Polyp Predicament: 2b-RAD reveals hidden dissimilarity in genus *Zoanthus* (Anthozoa: Hexacorallia)

By

Jeremy E. Smith

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of

Master of Science of Biology

Middle Tennessee State University

August, 2022

Thesis Committee:

Dr. Cole G. Easson, Chair

Dr. Jeff Leblond

Dr. Sarah Bergemann

Dr. Rebecca L. Seipelt-Thiemann

Abstract

Zoanthids (order Zoantharia, Gray 1832) are found in shallow reef systems throughout tropical oceans. These corals are also common in-home aquariums where their vibrant colors make them highly sought after. Over years of study these organisms have been grouped in a myriad of ways due to the complexity of identifying them morphologically. More recently, molecular techniques have been employed in zoanthid identification. Over the past twenty years researchers have produced many gene trees and used phylogenies to resolve zoanthid species relationships, but often produced contradictory results. In this study, we investigated 15 zoanthids that are popular within the aquarium trade, and one collected in the Florida Keys. Our aim was to identify potential linkages between morphology and genetics and determine how to best resolve species delimitations in the genus *Zoanthus*. 2b-RAD improved the ability to resolve species boundaries within the genus *Zoanthus* and a lack of consensus revealed a disconnect between molecular markers and morphology.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

ABBREVIATIONS

mtDNA	–	Mitochondrial DNA
nrDNA	–	Nuclear ribosomal DNA
COI	–	Cytochrome oxidase subunit I gene region
NGS	–	Next generation sequencing
2b-RAD	–	Type II restriction site-associated DNA genotyping
SNP	–	Single nucleotide polymorphism
PCR	–	Polymerase chain reaction
MLBS	–	Maximum likelihood bootstrap value
NMDS	–	Nonmetric multidimensional scaling
CCA	–	Canonical correspondence analysis
Fst	–	Fixation index
PTP	–	Poisson tree process
PIC	–	Phylogenetically independent contrast

Introduction

Nestled within the class Anthozoa is the order Zoantharia Gray, 1832 which consists of nine families containing 23 genera made up of 250 species (Low et al., 2016). These soft- bodied benthic corals, commonly referred to as zoanthids, are characterized by their internal mesenteric muscular structure and the incorporation of substrate and other materials into their mesoglea (excluding family Zoanthidae Rafinesque, 1815), with some (e.g., genus Palythoa Lamouroux, 1816) containing up to 45% of their overall weight as incorporated minerals (Reimer et al., 2010). Incorporating substrate into their tissue allows zoanthids to strengthen their polyp structure since they do not produce or rely on a calcium carbonate skeleton like stony corals, corals that produce calcium carbonate skeletons found in order Scleractinia also within class Anthozoa (Reimer et al., 2010). These soft corals are found in colonies of polyps throughout the planet’s oceans and are often found in shallow water coral reef systems. Many zoanthids associate with algal symbionts from the genera of Symbiodinaceae, though not all zoanthids possess these symbionts.

Zoanthids, specifically genera *Palythoa* and *Zoanthus*, are popular in the aquarium hobby trade due to their high variance of color and ease of husbandry. They are often labeled with elaborate common names in the aquarium trade and these names can vary greatly from one coral supplier to another. While variations of common names in the pet trade is typical, and often expected, it does not have the same rigor as scientific classification of organisms, i.e. taxonomy, and is thus less reliable. However, for zoanthid corals, morphological and phylogenetic classification can often lack consensus (Swain et al., 2016).

The field of zoanthid research has relied very heavily on a limited number of morphological characteristics, mainly microanatomical and internal differences, such as the location of the sphincter muscle in tissue layers or mesentery (an internal muscular structure found within soft corals that enables hydrostatic skeletons to function) counts to determine suborder, in the past (Reimer et al., 2010; Swain et al., 2016). Yet, the most commonly used characteristics for morphological identification used in the field are polyp column width and height, oral disc size, tentacle count, and color (Jaramillo et al., 2018; Reimer & Todd, 2009; Reimer, Foord, & Irei, 2012; Swain, 2016). However, this type of classification only loosely resolves taxonomic groups as Swain and Todd (2009) point out in their study where they state, "While identification to the generic level (*Zoanthus* or *Palythoa*) could be made with confidence, species level assignments often proved difficult." This difficulty is due to the high levels of phenotypic plasticity that are found in zoanthids and lack of qualitative characters that can be used to delineate species of zoanthids (Ryland & Lancaster, 2003).

However, modern technologies are allowing the field to progress in creating a more comprehensive and thorough understanding of the relatedness of these corals by combining morphological studies of tentacle musculature, mesentery arrangement, polyp size and diameter, and color morphology with genetic studies, or through robust genetic studies of multiple gene regions (Swain et al., 2016; Swain, 2018). As genetic research has grown and developed so too has our understanding of the relationships between difficult to identify zoanthids. Genetic analysis has helped to remove some of the burden of interpreting microanatomical structures, but it has brought its own suite of problems to overcome. Over the past twenty years researchers have produced many gene trees, including non-coding gene regions, and phylogenies to resolve the relationship of corals found within the order Zoantharia, but several gene trees have produced contradictory results (Reimer et al., 2006; Swain, 2018). In a review encompassing 107 zoanthid phylogenetic studies the vast majority, 94% specifically, relied solely on a single gene or fragment to explain the relatedness of species in the study (Swain, 2018). Though these studies attempted to create a species tree, the use of a single gene does not produce a species tree but instead produces a gene tree from which only conclusions on the gene's evolution, not the organism's, can be drawn (Swain, 2018).

More recently, mitochondrial markers were thought to be the ideal regions for barcoding organisms, using short segments of DNA to identify a specimen, around the world (Hebert et al., 2003). Though advancements in genetics have shown that a single universal barcode is not accurate, mitochondrial markers are still very useful in identifying organisms (Will et al., 2005). Moritz, Dowling, and Brown (1987) outlined the structure, versatility, and usefulness of mtDNA gene regions. Mitochondrial genes lack introns (a non-coding region) and the mitochondrial genome has varying divergence rates, which both lend to their usefulness in phylogenetic investigations (Moritz, Dowling, Brown; 1987). Early work on zoanthids focused on gene regions found within the mitochondrial genome that were thought to evolve in ways similar to higher metazoans, but this is not entirely accurate (Shearer et al, 2002). Cnidarian 16S mtDNA evolves at a rate similar to other organisms, however, the cytochrome oxidase subunit I (COI) region evolves drastically slower than higher metazoans (Shearer et al., 2002; Huang et al., 2008). The COI subunit in cnidarian evolution is nearly stagnant, as observed by Shearer et al. (2002) and Huang et al. (2008), yet several studies have utilized this region to resolve relationships of zoanthids (reviewed in Swain, 2018). Disappointingly, of the 100 attempted species trees utilizing a single gene region in the Swain (2018) review, 30% used only COI mtDNA to determine zoanthid relationships.

Some studies have used a combination of the highly conserved regions such as 16S mtDNA and COI mtDNA, the hypervariable internal transcribed spacer region (ITS nrDNA), or a combination of two or all of these (Reimer et al., 2006; Reimer et al., 2007a, Reimer et al., 2007b; Reimer & Todd, 2009; Sinniger et al, 2010). The most problematic combination DNA in many zoanthid studies are 16S mtDNA and COI mtDNA, both of which are not capable of accurately resolving relationships below the family level. Despite this, several studies have produced phylogenies attempting to resolve deep evolutionary relationships using only 16S mtDNA and COI mtDNA (reviewed in Swain, 2018). Since few studies have thoroughly examined zoanthid genetics, these faulty studies have had a disproportionate impact on our view of zoanthid relationships (Reimer et al., 2006; Reimer & Todd, 2009).

For example, Santos et al. (2016) utilized only sequence data from 16S mtDNA and COI mtDNA to identify and resolve the evolutionary relatedness of Brazilian coastal zoanthids. In this study they observed identical genetic identity among several samples and low reliability as measured by bootstrap values, a measurement of the confidence in phylogenetic results. These problems lessened their ability to identify and support relationships between specimens (Santos et al., 2016). They concluded that further tools are needed to identify zoanthids to the species level. The Santos et al. (2016) study underlines the risk of using highly conserved genetic regions to attempt to delineate species. To address this lack, researchers have strongly suggested zoanthid taxonomic methodology is in dire need of revision (Reimer et al., 2006 & 2010; Swain, 2018).

To revise our understanding of zoanthid taxa we need to revise the toolbox we utilize. Advancements in molecular technology have broadened the options researchers have to obtain greater amounts of data and information. Advancements in next generation sequencing (NGS) have developed a plethora of different techniques for researchers to adapt to their organisms of choice. One of these methods is 2b-RAD, a next generation sequencing method that is used to produce a large collection of short segments of sequences that are cut and amplified from across the genome (Wang et al., 2012). This method is a restriction site-associated DNA (RAD) genotyping method that sequences only short segments of DNA and reduces the complexity of genome analysis while allowing for the collection of large quantities of single nucleotide polymorphism (SNPs) data (Wang et al., 2012). The product of 2b-RAD is not a partial or whole genome assembly but a library of many DNA fragments that allows for a more informative analysis of these organisms to be created, when compared to only a handful of genetically informative sites produced through Sanger sequencing of conserved gene regions (Wang et al, 2012). SNPs data has become one of the ideal genetic data types to utilize, but for species with large genomes it can be obstructively costly to collect SNP data across an entire genome. 2b-RAD, however, allows for SNP data collection from across the genome while reducing the cost and allowing for reproducibility (Wang et al., 2012). It has also been shown to be powerful to identify relationships within non-model organisms where reference genomes are not available (Barbanti et al., 2020; Kelly et al., 2020; Wang et al. 2012). In Barbanti et al. (2020) 2b-RAD sequencing was used on non-model organisms with genomes of varying sizes using DNA of varying quality. They produced guidelines that optimize the cost and product quality of samples processed using the 2b-RAD method and underline its usefulness in non-model population genome studies (Barbanti et al., 2020). 2b-RAD sequencing has been used in this study because it allows for genome-wide analysis of SNPs within zoanthids, a non-model organism that lacks a reference genome (Wang et al., 2012).

The goal of this study is to identify a more powerful method for identifying similar zoanthid species along with producing data that will provide information and incentive for future studies to utilize NGS methods to increase our understanding of the relationships within these organisms. In this project it is expected that utilizing NGS, specifically 2b-RAD, will create a more resolved understanding of the relationship between the representative species included in this study. It is also expected that utilizing 2b-RAD will result in more robust, well-supported, and powerful results when compared to those obtained through Sanger sequencing.

Methods

Zoanthid sampling

Zoanthids from the genus *Zoanthus* were purchased from the aquarium trade. Corals of a similar color morphology (e.g., green oral disc), but not exact color type (green banded oral disc vs solid green oral disc), were preferred so that accurate species delimitation using SNPs will aid in identification of cryptic species that have varieties within similar yet different color morphologies, but a variety of color morphs were collected for comparison.

Fifteen large zoanthid colonies were purchased and grown, and a sixteenth was collected from the wild within the Florida Keys (Figure 1). Corals were reared in a 165-gallon closed aquarium system at a temperature of 78 – 80 degrees Fahrenheit, salinity was maintained at 35 particles per thousand (ppt), and a lighting regime with a twelve-hour photo period was used. These colonies possess visible differences in color and morphology (Figure 1). From each colony 6-8 polyps (0.5-1cm diameter) were collected for analysis (Figure 1).

