Universal target-enrichment baits for anthozoan (Cnidaria) phylogenomics: New approaches to long-standing problems

Andrea M. Quattrini
Brant C. Faircloth
Luisa F. Dueñas
Tom C.L. Bridge
Mercer R. Brugler

See next page for additional authors

Follow this and additional works at: https://digitalcommons.uri.edu/bio_facpubs

The University of Rhode Island Faculty have made this article openly available. Please let us know how Open Access to this research benefits you.

This is a pre-publication author manuscript of the final, published article.

Terms of Use
This article is made available under the terms and conditions applicable towards Open Access Policy Articles, as set forth in our Terms of Use.
Authors
Andrea M. Quattrini, Brant C. Faircloth, Luisa F. Dueñas, Tom C.L. Bridge, Mercer R. Brugler, Iván F. Calizto-Botía, Danielle M. DeLeo, Sylvain Forêt, Santiago Herrera, Simon M.Y. Lee, David J. Miller, Carlos Prada, Gandhi Rádis-Baptista, Catalina Ramirez-Portilla, Juan A. Sánchez, Estefanía Rodríguez, and Catherine S. McFadden
Universal target-enrichment baits for anthozoan (Cnidaria) phylogenomics: New approaches to long-standing problems


¹Department of Biology, Harvey Mudd College, 1250 N Dartmouth Ave, Claremont, CA 91711, USA
²Department of Biological Sciences and Museum of Natural Science, Louisiana State University, Baton Rouge, LA 70803, USA
³Departamento de Ciencias Biológicas-Facultad de Ciencias, Laboratorio de Biología Molecular Marina (BIOMMAR), Universidad de los Andes, Carrera 1 No. 18A – 12, (J 212, J308 Lab), Bogotá, Colombia
⁴Queensland Museum Network, 70-102 Flinders St, Townsville, Queensland 4810, Australia
⁵Australian Research Council Centre of Excellence for Coral Reef Studies, James Cook University, Townsville, Queensland 4811, Australia

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/1755-0998.12736

This article is protected by copyright. All rights reserved.
Abstract

Anthozoans (e.g., corals, anemones) are an ecologically important and diverse group of marine metazoans that occur from shallow to deep waters worldwide. However, our understanding of the evolutionary relationships among the ~7500 species within this class is hindered by the lack of phylogenetically informative markers that can be reliably sequenced.
across a diversity of taxa. We designed and tested 16,306 RNA baits to capture 720 Ultraconserved Element loci and 1,071 exon loci. Library preparation and target enrichment was performed on 33 taxa from all orders within the class Anthozoa. Following Illumina sequencing and Trinity assembly, we recovered 1,774 of 1,791 targeted loci. The mean number of loci recovered from each species was 638 ± 222, with more loci recovered from octocorals (783 ± 138 loci) than hexacorals (475 ±187 loci). Parsimony informative sites ranged from 26-49% for alignments at differing hierarchical taxonomic levels (e.g., Anthozoa, Octocorallia, Hexacorallia). The percent of variable sites within each of three genera (Acropora, Alcyonium, and Sinularia) for which multiple species were sequenced ranged from 4.7-30%. Maximum likelihood analyses recovered highly resolved trees with topologies matching those supported by other studies, including the monophyly of the order Scleractinia. Our results demonstrate the utility of this target-enrichment approach to resolve phylogenetic relationships from relatively old to recent divergences. Re-designing the baits with improved affinities to capture loci within each sub-class will provide a valuable toolset to address systematic questions, further our understanding of the timing of diversifications, and help resolve long-standing controversial relationships in the class Anthozoa.

