots, where in the pre-exposure time point, 29.9% of the microbiome consisted of 7 ASVs with more than 3% relative abundance, and these ASVs increased to 48.2% of the microbiome in the final time point (Figure 21A). Twelve microbial orders of 3% relative abundance or more made up 69% of the microbiome in the pre-exposure time point, which increased to 77.7% in the final time point (Task 4 Figure 21B). Samples were grouped by disease time points (pre-exposure, early exposure, initial disease, and final) to visualize microbiome shifts in response to SCTLD exposure (LDA; Figure 22). The analysis highlights that the microbiomes in the pre-disease exposure time point were distinct compared to those at the final time point. Differential abundance resulted in 18 bacteria significantly increasing in abundance in the final time point relative to the pre-exposure time point, with a *Vibrionales* ASV having the highest log fold change (lfc) in abundance in the final time point (+3.79 lfc). The remaining 26 bacteria significantly decreased in abundance in the final time point relative to pre-SCTLD exposure, with a *Rhodospirillales* ASV having the greatest log fold change in pre-exposure samples (-1.7 lfc; Figure 23A). Notably, 5 *Vibrionales* spp. had significantly higher abundance in the final SCTLD time point relative to pre-exposure conditions. By contrast, 4 *Rhizobiales* spp. had a significant decrease in abundance in response to SCTLD transmission at the final time point. *Rhodospirillales* sp. was the only bacteria that was significantly more abundant (+2.61 lfc) in resistant genotypes compared to highly susceptible genotypes in the healthy control state (Figure 23B).

3.2.2. Environmentally introduced and co-evolved microbial signatures of resilience

From the 322 ASVs analyzed, 157 were considered lineage-specific and 14 were considered highly variable ($p<0.1$) (Figure 24; see associated DataOne project for full metadata; section 2.4.). Of the 14 highly variable bacteria, 4 were significantly differently abundant between pre-exposure and final time points (Figure 23A). The *Vibrionales* ASV with the highest log fold change was classified as highly variable by the EVE analysis, which indicates this ASV's abundance patterns, along with the other highly variable bacteria, are treatment-mediated rather than mediated by co-evolutionary adaptations unique to *O. faveolata* genotypes. By contrast, the *Rhodospirillales* ASV that was significantly enriched in resistant genotypes (Figure 23B) was identified as a lineage-specific ASV – suggestive of coevolution.

3.3. Determine susceptibility hierarchy and transmission risk among restoration *O. faveolata* genotypes

3.3.1. Transmission experiment susceptibility

The four transmission studies conducted in 2022 exposed 170 putative genotypes of *O. faveolata* to SCTLD. Disease signs were observed after only one day of exposure in all experiments, and only a small portion of individuals exposed during each study remained

healthy in the approximately two-week long experiments (Figure 25). Relative risks among the four SCTLD experiments were quantified using the binomial likelihood distribution model (Figure 26). All experiments had similar significantly high disease risks, as all experiment credible intervals were greater than one and overlapped with each other (Figure 26).

WGS determined 154 multi-locus genotypes from the 170 putative genotypes in the transmission experiments. Hierarchical clustering of the three disease metrics (resistance, transmission, and progression) categorized approximately 2/3 of the genotypes tested as either highly susceptible or susceptible (Figure 27). These types of genotypes showed low levels of resistance (most/all replicates developed disease signs), fast rates of transmission, and high progression rates. However, the susceptible genotypes had slower transmission rates compared with the highly susceptible genets. The intermediate genotypes had higher resistance metrics (very few replicates showed disease signs) compared with the highly susceptible and susceptible genets, similar levels of transmission rates compared with the susceptible genotypes, and slower progression rates. The resistant genotypes were characterized by having high resistance and slow transmission and progression rates. The 1/1 resistant genotypes had complete resistance (the one replicate exposed did not develop disease signs) and therefore had no transmission or progression rates to quantify.

Since having only 1 or 2 replicates of a single genotype exposed is less than ideal to interpret susceptibility of those genets, the data were analyzed again using the same approach, but with only those genotypes with 3 or more replicates (Figure 28). This approach yielded comparably similar results with the majority of genotypes being highly susceptible or susceptible, but much fewer actually within the susceptible category. The number of intermediate genotypes stayed similar (~20 different genets) and there were only 3 genotypes that were identified within the resistant grouping – and none of these genotypes had all replicates remain healthy (i.e., 1/3 or more replicates showed signs of disease). These results highlight the high level of susceptibility this species has when exposed to SCTLD, which was also evident within field surveys of the natural reef community during the initial phases of this disease outbreak.

3.3.2. Outplant susceptibility and survivorship

Relative risks of 8 WGS-assigned *O. faveolata* genotypes used in the field study from 2018–2019 were quantified using the Poisson-Gamma model (Figure 29). 3 genotypes had lower disease risk within the field study: WGS-53, WGS-50, and WGS-154 (Figure 29) and 2 out of those 3 also were identified as resistant within the transmission experiments (identified in green; Figure 29). 1 genotype (WGS-154), however, was highly susceptible within the transmission experiment (identified in red; Figure 29), but had a significantly lower relative risk. 1 genotype (WGS-58) showed slightly reduced risk within the field study and was similarly categorized as intermediate within the transmission experiments (identified in blue; Figure 29). Finally 4 genotypes trended towards high levels of risk to SCTLD within the field study (WGS-18, WGS-12, WGS-155, and WGS-152), all of which were highly susceptible to SCTLD within the transmission experiments as well (identified

in red; Figure 29). These results suggest high levels of convergence between identifying susceptibility within lab- and field-based settings.

Additionally, the Mote's outplant survival data associated with the genotypes utilized in the present study were assessed to determine if susceptibility to SCTLD could predict likelihood of survival when these corals were outplanted onto reefs within the Lower Florida Keys. This dataset consisted of 722 outplants from 29 different outplanting events from 2018–2023 at 16 reefs. There was substantial variation in the one-year survival rates of outplanted corals among genotypes, and many cases were not in congruence with susceptibility groupings (Figure 30 top). For example, highly susceptible genotypes had 100% survival in some cases, whereas resistant genotypes at times had some of the lowest survival rates near 60%. When these data were averaged by susceptibility grouping, there was no indication that SCTLD susceptibility was driving outplant survival rates (Figure 30 bottom), although further study could be warranted. Within this outplant survival dataset, there was no way to control the amount of replicates among susceptibility groups, and the outplanting events occurred over several years and within numerous different habitat types.

4. DISCUSSION

4.1. We developed the most complete and up-to-date genome and transcriptome assemblies of any Caribbean coral species

Here, we have shown that PacBio circular consensus (CCS) long-reads dramatically improve the genome resource for *Orbicella faveolata* (Young et al., 2024). The previous genome assembly for *O. faveolata* utilized short-read sequencing (Prada et al., 2016) on Illumina HiSeq and MiSeq platforms. While these methods were cutting-edge at the time, they came with major limitations in generating highly contiguous assemblies due to difficulties in assembling highly repetitive regions of the genome. The long-read technology utilized here can span these entire repetitive regions, resulting in fewer contigs and a more contiguous assembly. This was evident from our final assembly of 51 scaffolds which was nearly 40 times fewer than the 1,932 scaffolds present in the previous short-read *O. faveolata* assembly. The benefits of the long-read methods were also observed in improved N50 (long-read: 40,246,328 versus short-read: 4,771,691), L50 (long-read: 7 versus short-read: 124), and BUSCO completeness (single-copy and duplicated, long-read: 93.6% versus short-read: 85.3%). There were similarities between the short- and long-read assemblies for *O. faveolata*, specifically for GC content (long-read: 39.49%, short-read: 38.5%), overall genome length (long-read: 494,730,336 bp, short-read: 485,548,939 bp) and a ploidy of two. We also saw an improvement in BUSCO completeness in the identified protein-coding genes between the previous short-read and the new long-read assembly, with an increase from 87.2% to 95.1%. These results clearly show how long-read methodologies can greatly improve on older genomic resources generated using short-read methodologies.

We also evaluated our long-read assembly to other publicly available stony coral genome assemblies generated using long-read methodologies. Despite only using high fidelity

(HiFi) reads for our assembly, we obtained comparable completeness and contiguity to assemblies that also incorporated secondary scaffolding techniques (e.g., Hi-C and optical mapping). Our study has therefore shown that it is possible to obtain a highly complete and contiguous genome resource for a coral species without auxiliary methods, and with continued advancements in long-read methods, it will be possible to generate chromosomal assemblies with only one sequencing method. At present, utilizing these auxiliary methods are still advantageous, however, as it can further reduce contig number and generate chromosomal-level genome assemblies. Our identified protein-coding gene completeness using BUSCO was also comparable to other long-read coral genomes, again indicating that genome resources generated with just HiFi reads are of comparable quality to genomes generated using auxiliary methods. OrthoFinder analysis corroborated this with 93% of our protein-coding genes assigned to orthogroups with other coral species with long-read genomes available, and only 5.9% species-specific to *O. faveolata*. This again shows that our gene prediction and annotation pipeline was of similar quality to other methods used for other long-read coral genomes. Future work should look to expand on the OrthoFinder results, with incorporation of more long-read genomes as they become available, allowing identification of core coral gene function as well as species-specific processes.

4.2. We did not find compelling evidence of genomic signatures of disease resistance in *Orbicella faveolata*

Despite using the best-possible sequencing approach at present (whole-genome sequencing, or WGS) to analyze genetic relationships to disease susceptibility, we were unable to find any strong links between single-nucleotide polymorphisms (SNPs) in the *O. faveolata* genome and quantified SCTLD resistance metrics (see section 4.4. for the latter). This is similar to the only other study at present examining genetic basis of SCTLD resistance in this species (Klein et al., 2024), although the previous study could not completely disentangle potential impacts of relatively low genome sequencing coverage due to restriction site-associated sequencing approaches (2bRAD) from any potential genetic mechanisms impacting susceptibility. Their study also examined corals in a field-based approach, which may have been affected by non-disease-associated mortality or variable levels of disease exposure relative to a lab-based study as in this current project. While it is tempting to conclude that there is not a genomic basis of SCTLD resistance in *O. faveolata*, there is more research needed to be done. For example, the genome-wide association studies (GWAS) will also be conducted with SNPs associated with the coral's endosymbiotic algae (Symbiodiniaceae), to investigate whether there are genetic relationships between symbiont genomes and observed resistance metrics. This is particularly important given the hypothesis that SCTLD may target the algal symbionts rather than the coral host (Dennison et al., 2021; Karp et al., 2023; Work et al., 2021). And while it is beyond the scope of this project, posttranslational modifications to the genome (i.e., epigenetics) may also play a role in resistance/susceptibility to disease, as these processes have been demonstrated to be important in the response of *Acropora cervicornis* to other environmental perturbations (Rodriguez-Casariego et al., 2018; Rodríguez-Casariego et al., 2020).

Interestingly, genetic structure also did not show any strong correlations with Mote's provenance data, including original reef sampling location, sample type (sexual recruits versus corals of opportunity), or algal symbiont assemblages. The latter can be explained by Mote's nursery infrastructure, which tends to promote *Durusdinium*-dominance over other symbiont genera (Gantt et al., 2023). Yet, the presence of sampling location across all observed genetic clusters suggests that ancestral lineages are present throughout the Florida Keys. Regardless, the genetic divergence among the 5 genetic lineages is likely low, and may not have contributed strongly to patterns of adaptation among lineages. Studies using population genetics approaches with *O. faveolata*, including in the Florida Keys where Mote's genotypes were originally sourced, generally suggest low levels of genetic differentiation among populations, and even sites with high levels of clonality (Manzello et al., 2019; Rippe et al., 2017). Our study found clone rates to be remarkably similar between sequencing approaches (WGS: 12.6%, 2bRAD: 12.7%), indicating a relatively high degree of clonality. Some of these clones were to be expected from sampling all putative genotypes in Mote's nurseries, particularly for individuals whose metadata were lost over time, but selection of samples for sequencing was done to attempt to minimize accidental sequencing of clonal genotypes. Therefore, we predict that a large proportion of our observed clones were in fact true clones rather than propagation of fragments from a single parent colony, corroborating field-based studies of genetic structure with this species.

The lack of strong genetic differentiation among samples may also in part explain why GWAS was unsuccessful in linking observed SCTLD resistance to regions of the *O. faveolata* genome. Many of the samples were highly related to one another due to their similar origins and lack of genetic structure; indeed, the highest genetic distance among samples was 15.8% for WGS samples and 18.7% for 2bRAD samples. This made clonal detection difficult (as many individuals were highly related, such as siblings, and therefore had similar genetic distance as sequenced duplicates/triplicates), and it may also have reduced the ability of GWAS to detect genotype-phenotype associations. Perhaps these analyses would be better able to distinguish genetic mechanisms of adaptation relative to disease susceptibility across divergent populations, such as was recently conducted for *Acropora cervicornis* across Florida and Panama (Vollmer et al., 2023).

Nonetheless, both sequencing approaches used in this study (WGS and 2bRAD) were extremely useful in providing high-resolution genotype identities for all of Mote's nursery *O. faveolata*. This is a priority for maintenance of high genetic diversity in propagation and outplanting efforts, particularly for samples that had previous genotype identities based on outdated genotyping methods (e.g., microsatellites) or those with missing metadata. Mote's coral nursery and restoration projects are a massive undertaking, involving hundreds of scientists and early-career researchers over the years; therefore it is inevitable that some provenance data will be lost. Capturing all of the genetic data here will provide a baseline of genotypic information with which to inform all restoration efforts going forward. Genotyping methods are also important in the identification of cryptic morphologies that confound species delineations – the recent study by Klein et al., (2024) identified several individuals of *O. franksi* in their sample set that was only supposed to contain its congener

O. faveolata. Fortunately, our analyses did not find any evidence of mis-identified species, but genotyping efforts such as these are necessary to rule out 'contamination' of restoration stocks by non-target species that do not contribute to genetic diversity of the target species.

The project also provided the unique opportunity to directly compare WGS and 2bRAD sequencing approaches. We can make several recommendations regarding the applicability of WGS versus 2bRAD approaches in coral genomics studies: 1) Both approaches take approximately the same amount of time in terms of library preparation, but WGS costs more than three times as much ($175 versus $48 per sample at the time of sequencing for this project), and requires three orders of magnitude more hard drive space (Tb versus Gb) and high-performance computing time/power. For example, genotyping of WGS samples with GATK took approximately a month straight of computing time, while 2bRAD samples took several hours. 2) For applications requiring the ability to distinguish among distinct genetic lineages (i.e., genotyping), 2bRAD is sufficient. 3) For applications requiring genome-wide association studies (GWAS) among genetic lineages and physiological/phenotypic data, 2bRAD will not provide sufficient depth of sequencing across the entirety of the genome. 4) For applications requiring examination of other genetic markers beyond SNPs (such as insertions/deletions, aka indels), WGS is required as 2bRAD only targets SNP loci. 5) For identification of dominant symbiont genera, 2bRAD is sufficient and generally matches the results of targeted symbiont assemblage profiling using symbiont gene markers such as ITS2 (Eckert et al., 2020). WGS may provide more data than 2bRAD on background symbiont genera, but ITS2 is considered the gold standard approach. For both approaches, however, sequenced duplicates/triplicates are required for distinguishing true clonal genotypes, particularly for datasets such as these when overall genetic divergence is low. Additionally, the use of technical replicates here allowed the comparison of algal symbiont profiling between WGS and 2bRAD approaches, as well as among sequenced replicates from the same sequencing approach. While both approaches resulted in the same dominant symbiont genus for all technical replicates (*Durusdinium*), there was generally a higher abundance of background genera in WGS samples, likely due to increased sequencing depth. Ultimately, the most appropriate sequencing approach is therefore highly dependent on the needs of the project and the underlying scientific questions.

4.3. Microbial communities exhibited signs of co-evolution with resistant coral genotypes

The microbiomes of non-disease-exposed *O. faveolata* were highly diverse, with an even abundance and no particularly dominant bacteria. This is evident by our alpha and beta diversity analysis of healthy corals through time. However, in response to SCTLD exposure and transmission, the microbiome began to deviate from this stability and show a trend with slight decreases in alpha-diversity and increases in beta-diversity. At the amplicon sequence variant (ASV)-level, *O. faveolata* samples that developed SCTLD lesions had a significant increase in 5 *Vibrionales* ASVs and a significant decrease in 4 *Rhizobiales* ASVs relative to pre-exposure corals. However, *Vibrionales* ASVs and *Rhizobiales* ASVs were not significantly differentially abundant in non-disease-exposed

corals of susceptible genotypes relative to those of resistant genotypes. This indicates that these bacterial groups are not co-evolved indicators of SCTLD resistance or vulnerability in *O. faveolata* during normal conditions, but are still significantly associated with the microbial shift that occurs in *O. faveolata* throughout the timeline of initial SCTLD exposure to the development of visible lesions. Additionally, the expression variance and evolution (EVE) analysis identified 2 of these *Vibrionales* ASVs as highly variable in their abundance and not mediated by genotype evolutionary adaptation, but rather their abundance was mediated by the coral's environment.