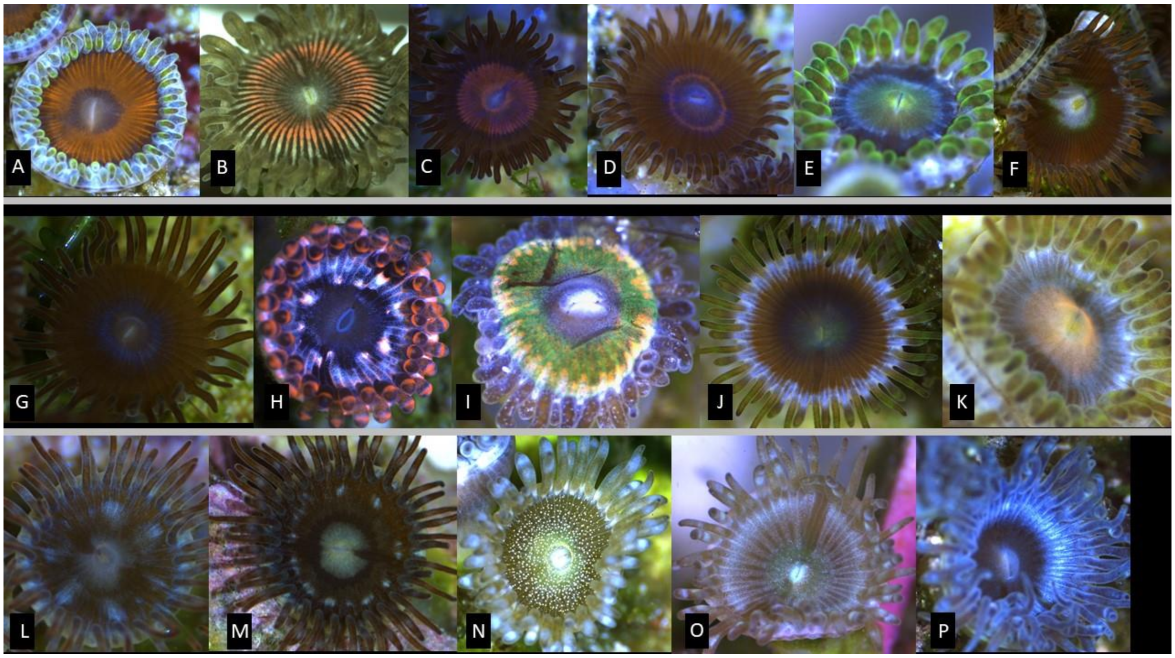


Figure 1 - Sample ID, Common Name, Scientific Name; JES201-A, Eagle Eye, *Zoanthus sansibaricus*; JES202-B, Unknown, *Zoanthus aff. pulchellus*; JES203-C, Gumball Machine Cracker, *Zoanthus sansibaricus*; JES204-D, Pink Panther, *Zoanthus sansibaricus* ; JES205-E, Dragon Eye, *Zoanthus sansibaricus*; JES206-F, Goblins on Fire, *Zoanthus sansibaricus*; JES207-G, Obsidian, *Zoanthus sansibaricus*; JES208-H, Armor of God, *Zoanthus kuroshio*; JES209-I, Sunny D, *Zoanthus cf. kuroshio*; JES2010-J, Dragon Eye, *Zoanthus sansibaricus*; JES2011-K, Green Bay Packers, *Zoanthus sansibaricus*; JES2012-L, Ultra Horizon, *Zoanthus sansibaricus*; JES2013-M, Venom X, *Zoanthus sansibaricus*; JES2014-N, Mohawk, *Zoanthus aff. pulchellus*; JES2015-O, Unknown, *Zoanthus aff. pulchellus*; JES2016-P, Field Collected, *Zoanthus aff. sociatus*. Scientific name determined through comparing ITS rDNA to available sequences in NCBI BLAST database.

Morphological data collection – When possible, multiple polyps (>5) from each sample colony were imaged and used to collect data that covers a range of possible morphological characteristics. Zoanthid polyp samples were imaged using a Leica DMC 4500 microscope (Leica Microsystems). Polyps were contained in a large view glass containing water from the rearing aquarium. Multiple measurements for polyp tentacle length, oral disc size, mouth width, and tentacle number were collected. Measurements were made inside the Leica imaging software to maintain accurate measurements across images. Data was compiled and averages for each morphological characteristic were calculated. Coloration metrics of zoanthid polyps were collected by first processing images using GIMP (GNU Image Manipulation Program v2.10.30) where image backgrounds were removed, and a collage reference image of all samples was assembled. The images were then processed in R (v 4.1.2) using the recolorize package (v. 0.1.0; Weller, 2021) where color proportions were collected (Appendix 8). The collected coloration data consisted of bins containing the colors black, brown, blue, white, green, yellow, and red which were considered to represent all colors present in the sample colonies. Color proportions were transformed into z-scores and assimilated into a table containing other morphological metrics.

Zoanthid DNA Sequencing and Alignment - DNA was extracted using the Thermofisher GeneJET Plant Genomic DNA Purification Mini Kit and the included protocol. 8mm steel beads were used to physically disrupt tissue using a QIAGEN TissueLyser II for 30 seconds at 20Hz. Extracted DNA was amplified through PCR using Platinum Hotstart PCR Master Mix (2x) (ThermoFisher) following the manufacturer's suggested protocol. The following primer sets were used to amplify 16S mtDNA, 12S mtDNA, and ITS nrDNA respectively: 16Sant0a (5' GAA GTA GGC TTG GAG CCA GCC A 3') and 16SbmoH (5' CGA ACA GCC AAC CCT TGG 3'), 12S1a (5' TAA GTG CCA GCM GAC GCG GT 3') and 12S3r (5' ACG GGC NAT TTG TRC TAA CA 3'), ITS-Zoan-f (5' CCT GAT CAT TTA GAG GGA GT 3') and ITS-Zoan-r (5' CGG AGA TTT CAA ATT TGA GCT 3') (Reimer et al., 2007; Sinniger et al., 2005; Sinniger et al., 2010). 16S mtDNA amplification was carried out using the protocol: 2 min at 94°C; then 35 cycles of: 1 min at 94°C, 30 sec at 52°C, 90 sec at 72°C; followed by a final elongation for 2 min at 72°C. 12s mtDNA amplification was performed using thermocycler protocol: 2 min at 94°C; then 40 cycles: 30 sec at 94°C, 1 min at 52°C, 2 min at 72°C; followed by a final 5 min elongation at 72°C (Sinniger et al., 2005). Finally, ITS nrDNA was amplified through the protocol: 2 min at 94°C; then 35 cycles of: 1 min at 94°C, 1 min at 50°C, and 2 min at 72°C; with a final elongation of 10 min at 72°C (Reimer et al., 2007).

The 16S mtDNA and 12S mtDNA sequences were aligned using both Multiple Sequence Alignment (MUSCLE) for automatic alignment within Geneious 10.2.6 (https://www.geneious.com) then manual alignment in Mesquite (v 3.6.1) (Edgar, 2004; Maddison and Maddison, 2019). Automated alignment of the ITS sequences were performed using PRANK (v.170427) followed by manual alignment in Mesquite (v 3.6.1) (Löytynoja and Goldman, 2008; Maddison and Maddison, 2019). Manual alignment of ITS sequences was necessary due to the hyper-variability of this gene region in zoanthids and the typical creation of over-aligned sequence regions leading to false homology in sequence data when ITS sequences are processed in automated alignment software (Swain, 2018). Once sequences were aligned, they were assembled into a concatenated sequence totaling 2,981 bases and partitioned into three separate gene sections of 16S mtDNA, 12s mtDNA, and ITS rDNA.

Symbiodinium 2b-RAD preparation - To ensure accurate sequencing of zooxanthellae symbiont DNA, cultures from each zoanthid colony were prepared. Using a glass Pasteur pipette drawn out to a fine point with a flame, zoanthid symbionts were physically collected from within the polyp. Collected material was placed on a microscope slide where another sterilized and drawn-out Pasteur pipette was used to specifically collect only zooxanthellae symbionts to avoid contamination of cultures. Collected symbionts were placed into culture flasks containing 60ml of F/2 liquid growth media. Cultures were grown for several weeks. Once grown to sufficient cell density, samples from the cultures were collected and DNA was extracted using a DNeasy Powersoil DNA extraction kit (QIAGEN). DNA was then processed following the procedure outlined below for 2b-RAD sequencing.

Symbiont DNA sequencing and alignment - Zoanthid symbiont DNA was extracted from zoanthid polyp tissue using the Thermofisher GeneJET Plant Genomic DNA Purification Mini Kit. Once DNA was extracted zooxanthellae specific primers were used to amplify only the zooxanthellae ITS2 nrDNA region for barcoding. The primer set ITSinfor2 (5’GAATTGCAGAACTCCGTG 3’) and ITS2CLAMP (5’CGCCCGCCGCGCCCCGCGCCCGTCCCGCCGCCCCCGCCCGGGATCCATATGCTTAAGTTCAGCGGGT 3’) were used to amplify the region (LaJeunesse and Trench, 2000). A touchdown thermocycler protocol was used: 3 min at 92°C followed by, then 35 cycles consisting of an initial annealing temperature of 62°C for 40 sec decreasing by 1°C every two cycles for 20 cycles until being held at 52°C for 15 more cycles with a 30 sec elongation step at 72°C in each cycle, final elongation was for 5min at 72°C (LaJeunesse and Trench, 2000). Amplified segments were sent to Molecular Cloning Laboratory (mclab.com; San Francisco, California) for sequencing. Recovered sequences were aligned using MUSCLE automated alignment (Edgar, 2004). The final aligned sequences were a maximum of 365 bases in length. Once aligned, sequences were compared to those found in NCBI's BLAST database to identify their respective clade.

Phylogenetic tree - A phylogenetic tree and gene tree for Sanger sequences derived markers from zoanthid DNA and symbiont DNA, respectively, were generated and support values were collected using a Maximum Likelihood (ML) analysis with ML bootstrap (MLBS) values generated through RAxML-HPC2 v8.2.12 (Stamatakis, 2006; Stamatakis et al., 2008). Maximum likelihood analysis was run with 1000 replications and 1000 MLBS replications to estimate support for nodes (Stamatakis, 2006). Tree topology was visualized using FigTree v1.4.4 (Rambaut 2012) and MLBS values were added. A phylogenetic tree was constructed to allow for the collection of a distance-based matrix to compare with trees using distance-based algorithms with data generated using 2b-RAD.

2b-RAD Sequencing – Extracted DNA was prepared for 2b-RAD genotyping by following the methods of Wang et al, (2012) with a few alterations outlined below. First, genetic DNA concentrated by ethanol precipitation and was digested in a master mix consisting of 10x Buffer R, 150 µM SAM, ALFI (2U), and nuclease-free water (NFW) for four hours at 37°C then the enzyme was inactivated by heating to 65°C for 20 min to produce fragments of equal length (36bp) (Wang et al., 2012). Next, adaptors were ligated to the digested sequence fragments using a ligation master mix of 10 mM ATP, 10x T4 ligase buffer, Adaptor 1 (2 µM) (5ILL_RG + anti_ILL), Adaptor 2 (2 µM) (3ILL_YG + anti_ILL), T4 DNA ligase, BSA (20mg/ml^{-1}), and NFW for a minimum of 16 hours at 4°C (Kelly et al., 2020) .

The ligated fragments were then amplified and barcoded by PCR with the master mix 10 mM(each) dNTP, 10 µM Ill-Lib1, 10 µM Ill-Lib2, 5x Q5 Buffer, BSA (20 mg/ml^{-1}), Q5 DNA Polymerase, NFW, and a unique forward and reverse barcode primer for each sample. Amplification of samples followed a thermocycler protocol of 94°C for 2 min; 30 cycles of 98°C for 5 sec, 60°C for 20 sec, 72°C for 10 sec; followed by a final elongation for 72°C for 10 min (Kelly et al., 2020).

Once fragments were amplified, they were gel purified to isolate target bands to remove primer dimers and residual DNA. The gel bands were sliced from the gel block, cleaned, and then sequenced using an Illumina MiSeq sequencing platform. Recovered sequences were used to construct a library.

Assembly of 2b-RAD sequences – 2b-RAD raw single end read quality was first assessed using FastQC (v0.11.9). Once read quality was checked, adapters were trimmed from sequences using cutadapt (v2.8) and sequences of only 36 base pairs were retained. Trimmed sequences were again checked for quality using FastQC (v0.11.9). Once sequence quality was checked they were loaded into Stacks (v2.59). A population map was created denoting each sample as its own population due to the unknown nature of their true population relationship, but with the assumption that they did not come from the same population. This population map was uploaded into Stacks. Once both the population map and trimmed single end reads are uploaded to Stackss a de novo reference genome was constructed and the pipeline ran using the denovo_map.pl command. The output for this command was then exported from Stacks using the “population” command where multiple file types were exported using alternate command options (Appendix 1; Stacks (v2.59)). The Fixation index (Fst) was also calculated between each sample and exported as a matrix showing sample divergence per comparison.

Nonmetric Multidimensional Scaling – Output files from Stacks were imported into R (v 4.1.2) where VEGAN (v2.5.7; Oksanen et al., 2020) was used to produce a plot using a nonmetric multidimensional scaling model (NMDS) of the 2b-RAD sequence data (Appendix 3). Maximum likelihood trees were also imported into RStudio where they were transformed into distance matrices using the ape package (v5.6-1). Vegan (v.2.5.7) was used to produce NMDS plots of the phylogeny and gene tree above, along with the morphological and color data table.

Population estimate: – A file formatted for use in Structure (v2.3.4) was output from Stacks (v2.59) and used to identify the number of populations present in the samples of this study (Pritchard et al., 2000). Due to the true number of populations being initially unknown, multiple assumed populations were tested and compared varying from 2 populations to 14 populations with a 10,000 burn-in period and 100,000 replications. Optimal population number was estimated by comparing estimated Ln probability of data in each assumed population estimate and using the output with the highest probability.