Introduction

Anthozoan cnidarians play critical roles in many marine ecosystems. The class contains ~7,500 extant species (i.e., soft corals, sea fans, stony corals, black corals, and anemones) that live worldwide in a variety of marine habitats—from tropical shallow waters to the cold, deep sea (Daly et al. 2007). Classification of Anthozoa has traditionally been based on morphological characters such as skeletal morphology, colony organization and soft-tissue anatomy of the polyps (Daly et al. 2007), including the arrangement of internal mesenteries (Fautin and Mariscal 1991). Long-standing views have recognized the anthozoan sub-classes Octocorallia and Hexacorallia as reciprocally monophyletic (Daly et al. 2007), a view also supported by recent phylogenomic analyses of 10s to 100s of genes (Zapata et al. 2015; Pratlong et al. 2017). Within each sub-class, however, molecular phylogenetic studies have revealed widespread homoplasy in morphological characters and widespread polyphyly at the ordinal, sub-ordinal, family, and
Consequently, deep flaws exist in our understanding of the phylogenetic relationships among and within anthozoan orders. Attempts to resolve the deep phylogenetic relationships among anthozoans using molecular data have largely been unsuccessful due to relatively slow evolutionary rates of mitochondrial genomes (Shearer et al. 2002; Hellberg 2006; Huang et al. 2008; Forsman et al. 2009), lack of signal in rDNA (Bernston et al. 2001; Daly et al. 2003) and difficulty in identifying and developing PCR primers for single-copy nuclear genes that can be amplified across the entire class (McFadden et al. 2011).

Within most anthozoan orders, there is also a lack of phylogenetic resolution at the species level. This may be due to incomplete lineage sorting in gene trees, insufficient data due to the small number of currently available markers, hybridization, and/or lack of morphological synapomorphies in taxonomy (McFadden et al. 2010, 2011, 2017; Prada et al. 2014; Rodríguez et al. 2014; Grajales and Rodríguez 2016; Daly et al. 2017). Currently available markers are insufficient at resolving species boundaries for the majority of anthozoans. For octocorals, an extended mitochondrial barcode (COI+igr1+mtMutS) has proven useful for revealing cryptic species and delimiting species boundaries within some clades; however, the divergence criterion proposed (McFadden et al. 2011) to elucidate these boundaries is low (>0.5% p-distance) and often no genetic divergence is observed among congeneric species (McFadden et al. 2011, Dueñas et al. 2014, Pante et al. 2015). The low genetic variability in the mitochondrial genome has been attributed to a unique mismatch repair enzyme (mtMutS) that potentially repairs mutations (Bilewitch and Degnan 2011) thereby causing reduced mitochondrial sequence variation in octocorals when compared to other metazoans (Shearer et al. 2002). Mitochondrial sequence variation is also low in the hexacorals (Hellberg et al. 2006; Daly et al. 2010), creating difficulties in resolving species boundaries using traditional mitochondrial barcodes (i.e., COI, Hebert et al. 2003; Shearer and Coffroth 2008). Although several studies have resolved species boundaries using a nuclear ITS marker (e.g., Medina et al. 1999; Pinzon and LaJeunesse 2011), using ITS poses problems as it is not a single-locus marker (Vollmer and Palumbi 2004) and there are often high levels of intra-specific variation (Van Oppen et al. 2000). Methods that allow for collecting and analyzing numerous loci across shallow and deep levels of divergence are sorely needed.

NGS-based methods that have been developed to enable the capture of large numbers of
homologous loci in large-scale phylogenetic studies include amplicon sequencing, restriction site-associated DNA (RADseq) methods, transcriptome sequencing and target enrichment of genomic DNA (see McCormack et al. 2013a). Although RADSeq is an effective approach for species-level phylogenetics and species delimitation within anthozoan genera (e.g., Combosch and Vollmer 2015; Pante et al. 2015; Herrera and Shank 2016; McFadden et al. 2017; Johnston et al. 2017), using RADseq to address deeper-level relationships is not feasible due to locus dropout (Althoff et al. 2007; McCormack et al. 2013a). Transcriptomic data have been used to reconstruct deep relationships within Cnidaria (Zapata et al. 2015; Pratlong et al. 2017), but the need for RNA limits the use of this method to taxa for which fresh material can be collected and preserved appropriately. Alternatively, target enrichment of ultraconserved elements (UCEs) (Faircloth et al. 2012) has proven robust in inferring species histories of both vertebrates [e.g., fishes (Faircloth et al. 2013), birds (McCormack et al. 2013b), reptiles (Crawford et al. 2012), and mammals (McCormack et al. 2012)] and invertebrates [e.g., arachnids (Starrett et al. 2016), hymenopterans (Branstetter et al. 2017), and coleopterans (Baca et al. 2017)] across shallow to deep timescales. UCEs occur in high numbers throughout genomes across the tree of life, including Cnidaria (Ryu et al. 2012), making them easy to identify and align among divergent species (Faircloth et al. 2012). As the name implies, UCEs are highly conserved regions of the genome, but the flanking regions surrounding UCEs are more variable and phylogenetically informative (Faircloth et al. 2012). Some advantages of using target enrichment of UCEs include that 100s to 1000s of loci can be sequenced at a relatively low cost from a wide range of taxa (Faircloth et al. 2012); they can be generated from 100 year old, formalin-preserved museum specimens and specimens with degraded DNA (McCormack et al. 2016; Ruane and Austin 2017); and they have proven useful at resolving evolutionary questions across both shallow and deep time scales (Smith et al. 2013; McCormack et al. 2013b; Manthey et al. 2016). Similar approaches using target-enrichment of coding regions, or exon capturing (Bi et al. 2012; Ilves and López-Fernández 2014; Hugall et al. 2016), have also proven valuable in phylogenomics.