In this study, we identified putative microbial biomarkers that were associated with resistance to SCTLD, by detecting bacteria with co-evolutionary signatures linked to more resistant coral genotypes (see section 4.4. below). Among non-disease-exposed samples, an uncultured *Rhodospiralles* ASV had a significant co-evolutionary signature that was also significantly more abundant in resistant coral genotypes relative to highly susceptible genotypes. Our results suggest that this *Rhodospirillales* ASV is a core microbiome member that may serve integral functional contributions to the host or microbiome such as a bioremediator or first-line defense. *Rhodospirillales* consists of bacteria with diverse metabolic capabilities and is biologically relevant among healthy samples in a previous SCTLD field study (Rosales et al., 2023). For a co-evolutionary relationship to develop, the functional contributions of a bacterial group must survive historical environmental change, competition, and selection, which may contribute to resistance. By contrast, bacteria with parasitic properties must elude eradication and can also develop as co-evolved bacteria associated with host vulnerability. By contrast, *O. faveolata* did not possess any bacteria in non-disease-exposed conditions which made it more vulnerable to disease incidence. While our data suggest that this *Rhodospirillales* ASV is associated with resilience, the absence of this co-evolved bacteria is associated with highly susceptible genotypes. The functional and network connectivity of this bacteria is worth further exploration to better understand the mechanistic contributions between microbe and host that are significantly associated with SCTLD disease fate. *Rhodospirillales* offers a concise target to assess SCTLD susceptibility risk in *O. faveolata* conservation and restoration efforts.

4.4. Susceptibility metrics quantified in lab-based disease transmission experiments predicted field-based restoration outcomes

The hierarchical clustering approach, which utilized several disease susceptibility metrics simultaneously including disease development (resistance), days to disease (transmission), and disease progression rates, provided four different groupings of susceptibility. These four groupings included highly susceptible, susceptible, intermediate susceptible, and resistant groups of genotypes. Highly susceptible genotypes were characterized by having all replicates develop signs of SCTLD that occurred shortly after exposure and progressed quickly. Susceptible genotypes generally had slower transmission rates (i.e., they developed disease signs later). Most genotypes were either highly susceptible or susceptible to SCTLD, whereas only a handful of genotypes were identified as intermediately susceptible. Intermediate genotypes generally had some resistance (~1/3

replicates remained healthy), and had slower transmission and progression rates. A total of 12 genotypes out of the 154 genotypes were resistant to SCTLD; however, an additional 14 genotypes that only had one replicate exposed that remained healthy could also be considered resistant. This level of disease resistance within a population (~10%) is similar to outcomes associated with population-level resistance to other types of coral diseases (Muller et al., 2018; Vollmer & Kline, 2008).

When comparing the susceptibility groupings from the experimental results to the same genotypes outplanted within the invasion and epidemic zone as SCTLD moved through the Lower Florida Keys, the susceptibility of these genotypes showed comparatively similar outcomes. These results suggest that the lab-based experimental results, which likely included a high dose of exposure to SCTLD disease agents due to the enclosed tank conditions during the exposure, potentially mimicked field-based, 'real-world' scenarios during the invasion and epidemic periods of the disease outbreak. Additionally, significantly low relative risks of 2 outplant genotypes (WGS-53 / OF2 & WGS-50 / OF126) that only had 1 replicate exposed that also remained healthy (e.g., 1/1 resistant) support the ability of controlled tank experiments to predict disease resistance even with low replication. Our results suggest that particular *O. faveolata* genotypes showing high levels of resistance within tank exposures may be good candidates for restoration planning if encouraging low disease mortality is a priority within outplant designs. These corals could also be utilized within a managed breeding program to help encourage disease resistance within restored populations, but only if the traits governing disease resistance are actually heritable. To date, that outcome is unknown. However, when comparing the overall susceptibility of genotypes to SCTLD to long-term outplant survival, there does not appear to be an association. Therefore, SCTLD may not be a main driver of long-term or even short-term survival of outplanted *O. faveolata,* at least not within the Lower Florida Keys.

4.5. Management recommendations

Through our four-year investigation into SCTLD resistance in *O. faveolata* (this study; Muller et al., 2023), we have quantitatively determined that there are inherently disease-resistant coral genotypes within Mote's nurseries. Further, the metrics developed in a lab-based setting to assess SCTLD resistance and susceptibility translate well to a field-based outplanting and monitoring program. This knowledge can be operationalized to target the propagation of known SCTLD-resistant genotypes for large-scale production. Coupled with the high-resolution genotypic data of all *O. faveolata* genotypes in Mote's restoration pipeline collected as part of this project, we now have the tools to 1) strategically incorporate disease-resistant traits into restoration efforts, 2) maintain high genetic diversity within nurseries and outplanting sites, 3) provide high-quality genetic and provenance data to restoration partners during broodstock exchanges, and 4) develop a platform for the evaluation of SCTLD resistance in additional *O. faveolata* genotypes beyond the scope of this study. The development of near-complete genome and transcriptome assemblies for this species will aid in all future studies investigating genetic, transcriptomic, proteomic, metabolomic, and epigenetic mechanisms of resilience, by

providing a well-annotated reference upon which to conduct 'omics analyses. This resource is already publicly available (Young et al., 2024), and represents the best genetic repository for any Caribbean coral species. Second, the identification of co-evolved microbial partners related to SCTLD exposure and transmission is a particularly exciting development, as it goes beyond previous studies that determined disease-associated bacterial taxa. *Rhodospirillales* in particular represents a high-priority bioindicator taxa, which can be quickly and cheaply quantified across a large swath of *O. faveolata* genotypes (such as with a targeted qPCR assay that requires no off-site and costly sequencing), or potentially for other coral species affected by SCTLD. Here, we provide a strong foundation for future research efforts into coral resilience to SCTLD, and inform management strategies to maximize success of adaptive conservation and restoration efforts. Taken together, the tools employed in this study to investigate SCTLD resistance and associated molecular mechanisms are extremely valuable in advancing our understanding of coral resilience in response to the most severe disease outbreak ever recorded.

5. REFERENCES

Aeby, G. S., Ushijima, B., Campbell, J. E., Jones, S., Williams, G. J., Meyer, J. L., Häse, C., & Paul, V. J. (2019). Pathogenesis of a Tissue Loss Disease Affecting Multiple Species of Corals Along the Florida Reef Tract. *Frontiers in Marine Science*, *6*. https://doi.org/10.3389/fmars.2019.00678

Andrews, S. (2010). *FastQC: A Quality Control Tool for High Throughput Sequence Data* [Computer software]. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Auwera, G. V. der, & O'Connor, B. (2020). *Genomics in the Cloud: Using Docker, GATK, and WDL in Terra* (1st edition). O'Reilly Media.

Avila-Magaña, V., Kamel, B., DeSalvo, M., Gómez-Campo, K., Enríquez, S., Kitano, H., Rohlfs, R. V., Iglesias-Prieto, R., & Medina, M. (2021). Elucidating gene expression adaptation of phylogenetically divergent coral holobionts under heat stress. *Nature Communications*, *12*(1), 5731. https://doi.org/10.1038/s41467-021-25950-4

Baselga, A., & Orme, C. D. L. (2012). betapart: An R package for the study of beta diversity. *Methods in Ecology and Evolution*, *3*(5), 808–812. https://doi.org/10.1111/j.2041-210X.2012.00224.x

Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., Alexander, H., Alm, E. J., Arumugam, M., Asnicar, F., Bai, Y., Bisanz, J. E., Bittinger, K., Brejnrod, A., Brislawn, C. J., Brown, C. T., Callahan, B. J., Caraballo-Rodríguez, A. M., Chase, J., … Caporaso, J. G. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nature Biotechnology*, *37*(8), 852–857. https://doi.org/10.1038/s41587-019-0209-9

Cheng, H., Concepcion, G. T., Feng, X., Zhang, H., & Li, H. (2021). Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nature Methods*, *18*(2), Article 2. https://doi.org/10.1038/s41592-020-01056-5

Coombe, L., Li, J. X., Lo, T., Wong, J., Nikolic, V., Warren, R. L., & Birol, I. (2021). LongStitch: High-quality genome assembly correction and scaffolding using long

reads. *BMC Bioinformatics*, *22*(1), 534. https://doi.org/10.1186/s12859-021-04451-7

Cornwall, C. E., Comeau, S., Kornder, N. A., Perry, C. T., van Hooidonk, R., DeCarlo, T. M., Pratchett, M. S., Anderson, K. D., Browne, N., Carpenter, R., Diaz-Pulido, G., D'Olivo, J. P., Doo, S. S., Figueiredo, J., Fortunato, S. A. V., Kennedy, E., Lantz, C. A., McCulloch, M. T., González-Rivero, M., … Lowe, R. J. (2021). Global declines in coral reef calcium carbonate production under ocean acidification and warming. *Proceedings of the National Academy of Sciences*, *118*(21), e2015265118. https://doi.org/10.1073/pnas.2015265118

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., Durbin, R., & 1000 Genomes Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M., & Li, H. (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, *10*(2), giab008. https://doi.org/10.1093/gigascience/giab008

Dennison, C. E., Karp, R. F., Weiler, B. A., Goncalves, A., del Campo, J., Rosales, S. M., Traylor-Knowles, N., & Baker, A. C. (2021). *The role of algal symbionts (genus Breviolum) in the susceptibility of corals to Stony Coral Tissue Loss Disease in South Florida*. Florida DEP. Miami FL.

Dobbelaere, T., Muller, E. M., Gramer, L. J., Holstein, D. M., & Hanert, E. (2020). Coupled Epidemio-Hydrodynamic Modeling to Understand the Spread of a Deadly Coral Disease in Florida. *Frontiers in Marine Science*, *7*. https://doi.org/10.3389/fmars.2020.591881

Eckert, R. J., Reaume, A. M., Sturm, A. B., Studivan, M. S., & Voss, J. D. (2020). Depth Influences Symbiodiniaceae Associations Among Montastraea cavernosa Corals on the Belize Barrier Reef. *Frontiers in Microbiology*, *11*. https://doi.org/10.3389/fmicb.2020.00518

Emms, D. M., & Kelly, S. (2019). OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biology*, *20*(1), 238. https://doi.org/10.1186/s13059-019-1832-y

Enochs, I. C., Studivan, M. S., Kolodziej, G., Foord, C., Basden, I., Boyd, A., Formel, N., Kirkland, A., Rubin, E., Jankulak, M., Smith, I., Kelble, C. R., & Manzello, D. P. (2023). Coral persistence despite marginal conditions in the Port of Miami. *Scientific Reports*, *13*(1), 6759. https://doi.org/10.1038/s41598-023-33467-7

Frichot, E., & François, O. (2015). LEA: An R package for landscape and ecological association studies. *Methods in Ecology and Evolution*, *6*(8), 925–929. https://doi.org/10.1111/2041-210X.12382

Gantt, S. E., Keister, E. F., Manfroy, A. A., Merck, D. E., Fitt, W. K., Muller, E. M., & Kemp, D. W. (2023). Wild and nursery-raised corals: Comparative physiology of two framework coral species. *Coral Reefs*, *42*(2), 299–310. https://doi.org/10.1007/s00338-022-02333-9

Grønvold, L. (2021). *Evemodel · GitLab* [Computer software]. https://gitlab.com/sandve-

lab/evemodel
Hawthorn, A. C., Dennis, M., Kiryu, Y., Landsberg, J., Peters, E., & Work, T. M. (2024). Stony coral tissue loss disease (SCTLD) case definition for wildlife. In *Techniques and Methods* (19-I1). U.S. Geological Survey. https://doi.org/10.3133/tm19I1
Hoegh-Guldberg, O., Poloczanska, E. S., Skirving, W., & Dove, S. (2017). Coral Reef Ecosystems under Climate Change and Ocean Acidification. *Frontiers in Marine Science*, *4*. https://doi.org/10.3389/fmars.2017.00158
Hoegh-Guldberg, O., Skirving, W., Dove, S. G., Spady, B. L., Norrie, A., Geiger, E. F., Liu, G., De La Cour, J. L., & Manzello, D. P. (2023). Coral reefs in peril in a record-breaking year. *Science*, *382*(6676), 1238–1240. https://doi.org/10.1126/science.adk4532
Jackson, J., Donovan, M., Cramer, K., & Lam, V. (2014). *Status and trends of Caribbean coral reefs: 1970-2012*. Global Coral Reef Monitoring Network. https://pubs.usgs.gov/publication/70115405
Kamvar, Z. N., Tabima, J. F., & Grünwald, N. J. (2014). Poppr: An R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ*, *2*, e281. https://doi.org/10.7717/peerj.281
Karp, R. F., Dennison, C. E., Kron, N. S., & Baker, A. C. (2023). *The use of algal symbiont cultures (Family Symbiodiniaceae) as model systems to study stony coral tissue loss disease (SCTLD): Use of fractionated disease isolates to help with pathogen identification* (p. 27). Florida Department of Environmental Protection. Tallahassee, Florida.
Klein, A. M., Sturm, A. B., Eckert, R. J., Walker, B. K., Neely, K. L., & Voss, J. D. (2024). Algal symbiont genera but not coral host genotypes correlate to stony coral tissue loss disease susceptibility among Orbicella faveolata colonies in South Florida. *Frontiers in Marine Science*, *11*. https://doi.org/10.3389/fmars.2024.1287457
Kramer, P. R., Roth, L., & Lanj, J. (2019). *Map of stony coral tissue loss disease outbreak in the Caribbean*. AGRRA. https://www.agrra.org/coral-disease-outbreak/
Krueger, F. (2024). *FelixKrueger/TrimGalore* [Perl]. https://github.com/FelixKrueger/TrimGalore (Original work published 2016)
Lane, D. R., Ready, R. C., Buddemeier, R. W., Martinich, J. A., Shouse, K. C., & Wobus, C. W. (2013). Quantifying and Valuing Potential Climate Change Impacts on Coral Reefs in the United States: Comparison of Two Scenarios. *PLOS ONE*, *8*(12), e82579. https://doi.org/10.1371/journal.pone.0082579
Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 357–359. https://doi.org/10.1038/nmeth.1923
Lin, H., & Peddada, S. D. (2024). Multigroup analysis of compositions of microbiomes with covariate adjustments and repeated measures. *Nature Methods*, *21*(1), Article 1. https://doi.org/10.1038/s41592-023-02092-7
MacKnight, N. J., Dimos, B. A., Beavers, K. M., Muller, E. M., Brandt, M. E., & Mydlarz, L. D. (2022). Disease resistance in coral is mediated by distinct adaptive and plastic gene expression profiles. *Science Advances*, *8*(39), eabo6153.

https://doi.org/10.1126/sciadv.abo6153
Manni, M., Berkeley, M. R., Seppey, M., Simão, F. A., & Zdobnov, E. M. (2021). BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. *Molecular Biology and Evolution*, *38*(10), 4647–4654. https://doi.org/10.1093/molbev/msab199
Manzello, D. P., Matz, M. V., Enochs, I. C., Valentino, L., Carlton, R. D., Kolodziej, G., Serrano, X., Towle, E. K., & Jankulak, M. (2019). Role of host genetics and heat-tolerant algal symbionts in sustaining populations of the endangered coral Orbicella faveolata in the Florida Keys with ocean warming. *Global Change Biology*, *25*(3), 1016–1031. https://doi.org/10.1111/gcb.14545
Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMB.Net Journal*, *17*(1), 10–12. https://doi.org/10.14806/ej.17.1.200.
Micheli, F., Mumby, P. J., Brumbaugh, D. R., Broad, K., Dahlgren, C. P., Harborne, A. R., Holmes, K. E., Kappel, C. V., Litvin, S. Y., & Sanchirico, J. N. (2014). High vulnerability of ecosystem function and services to diversity loss in Caribbean coral reefs. *Biological Conservation*, *171*, 186–194. https://doi.org/10.1016/j.biocon.2013.12.029
Mikheenko, A., Prjibelski, A., Saveliev, V., Antipov, D., & Gurevich, A. (2018). Versatile genome assembly evaluation with QUAST-LG. *Bioinformatics*, *34*(13), i142–i150. https://doi.org/10.1093/bioinformatics/bty266
Mikheenko, A., Valin, G., Prjibelski, A., Saveliev, V., & Gurevich, A. (2016). Icarus: Visualizer for de novo assembly evaluation. *Bioinformatics*, *32*(21), 3321–3323. https://doi.org/10.1093/bioinformatics/btw379
Muller, E. M., Bartels, E., & Baums, I. B. (2018). Bleaching causes loss of disease resistance within the threatened coral species Acropora cervicornis. *eLife*, *7*, e35066. https://doi.org/10.7554/eLife.35066
Muller, E. M., Sartor, C., Alcaraz, N. I., & van Woesik, R. (2020). Spatial Epidemiology of the Stony-Coral-Tissue-Loss Disease in Florida. *Frontiers in Marine Science*, *7*. https://doi.org/10.3389/fmars.2020.00163
Muller, E. M., Williams, S., Sirotzke, S. M., Studivan, M., Young, B. D., Traylor-Knowles, N., Isma, L. M., Andrade Rodriguez, N., Rosales, S. M., & MacKnight, N. J. (2023). *Strategic integration of stony coral tissue loss disease resistance into coral reproduction and restoration practices to recover Florida's Coral Reef* (p. 37). Florida Department of Environmental Protection. Dania Beach.
NOAA. (2018). *Stony coral tissue loss disease case definition* (p. 10). NOAA. Silver Spring, MD.
Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'hara, R., Simpson, G. L., Solymos, P., Stevens, M. H. H., & Wagner, H. (2013). Package 'vegan'. *Community Ecology Package, Version*, *2*(9), 1–295.
Palacio-Castro, A. M., Enochs, I. C., Besemer, N., Boyd, A., Jankulak, M., Kolodziej, G., Hirsh, H. K., Webb, A. E., Towle, E. K., Kelble, C., Smith, I., & Manzello, D. P. (2023). Coral Reef Carbonate Chemistry Reveals Interannual, Seasonal, and Spatial Impacts on Ocean Acidification Off Florida. *Global Biogeochemical*