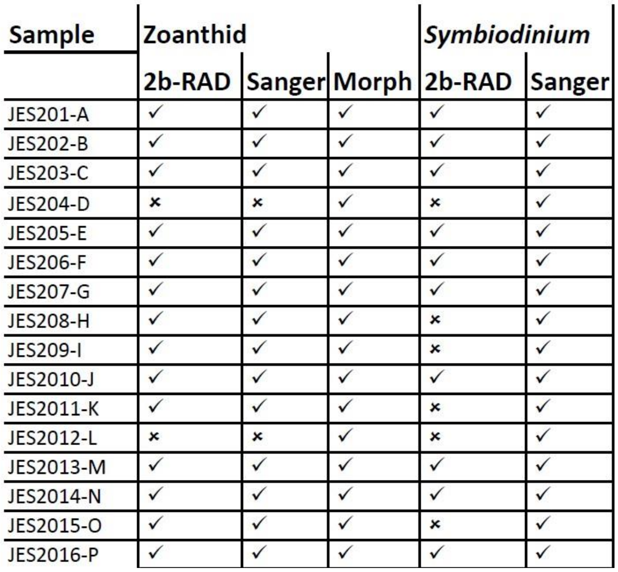
2b-RAD Bayesian Phylogenetic Tree - 2b-RAD sequence data was imported into BEAST (v2.6.7) from Stacks (v2.59) (Bouckaert et al., 2019). Stacks produced a Phylip format output consisting of 6,196 informative sites across the sequenced 2b-RAD data. A Bayesian tree was produced using the General Time Reversible (GTR) model, random local clock, and birth-death model. The program ran for 10,000,000 replications with a 100,000 burn-in and sampling every 1,000 replicates.

Species delimitation - To estimate species boundaries within our molecular markers, Bayesian and ML trees were processed using a Poisson Tree Process (PTP) model (Zhang et al., 2013). Trees were run with 500,000 Markov chain Monte Carlo (MCMC) iterations, a 10% burn-in, and 100 thinning. Likelihood trace plots were visualized to identify convergence of samples. Outputs were downloaded and used to identify species delimitations.

Data comparison - Sequence data from both 2b-RAD and Sanger sequencing along with morphological data were imported into R (v4.1.2). All data sets were transformed into distance matrices and compared through Mantel tests using Euclidean distance as the method in the ape package (v5.6-1) (Appendix 2). In vegan (v2.5.7; Oksanen et al., 2020) a canonical correspondence analysis (CCA) was performed between morphological data and zoanthid 2b-RAD data (Appendix 7). Also, the two datasets were compared using a stepwise CCA to determine if any single variable within the morphological data explained the observed variance (Appendix 7).

Samples used in each analysis varied and are outlined in Table 1. To keep datasets comparable, any sample that was not present in one of the datasets was treated as missing data.

Table 1 - Samples used in each analysis.



Sample	Zoanthid			*Symbiodinium*	
	2b-RAD	Sanger	Morph	2b-RAD	Sanger
JES201-A	✓	✓	✓	✓	✓
JES202-B	✓	✓	✓	✓	✓
JES203-C	✓	✓	✓	✓	✓
JES204-D	×	×	✓	×	✓
JES205-E	✓	✓	✓	✓	✓
JES206-F	✓	✓	✓	✓	✓
JES207-G	✓	✓	✓	✓	✓
JES208-H	✓	✓	✓	×	✓
JES209-I	✓	✓	✓	×	✓
JES2010-J	✓	✓	✓	✓	✓
JES2011-K	✓	✓	✓	×	✓
JES2012-L	×	×	✓	×	✓
JES2013-M	✓	✓	✓	✓	✓
JES2014-N	✓	✓	✓	✓	✓
JES2015-O	✓	✓	✓	×	✓
JES2016-P	✓	✓	✓	✓	✓

Results

Zoanthid 2b-RAD Data

To produce a high resolution dataset of the diversity and divergence of the 16 samples used in this study, the 2b-RAD method generated 46,776 loci composed of 1,683,948 bases were recovered containing 13,402 variant sites. Three individual samples and four distinct groups were found within the NMDS plot (Figure 3) and F_{st} table estimates (Table 3). The Bayesian tree (Figure 2) resolved three individual samples, one with low support (JES202-B, BPP = 0.5031), and five monophyletic clades. Using a cutoff F_{st} value of 0.45, our data suggest two of these clades are closely related and can reliably be grouped into a single clade (Figure 2, blue). One sample (JES206-F) possessed F_{st} values that placed it between two groups (Figure 2, blue and pink) suggesting it remains unresolved. Groupings within the phylogeny (Figure 2) are reflected in the NMDS plot (Figure 3) by clustering of points that fall into these distinct clades.

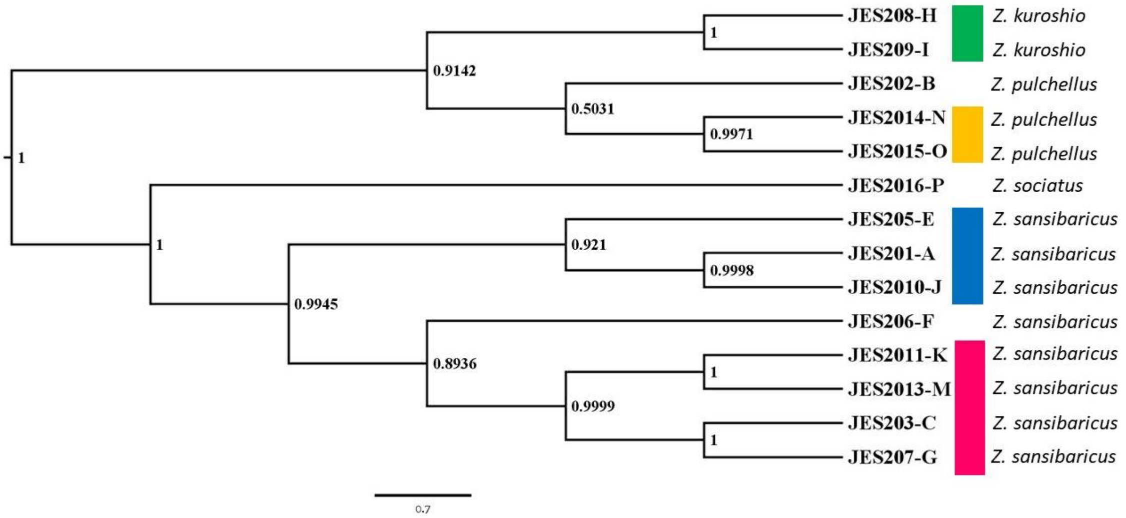


Figure 2 - Bayesian tree created using zoanthid 2b-RAD sequence data. Percentages represent Bayesian posterior probability with 100,000 burn-in and 10,000,000 replications. In the figure, color bars represent groupings determined by mean sample F_{st}, cutoff .45, calculated in Stacks (v2.59). Color bars placed next to identified clades.

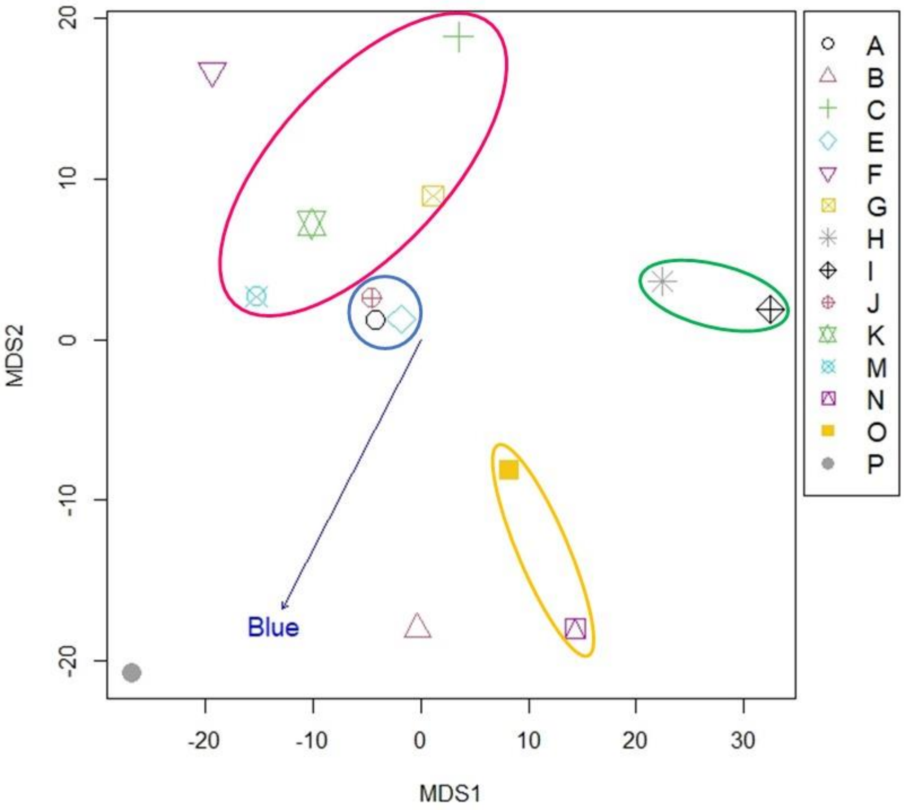


Figure 3 –NMDS of 2b-RAD sequence data. Ellipses denote groupings of F_{st} values below 0.45 for all samples in that group. Vector arrow correlates to the color blue from morphological data described below correspond to ellipses placed in figures 2 and 5 along with clades found in the ML tree (Figure 6). Species name identified through ITS nrDNA barcoding.

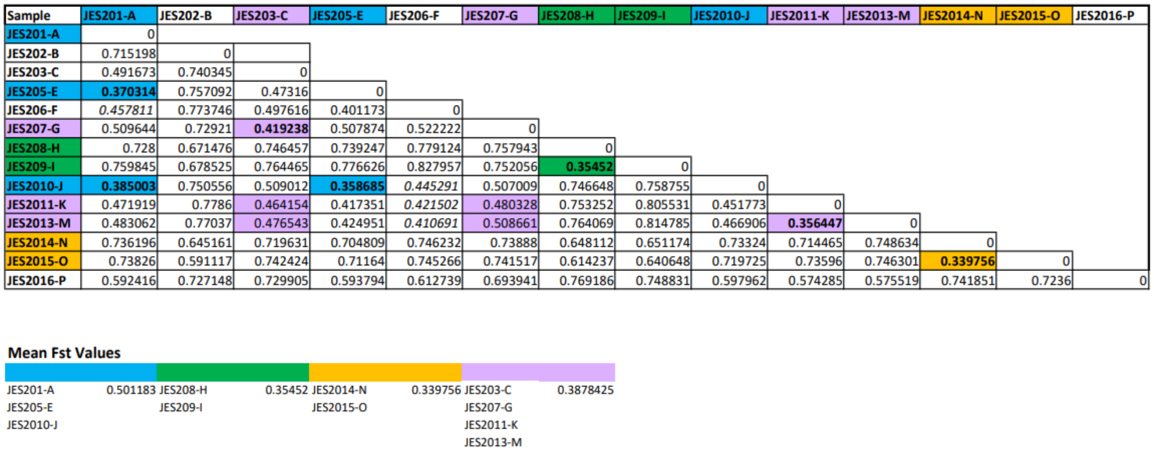
Results from the Poisson Tree Process (PTP) model show samples separated into 8 species (Table 2). Species delimitations are determined by support values, which vary across the dataset. Sample JES2016-P and JES206-F have high Bayesian posterior probability (0.967 and 0.768, respectively) suggesting high resolution between these samples and others in the dataset. Samples JES205-E (0.605) and JES202-B (0.565) separated from the remaining dataset, but with lower support values. All other samples had very low support (BPP < 0.50).

Table 2 - Species delimitations based on Poisson Tree Process (PTP) model of 2b-RAD sequence data (Zhang et al., 2013). Calculations based on 500,000 Markov chain Monte Carlo, 10% burn-in, and 100 thinning as inputs.

Species Group	Samples	Support
1	JES2016-P	96.7%
2	JES206-F	76.8%
3	JES2011-K, JES2013-M, JES203-C, JES207-G	21.6%
4	JES205-E	60.5%
5	JES201-A, JES2010-J	33.2%
6	JES202-B	56.5%
7	JES2014-N, JES2015-O	30.7%
8	JES208-H, JES209-I	46.6%

F_{st} values show similar patterns of separation to the PTP model. The F_{st} cutoff value was set at 0.45 due to groupings identified using this value being reflected in other datasets JES2016-P had a mean F_{st} of 0.677, while JES202-B had a high average F_{st} value of 0.717 making it highly divergent from other samples (Figure 3). Mean F_{st} values for each group are outlined in Table 2. Euclidean distances (generated to produce Figure 3) among samples were congruent with F_{st} and PTP model values. The mean Euclidean distance for all samples is 49.124, however the mean Euclidean distance for JES202-B, JES206-F, and JES2016-P were 51.252, 51.861, and 55.556, respectively, suggesting these samples were dissimilar from others in the 2b-RAD distance matrix.

Table 3 - Fixation index (F_{st}) table produced by STRUCTURE (v2.3.4). F_{st} cutoff 0.45 used to separate and group samples into related populations. Colors coordinate with groupings found in Figures 2 and 3.