We used all available genomes and transcriptomes to design a set of target-capture baits for enriching both UCEs and exons for use in anthozoan phylogenetics. Herein, we discuss how loci were targeted and baits were designed. Using an in silico analysis, we demonstrate that these loci recover the established sub-class and ordinal relationships among anthozoans. Finally, we test the utility of these baits in vitro using 33 species from across both sub-classes of Anthozoa.
Materials and Methods
Preparation of Genomes and Transcriptomes

Genomic and transcriptomic data were gathered from various sources for use in bait design and in silico testing (Table S1). All data were masked for repetitive regions, retroelements, small RNAs, and transposons using RepeatMasker open-4.0 (Smit et al. 2015). The N50 was calculated for each genome using stats.sh in the BBTools package (Bushnell 2015). We then constructed 2bit files for all genomes and transcriptomes (faToTwoBit, BLAT Suite, Kent 2002) and simulated 100 bp paired reads from each genome and transcriptome using the program art_illumina (Huang et al. 2012) in order to map simulated reads back to the genomes. All programs and parameters used for the entire workflow can be found in Supplemental File 1.

Identification of UCE Loci and Bait Design

We used the open-source program PHYLUCE (Faircloth 2016) and followed the workflow in the online tutorial (http://phyluce.readthedocs.io/en/latest/tutorial-four.html), with a few modifications to identify conserved regions and design baits to target these regions for downstream next-generation sequencing (Faircloth 2017). We first aligned an average of 34 million, 100 bp simulated-reads from each of the four exemplar taxa, Acropora digitifera, Exaiptasia pallida, Renilla muelleri, and Pacifigorgia irene, to a base genome, Nematostella vectensis. Nematostella vectensis (‘nemve’) was chosen as the base genome for the primary bait design because it is one of the most well-assembled and annotated anthozoan genomes. We used stampy v. 1 (Lunter and Goodson 2011), with a substitution rate set at 0.05, to map conserved regions of each read-simulated genome to the base genome. Across all taxa, 0.6 to 1.8% of the reads mapped to the nemve genome. The resulting alignment file was transformed from SAM format into BAM format (samtools, Li et al. 2009) and then transformed into a BED formatted file (BEDtools, Quinlan and Hall 2010). These BED files were sorted by scaffold/contig and then by position along that scaffold/contig. We then merged together the alignment positions in each file that were close (<100 bp) to one another using bedtools. In addition, sequences that included masked regions (>25%) or ambiguous (N or X) bases or were too short (<80 bp) were removed using phyluce_probe_strip_masked_loci_from_set. These steps resulted in BED files containing regions of conserved sequences shared between nemve and each of the exemplar taxa for further
analysis. An SQLite table was created using phyluce_probe_get_multi_merge_table, and included 70,312 loci that were shared between pairs of taxa.

We queried the SQLite table and output a list of 1,794 conserved regions found in nemve and the other four exemplar taxa using phyluce_probe_query_multi_merge_table. This list plus phyluce_probe_get_genome_sequences_from_bed was used to extract the conserved regions from the nemve genome. These regions were buffered to 160 bp by including an equal amount of 5' and 3' flanking sequence from the nemve genome. Another filter was performed at this stage to remove sequences < 160 bp, sequences with > 25% masked bases, or sequences with ambiguous bases. A temporary set of sequence capture baits was designed from the loci found in this final FASTA file. Using phyluce_probe_get_tiled_probes, we designed the bait set by tiling two 120 bp baits over each locus that overlapped in the middle by 40 bp (3X density). This temporary set of baits was screened to remove baits with >25% masked bases or high (>70%) or low (<30%) GC content. Any potential duplicates were also removed using phyluce_probe_easy_lastz and phyluce_probe_remove_duplicate_hits_from_probes_using_lastz. Bait sequences were considered duplicates if they were ≥50% identical over ≥50% of their length.