Cycles, *37*(12), e2023GB007789. https://doi.org/10.1029/2023GB007789
Palmer, J. M., & Stajich, J. (2020). *Funannotate v1.8.1: Eukaryotic genome annotation* [Computer software]. Zenodo. https://doi.org/10.5281/zenodo.4054262
Paradis, E., & Schliep, K. (2019). ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, *35*(3), 526–528. https://doi.org/10.1093/bioinformatics/bty633
Prada, C., Hanna, B., Budd, A. F., Woodley, C. M., Schmutz, J., Grimwood, J., Iglesias-Prieto, R., Pandolfi, J. M., Levitan, D., Johnson, K. G., Knowlton, N., Kitano, H., DeGiorgio, M., & Medina, M. (2016). Empty Niches after Extinctions Increase Population Sizes of Modern Corals. *Current Biology*, *26*(23), 3190–3194. https://doi.org/10.1016/j.cub.2016.09.039
Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., de Bakker, P. I. W., Daly, M. J., & Sham, P. C. (2007). PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *The American Journal of Human Genetics*, *81*(3), 559–575. https://doi.org/10.1086/519795
R Core Team. (2020). *R: A language and environment for statistical computing. R Foundation for Statistical Computing* [R]. https::www.R-project.org/
Rippe, J. P., Matz, M. V., Green, E. A., Medina, M., Khawaja, N. Z., Pongwarin, T., Pinzón C., J. H., Castillo, K. D., & Davies, S. W. (2017). Population structure and connectivity of the mountainous star coral, Orbicella faveolata, throughout the wider Caribbean region. *Ecology and Evolution*, *7*(22), 9234–9246. https://doi.org/10.1002/ece3.3448
Rodriguez-Casariego, J. A., Ladd, M. C., Shantz, A. A., Lopes, C., Cheema, M. S., Kim, B., Roberts, S. B., Fourqurean, J. W., Ausio, J., Burkepile, D. E., & Eirin-Lopez, J. M. (2018). Coral epigenetic responses to nutrient stress: Histone H2A.X phosphorylation dynamics and DNA methylation in the staghorn coral Acropora cervicornis. *Ecology and Evolution*, *8*(23), 12193–12207. https://doi.org/10.1002/ece3.4678
Rodríguez-Casariego, J. A., Mercado-Molina, A. E., Garcia-Souto, D., Ortiz-Rivera, I. M., Lopes, C., Baums, I. B., Sabat, A. M., & Eirin-Lopez, J. M. (2020). Genome-Wide DNA Methylation Analysis Reveals a Conserved Epigenetic Response to Seasonal Environmental Variation in the Staghorn Coral Acropora cervicornis. *Frontiers in Marine Science*, *7*. https://doi.org/10.3389/fmars.2020.560424
Rohlfs, R. V., & Nielsen, R. (2015). Phylogenetic ANOVA: The Expression Variance and Evolution Model for Quantitative Trait Evolution. *Systematic Biology*, *64*(5), 695–708. https://doi.org/10.1093/sysbio/syv042
Rosales, S. M., Huebner, L. K., Evans, J. S., Apprill, A., Baker, A. C., Becker, C. C., Bellantuono, A. J., Brandt, M. E., Clark, A. S., del Campo, J., Dennison, C. E., Eaton, K. R., Huntley, N. E., Kellogg, C. A., Medina, M., Meyer, J. L., Muller, E. M., Rodriguez-Lanetty, M., Salerno, J. L., … Voss, J. D. (2023). A meta-analysis of the stony coral tissue loss disease microbiome finds key bacteria in unaffected and lesion tissue in diseased colonies. *ISME Communications*, *3*(1), 1–14. https://doi.org/10.1038/s43705-023-00220-0
Scrucca, L., Fop, M., Murphy, T., Brendan, & Raftery, A., E. (2016). mclust 5:

Clustering, Classification and Density Estimation Using Gaussian Finite Mixture Models. *The R Journal*, *8*(1), 289. https://doi.org/10.32614/RJ-2016-021

Storlazzi, C. D., Reguero, B. G., Yates, K. K., Cumming, K. A., Cole, A., Shope, J. B., L, C. G., Zawada, D. G., Arsenault, S. R., Fehr, Z. W., Nickel, B. A., & Beck, M. W. (2021). Rigorously valuing the impact of projected coral reef degradation on coastal hazard risk in Florida. In *Open-File Report* (2021–1055). U.S. Geological Survey. https://doi.org/10.3133/ofr20211055

Studivan, M. S., Baptist, M., Molina, V., Riley, S., First, M., Soderberg, N., Rubin, E., Rossin, A., Holstein, D. M., & Enochs, I. C. (2022). Transmission of stony coral tissue loss disease (SCTLD) in simulated ballast water confirms the potential for ship-born spread. *Scientific Reports*, *12*(1), 19248. https://doi.org/10.1038/s41598-022-21868-z

Studivan, M. S., Rossin, A. M., Rubin, E., Soderberg, N., Holstein, D. M., & Enochs, I. C. (2022). Reef Sediments Can Act As a Stony Coral Tissue Loss Disease Vector. *Frontiers in Marine Science*, *8*. https://doi.org/10.3389/fmars.2021.815698

Sturtz, S., Ligges, U., & Gelman, A. (2005). R2WinBUGS: A Package for Running WinBUGS from R. *Journal of Statistical Software*, *12*, 1–16. https://doi.org/10.18637/jss.v012.i03

van Woesik, R., & Randall, C. J. (2017). Coral disease hotspots in the Caribbean. *Ecosphere*, *8*(5), e01814. https://doi.org/10.1002/ecs2.1814

Venables, W. N., & Ripley, B., D. (2013). *Modern applied statistics with S-PLUS*. Springer Science & Business Media.

Vollmer, S. V., & Kline, D. I. (2008). Natural Disease Resistance in Threatened Staghorn Corals. *PLOS ONE*, *3*(11), e3718. https://doi.org/10.1371/journal.pone.0003718

Vollmer, S. V., Selwyn, J. D., Despard, B. A., & Roesel, C. L. (2023). Genomic signatures of disease resistance in endangered staghorn corals. *Science*, *381*(6665), 1451–1454. https://doi.org/10.1126/science.adi3601

Vries, A. de. (2024). *Andrie/ggdendro* [R]. https://github.com/andrie/ggdendro (Original work published 2011)

Walton, C. J., Hayes, N. K., & Gilliam, D. S. (2018). Impacts of a Regional, Multi-Year, Multi-Species Coral Disease Outbreak in Southeast Florida. *Frontiers in Marine Science*, *5*. https://doi.org/10.3389/fmars.2018.00323

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis* [R]. https://ggplot2.tidyverse.org.

Work, T. M., Weatherby, T. M., Landsberg, J. H., Kiryu, Y., Cook, S. M., & Peters, E. C. (2021). Viral-Like Particles Are Associated With Endosymbiont Pathology in Florida Corals Affected by Stony Coral Tissue Loss Disease. *Frontiers in Marine Science*, *8*. https://doi.org/10.3389/fmars.2021.750658

Young, B. D., Williamson, O. M., Kron, N. S., Andrade Rodriguez, N., Isma, L. M., MacKnight, N. J., Muller, E. M., Rosales, S. M., Sirotzke, S. M., Traylor-Knowles, N., Williams, S. D., & Studivan, M. S. (2024). Annotated genome and transcriptome of the endangered Caribbean mountainous star coral (Orbicella faveolata) using PacBio long-read sequencing. *BMC Genomics*, *25*(1), 226. https://doi.org/10.1186/s12864-024-10092-w

6. TABLES

Table 1. Mean Symbiodiniaceae abundance combined across all experimental (WGS) and nursery (2bRAD) putative genotypes.

Symbiodiniaceae Genus	Mean Relative Abundance	Standard Error of the Mean (±)
Symbiodinium	3.4%	0.2%
Breviolum	1.5%	0.4%
Cladocopium	5.9%	0.5%
Durusdinium	89.2%	0.8%

Alt text: Table of mean relative abundance of algal symbionts (Symbiodiniaceae) across four genera from genotyped samples, with standard error of the mean. Underlying data available on the SCTLD DataOne portal (urn:uuid:f1c6f769-e7aa-464f-a046-504782f402cd), processed using scripts available on GitHub (https://github.com/mstudiva/SCTLD-resistance-genomics).

Table 2. A Kruskal-Wallis test was performed on the diversity indices. "$\chi 2$" indicates the Chi-square test, "*p*" indicates the *p*-value.

Kruskal-Wallis	Treatment			Disease Outcome			Disease Time point		
Index	$\chi 2$	df	p	$\chi 2$	df	p	$\chi 2$	df	p
Shannon	17.413	1	<1e-5	40.372	2	<1e-5	15.718	3	0.001296
Pielou	11.537	1	0.0006821	31.426	2	<1e-5	17.516	3	0.0005534
Simpson	12.358	1	0.0004392	32.815	2	<1e-5	15.517	3	0.001424
Beta	11.166	1	0.0008329	27.786	2	<1e-5	6.3206	3	0.09701

Alt text: Table of statistical outputs from Kruskal-Wallis tests conducted on four microbial diversity metrics by disease transmission experiment treatment, disease outcome, and sampling time point. Underlying data available on the SCTLD DataOne portal (urn:uuid:f1c6f769-e7aa-464f-a046-504782f402cd), processed using scripts available on GitHub (https://github.com/nmacknight/Ofav.16s.SCTLD).

7. FIGURES

A)

Measure	*Orbicella faveolata* (long-read)	*Orbicella faveolata* (short-read)	*Montipora capitata* (V1) [1]	*Montipora capitata* (V3) [2]	*Pocillopora acuta* [2]	*Porites compressa* [2]	*Pocillopora meandrina* [2]	*Pocillopora effuse* [3]	*Porites lobata* [3]	*Acropora tenuis* [4]	*Acropora digitfera* [5]	*Acropora millepora* [6]
# Scaffolds	51	1933	3043	1699	474	608	212	252	1098	614	955	854
Longest Scaffold	40,246,328	4,771,691	3,468,816	69,151,359	16,633,824	18,348,340	21,651,136	11,895,822	3,468,816	13,556,973	7,632,824	39,361,238
N50	33,295,526	1,162,446	540,623	47,716,837	3,166,945	4,003,051	10,024,633	4,753,879	540,623	2,836,608	1,856,312	19,840,543
L50	7	124	467	7	13	7	9	23	467	46	69	9
N's/100kbp	1.02	26,684	0	18.07	0	0	0	0	0	0	425.34	8.16
GC content	39.49%	38.99%	39.63%	39.66%	38.03%	39.66%	39.06%	37.99%	39.63	39.07	39.04%	39.06%

B)

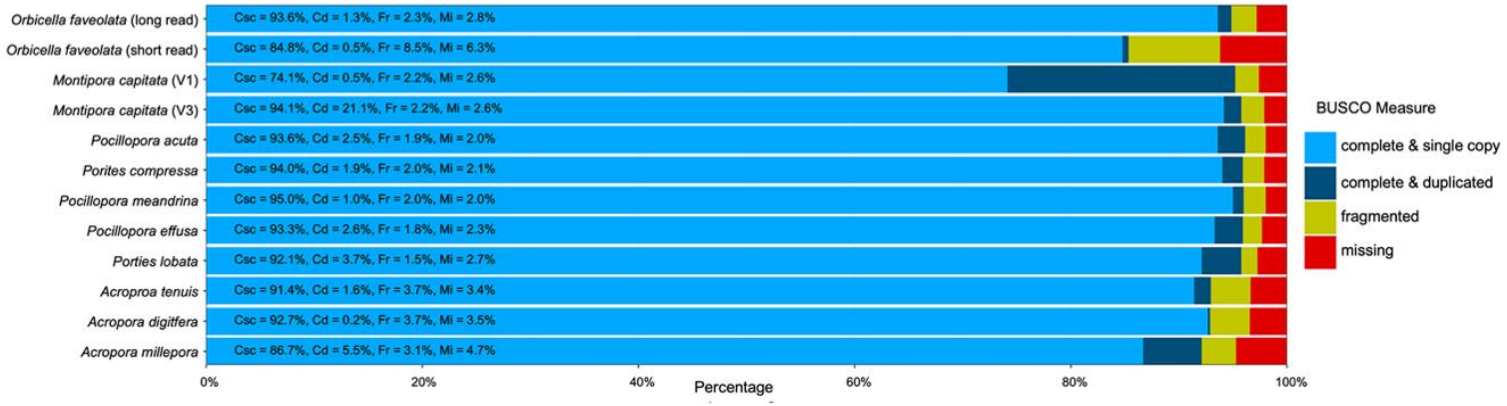


Figure 1. Quast and BUSCO analysis results of long-read stony coral genomes. *A*) Results from Quast analysis of our *de-novo* assembly, previous short read *Orbicella faveolata* assembly, and all publicly available long-read stony coral genomes. *B*) Results of BUSCO analysis using our *O. faveolata de-novo* assembly, the previous short-read *O. faveolata* assembly, and all publicly available long-read stony coral genomes with the metazoa_odb10 database. Completeness is split into single copy (light blue) and duplicated (dark blue). Fragmented = yellow, missing = red. Percentages for each metric are present in each bar: Csc = complete and single copy, Cd = complete and duplicated, Fr = fragmented, M = missing. For both (A) and (B) "*Orbicella faveolata* (short-read)" is the previously assembled short-read genome, and "*Orbicella faveolata* (long-read)" is the *de-novo* assembly using PacBio HiFi reads. This figure is published in Young et al. (2024).

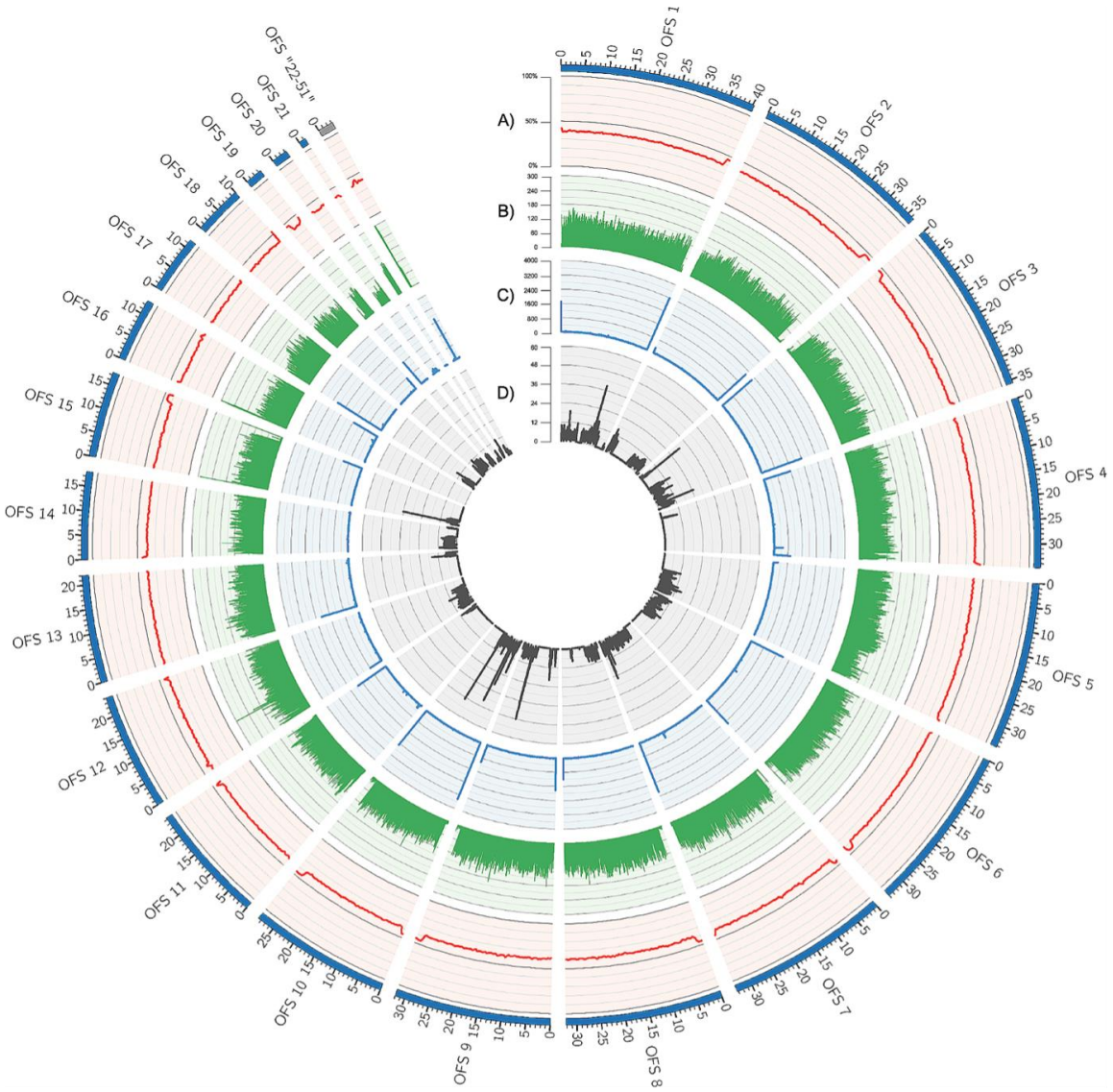


Figure 2: Visualization of scaffolded genome assembly of *Orbicella faveolata*. *A*) GC content calculated with a sliding window of 50,000 base pairs (bp). Y-axis shows the percentage calculated for GC content over each 50,000 bp sliding window. *B*) Repeat content plotted using a sliding window of 50,000 base pairs and the gff output file from RepeatMasker. Y-axis shows counts of repetitive regions for each sliding window of 50,000 base pairs. *C*) Telomeric repeats generated with a sliding window of 50,000 base pairs and the repeat pattern of "TTAGGG". Y-axis shows the counts of the telomeric repeat for each sliding window of 50,000 base pairs. Telomeric repeats can be identified by peaks at either the start or end of each scaffold. *D*) Gene density calculated with a sliding window of 50,000 base pairs and the "gene" identifiers from the gff file generated from funannotate::annotate. Y-axis shows the counts of genes for each sliding window of 50,000 base pairs. This figure is published in Young et al. (2024).