Sample	**JES201-A**	**JES202-B**	**JES203-C**	**JES205-E**	**JES206-F**	**JES207-G**	**JES208-H**	**JES209-I**	**JES2010-J**	**JES2011-K**	**JES2013-M**	**JES2014-N**	**JES2015-O**	**JES2016-P**
JES201-A	0													
JES202-B	0.715198	0												
JES203-C	0.491673	0.740345	0											
JES205-E	**0.370314**	0.757092	0.47316	0										
JES206-F	*0.457811*	0.773746	0.497616	0.401173	0									
JES207-G	0.509644	0.72921	**0.419238**	0.507874	0.522222	0								
JES208-H	0.728	0.671476	0.746457	0.739247	0.779124	0.757943	0							
JES209-I	0.759845	0.678525	0.764465	0.776626	0.827957	0.752056	**0.35452**	0						
JES2010-J	**0.385003**	0.750556	0.509012	**0.358685**	*0.445291*	0.507009	0.746648	0.758755	0					
JES2011-K	0.471919	0.7786	0.464154	0.417351	*0.421502*	0.480328	0.753252	0.805531	0.451773	0				
JES2013-M	0.483062	0.77037	0.476543	0.424951	*0.410691*	0.508661	0.764069	0.814785	0.466906	**0.356447**	0			
JES2014-N	0.736196	0.645161	0.719631	0.704809	0.746232	0.73888	0.648112	0.651174	0.73324	0.714465	0.748634	0		
JES2015-O	0.73826	0.591117	0.742424	0.71164	0.745266	0.741517	0.614237	0.640648	0.719725	0.73596	0.746301	**0.339756**	0	
JES2016-P	0.592416	0.727148	0.729905	0.593794	0.612739	0.693941	0.769186	0.748831	0.597962	0.574285	0.575519	0.741851	0.7236	0

Mean Fst Values

JES201-A	0.501183	JES208-H	0.35452	JES2014-N	0.339756	JES203-C	0.3878425
JES205-E		JES209-I		JES2015-O		JES207-G	
JES2010-J						JES2011-K	
						JES2013-M	

Zoanthid Sanger Sequencing

To directly compare the samples used in this study to those of previous studies Sanger sequencing was used. This method recovered sequence lengths totaling 2,981 base pairs which contained 160 informative sites throughout the concatenated sequences of 12S, 16S and ITS regions. The recovered maximum likelihood (ML) tree (Figure 4) from zoanthid DNA produced five supported clades (MLBS = 100; Figure 4). Nearly all nodes within the largest clade were not well supported (MLBS < 75), however the second largest clade had high internal support values (MLBS > 75; Figure 4). Two clades contained a single sample (JES203-C, and JES2016-P), all with an MLBS of 100.

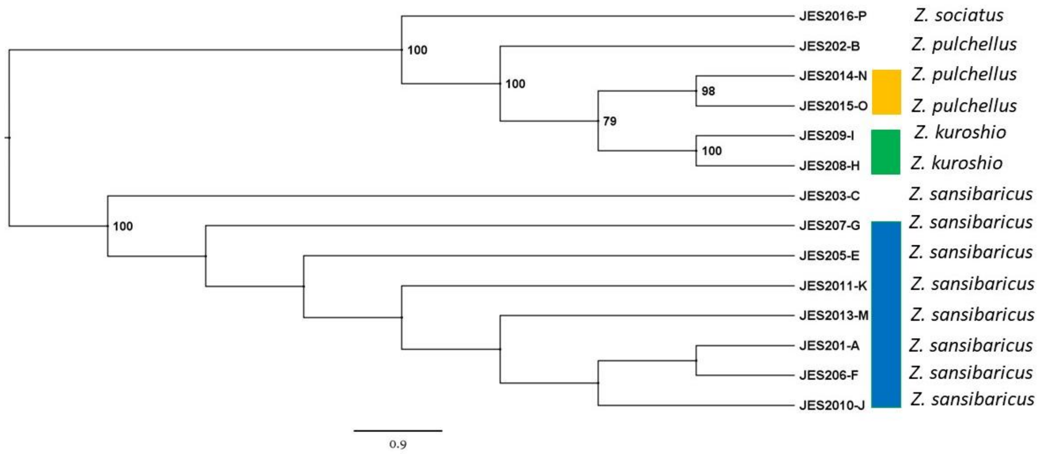


Figure 4 - Maximum likelihood phylogenetic tree with scale bar produced from concatenated 16S mtDNA, 12S mtDNA, and ITS rDNA. Tree produced through RAxML using 1000 replications and 1000 ML bootstrap replications to estimate support for nodes (Stamatakis, 2006). MLBS values < 75 have been removed from the tree. Species name identified through ITS nrDNA barcoding with samples possessing 100% identity being considered identical.

Samples were placed into five delimited species by the PTP model (Table 4). All species groups possessed Bayesian posterior support values greater than 98%. JES2016-P again is placed into its own species bin with 100% support. Species group 3 was the largest species bin with samples A, C, E, F, G, J, K, and M (Table 4) being placed in it.

Table 4 - Species delimitations based on Poisson Tree Process (PTP) model of Sanger sequence data (Zhang et al., 2013). Calculations based on 500,000 Markov chain Monte Carlo iterations, 10% burn-in, and 100 thinning as inputs.

Species Group	Samples	Support	Identity based on ITS nrDNA
1	JES2016-P	100%	*Zoanthus sociatus*
2	JES209-I, JES208-H	99.8%	*Z. kuroshio*
3	JES205-E, JES203-C, JES207-G, JES2011-K, JES2013-M, JES201-A, JES206-F, JES2010-J	98.6%	*Z. sansibaricus*
4	JES202-B	100%	*Z. pulchellus*
5	JES2014-N, JES2015-O	98%	*Z. pulchellus*

Within the NMDS (Figure 5) samples JES202-B and JES2016-P grouped independently of the other samples, as observed in the ML phylogeny (Figure 4). Clades found within the phylogeny (Figure 4) related to clustering observed in NMDS plot (Figure 5), these clusters reflected the grouping of samples found in the clades. Samples from within the largest clade of figure 4 group near each other, yet samples A, F, J, K, and M all group together while C, E, and J group nearby. The mean Euclidean distance for all samples was 0.556 while samples JES202-B and JES2016-P had a mean distance of 0.676 and 0.633 respectively.

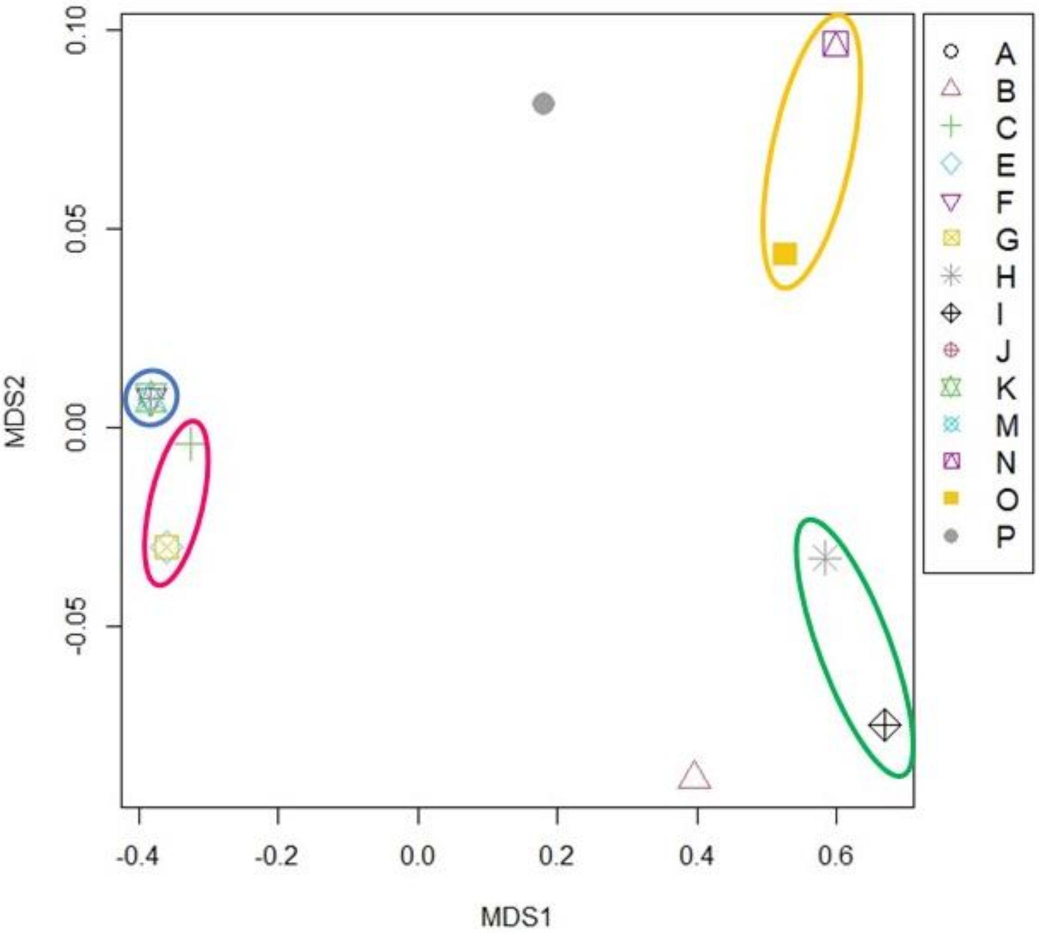


Figure 5 - NMDS of zoanthid Sanger sequencing data. Ellipses reflect groupings found in 2b-RAD Bayesian tree (Figure 2).

Morphological data

By using four commonly used morphological characteristics and the proportion of seven colors across all samples we were better able to understand the relationship of morphology to genetic data. A NMDS plot was used to visualize dissimilarity in morphological features across samples (Figure 6).

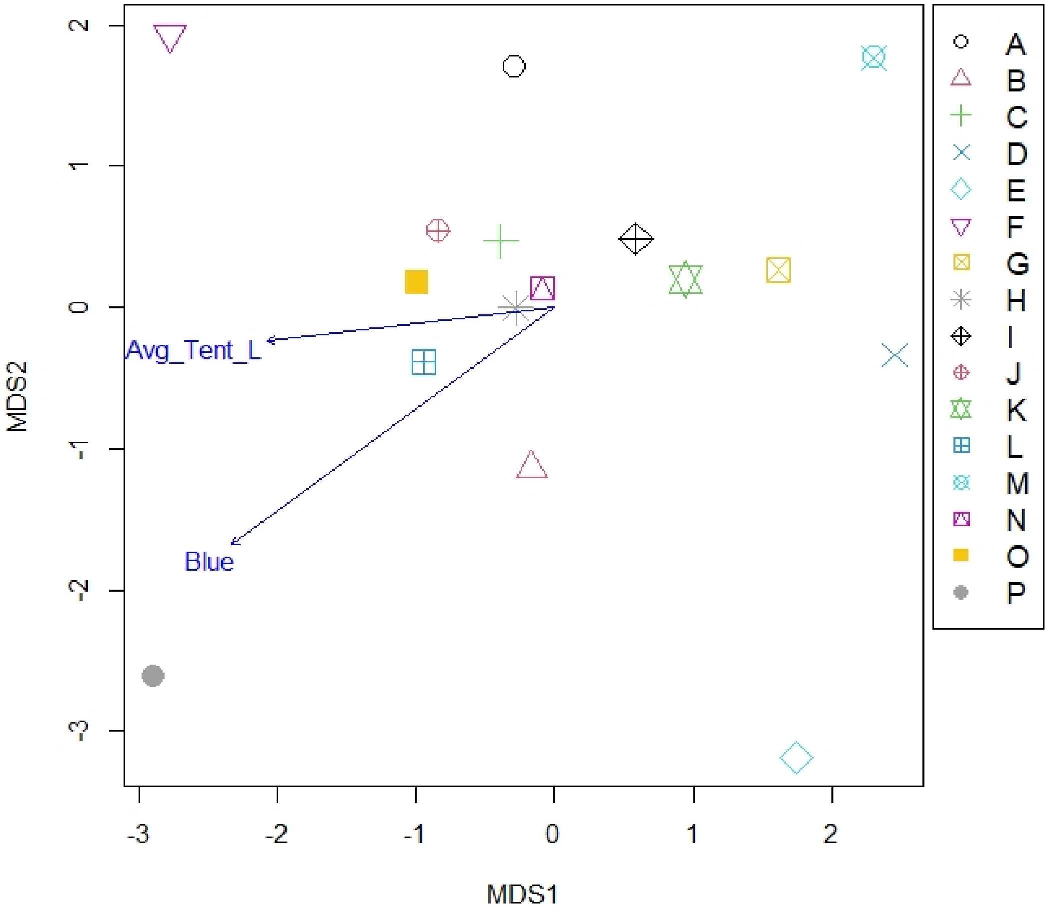


Figure 6 - NMDS of zoanthid morphological data. Vectors depict significant morphological characteristics from phylogenetically independent contrast (PIC) analysis and show factors driving separation within the samples.

Samples varied across measurements with mouth width ranging from 0.25mm - 0.54mm ($\bar{x}$ = 0.38mm), tentacle length 0.37mm - 1.35mm ($\bar{x}$ = 0.80mm), and oral disc diameter 1.56mm - 3.86mm ($\bar{x}$ = 2.37mm) (Appendix 5). The number of tentacles each sample possessed also varied from 42 tentacles to 64 ($\bar{x}$ = 50). Coloration also varied across samples, black 0% - 97% ($\bar{x}$ = 21%), brown 0% - 91% ($\bar{x}$ = 41%), blue 0% - 100% ($\bar{x}$ = 10%), white 0% - 18% ($\bar{x}$ = 5%), green 0% - 100% ($\bar{x}$ = 10%), yellow 0% - 22% ($\bar{x}$ = 4%), red 0% - 63% ($\bar{x}$ = 10%) (Appendix 6).