The temporary bait set (2,131 baits, targeting 1,787 loci) was aligned back to nemve and the four exemplar taxa using phyluce_probe_run_multiple_lastzs_sqlite, with an identity value of 70% (the minimum sequence identity for which a bait could be an accepted match to the genome) and a minimum coverage of 83% (default value). From these alignments, baits that matched multiple loci were removed. We then extracted 180 bp of the sequences from the alignment files and input the data into FASTA files using phyluce_probe_slice_sequence_from_genomes. A list containing 710 loci found in at least three of the taxa was created. Based on this list of 710 loci, the anthozoan UCE bait set was re-designed to target these 710 loci using phyluce_probe_get_tiled_probe_from_multiple_inputs, nemve, and the four exemplar genomes. Using this script, 120-bp baits were tiled (3X density, middle overlap) and screened for high (>70%) or low (<30%) GC content, masked bases (>25%), and duplicates. This bait set included a total of 5,459 non-duplicated baits targeting 710 anthozoan loci. All above methods were repeated to produce additional octocoral-specific baits and capture octocoral-specific loci. We repeated the above analyses using R. muelleri as the base genome and P. irene, Paragorgia stephencairnsi, and Antillogorgia bipinnata as the exemplar taxa to add 1,317 baits targeting an additional 168 UCE loci to the dataset.
Identification of Exon Loci and Bait Design

To design baits to target exon regions, the above methods were repeated using available transcriptome data. An average of 7 million reads from five exemplar transcriptome-enabled taxa (A. digitifera, Cerianthidae, Edwardsiella lineata, Gorgonia ventalina, and Paramuricea sp.) were simulated and 1.1 to 15.3% of these reads per species were aligned to the nemve transcriptome. After we converted the alignments to BED files, merged overlapping reads, and filtered data for short loci and repetitive regions, 44,215 conserved sequences were added to an SQLite database. We queried this database and selected 3,700 loci that were found in nemve and the additional five exemplar taxa. Following a second screening for masked regions, high/low GC content, and duplicates, a temporary exon bait set (5,661 baits) targeting 3,633 exon loci was designed. The temporary baits were re-aligned to the transcriptomes of nemve and the additional five exemplar anthozoans to ensure we could locate the loci. A set of 906 loci that were found in nemve and the additional five exemplar anthozoans were added to an SQLite database. We re-designed the exon bait set to target these 906 exon loci using phyluce_probe_get_tiled_probe_from_multiple_inputs, nemve, and the five exemplar transcriptomes. This bait set included a total of 8,080 non-duplicated baits targeting 906 loci across all anthozoans. To add more octocoral-specific baits and loci, we then repeated the above analyses with Paramuricea sp. as the base transcriptome and Anthomastus sp., Corallium rubrum, Eunicea flexuosa, G. ventalina, Keratoisidinae sp., and Nephyigorgia sp. as the exemplar taxa to add 4,914 baits targeting an additional 407 loci to the dataset.

Final Bait Screening

All of the bait sets designed with various sets of data as described above (see Table S1) were screened against one another to remove redundant baits (>50% identical over >50% of their length), allowing us to create a final non-duplicated Anthozoa bait set. We also screened these baits (70% identity, 70% coverage) against the Symbiodinium minutum genome by using phyluce_probe_run_multiple_lastzs_sqlite and phyluce_probe_slice_sequence_from_genomes and removed loci that matched the symbiont. Bait names in the final bait FASTA file begin with ‘uce-’ if designed using genomes to target UCEs and ‘trans-’ if designed using transcriptomes to target exons.
In Silico Test