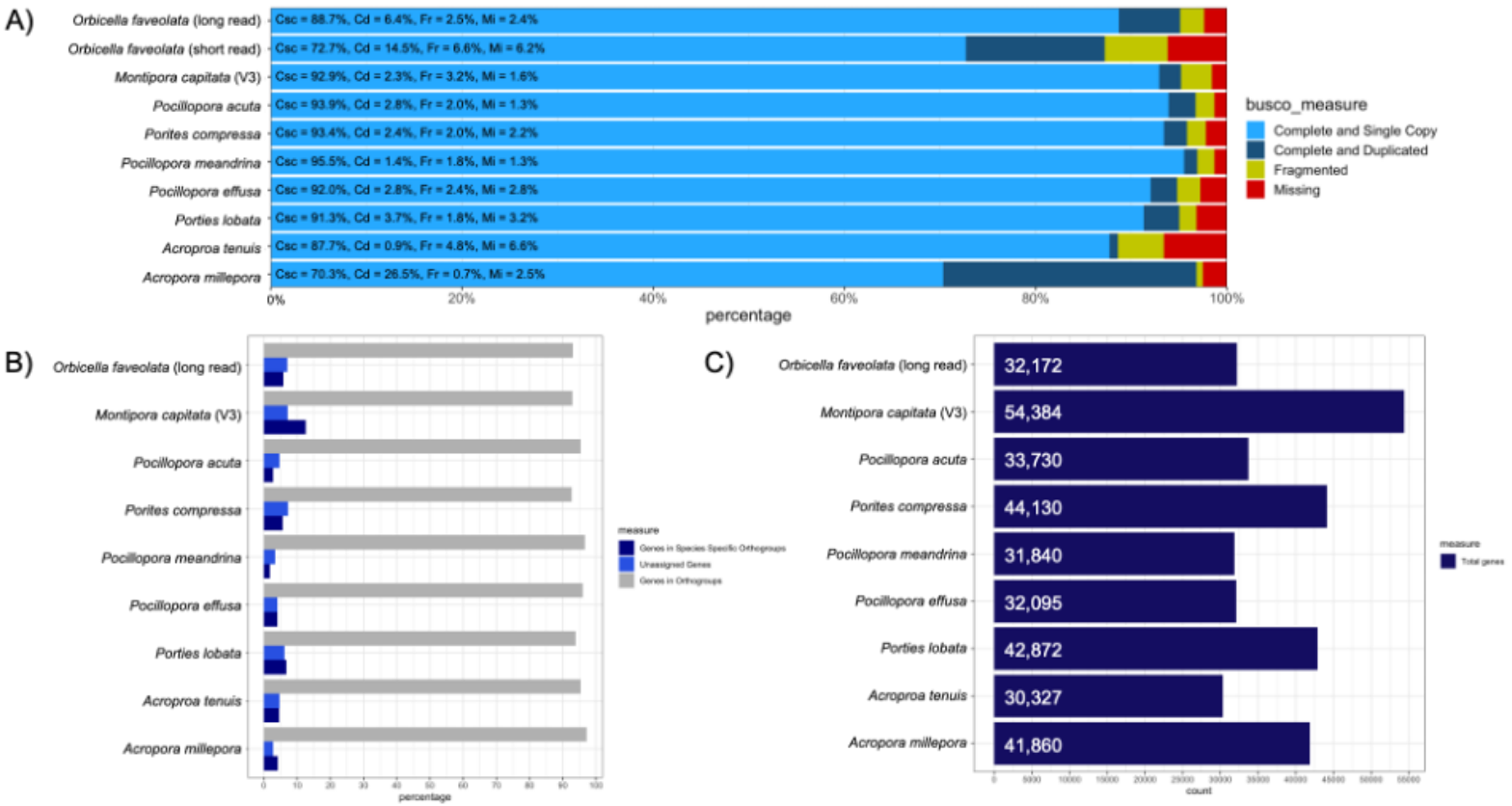


Figure 3: Results from the analysis of BUSCO and OrthoFinder on the protein coding genes from our *de-novo* assembly, previous *O. faveolata* reference genome, and other long-read coral genomes. A) Results of BUSCO (database = metazoa_odb10) analysis on the protein coding genes on our *de-novo* assembly, the previous *O. faveolata* reference genome, and other long-read coral genomes. Completeness is split into single-copy (light blue) and duplicated (dark blue). Fragmented = yellow, Missing = red. Percentages for each metric are present in each bar: Csc = complete and single-copy, Cd = complete and duplicated, Fr = fragmented, M = missing. "*Orbicella faveolata* (short read)" is the previously assembled short-read genome, and "*Orbicella faveolata* (long read)" is our *de-novo* assembly using PacBio HiFi reads. B) Results from OrthoFinder analysis between our *de-novo* assembly and other publicly available coral long-read genomes. C) Total number of protein-coding genes present in coral long-read genomes used in OrthoFinder analysis. Number within the bar shows the total number of protein coding genes present in each long-read genome assembly. This figure is published in Young et al. (2024).

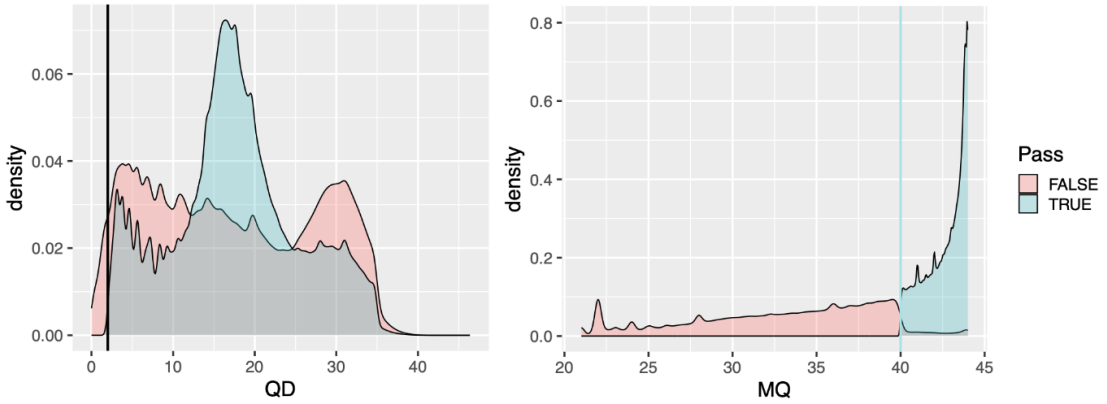


Figure 4. Density of WGS SNP calls across QualByDepth (QD), a metric of the confidence that a particular variant locus is high-quality (i.e., real, and not a false positive) before and after filtering at a QD threshold of 2. Two peaks at QD values of ~17 and ~30 represent homozygous and heterozygous variants, respectively. (right) Density of WGS SNPs calls across MappingQuality (MQ), the mapping quality of the locus to the genome, before and after filtering at an MQ threshold of 40.

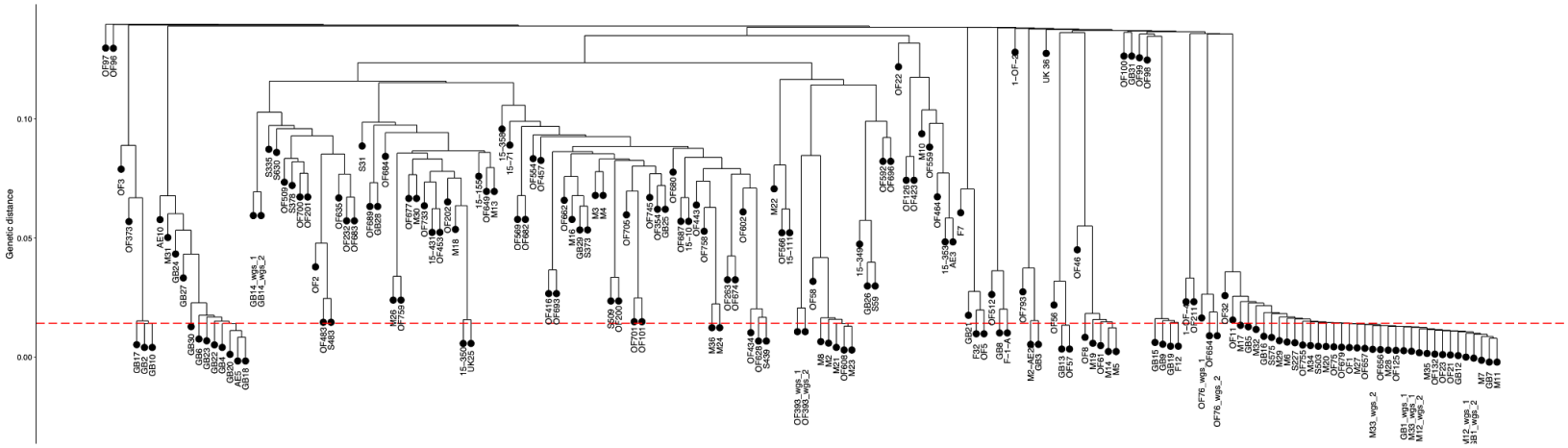


Figure 5. Dendrogram of WGS samples (experimental genotypes) based on genetic distance. Horizontal dashed line indicates the clonal threshold used in clonal genotype detection. Any branches below the threshold are considered clonal genotypes.

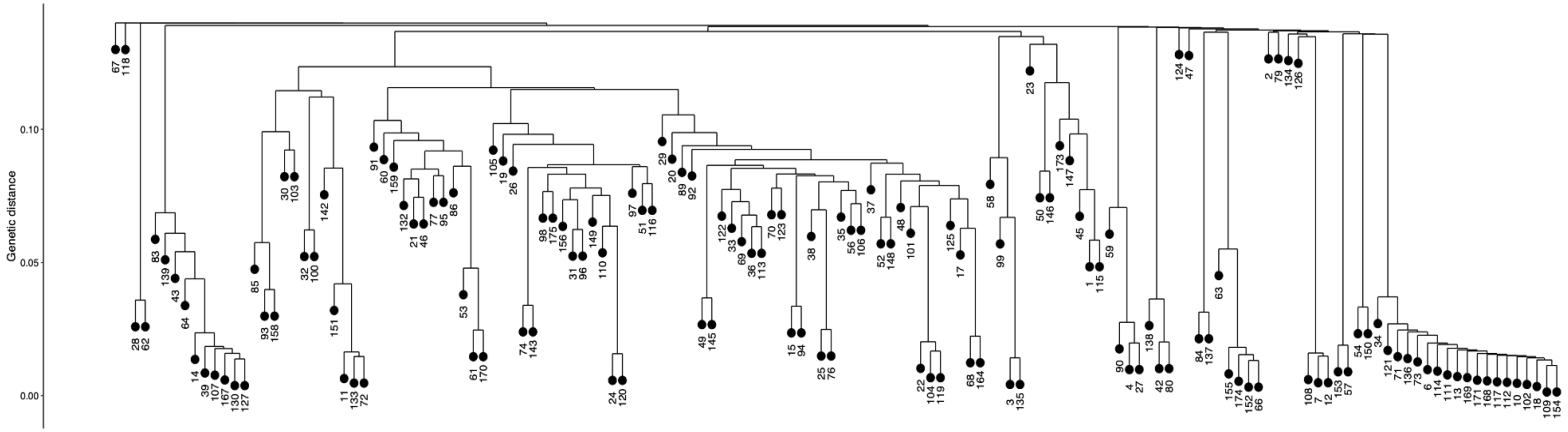


Figure 6. Dendrogram of unique WGS multi-locus genotypes (experimental genotypes) based on genetic distance following removal of clonal genotype samples.

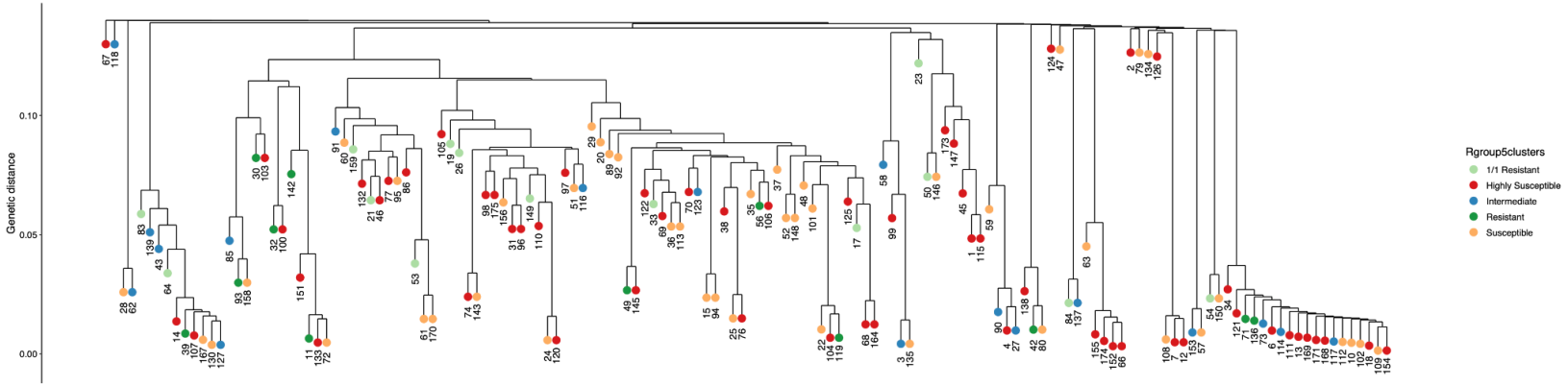


Figure 7. Dendrogram of unique WGS multi-locus genotypes (experimental genotypes) based on genetic distance. Colors indicate disease susceptibility hierarchy (see section 3.3.1.).

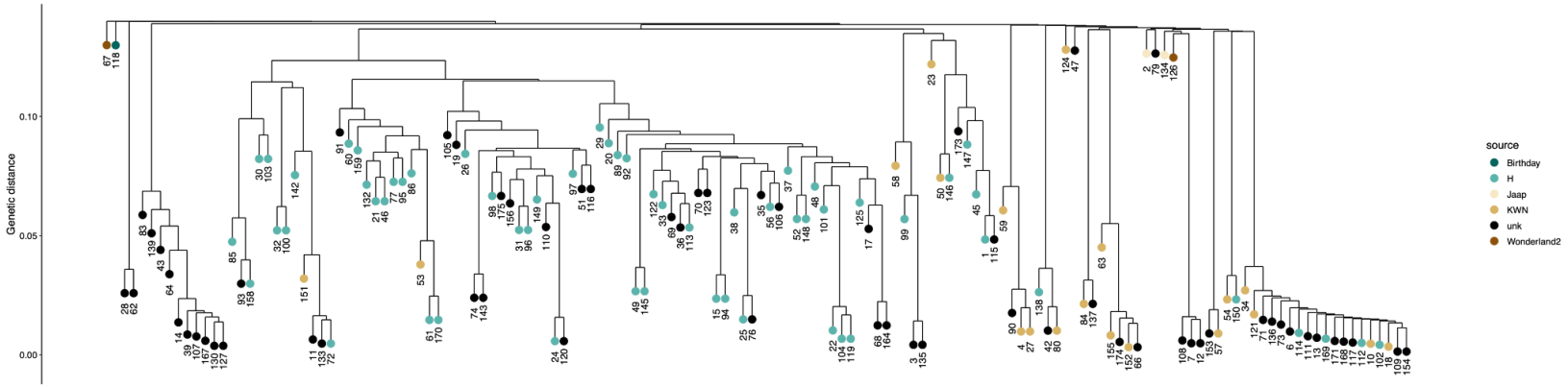


Figure 8. Dendrogram of unique WGS multi-locus genotypes (experimental genotypes) based on genetic distance. Colors indicate original sampling location. KWN = Key West nursery, H = Horseshoe, unk = unknown