Data comparisons

2b-RAD and Sanger sequencing data.

Zoanthid Sanger Sequencing data was positively correlated with zoanthid 2b-RAD sequencing data (Mantel test; r = 0.5807, p = 0.001, permutations = 999; Table 5). Qualitative comparisons between phylogenies show similar topologies, but greater supports in the 2b-RAD dataset.

Table 5 - Results from Mantel tests performed on data sets. Permutations = 999 for all comparisons

Comparison	r	p
Zoanthid		
2b-RAD - Sanger seq.	0.58	0.001***
2b-RAD - Morphology	-0.08505	0.65
Sanger seq. - Morphology	-0.104	0.833
Symbiodinium		
2b-RAD - Sanger seq.	-0.1409	0.616
Zoanthid 2b-RAD - Symbiodinium 2b-RAD	-0.02413	0.51
Zoanthid morphology - Symbiodinium 2b-RAD	-0.00329	0.49

2b-RAD and morphology

Overall zoanthid 2b-RAD sequencing data dissimilarity did not correlate with morphological dissimilarity in our data (Mantel test; $r = -0.085$, $P = 0.65$; Table 5). To understand if individual morphological characteristics corresponded to 2b-RAD data a canonical correspondence analysis was conducted. In the analysis all variables only explained 12.8% of the variance seen in the 2b-RAD data (adjusted $r^2 = 0.12821$, $p = 0.026$, permutations = 999). Only a single variable explained a significant portion of the variance. The color blue explained 4% of variance in the 2b-RAD data (adjusted $r^2 = 0.04014$, $p = 0.028$, permutations = 999), but no other individual variable was significant ($P < 0.05$). The significance of the color blue may be driven by a single sample

(JES2016-P), which was 100% blue (highest proportion of blue in other samples was 20%) and highly divergent in the 2b-RAD data (Figure 3). This variable was mapped onto Figure 3 and the resulting vector corresponds to the outgrouping of sample JES2016-P which possessed a solid blue coloration (Figure 1).

To understand the relationship of individual morphological variables to phylogenetic relationships among corals, characteristics were mapped to the Bayesian tree and tested for phylogenetic signal (Figure 7). Among all the morphological variables, only the average tentacle length (Blomberg's K = 0.86, P = 0.03) and blue coloration of the polyp (K = 0.99, P = 0.05) showed a significant phylogenetic signal based on phylogenetically independent contrasts (PIC, Figure 7).

Overall zoanthid sanger sequencing dissimilarity did not correlate with morphological dissimilarity in our data (Mantel test; r = -0.104, P = 0.83).

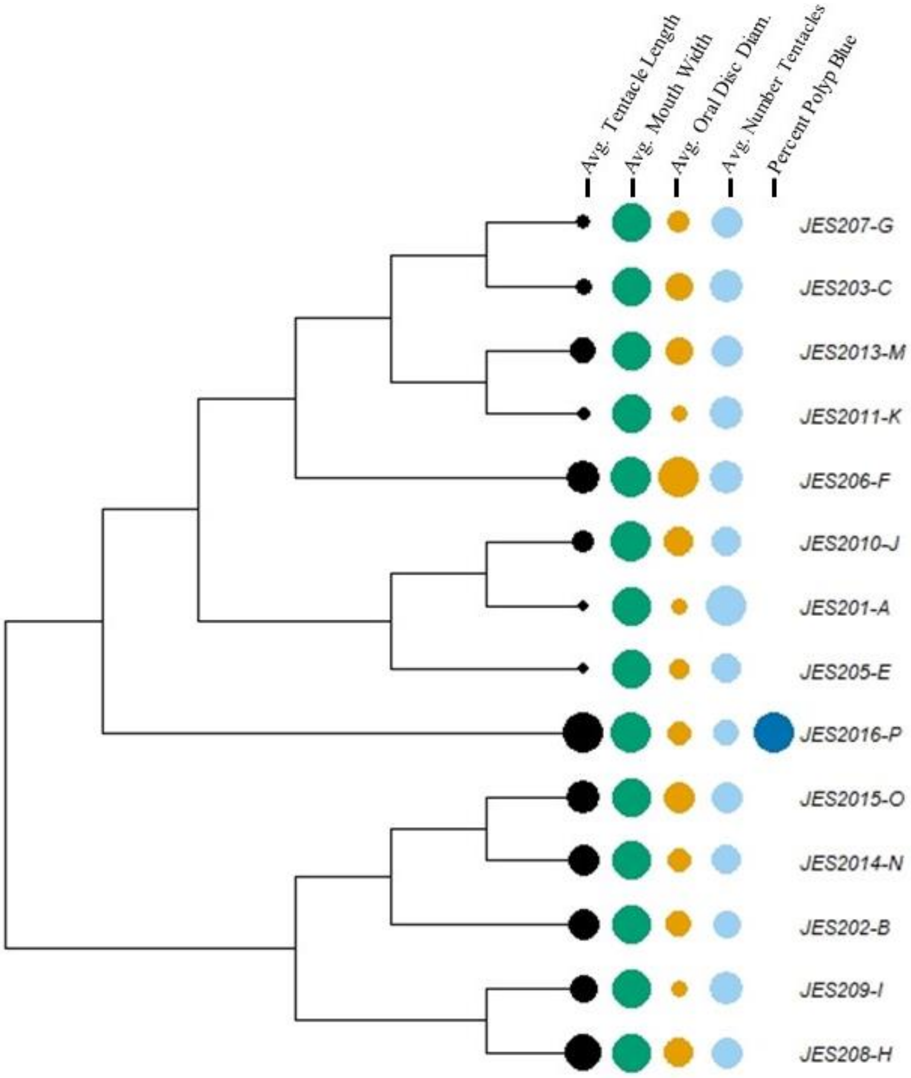


Figure 7 - Bayesian phylogeny with morphological characteristics added to tip labels. Plot produced in R package picante (v 1.8.2). Tip labels (left to right): Average tentacle length, average mouth width, average oral disc width, average number of tentacles, percent polyp color blue. Larger circles indicate higher values for each trait.

Table 6 - Phylogenetically independent contrast (PIC) scores. Produced in picante v 1.8.2. Significant factors are shown in bold font

Characteristic	K statistic	p
Avg. Mouth Width	0.388	0.6
Avg. Tentacle Length	**0.858**	**0.025**
Avg. Oral Disc Diam.	0.337	0.712
Avg. Number of Tentacles	0.411	0.573
Black	0.355	0.524
Brown	0.358	0.374
Blue	**0.987**	**0.051**
White	0.420	0.398
Green	0.556	0.164
Yellow	0.559	0.149
Red	0.589	0.192

Symbiont 2b-RAD Data

A total of 11 symbiont cultures produced 36,488 loci composed of 1,313,568 bases containing 5,137 variant sites. The resulting NMDS displayed partial grouping, but six samples did not show any relationship. Samples G, H, J, M, and P were all grouped together while all other samples did not group with any other samples. Samples A, B, C, E, F, and N were isolated around distinct boundaries of the NMDS plot (Figure 8). The data set used to produce the NMDS was the smallest of any dataset (11 samples in total).

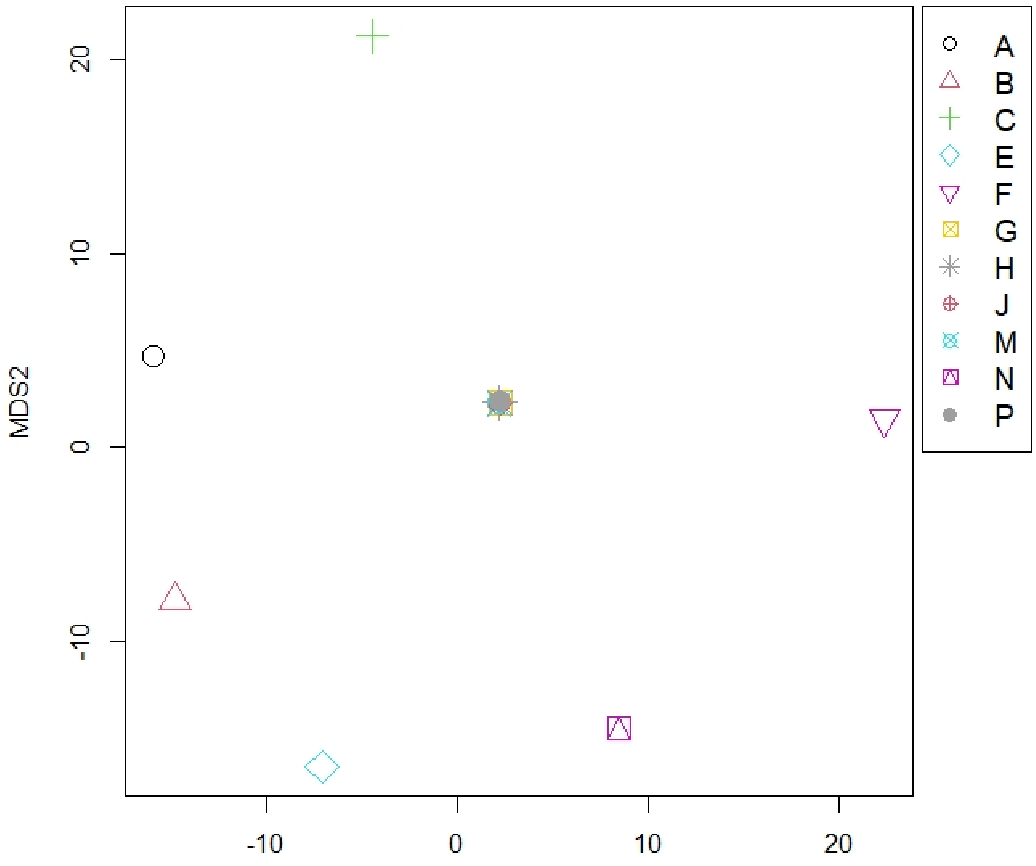


Figure 8 - NMDS of *Symbiodinium 2b-RAD sequence data.*

In the Bayesian phylogenetic tree (Figure 9) several samples formed pairings that were well supported, and one clade possessed three samples (A, B, and C). Pairings with high support, Bayesian posterior probability > 0.95, included E and J, and G and M. One pairing (N and M) had lower support with a BPP < 0.95 (0.87). Two individual samples (Zoo203-C and Zoo206-F) grouped independently of the clades with low support (BPP < 0.46).

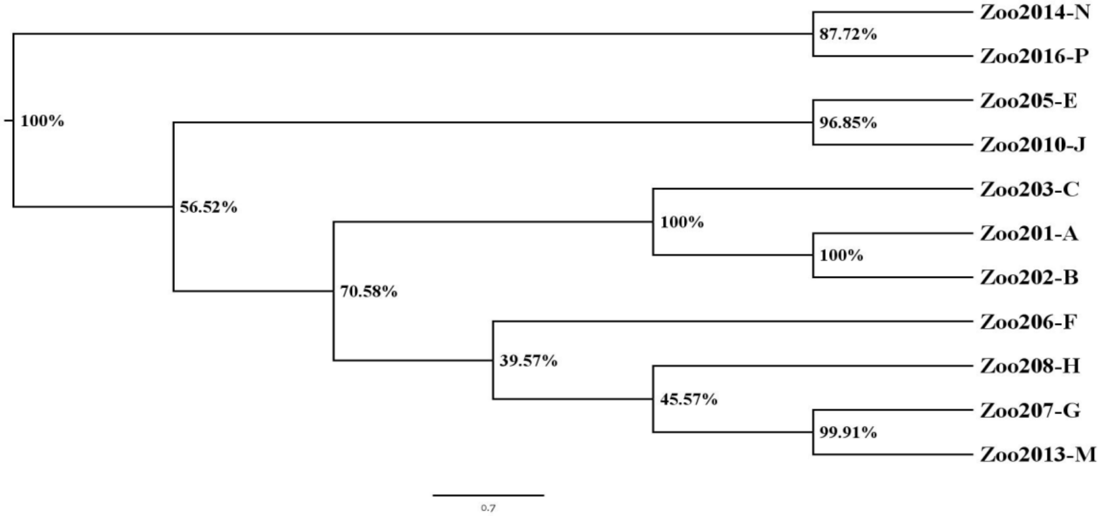


Figure 9 – Bayesian tree created using symbiont 2b-RAD sequence data. Percentages represent Bayesian posterior probability with 100,000 burn-in and 10,000,000 replications.