In silico tests were performed to check how well the designed baits aligned to existing genomes and transcriptomes. First, phyluce_probe_run_multiple_lastzs_sqlite was used to align the UCE baits to the nine 2-bit formatted genomes and an outgroup genome (Hydra magnipapillata) and the exon baits to the 24 2-bit formatted transcriptomes (Table S1). An identity value of 50% was chosen for alignments (following the PHYLUCE tutorial). For each bait test, the matching FASTA data were sliced out of each genome or transcriptome, plus 200 bp of 5' and 3' flanking regions, using phyluce_probe_slice_sequence_from_genomes. This resulted in an average of 429 ± 178 SD (44 to 599 per species) UCE loci and 497 ± 230 SD (206 to 857) exon loci per anthozoan species (Table 1). To do a final screen for duplicates, loci were matched back to the baits using phyluce_assembly_match_contigs_to_probes, with a minimum coverage of 67% and minimum identity of 80% (default values following the PHYLUCE tutorial). Here, an average of 355 ± 166 SD (25 to 529 per species) non-duplicate UCE loci and 354 ± 210 SD (106 to 670) non-duplicate exon loci were recovered per anthozoan species (Table 1). Each locus was exported into a FASTA file and aligned with MAFFT (Katoh et al. 2002) using phyluce_align_seqcap_align with default parameters.

The resulting alignments were trimmed internally using GBlocks (Castresana 2000, Talavera and Castresana 2007) using phyluce_align_get_gblocks_trimmed_alignments_from_untrimmed with default parameters. Two final datasets were then created using phyluce_align_get_only_loci_with_min_taxa, in which all locus alignments contained at least 4 of the 10 taxa for the genome data and 9 of the 24 taxa for the transcriptome data. We then concatenated the resulting alignments into separate supermatrices; one containing UCE loci from 10 genome-enabled taxa and the other containing exon loci from the 24 transcriptome-enabled taxa. Maximum likelihood (ML) inference was conducted on each supermatrix using RAxML v8 (Stamatakis 2014). This analysis was carried out using rapid bootstrapping, which allows for a complete analysis (20 ML searches and 200 bootstrap replicates) in one step. We also conducted a Bayesian inference (10 million generations, 35% burnin) using ExaBayes (Aberer et al. 2014). An extended majority rule consensus tree was produced. A General-Time Reversible model of
nucleotide substitution with a gamma distributed rate variation (GTRGAMMA) was used in both ML and Bayesian analyses.

**In Vitro Test**

Following the *in silico* test, the list of designed baits was sent to MYcroarray for synthesis. MYcroarray further screened and removed baits that either had repetitive elements or the potential to cross-hybridize (0.007% total baits removed). We then tested the bait set on 33 anthozoan specimens (Table 2), with both sub-classes and all major orders and sub-orders (for Octocorallia) represented. DNA from these specimens included recent extractions from tissue that had been stored frozen (in liquid nitrogen) for 25 yrs or in 95% EtOH for up to 10 yrs, as well as extractions that had been stored frozen (-20 °C) for 10 yrs (see Table S2).

DNA was extracted using a Qiagen DNeasy Blood & Tissue kit, Qiagen Gentra Kit, or a CTAB extraction protocol (McFadden *et al.* 2006). DNA quality was assessed using a Nanodrop spectrophotometer, with 260/280 ratios ranging from 1.8-2.1 and 260/230 ratios ranging from 1.4-3.2. The initial concentration of each sample was measured with a Qubit 2.0 fluorometer. For the majority of samples, we then sheared approximately 600 ng DNA (10 ng per μL) to a target size range of 400-800 bp using sonication (Q800R QSonica Inc. Sonicator). For eight samples (Table 2), we sheared 35 μL (115-372 ng, average 217 ng) of EDTA-free DNA using enzymes from the Kapa HyperPlus (Kapa Biosystems) library preparation kit. These samples were mixed on ice with 5 μL of Kapa Frag buffer and 10 μL of the Kapa Frag enzyme and put on a pre-cooled (4°C) thermocycler prior to incubation for 10-15 min at 37 °C to achieve a target size range of 400-800 bp. After shearing, DNA was run out on a 1% agarose gel (120V, 60 min). Small DNA fragments were removed from each sample (250 ng DNA) using a generic SPRI substitute (Rohland and Reich 2012; Glenn *et al.* 2016) bead cleanup (3X). DNA was re-suspended in 25 μL double-distilled water (ddH20).