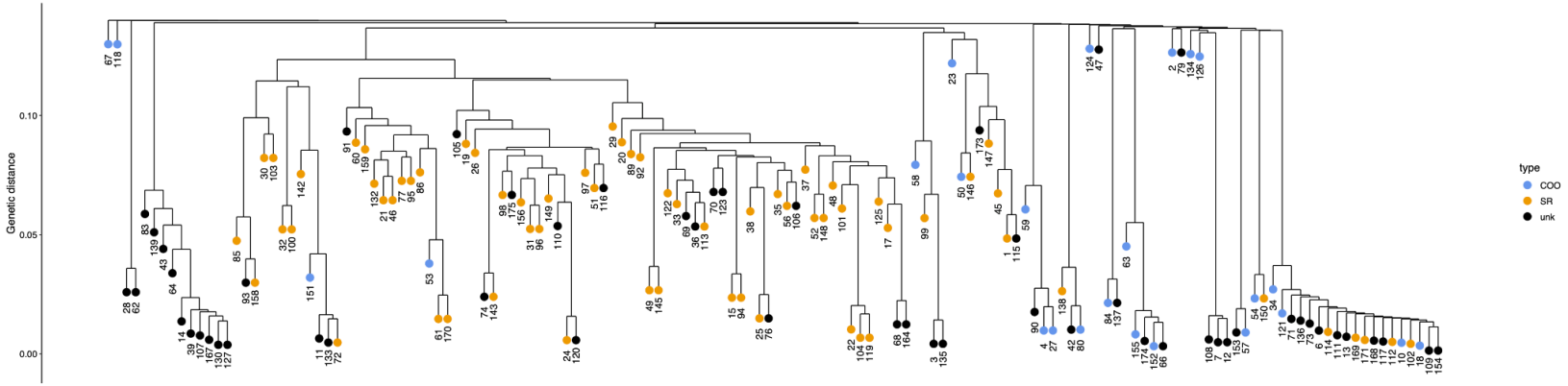


Figure 9. Dendrogram of unique WGS multi-locus genotypes (experimental genotypes) based on genetic distance. Colors indicate coral type. COO = coral of opportunity, SR = sexual recruit, unk = unknown

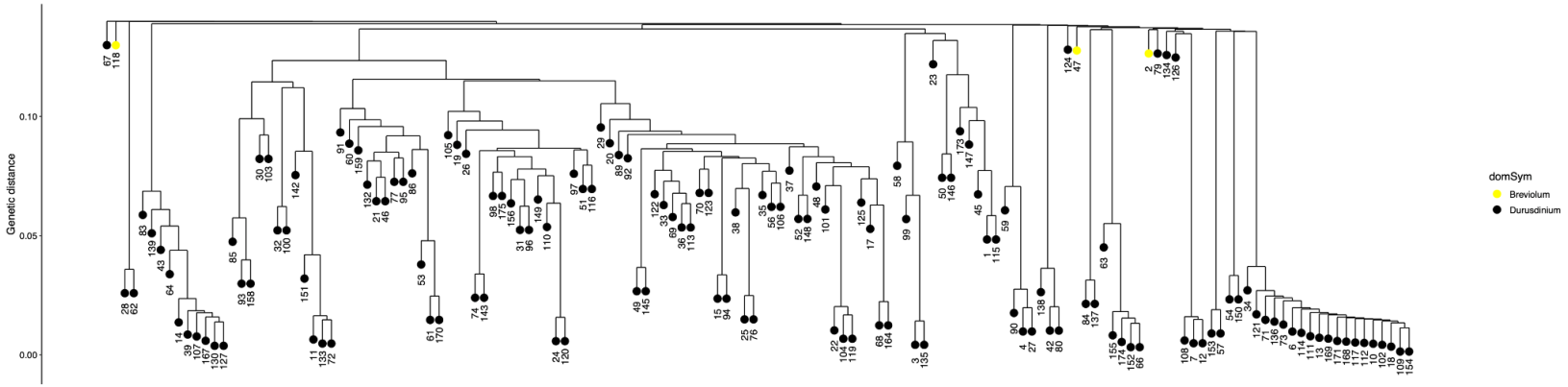


Figure 10. Dendrogram of unique WGS multi-locus genotypes (experimental genotypes) based on genetic distance. Colors indicate dominant Symbiodiniaceae genus based on read alignment (see section 3.1.4.).

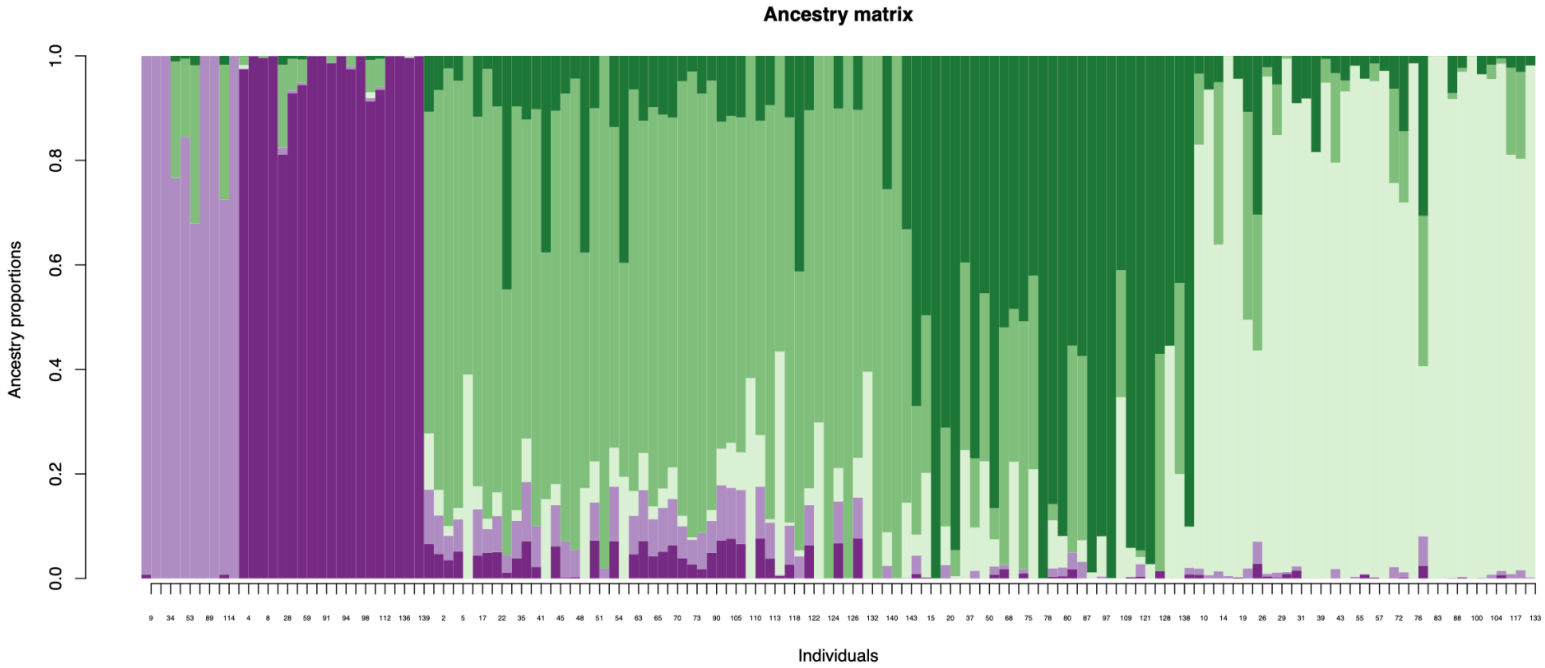


Figure 11. Structure bar plot denoting the relative likelihood of assignment of each WGS sample (column) to five ancestral populations, denoted by color.

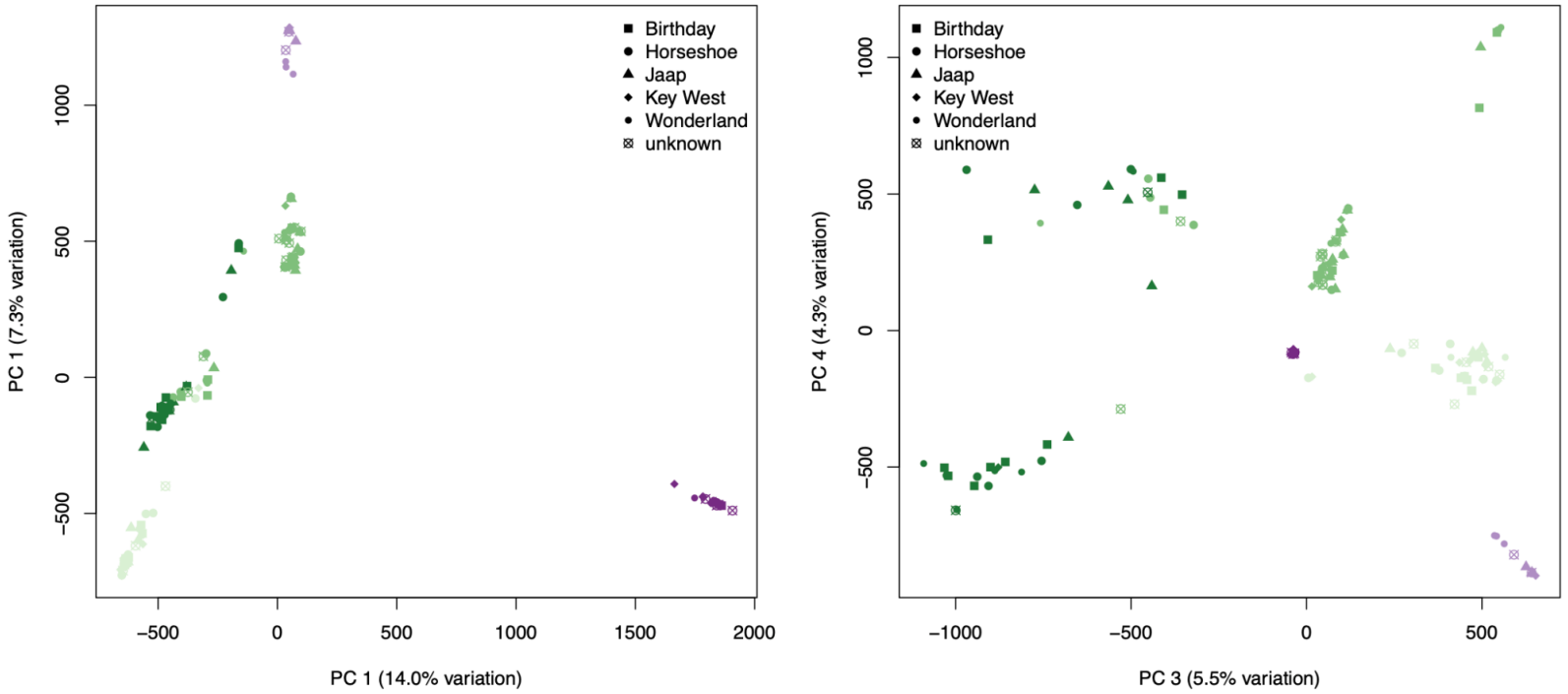


Figure 12. Principal component analysis (PCA) of WGS samples, where colors correspond to ancestral populations determined by structure analysis, and shapes correspond to sampling locations in the Florida Keys. Axes labels denote the amount of model variation explained by the respective axis.

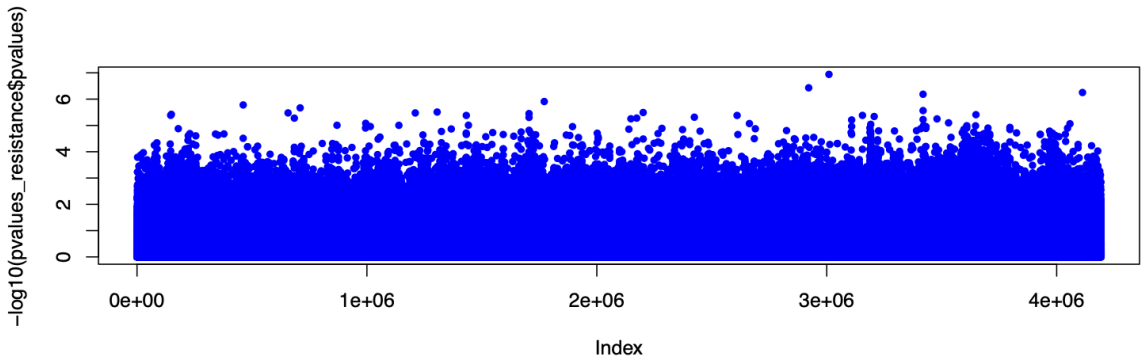


Figure 13. Manhattan plot of potential relationships between genome regions (SNPs; Index) and the proportion of healthy replicates per genotype (resistance). A -log p value threshold of 8 is considered significant.

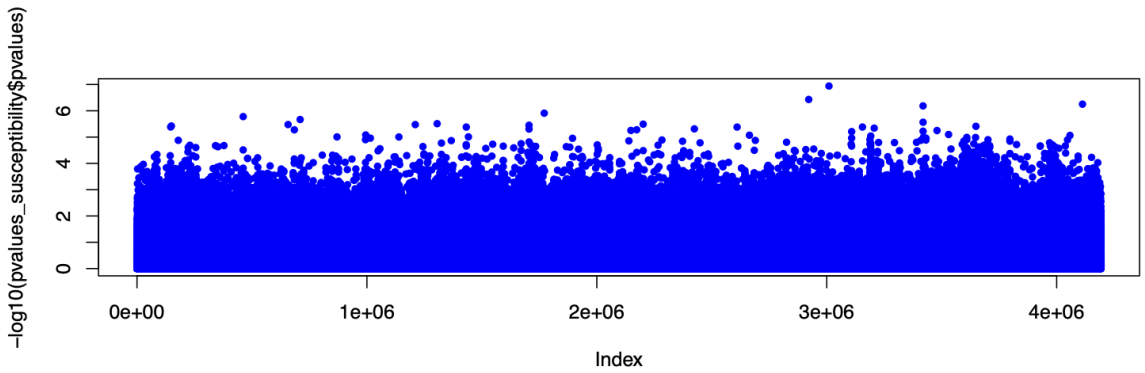


Figure 14. Manhattan plot of potential relationships between genome regions (SNPs; Index) and the proportion of diseased replicates per genotype (susceptibility). A -log p value threshold of 8 is considered significant.

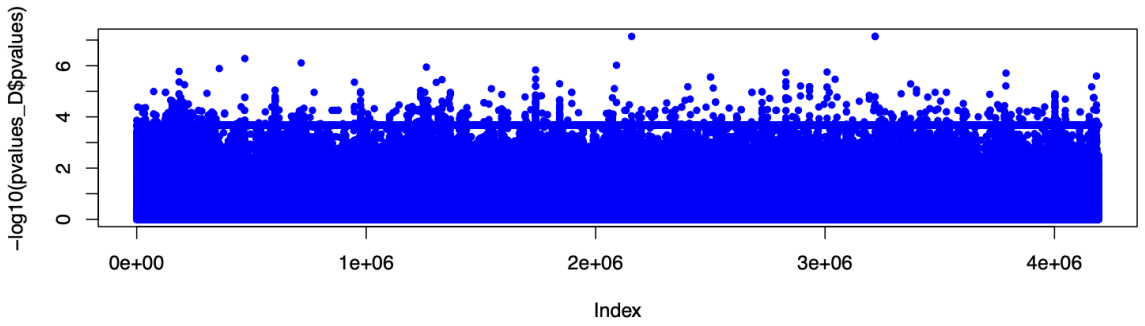


Figure 15. Manhattan plot of potential relationships between genome regions (SNPs; Index) and the proportion of *Durusdinium* abundance (*Durusdinium*). A -log p value threshold of 8 is considered significant.

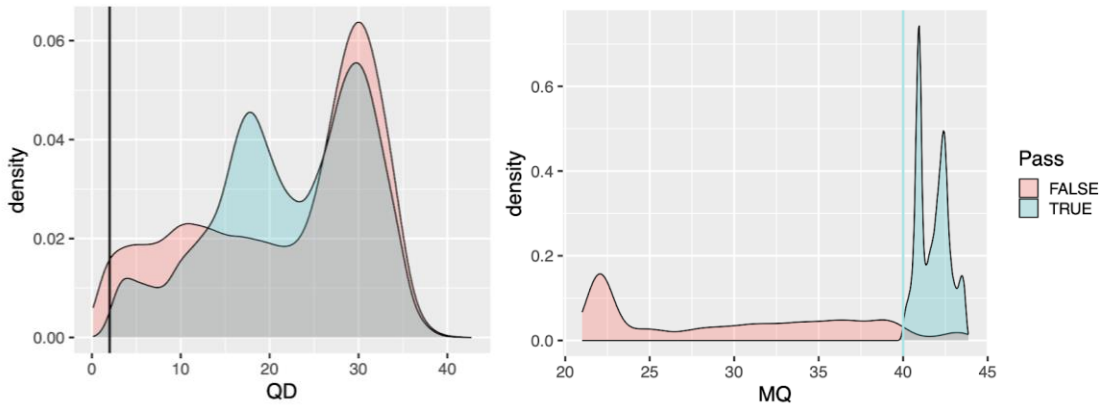


Figure 16. Density of 2bRAD SNPs calls across QualByDepth (QD), a metric of the confidence that a particular variant locus is high-quality (i.e., real, and not a false positive) before and after filtering at a QD threshold of 2. Two peaks at QD values of ~17 and ~30 represent homozygous and heterozygous variants, respectively. (right) Density of 2bRAD SNPs calls across MappingQuality (MQ), the mapping quality of the locus to the genome, before and after filtering at an MQ threshold of 40.