Symbiont Sanger Sequencing

An ITS-2 gene tree (Figure 10) was produced for symbiont DNA as only a single gene was used in the analysis. Recovered sequence length for the ITS-2 nrDNA barcoding region was ~350bp in length containing 16 variant sites (SNPs). When recovered sequences were compared to the NCBI database using the nucleotide basic alignment and search tool (BLAST), all the samples were found to belong to the

Symbiodinium Clade C (Altschul et al, 1990). Symbiodinium Clade C is a legacy name that is now recognized as the genus *Cladocopium* (LaJeunesse et al., 2018). Henceforth the symbionts in this study will be referred to accordingly.

In the produced maximum likelihood tree (Figure 10) two supported clades (MLBS ≥ 85) and one poorly supported clade (MLBS < 85) were recovered with one containing a single sample (Zoo201) with high support (MLBS value = 91; Figure 10). The only other supported clade, containing 9 samples, had high support of (MLBS = 85, Figure 10). Samples outside of these two clades grouped together but possessed low internal support, suggesting a lack of dissimilarity in this clade. Interestingly, one sample (Zoo2016-P) was not grouped with any other sample and was rooted at the base of the tree, however this node did not possess significant support.

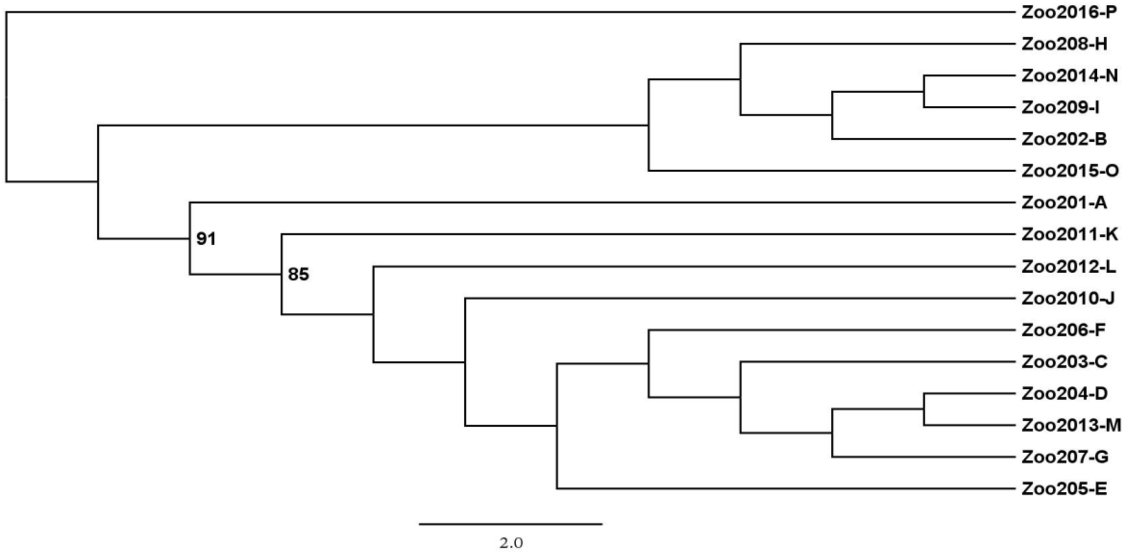


Figure 10 – Maximum likelihood gene tree produced from ITS-2 nrDNA. Tree produced through RAxML using 1000 replications and 1000 ML bootstrap replications to estimate support for nodes (Stamatakis, 2006). MLBS values less than 75 have been removed from the tree.

The resulting NMDS (Figure 11) revealed poorly differentiated and tightly associated groups. Samples A, E, and P were the only ones visually discernible outside of three groups also found in the plot.

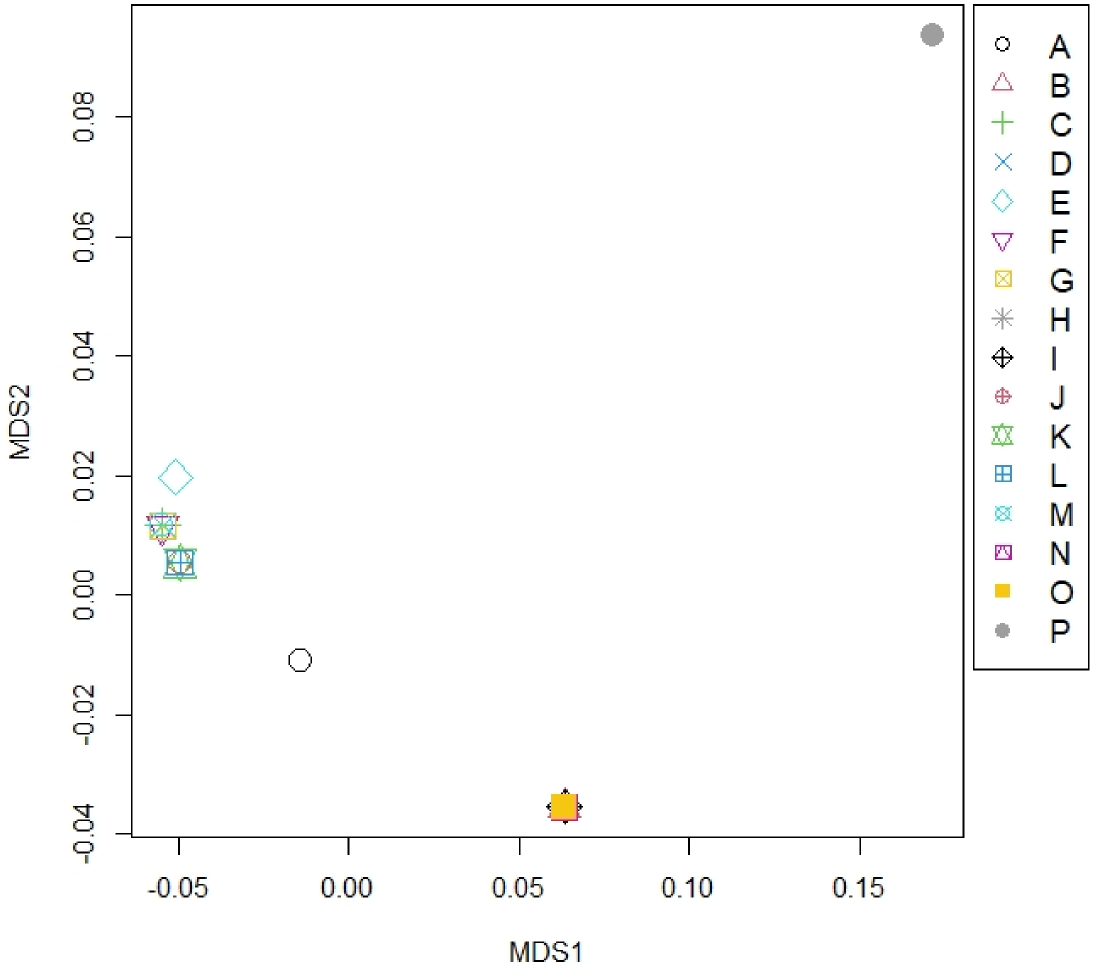


Figure 11 – NMDS of *Symbiodinium* ITS-2 nrDNA Sanger sequencing data.

Cladocopium sp. data comparisons

2b-RAD and Sanger sequencing correlation data of *Cladocopium sp.* did not produce a significant result (Mantel r = -0.149, p = 0.602). The recovered sequences did not correlate to the zoanthid 2b-RAD, Sanger sequencing, nor morphology data sets

(Table 5). Samples that were present in both the zoanthid and *Symbiodinium* Bayesian trees (Figures 2 and 9) lacked similar topology. It should be noted that figure 9 is the only figure to group sample JES2016-P/Zoo2016-P with any other sample. However, this grouping reflects a similarity in the symbiont and not the host genome.

Discussion

Zoanthid corals are important members of coral reef communities worldwide, and popular organisms within the aquarium trade. The current study revealed a high level of genetic diversity within several common zoanthids by leveraging 2b-RAD sequencing. Genetic diversity observed in 2b-RAD sequencing data was higher compared to Sanger sequencing data revealing cryptic diversity that has not previously been recognized. While these genetic data types were correlated, the greater resolution of 2b-RAD data allows for a more precise identification of species found both on reefs and within the aquarium trade. Within the aquarium trade, zoanthids are marketed based on their eye-catching and unique morphologies, yet the current study found no relationship between zoanthid genetics and a suite of common morphological characters. This result may thus have important implications for the importation of corals and other cnidarians within the aquarium trade. Additionally, this study shows that many common morphological characters used to identify zoanthids on natural reefs may not adequately capture true diversity within zoanthids and thus may overlook cryptic or at-risk rarer species.

2b-RAD sequencing can assist in accurately estimating the true genetic diversity in non-model species and thus has utility in determining the number of species of zoanthids in the world (Barbanti et al., 2020; Kelly et al., 2020). For example, Kelly et al.

(2020) identified diversity within cryptic marine sponge species, genus *Ircinia*, undergoing ecological divergence. Within our own dataset, we also observed hidden diversity in two of the most prolific zoanthid types in the aquarium trade are commonly named “Eagle Eye zoanthids” (Figure 1A) and “Dragon Eye zoanthids” (Figure 1, E and J). Morphological data indicate nearly identical morphologies for the two Dragon Eye Zoanthids (JES205-E & JES2010-J; Figure 6) but the 2b-RAD data show that the specimen marketed as Eagle-Eye Zoanthids (JES201-A) grouped within this clade (Figures 2 and 3) and was more similar to one of the Dragon Eye Zoanthids than they were to each other based on F_{st} values (Table 3). Conversely, our 2b-RAD data shows groupings among samples C, G, K, M (Figures 2 and 3). These corals are marketed in the aquarium trade as being different organisms (Figure 1), and they possess highly variable morphological characteristics (Appendix 5 & 6). Despite these similarities our data suggests they are closely related (Figure 2 and 3). Divergence in both genetics and morphology is clear in one sample, JES2016-P, which was collected in the Atlantic Ocean while most samples used in this study are assumed to be from the Pacific.

Fixation indexes (F_{st}) are commonly used in population genetics to define population boundaries based on gene flow between populations. Yet currently there does not appear to be a well-defined cut-off for F_{st} values produced by NGS data. Several studies within phylum Cnidaria have utilized F_{st} to define species boundaries in the past through Sanger sequencing, but all with varying degrees of F_{st} values (Calixto-Botía & Sánchez., 2017; Faleh et al., 2017; Russo et al., 1994). Interestingly a study on *Zoanthus coppingeri* revealed an F_{st} value (0.43) like that seen in this study across all their study sites (Burnett et al., 1995). However, adjustments may be needed to accurately compare

NGS data F_{st} values to that of sequences recovered through Sanger sequencing or allozyme data (Burnett et al. 1995). In studies on organisms outside of cnidarians, F_{st} values can also vary drastically and prevent the application of a universal cutoff across organisms (Cooke et al., 2020; Roesti et al., 2012; Smaragdov & Kudinov, 2020). Though there is not yet a standard F_{st} cutoff for zoanthids; a threshold of 0.45 is proposed as the cutoff for zoanthids used in this study (Figures 2, 3 and Table 3). This cutoff value reflects natural groupings observed in multiple data types and it creates a clear threshold for defining different genetic groupings within the data. By applying this cutoff value, we identified seven unique genetic groupings within the samples of this study.

A study on the sister genus *Palythoa* recently revealed species delimitation difficulties similar to those found in genus *Zoanthus* (Dudoit et al., 2021). In their study Dudoit et al. (2021) used a PTP model to parse species delimitations, however their resulting models varied, sometimes drastically, in species groupings. This variation in models was not found in the study presented here. Multiple models were run with all models converging on identical sample groupings. The difference in resolution may be due to the genetic compartments, i.e., nuclear DNA vs mitochondrial DNA, utilized in this study. Dudoit et al. (2021) focused on mitochondrial genetic data and found their sequence divergence reflected that seen in COI mtDNA divergence. As previously stated, COI mtDNA is often a very poor genetic region for delineating or identifying species within Cnidaria (Shearer et al., 2002). Dudoit et al. (2021) found their nearly complete mitochondrial genome data varied only by 0.24% to 1.56% and COI mtDNA divergence in studies to date only varied by 0% to 1.14% (Sinniger et al., 2005 & 2008; Reimer et al., 2006, 2007, & 2012). Our study used nuclear and mitochondrial DNA throughout our analyses which may influence our ability to consistently resolve similar species groupings using PTP models. Our study still is unable to identify exact species delimitation by using the PTP model alone. However, combining the results from multiple data types (Figures 2 & 3; Tables 2 & 3), it appears the monophyletic clade of *Zoanthus sansibaricus* consists of cryptic species based on F_{st} values and clustering in the NMDS.

Until recently Sanger sequencing has been the main method for defining species boundaries within zoanthids but with varying levels of support, resolution, and consensus (reviewed in Swain 2018). These studies have revealed approximately 14 verified and 8 unverified species globally (Reimer & Sinniger, 2022). Conversely the aquarium trade recognizes over 680 common names for polyps in genus *Zoanthus* specifically (zoanthids.com). This number may be partly influenced by the phenotypic plasticity of zoanthids, but also suggests that zoanthid species richness globally may be much higher than previously recognized. Our Sanger sequencing data separated samples into either four species, through ITS nrDNA barcoding, or five species (Table 4) and produced a ML tree consisting of two monophyletic clades (Figure 4). Though Sanger sequence data suggests four to five species, the NGS data assembled in this study supports seven species and one unresolved sample. This suggests that several collections in prior studies may have been novel species that were misidentified as *Zoanthus sansibaricus*. This misidentification could be possible in other collections and may be due to the lower resolution of prior molecular techniques.