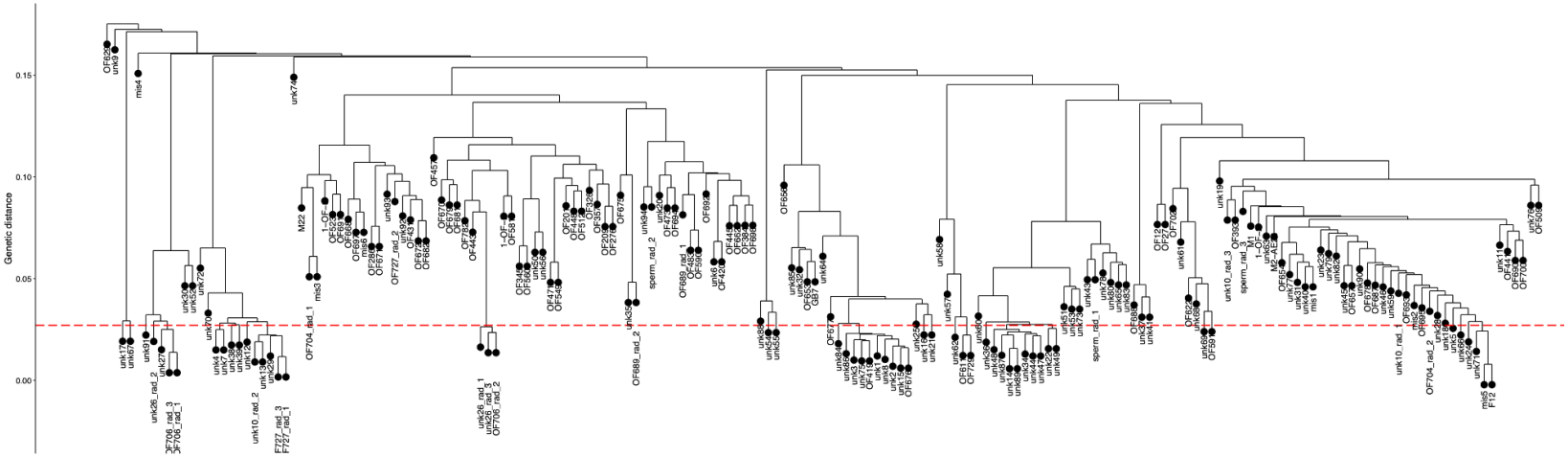


Figure 17. Dendrogram of 2bRAD samples (nursery genotypes) based on genetic distance. Horizontal dashed line indicates the clonal threshold used in clonal genotype detection. Any branches below the threshold are considered clonal genotypes.

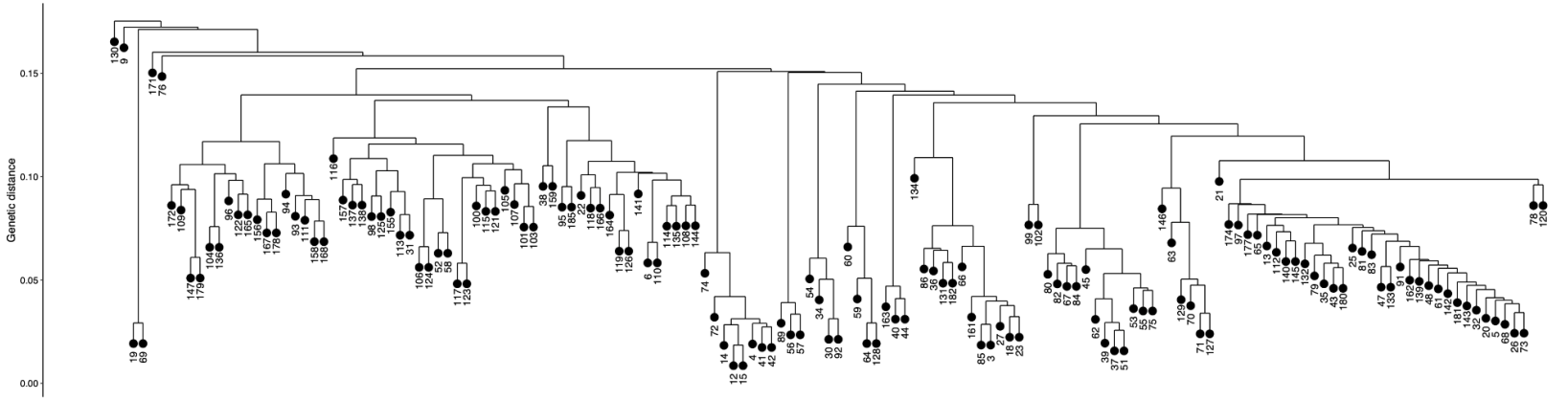


Figure 18. Dendrogram of unique 2bRAD multi-locus genotypes (nursery genotypes) based on genetic distance following removal of clonal genotype samples.

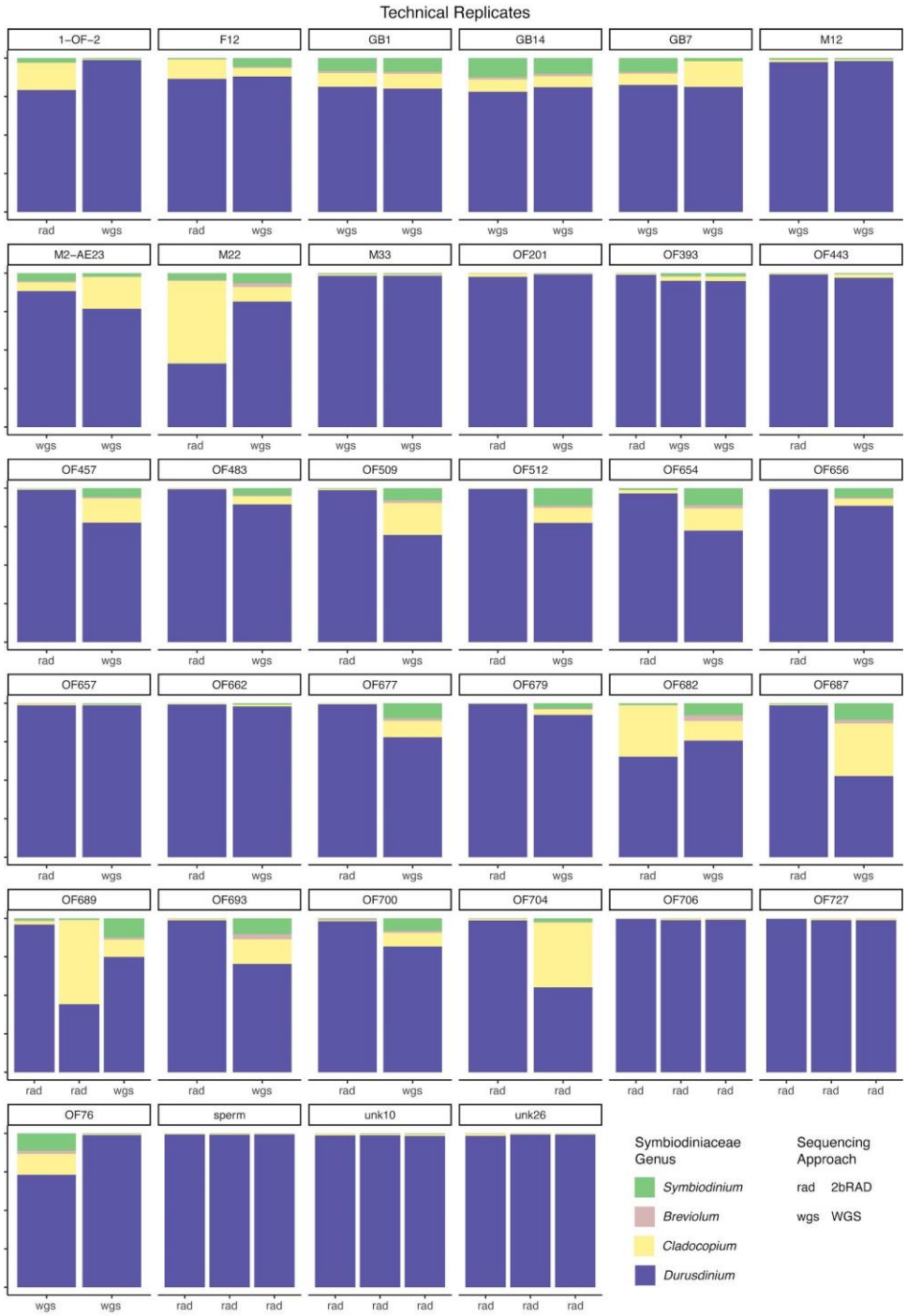


Figure 19. Relative abundance of Symbiodiniaceae genera across technical replicates of the experimental and nursery putative genotypes.

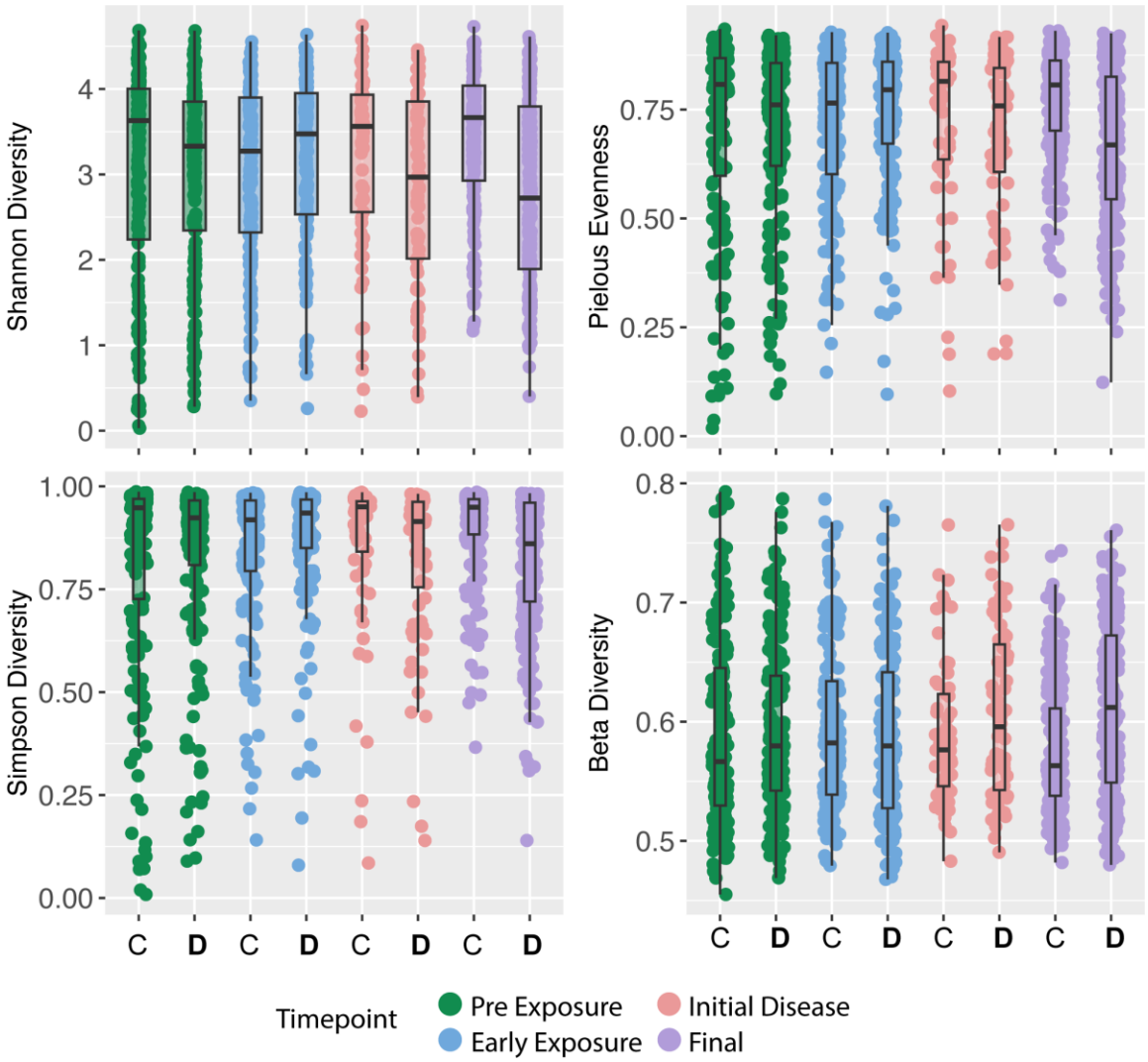


Figure 20. Diversity indices of *O. faveolata* across control and SCTLD-exposed four timepoints. The left panel shows alpha diversity metrics and the right beta-diversity.

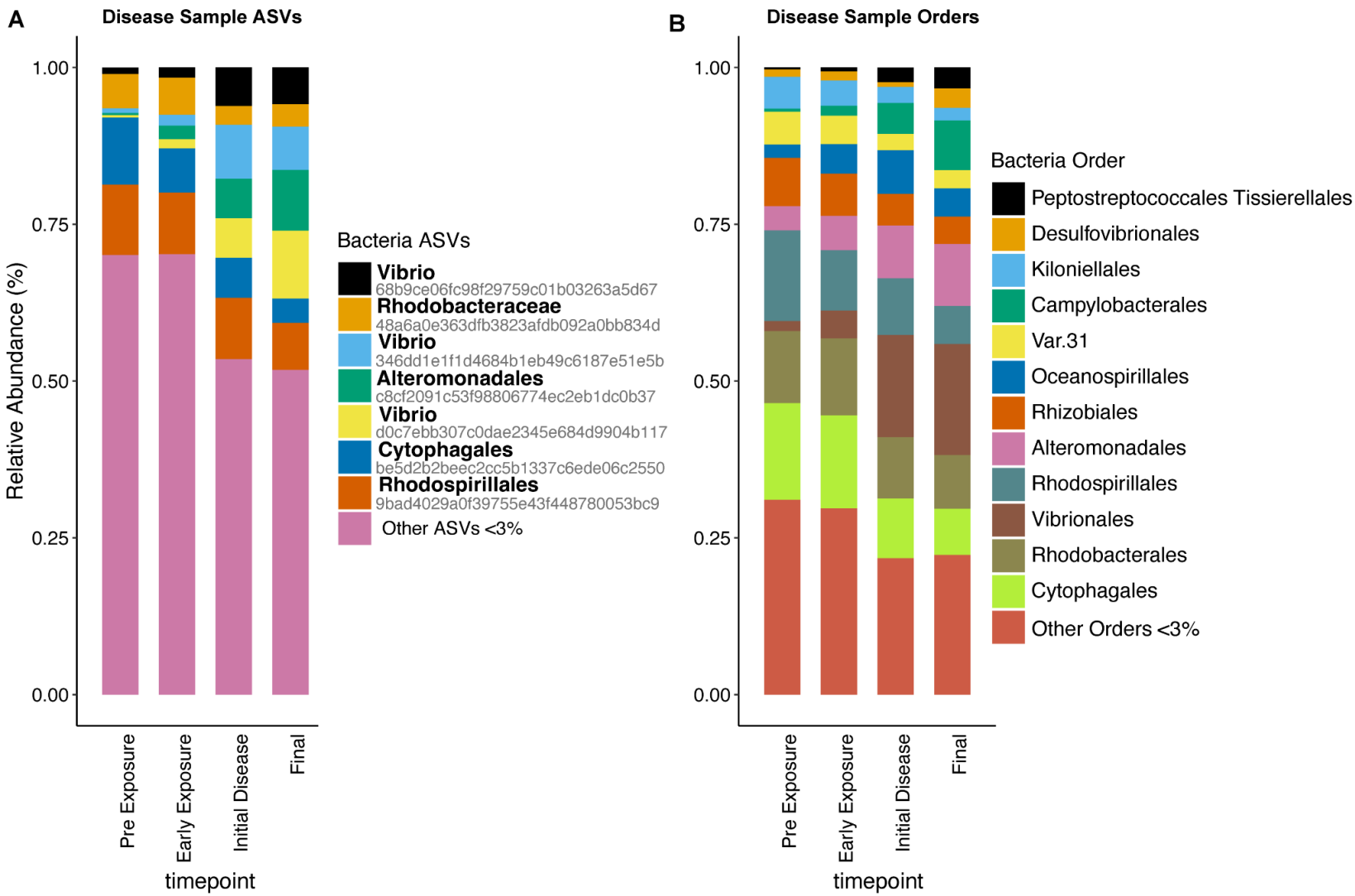


Figure 21. Relative abundance of SCTLD exposed samples at four time points (A) Stacked column graph of bacterial ASVs with a relative abundance greater than 3% in any time point. The legend includes the ASV's ID accompanied by their bacterial order. (B) Stacked column graph of bacterial Orders with a relative abundance greater than 3% in any time point. This demonstrates the shift in the most abundant ASVs and orders between disease timepoints.

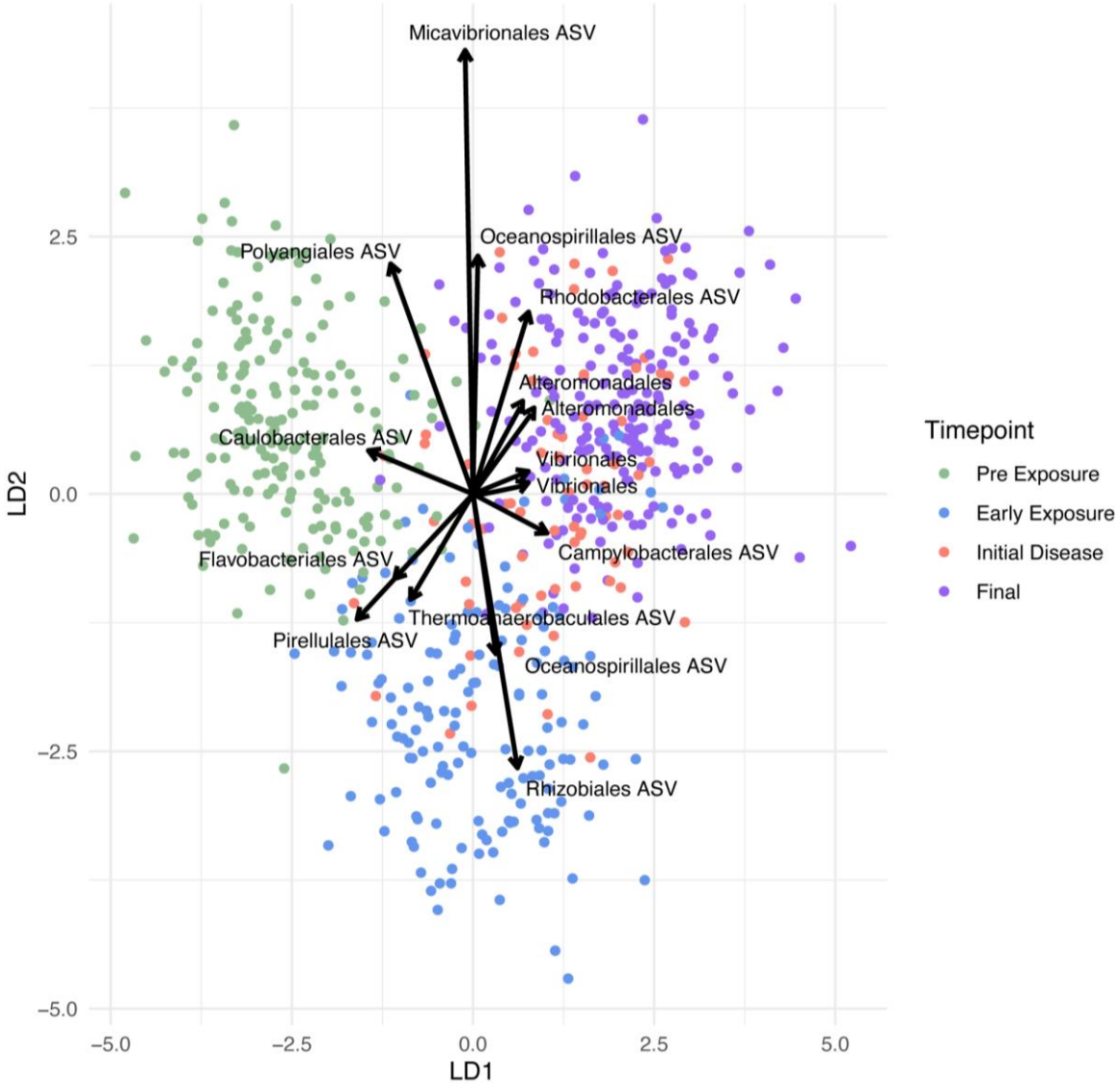


Figure 22. Linear discriminant analysis plot of *O. faveolata* fragments exposed to SCTLD and grouped by four time point.Vectors represent ASVs that are the most correlated to axes 1 and 2.

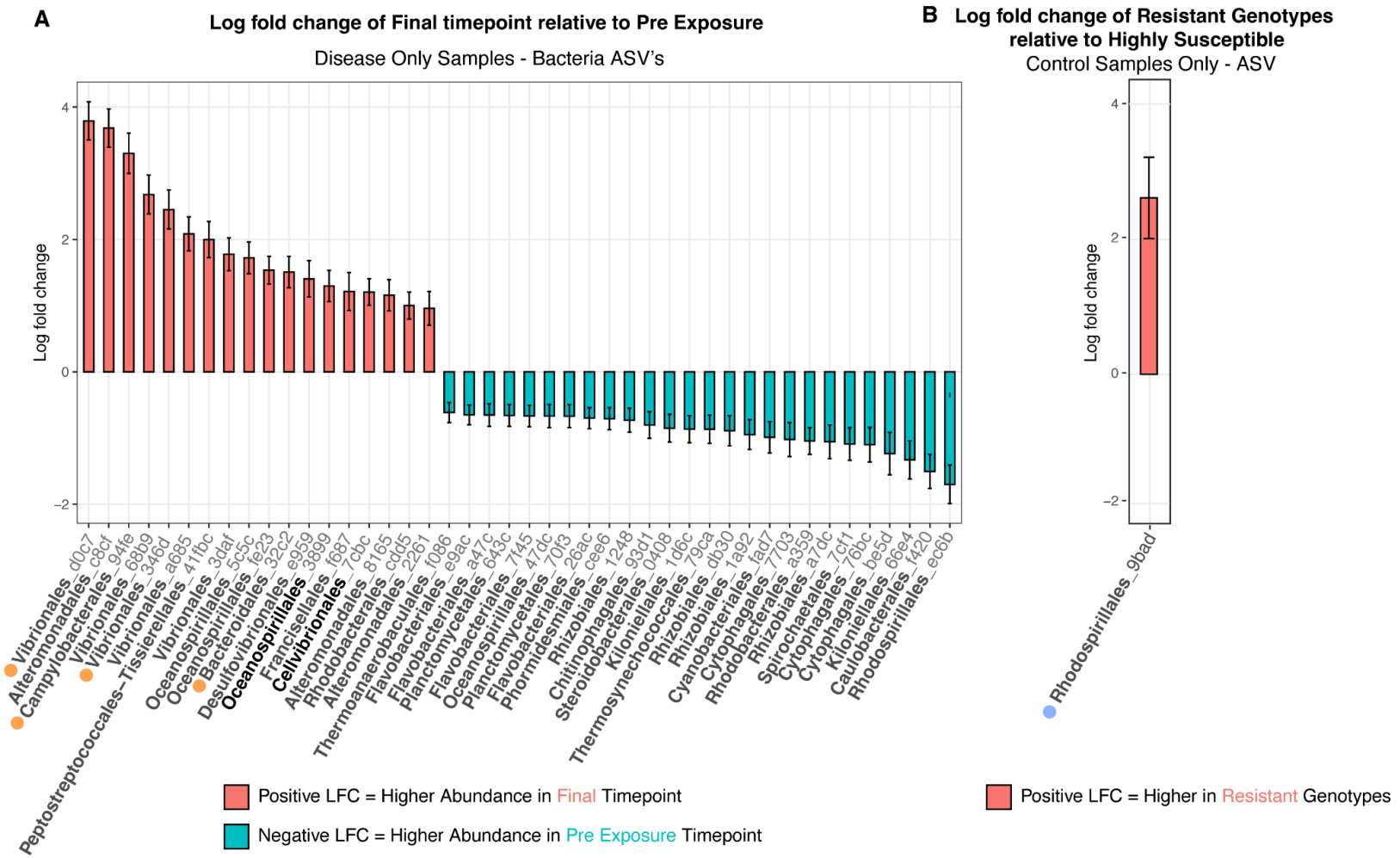


Figure 23. (A) Significant log-fold change of disease treatment samples between pre-exposure and final time point to demonstrate significant ASV-level shifts in response to SCTLD transmission calculated by ANCOM BCII with a Bonferroni correction. Orange dots indicate bacteria that have been classified as highly variable by the EVE analysis (B) Significant log-fold change of control samples between resistant genotypes relative to highly susceptible genotypes. ASVs are presented by their order name and an abbreviated ASV ID. The blue dot indicates *Rhodospirillales* has been classified as lineage-specific by the EVE analysis.

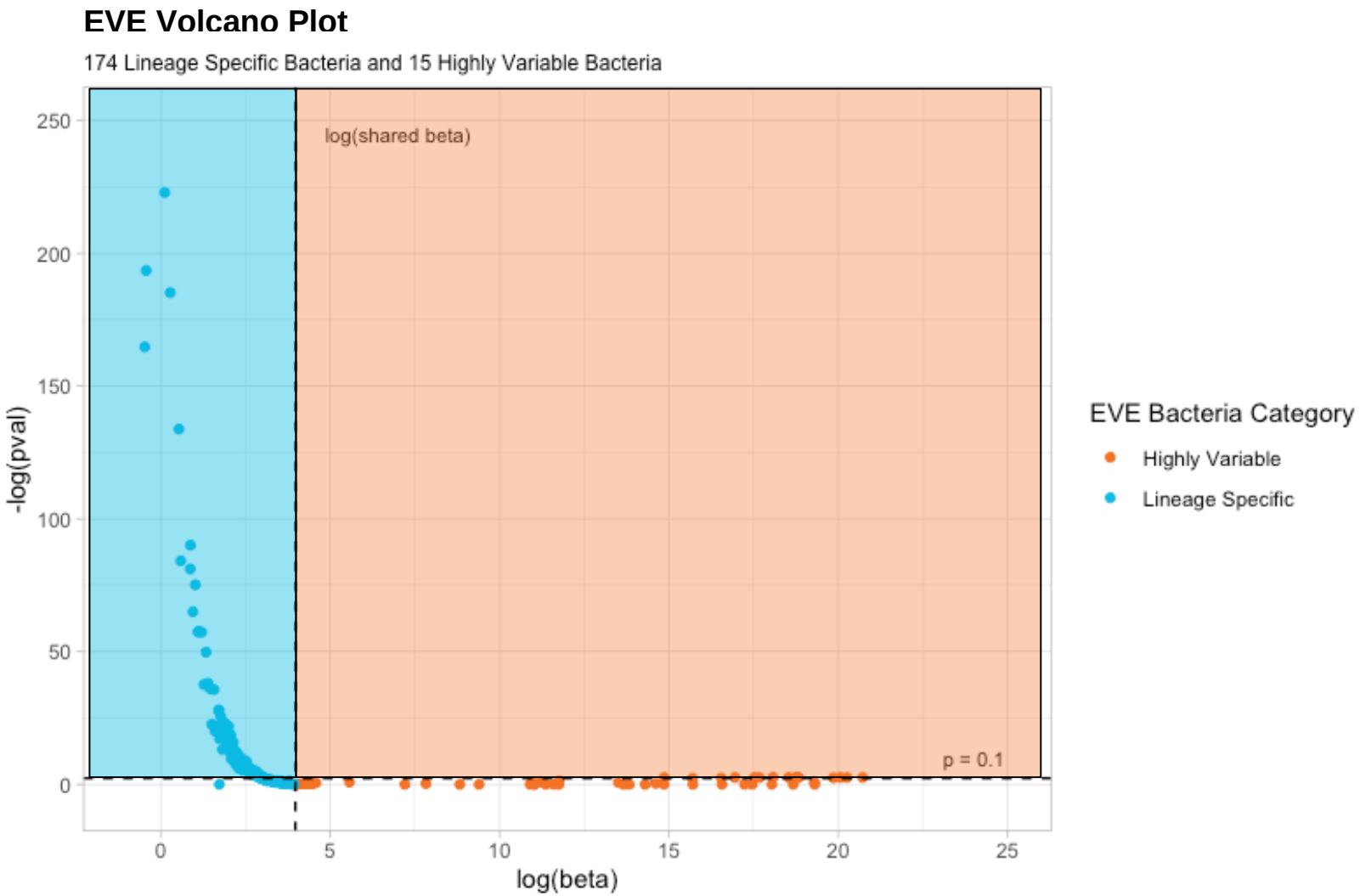


Figure 24. Results of the expression variance and evolution model that classified 171 bacteria as either lineage-specific (blue) or highly variable (orange). Each dot represents a bacteria ASV. The vertical line is the cutoff on whether the bacteria is categorized as lineage specific (blue dot) or highly variable (orange dot). The horizontal line indicates if that categorical assignment is significant ($p<0.1$).

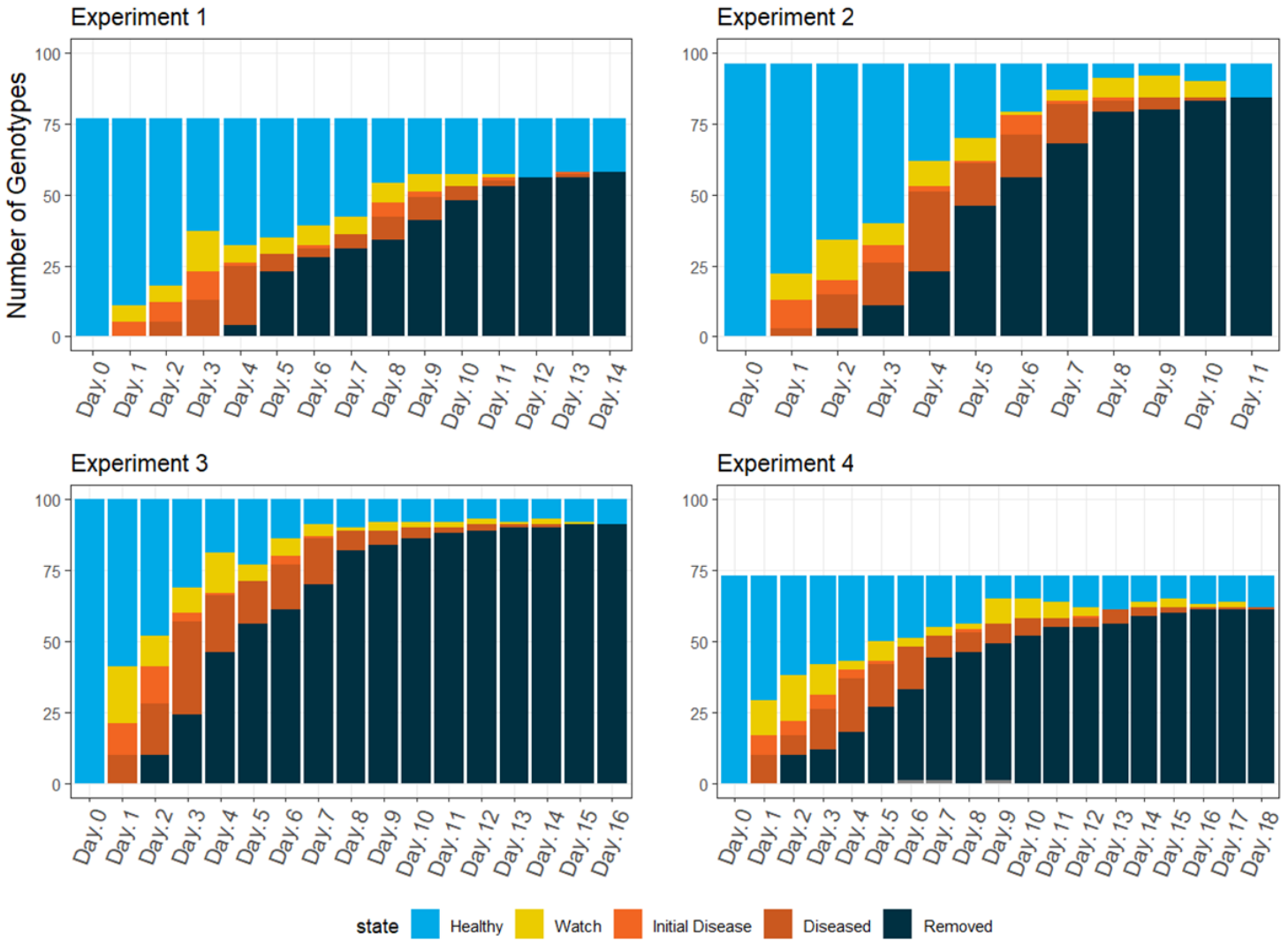


Figure 25. Disease incidence over time for all four transmission experiments visualized as the number of exposed individuals in one of five health categories: healthy, watch (stressed, but not clear disease signs), initial disease (first day of disease signs), diseased (clear disease signs), and removed (more than 10% tissue loss, sampled, and removed from the experiment).

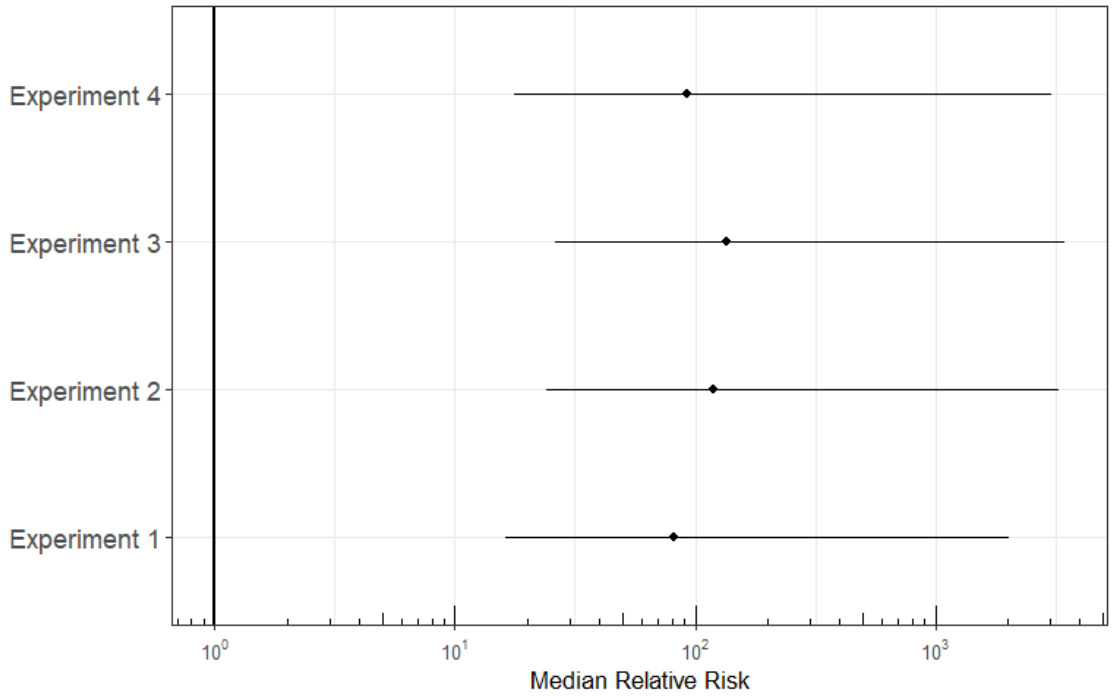


Figure 26. Caterpillar plot of the Bayesian relative risks on the log scale for *O. faveolata* fragments exposed to SCTLD in the four experiments conducted in 2022. Points are median relative risk values and lines denote the 95% credible intervals. Credible intervals entirely above (below) a relative risk of 1 (vertical black line at 10^0) (vertical black line at 100) indicate a significant increase (decrease) in disease risk after exposure to the risk. Credible intervals that include a value of 1 indicate no significant influence of exposure to the risk.