Morphological classification has been historically utilized to delineate one zoanthid species from another (Lwowsky, 1913; Ryland and Lancaster, 2003). In 1913 Lwowsky published an in-depth study of zoanthid morphology based on numerous collections made around the world. Though the study was based solely on morphology they still produced results with enough support to justify revising genus *Sidisia*. This revision moved many species from genus *Epizoanthus* into genus *Sidisia*, however this has since been reverted, generic delimitation was revised and *Epizoanthus* was resurrected (Lwowsky, 1913). This revision and reversion of genera may be due to the phenotypic plasticity of zoanthids or a disconnection between morphology used to identify corals and the genome on which evolution operates. Statistical analysis in the current study does not show any significant relationship between this study's genetic data, (NGS or Sanger sequencing) and morphology. A clear example of this decoupling can be observed when the morphology of samples JES201-A, JES205-E, JES2010-J (see Figure 1). These three samples possess drastically different color morphology and measured morphology characteristics, but they possess identical Sanger sequenced DNA segments and all matching *Zoanthus sansibaricus* in the NCBI database. This disconnection between morphology and genome has been observed in studies on the coral *Pocillopora spp*. (Johnston et al., 2020; Burgess et al., 2021), but how widespread this phenomena decoupling is within the phylum is currently unknown.

In a review, Swain (2016) expressed concern about molecular studies being disconnected from studies utilizing morphology. This study used an integrative approach that includes broadly applicable morphological characteristics combined with molecular techniques as the ideal method for characterizing and identifying zoanthids (Swain, 2016). However, since the data presented here suggests morphology and genetics are not correlated, it raises concern whether using a range of characteristics would weaken the ability of a study to identify and characterize zoanthid species. Utilizing morphology alone to discern unique species or populations may not reflect true evolutionary relationships, even when used in conjunction with molecular markers (Johnston et al., 2020; Burgess et al., 2021). Thus, an over reliance on morphology can lead to inaccurate conclusions of population structure and dynamics (Burgess et al., 2021).

Further analysis of the correlation between genome and morphology needs to be conducted to understand how future studies should approach the concept of utilizing a range of characteristics. Use of too many characteristics could misdirect the researcher while using too few methods may lead to poor and inaccurate resolution of species. Future studies should utilize 2b-RAD in conjunction with an in-depth histological study. By doing this we could understand whether utilizing a greater quantity of structural morphological characteristics, such as those used by Swain et al. (2016), identifies a correlation to the genome of zoanthids. Through using a more comprehensive set of morphological characteristics future studies could seek to unravel which are correlated with the genome and which are decoupled.

From our study it is not clear what drives morphological characteristics, such as zoanthid polyp coloration. Further molecular research into characteristics such as fluorescent proteins, could help increase our understanding of the correlation between genetics and morphology in these organisms. Also, further investigation into how phenotypic plasticity is influenced by the genome would aid our understanding of how and why we see so much variation in morphology in zoanthids. 2b-RAD, can help in this endeavor due to its nature of sampling small segments across the genome. A study with increased sample size could increase the quantity of color measurements and allow for the ability to identify if there is additional phylogenetic signal in polyp coloration. 2b-RAD can also allow for a survey of zoanthid genomes that could facilitate the identification of specific gene regions for targeted amplification that correlate to morphological characters.

Cladocopium sp.

Genetic data for Cladocopium sp.

Research on genetic variability in Symbiodinaceae in corals is prolific. This area is of great interest to coral reef scientists due to its potential implications in coral survival of climate change, disease, and other stressors. Studies within the genus *Zoanthus* have cataloged several different genera of Symbiodinaceae within their polyps. The identified zooxanthellae within genus *Zoanthus* have been genus *Symbiodinium*, genus *Breviolum* and genus *Cladocopium* (formerly clades A, B, & C respectively; Costa et al., 2008; Kamezaki et al., 2013; LaJeunesse et al., 2018; Rabelo et al., 2014; Reimer & Todd, 2009;). Several studies have found that species do not necessarily associate with only one genus of symbiont and both *Zoanthus sansibaricus* and *Z. sociatus* have been found to associate with genera *Symbiodinium* and *Cladocopium* (Kamezaki et al., 2013, Rabelo et al., 2014). The main driver behind the variation *Z. sansibaricus* appears to be host habitat depth with deeper colonies possessing genus *Cladocopium* instead of *Symbiodinium* which is more common in high-photic zones (Kamezaki et al., 2013, LaJeunesse, 2002).

Interestingly, studies that included *Z. sociatus* identified this species as most often being associated with genus *Symbiodinium* (Costa et al., 2013; LaJeunesse, 2002; Rabelo et al., 2014 & 2015). In the current study, however, only the genus *Cladocopium* was identified, and samples possessed nearly identical amplified ITS2 rDNA regions. 2b-RAD sequencing data provided much greater differentiation among the *Cladocopium sp.* samples, but this could largely be due to the sheer imbalance in sequencing information between the two techniques (365 bases in Sanger and 2,981 bases in 2b-RAD). Interestingly, these two techniques did not correlate with one another as observed in the host data.

The genome of *Cladocopium sp.* dwelling within the zoanthids does not correlate to any host molecular data collected. The disconnection of host and symbiont genetic datasets may be due to a lack of coevolution relating to transmission of symbionts between parent and larva, though this result is counter to several other studies. Johnston et al. (2022) have identified cophylogeny within host-symbiont relationships of *Cladocopium spp.* and suggest this relationship is due to the vertical transmission of their symbionts during broadcast spawning. Vertical transmission of *Cladocopium* symbionts occurs in around 20% of other scleractinian corals (Baird et al., 2009). Most scleractinian corals that undergo broadcast spawning collect their Symbiodinaceae horizontally (Baird et al., 2009). Zoanthids are considered to be broadcast spawners, except for *Isozoanthus giganteus* (Ryland, 1997), which may explain the disconnect in host-symbiont genome. Alternatively, local environmental factors may play a large role symbiont type as observed in previous studies (Kamezaki et al., 2013, Rabelo et al., 2014). In this case, generations within confined aquaria with standardized light, food, and flow may result in lower symbiont community diversity compared to wild caught zoanthids.

Ryland (1997) found that *Protopalythoa sp.*, a close relative to genus *Zoanthus,* are known to vertically transmit their symbionts to larva produced in mass spawning events, but since symbiont transmission in genus *Zoanthus* is poorly understood, further studies should be conducted to understand what mechanism underlies the disconnection between host and symbiont genomes. Given that the *Cladocopium sp.* present in the samples were nearly identical and the lack of correlation between 2b-RAD data and other data sets, it appears symbiont genetics alone is not a viable tool for parsing zoanthid species from one another in this study.

Conclusion

2b-RAD has proven to be a versatile, robust, and informative sequencing technique for identifying zoanthids. It allows for the identification of dissimilarities within species and thus can provide ways to identify and differentiate morphologically similar populations, previously unresolved relationships, and species that may be at risk of extinction. Our NGS data suggest there are eight distinct species groups, one of these (JES206-F) being a possible hybrid species, while genetic barcoding only identifies four species. Our data suggest two samples (JES202-B and JES2016-P) used in this study merit further investigation to determine if their dissimilarity to other species warrants formal description as new species. Additionally, we suggest further investigation into the relationship between genome and specific morphological characters to potentially improve resolving relationships between genetics and morphology. The impact zoanthids have on reef ecosystems requires further research to understand what niches they fill and what influence they have on the habitat. Powerful tools, such as 2b-RAD will aid in identifying what species and populations of zoanthids are present.

REFERENCES

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool (BLAST). *Journal of Molecular Biology*, *215*(3), 403–410.

Barbanti, A., Torrado, H., Macpherson, E., Bargelloni, L., Franch, R., Carreras, C., & Pascual, M. (2020). Helping decision making for reliable and cost-effective 2b-RAD sequencing and genotyping analyses in non-model species. *Molecular Ecology Resources*, *20*(3), 795–806.

Baird, A. H., Guest, J. R., & Willis, B. L. (2009). Systematic and biogeographical patterns in the reproductive biology of Scleractinian corals. *Annual Review of Ecology, Evolution, and Systematics*, *40*(1), 551–571.

Ben Faleh, A. R., Allaya, H., Armani, A., & Shahin, A. A. B. (2017). Significant genetic differentiation among meroplanktonic barrel jellyfish *Rhizostoma pulmo* (Cnidaria: Scyphozoa) in the Mediterranean Sea. *African Journal of Marine Science*, *39*(1), 1–8.

Burgess, S. C., Johnston, E. C., Wyatt, A. S., Leichter, J. J., & Edmunds, P. J. (2021). Response diversity in corals: Hidden differences in bleaching mortality among cryptic *Pocillopora* species. *Ecology*, *102*(6).

Burnett, W. J., Benzie, J. A., Beardmore, J. A., & Ryland, J. S. (1995). Patterns of genetic subdivision in populations of a clonal cnidarian, Zoanthus coppingeri, from the Great Barrier Reef. *Marine Biology*, *122*(4), 665–673.

Calixto-Botía, I., & Sánchez, J. A. (2017). A case of modular phenotypic plasticity in the depth gradient for the gorgonian coral antillogorgia bipinnata (cnidaria: Octocorallia). *BMC Evolutionary Biology*, *17*(1).

Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Lozupone, C. A., Turnbaugh, P. J., Knight, R. (2010). Global patterns of 16S16S rRNA diversity at a depth of millions of sequences per sample. *Proceedings of the National Academy of Sciences, 108*, 4516-4522.

Catchen, J., Hohenlohe, P., Bassham, S., Amores, A., & Cresko, W. (2013). Stacks: an analysis tool set for population genomics. *Molecular Ecology*.

Costa, C. F., Sassi, R., & Gorlach-Lira, K. (2008). Zooxanthellae genotypes in the coral Siderastrea stellata from coastal reefs in northeastern Brazil. *Journal of Experimental Marine Biology and Ecology*, *367*(2), 149–152.

Dudoit, ‘A., Santos, M. E., Reimer, J. D., & Toonen, R. J. (2021). Phylogenomics of Palythoa (Hexacorallia: Zoantharia): Probing species boundaries in a globally distributed genus. *Coral Reefs*.

Edgar, R. C. (2004). Muscle: Multiple sequence alignment with improved accuracy and speed. *BMC Bioinformatics*, *5*(113).

Gray, J.E. (1832). Synopsis of the contents of the British Museum. London (27th edition.). (4), 1-212.

Huang, D., Meier, R., Todd, P. A., & Chou, L. M. (2008). Slow mitochondrial COI sequence evolution at the base of the metazoan tree and its implications for DNA barcoding. *Journal of Molecular Evolution*, *66*(2), 167–174.

Jaramillo, K. B., Reverter, M., Guillen, P. O., McCormack, G., Rodriguez, J., Sinniger, F., & Thomas, O. P. (2018). Assessing the Zoantharian diversity of the Tropical Eastern Pacific through an integrative approach. *Scientific Reports*, *8*(1).

Johnston, E. C., Wyatt, A. S., Leichter, J. J., & Burgess, S. C. (2021). Niche differences in co-occurring cryptic coral species (*Pocillopora* spp.). *Coral Reefs*.

Johnston, E. C., Cunning, R., & Burgess, S. C. (2022). Cophylogeny and specificity between cryptic coral species (*Pocillopora* spp.) at Mo’orea and their symbionts (Symbiodiniaceae). *PrePrint*.

Kamezaki, M., Higa, M., Hirose, M., Suda, S., & Reimer, J. D. (2012). Different zooxanthellae types in populations of the Zoanthid *Zoanthus sansibaricus* along depth gradients in Okinawa, Japan. *Marine Biodiversity*, *43*(1), 61–70.

Kelly, J. B., Carlson, D., Low, J. S., Rice, T., & Thacker, R. W. (2020). The relationship between microbiomes and selective regimes in the sponge genus Ircinia. *Frontiers in Microbiology*.

LaJeunesse, T. C. (2002). Diversity and community structure of symbiotic dinoflagellates from Caribbean coral reefs. *Marine Biology*, *141*(2), 387–400.

LaJeunesse, T. C., Parkinson, J. E., Gabrielson, P. W., Jeong, H. J., Reimer, J. D., Voolstra, C. R., & Santos, S. R. (2018). Systematic revision of Symbiodiniaceae highlights the antiquity and diversity of coral endosymbionts. *Current Biology*, *28*(16).

Lamouroux, J. V. F. (1816). Histoire des polypiers coralligènes flexibles, vulgairement nommés zoophytes. Caen: Impr. de F. Poisson.

Low, M. E., Reimer, J. D., & Sinniger, F. (2016). The order Zoantharia Rafinesque, 1815 (Cnidaria, Anthozoa: Hexacorallia): supraspecific classification and nomenclature. *ZooKeys*, 641, 1–80.