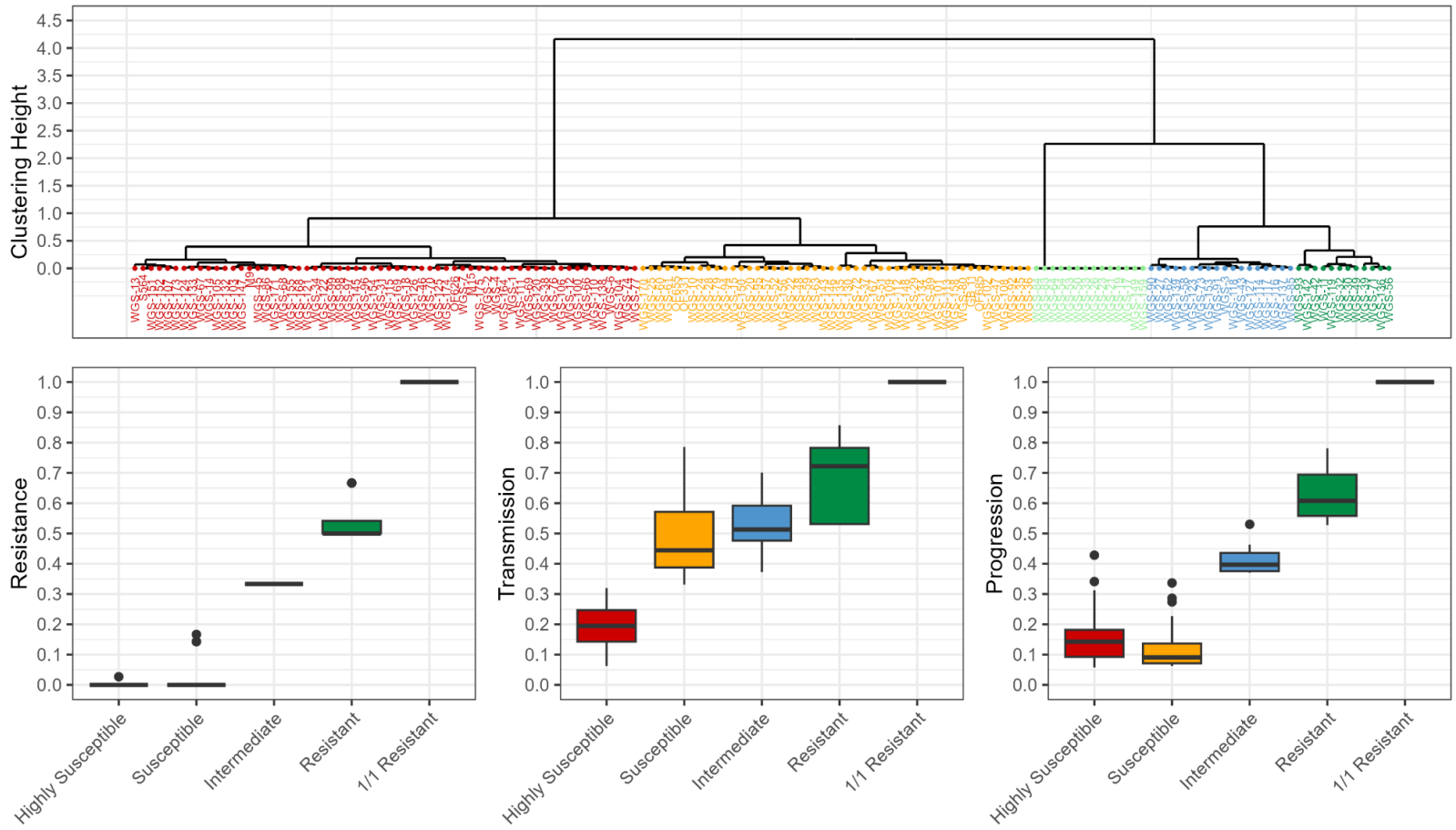


Figure 27. Results of the hierarchical clustering utilizing the susceptibility metrics measured from the four different stony coral tissue loss disease transmission experiments conducted on 154 different genotypes of *Orbicella faveolata*. The three different susceptibility metrics are shown (bottom) including the level of resistance (number of replicates that showed disease signs), rate of transmission (days to disease signs), and rate of progression (days until 10% of tissue loss). Higher values represent more resistant traits.

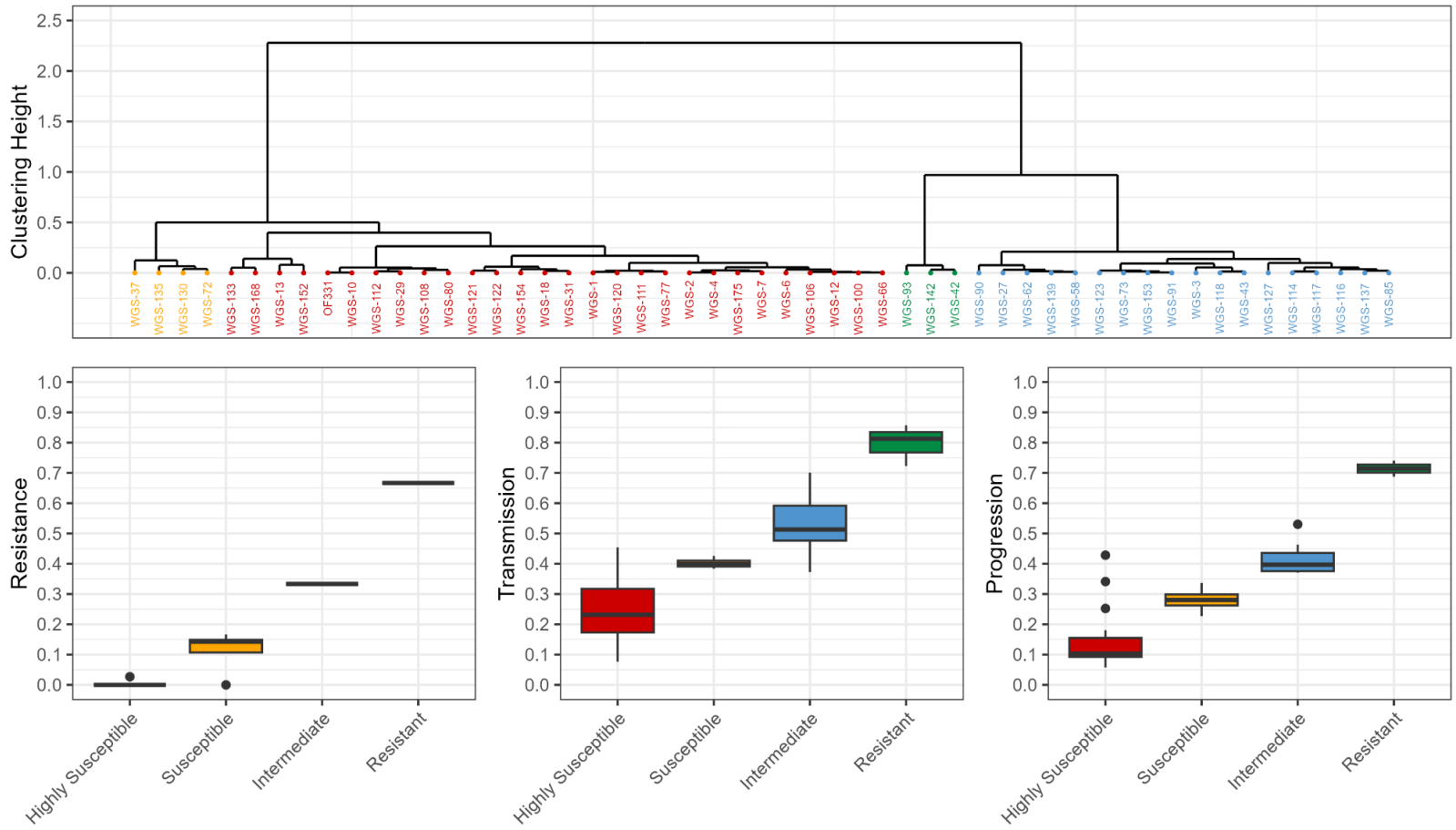


Figure 28. Results of the hierarchical clustering utilizing the susceptibility metrics measured from the four different stony coral tissue loss disease transmission experiments conducted on the 53 genotypes of *Orbicella faveolata* that had three or more replicates within the transmission experiments. The three different susceptibility metrics are shown (bottom) including the level of resistance (number of replicates that showed disease signs), rate of transmission (days to disease signs), and rate of progression (days until 10% of tissue loss). Higher values represent more resistant traits.

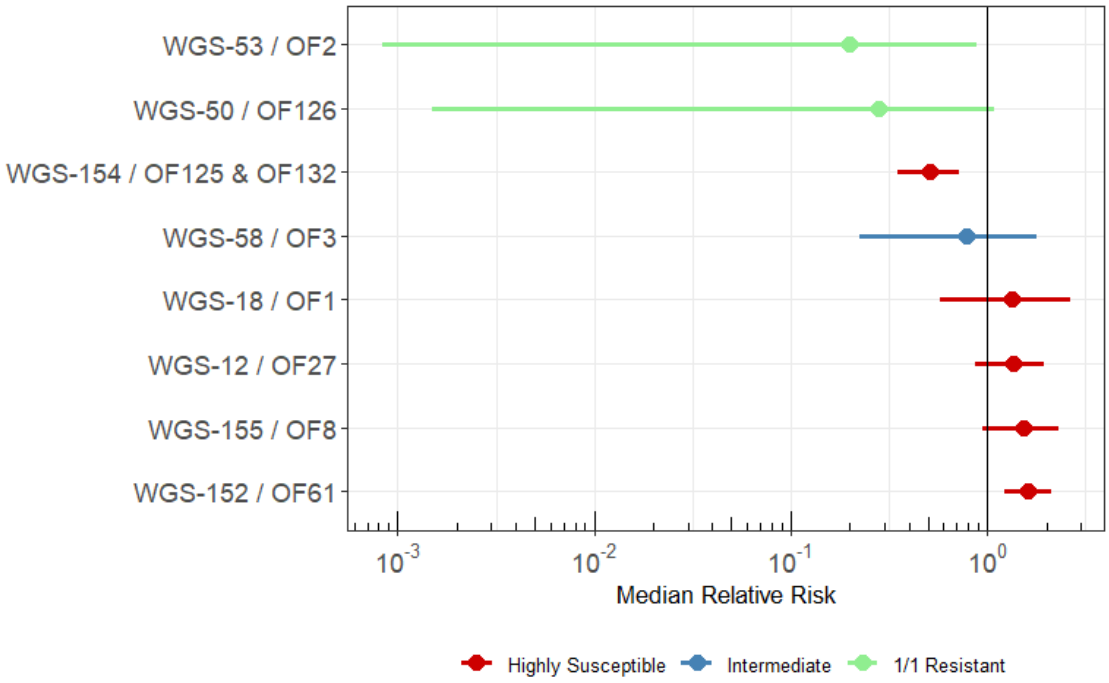
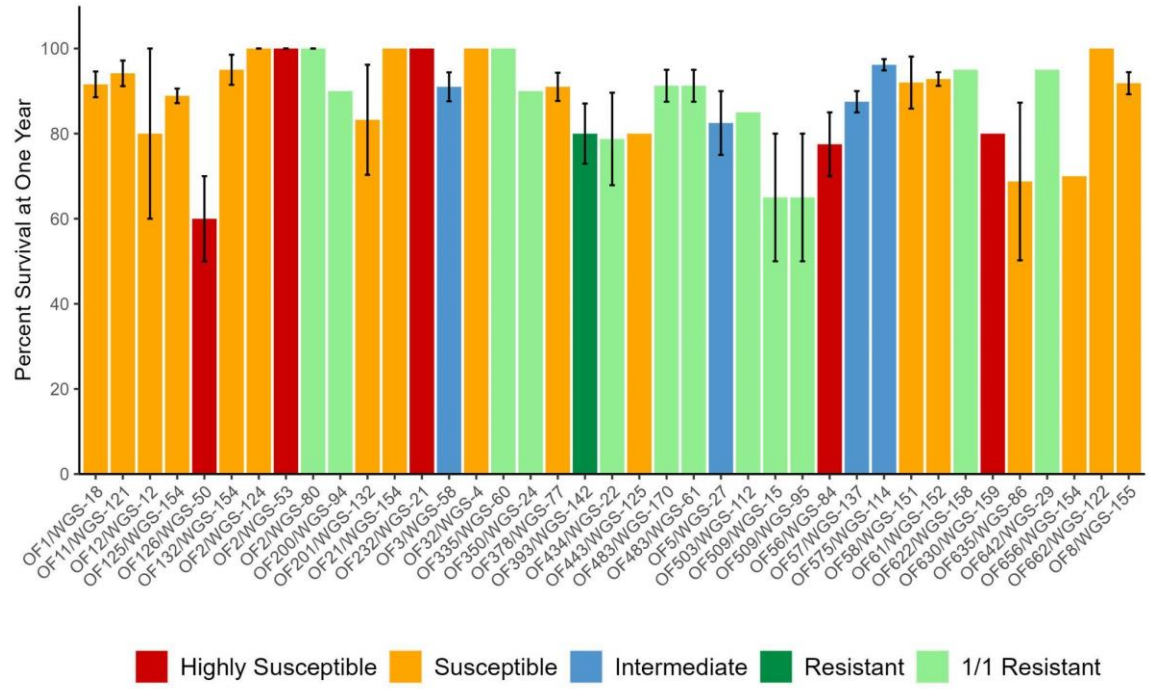


Figure 29. Caterpillar plot of the Bayesian relative risk analysis on the log scale. Relative risks for Mote Restoration *Orbicella faveolata* genotypes to SCTLD observed in the field study. Points are median relative risk values and lines denote the 95% credible intervals. Credible intervals entirely above (below) a relative risk of 1 (vertical black line at 10^0) (vertical black line at 100) indicate a significant increase (decrease) in disease risk after exposure to the risk. Credible intervals that include a value of 1 indicate no significant influence of exposure to the risk. Genotypes are colored by their susceptibility group determined by the hierarchical clustering analysis of the three disease metrics.



.

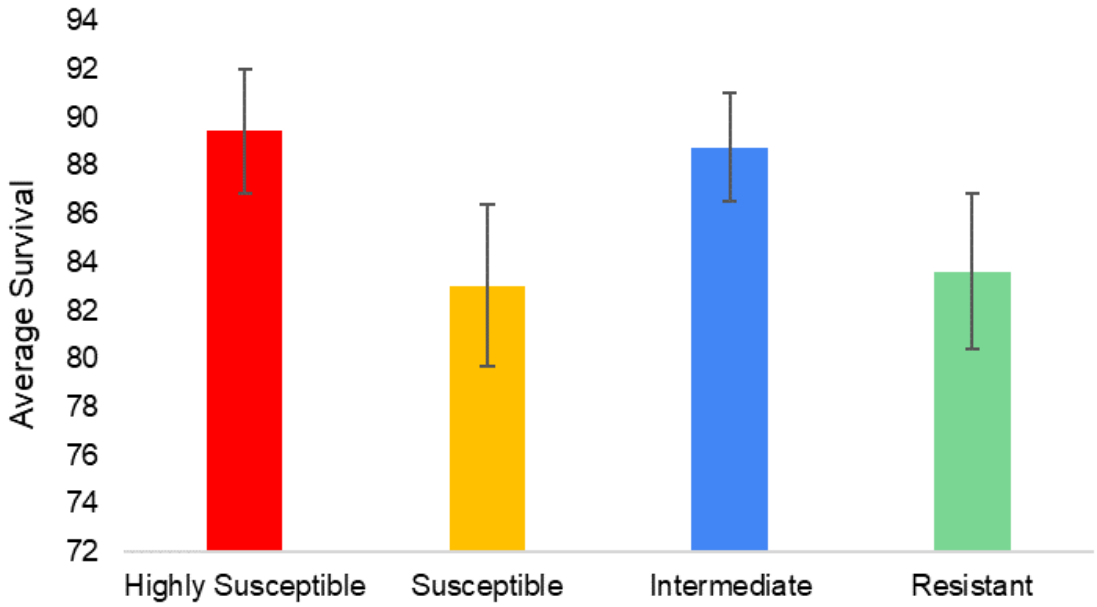


Figure 30. Average one-year survival rates of outplanted fragments of *Orbicella faveolata*. Top: data represented by different genotypes of *O. faveolata*. Bottom: data averaged by all susceptibility groups. Colors depict the SCTLD susceptibility grouping identified from the experimental transmission experiments. Error bars represent standard error of the mean.