Löytynoja Ari, & Goldman, N. (2008). Phylogeny-Aware Gap Placement Prevents Errors in Sequence Alignment and Evolutionary Analysis. *Science*, *320*(5883), 1632–1635.

Lwowsky, F. F. (1913). Revision der Gattung *Sidisia* Gray (*Epizoanthus auct*.). Ein Beitrag zur Kenntnis der Zoanthiden. *Zoologische Jahrbücher*, 557–614.

Maddison, W. P. & D.R. Maddison. (2019). Mesquite: a modular system for evolutionary analysis

Madeira, F., Park, Y. M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., & Lopez, R. (2019). The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Research*, *47*(W1).

Martin, M. (2011). CUTADAPT removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, *17*(1), 10.

Oksanen, J. A. R. I., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin R. P., O'Hara, R. B., Simpson, G. L., Solymos, P., Stevens, M.H., Szoecs, E., & Wagner, H. (2021). vegan: Community Ecology Package. R package version 2.5-7. 2020.

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000a). Inference of population structure using multilocus genotype data. *Genetics*, (155), 945–959.

Rabelo, E. F., Rocha, L. L., Colares, G. B., Bomfim, T. A., Nogueira, V. L., Katzenberger, M., Matthews-Cascon, H., & Melo, V. M. (2014). Symbiodinium diversity associated with zoanthids (Cnidaria: Hexacorallia) in northeastern Brazil. *Symbiosis*, *64*(3), 105–113.

Rabelo, E. F., Soares, M. de, Bezerra, L. E., & Matthews-Cascon, H. (2015). Distribution pattern of Zoanthids (Cnidaria: Zoantharia) on a Tropical Reef. *Marine Biology Research*, *11*(6), 584–592.

Rafinesque, C. S. (1815). Analyse de la nature, ou, Tableau de lunivers et des corps organises. *Palerme: Aux depens de lauteur*.

Rambaut A. (2012). FigTree 1.4.4. http://tree.bio.ed.ac.uk/software/figtree/

Reimer, J.; Sinniger, F. (2022). World List of Zoantharia. *Zoanthus* Lamarck, 1801.

Reimer, J. D., & Todd, P. A. (2009). Prelimiary Molecular Examination of Zooxanthellate Zoanthids (Hexacorallia: Zoantharia) and Associated Zooxanthellae (*Symbiodinium* spp.) Diversity in Singapore. *The Raffles Bulletin of Zoology*, (22), 103-120.

Reimer, J. D., Ono, S., Takishita, K., Tsukahara, J., & Maruyama, T. (2006). Molecular Evidence Suggesting Species in the Zoanthid Genera *Palythoa* and *Protopalythoa* (Anthozoa: Hexacorallia) Are Congeneric. *Zoological Science*, *23*(1), 87-94.

Reimer, J., & Fujii, T. (2010a). Four new species and one new genus of zoanthids (Cnidaria, Hexacorallia) from the Galapagos Islands. *ZooKeys*, 42, 1–36.

Reimer, J. D., Nakachi, S., Hirose, M., Hirose, E., & Hashiguchi, S. (2010b). Using Hydrofluoric Acid for Morphological Investigations of Zoanthids (Cnidaria: Anthozoa): A Critical Assessment of Methodology and Necessity. *Marine Biotechnology*, 12(5), 605–617.

Reimer, J. D., Takishita, K., Ono, S., & Maruyama, T. (2007a). Diversity and evolution in the zoanthid genus Palythoa (Cnidaria: Hexacorallia) based on nuclear ITS-rDNA. *Coral Reefs*, *26*(2), 399-410.

Reimer, J. D., Takishita, K., Ono, S., Tsukahara, J., & Maruyama, T. (2007b). Molecular Evidence Suggesting Interspecific Hybridization in *Zoanthus* spp. (Anthozoa: Hexacorallia). *Zoological Science, 24*(4), 346-359.

RStudio Team (2021). RStudio: Integrated Development Environment for R. RStudio, PBC, Boston, MA

Rhyne, A. L., Tlusty, M. F., Schofield, P. J., Kaufman, L., Morris, J. A., & Bruckner, A. W. (2012). Revealing the appetite of the Marine Aquarium Fish Trade: The volume and biodiversity of fish imported into the United States. *PLoS ONE*, *7*(5).

Russo, C. A., Solé-Cava, A. M., & Thorpe, J. P. (1994). Population structure and genetic variation in two tropical sea anemones (Cnidaria, Actinidae) with different reproductive strategies. *Marine Biology*, *119*(2), 267–276.

Ryland, J. S. (1997). Reproduction in Zoanthidea (Anthozoa: Hexacorallia). *Invertebrate Reproduction & Development*, *31*(1-3), 177–188.

Ryland, J. S., & Lancaster, J. E. (2003). Revision of methods for separating species of Protopalythoa (Hexacorallia : Zoanthidea) in the tropical West Pacific. *Invertebrate Systematics*, *17*(3), 407.

Shearer, T. L., van Oppen, M. J., Romano, S. L., & Wörheide, G. (2002). Slow mitochondrial DNA sequence evolution in the Anthozoa (Cnidaria). Molecular Ecology, 11(12), 2475–2487.

Sinniger, F., Reimer, J. D., & Pawlowski, J. (2010). The Parazoanthidae (Hexacorallia: Zoantharia) DNA taxonomy: Description of two new genera. *Marine Biodiversity, 40*(1), 57-70.

Stamatakis, A. (2006). RAxML-VI-HPC: Maximum Likelihood-based Phylogenetic Analyses with Thousands of Taxa and Mixed Models. *Bioinformatics*, 1–2.

Stamatakis, A., Hoover, P., & Rougemont, J. (2008). A Rapid Bootstrap Algorithm for the RAxML Web Servers. *Systematic Biology*, 57(5), 758–771.

Swain, T. D., & Swain, L. M. (2014). Molecular parataxonomy as taxon description: examples from recently named Zoanthidea (Cnidaria: Anthozoa) with revision based on serial histology of microanatomy. *Zootaxa*, 3796(1), 81.

Swain, T. D., Schellinger, J. L., Strimaitis, A. M., & Reuter, K. E. (2015). Evolution of anthozoan polyp retraction mechanisms: convergent functional morphology and evolutionary allometry of the marginal musculature in order Zoanthidea (Cnidaria: Anthozoa: Hexacorallia). *BMC Evolutionary Biology*, 15(1).

Swain, T. D., Strimaitis, A. M., Reuter, K. E., & Boudreau, W. (2016). Towards integrative systematics of Anthozoa (Cnidaria): evolution of form in the order Zoanthidea. *Zoologica Scripta*, 46(2), 227–244.

Swain, T. D. (2018). Revisiting the phylogeny of Zoanthidea (Cnidaria: Anthozoa): Staggered alignment of hypervariable sequences improves species tree inference. *Molecular Phylogenetics and Evolution*, 118, 1–12.

Wang, S., Meyer, E., Mckay, J. K., & Matz, M. V. (2012). 2b-RAD: A simple and flexible method for genome-wide genotyping. *Nature Methods*, 9(8), 808-810.

Weller, H. (2021, April 20). *Recolorize: Color-based image segmentation (for people with other things to do)*. Hannah Weller. Retrieved February 21, 2022.

Will, K. W., Mishler, B. D., & Wheeler, Q. D. (2005). The perils of DNA barcoding and the need for integrative taxonomy. Systematic Biology, 54(5), 844–851.

Zhang, J., Kapli, P., Pavlidis, P., & Stamatakis, A. (2013). A general species delimitation method with applications to phylogenetic placements. *Bioinformatics*, *29*(22), 2869–2876.

APPENDICES

Appendix 1 – 2b-RAD sequence processing protocol

https://github.com/Jes8x/M.S.-Thesis-Supplemental-Material/blob/ebab4957ee2c72311884efab985bbfc12fb3187f/Appendix%201_Guide%20to%20processing%202bRAD%20sequences.txt

Appendix 2 – Mantel test and genpop file conversion

https://github.com/Jes8x/M.S.-Thesis-Supplemental-Material/blob/ebab4957ee2c72311884efab985bbfc12fb3187f/Appendix%202_Genpop%20file%20conversion%20and%20Mantel%20tests.R

Appendix 3 – NMDS R code

https://github.com/Jes8x/M.S.-Thesis-Supplemental-Material/blob/ebab4957ee2c72311884efab985bbfc12fb3187f/Appendix%203_NMDS%20Script.R

Appendix 4 - Phylosignal R Code

https://github.com/Jes8x/M.S.-Thesis-Supplemental-Material/blob/ebab4957ee2c72311884efab985bbfc12fb3187f/Appendix%204_Phylo%20Signal.R

Appendix 5 – Morphology data

Sample		Mouth Width	Tentacle Length	Oral Disc Width	Number of Tentacles
JES201	Avg. ± Std dev.	0.337 ± 0.136	0.374 ± 0.103	1.645 ± 0.409	63 ± 1.414
	n	3	18	8	2
JES202	Avg. ± Std dev.	0.367 ± 0.005	1.073 ± 0.168	2.53 ± 0.105	44.333 ± 4.509
	n	3	6	5	3
JES203	Avg. ± Std dev.	0.43 ± 0.072	0.543 ± 0.093	2.68 ± 0.316	52.75 ± 2.816
	n	10	46	16	8
JES204	Avg. ± Std dev.	0.311 ± 0.05	0.503 ± 0.117	1.895 ± 0.234	51.778 ± 4.521
	n	11	28	20	9
JES205	Avg. ± Std dev.	0.322 ± 0.144	0.406 ± 0.068	1.906 ± 0.469	47.333 ± 4.163
	n	4	18	7	3
JES206	Avg. ± Std dev.	0.536 ± 0.053	1.106 ± 0.176	3.863 ± 0.353	53 ± 4.243
	n	7	26	7	2
JES207	Avg. ± Std dev.	0.368 ± 0.080	0.481 ± 0.122	2.21 ± 0.272	48.5 ± 3
	n	11	30	9	4
JES208	Avg. ± Std dev.	0.365 ± 0.031	1.233 ± 0.231	2.8 ± 0.251	49.5 ± 5.260
	n	4	13	10	4
JES209	Avg. ± Std dev.	0.402 ± 0.074	0.915 ± 0.156	1.607 ± 0.155	53 ± 4.243
	n	5	4	4	2
JES2010	Avg. ± Std dev.	0.44 ± 0.071	0.767 ± 0.156	2.865 ± 0.113	48
	n	2	21	4	2

JES2011	Avg. ± Std dev.	0.25 ± 0.028	0.422 ± 0.038	1.567 ± 0.300	52.667 ± 6.429
	n	2	12	4	3
JES2012	Avg. ± Std dev.	0.284 ± 0.038	0.578 ± 0.137	1.822 ± 0.151	45.667 ± 0.816
	n	7	36	13	6
JES2013	Avg. ± Std dev.	0.425 ± 0.007	0.903 ± 0.235	2.74 ± 0.044	50
	n	2	21	5	2
JES2014	Avg. ± Std dev.	0.363 ± 0.021	1.077 ± 0.174	2.26 ± 0.244	47 ± 1.414
	n	3	18	5	2
JES2015	Avg. ± Std dev.	0.372 ± 0.074	1.144 ± 0.134	3.123 ± 0.289	50
	n	5	9	6	1
JES2016	Avg. ± Std dev.	0.452 ± 0.056	1.354 ± 0.108	2.392 ± 0.310	42
	n	4	8	8	3

Appendix 6 – Zoanthid color proportions

	Black	**Brown**	**Blue**	**White**	**Green**	**Yellow**	**Red**
JES201	0%	36%	0%	1%	0%	0%	63%
JES202	39%	51%	0%	8%	0%	0%	2%
JES203	0%	45%	0%	2%	0%	0%	54%
JES204	90%	0%	8%	0%	0%	0%	2%
JES205	0%	0%	0%	0%	100%	0%	0%
JES206	0%	91%	0%	7%	1%	2%	0%
JES207	90%	0%	10%	0%	0%	0%	0%
JES208	0%	54%	0%	5%	0%	0%	41%
JES209	0%	0%	15%	16%	46%	22%	0%
JES2010	0%	76%	0%	6%	17%	0%	0%
JES2011	0%	62%	20%	0%	0%	18%	0%
JES2012	16%	82%	0%	3%	0%	0%	0%
JES2013	97%	3%	0%	0%	0%	0%	0%
JES2014	0%	65%	0%	18%	0%	17%	0%
JES2015	0%	84%	0%	4%	0%	12%	0%
JES2016	0%	0%	100%	4%	0%	0%	0%

Appendix 7 – Canonical correspondence analysis

https://github.com/Jes8x/M.S.-Thesis-Supplemental-Material/blob/80c37b5a29d54bbe941181fd9151c83a71cdd973/Appendix%207_Correspondence%20Analysis.R

Appendix 8 – recolorize R code

https://github.com/Jes8x/M.S.-Thesis-Supplemental-Material/blob/d486dce0e6935ba67300e80362dfadb55ab6cde8/Appendix%208_recolorize.R