Details of library preparation and target enrichment can be found in Supplemental File 2. Briefly, library preparation (Kapa Biosystems) was carried out on the majority of DNA samples (Table 2) using a Kapa Hyper Prep protocol. For the subset of the samples for which DNA was sheared using enzymes (Table 2), we followed the protocol in the Kapa Hyper Plus enzyme-shearing library preparation kit (Kapa Biosystems). Universal Y-yoke oligonucleotide adapters and custom iTru dual-indexed primers were used in library preparations (Glenn *et al.* 2016). For
target enrichment, the MYcroarray MyBaits were diluted in 1/2 (250 ng) of the standard (500 ng) MyBaits reaction, using 2.5 μL of the baits and 2.5 μL of ddH2O for all samples. Different bait strengths were tested on a set of six samples (Table 2): full bait strength (500 ng), 1/2 bait strength (250 ng), 1/4 bait strength (125 ng), and 1/8 strength (63 ng). One combined pool of all enriched libraries was sent to Oklahoma Medical Research Facility for sequencing on 2/3 of a lane of Illumina HiSeq 3000 (150bp PE reads).

Post-Sequencing Analyses

De-multiplexed Illumina reads were processed using PHYLUCE following the workflow in the online tutorial (http://phyluce.readthedocs.io/en/latest/tutorial-one.html/), with a few modifications (Suppl. File 1). The reads were first trimmed using the Illumiprocessor wrapper program (Faircloth 2012) with default values and then assembled using Trinity v. 2.0 (Haas et al. 2013). We also assembled the data using Abyss 2.0 (Simpson et al. 2009) with a kmer value of 31. UCE and exon bait sequences were then separately matched to the assembled contigs (70% identity, 70% coverage) using phyluce_assembly_match_contigs_to_probes to locate the loci. Loci were then extracted using phyluce_assembly_get_match_counts and phyluce_assembly_get_fasstas_from_match_counts, exported into separate FASTA files and aligned with default parameters using phyluce_align_seqcap_align, which uses MAFFT. Loci were internally trimmed with GBlocks using phyluce_align_get_gblocks_trimmed_alignments_from_untrimmed with default parameters.

Data matrices of locus alignments were created using phyluce_align_get_only_loci_with_min_taxa, in which each locus had either 25% or 50% species occupancy. Concatenated locus alignments consisted of exon loci only, UCE loci only, and all loci. The number of parsimony informative sites was calculated for each alignment across various taxonomic datasets. The script phyluce_align_get_informative_sites was used on the following taxonomic datasets: Anthozoa+genome+outgroup (33 taxa used in in vitro test, plus nine genome-enabled taxa and the outgroup H. magnipapillata), Anthozoa (33 taxa used in in vitro test), Hexacorallia only (17 taxa used in in vitro test), and Octocorallia only (16 taxa used in in vitro test). The total number of variable sites, total number of parsimony informative sites and number of parsimony informative sites per locus were calculated. We also calculated the total number of variable sites and the number of variable sites per locus for alignments.
containing species in each of three genera: *Acropora* (*A. digitifera, A. millepora, A. muricata*),
*Alcyonium* (*A. acaule, A. digitatum, A. haddoni*), and *Sinularia* (*S. slieringsi, S. lochmodes, S.
*maxima*). For the three *Acropora* species, we used loci from one target-capture enrichment
sample and from the two *Acropora* genomes that were available.

ML inference was conducted on each alignment (exon loci only, UCE loci only, and all
loci) for the Anthozoa+genome+outgroup taxon set using RAxML v8. This analysis was carried
out using rapid bootstrapping, which allows for a complete analysis (20 ML searches and 200
bootstrap replicates) in one step. We also conducted a Bayesian analysis (10 million generations,
35% burnin) on the 25 and 50% all-loci datasets using ExaBayes (Aberer *et al.* 2014). An
extended majority rule consensus tree was produced. A GTRGAMMA model was used in both
ML and Bayesian analyses.

**Results**

**Identification of Loci and Bait Design**

A total of 16,306 baits were designed to capture 1,791 anthozoan loci with 4 to 10 baits
targeting each locus. The principal UCE bait set included 5,513 baits designed to target 720 loci.
The principal exon bait set included 10,793 baits to target 1,071 loci. Four loci that matched
genomic regions in *Symbiodinium minutum* were removed from the dataset. These loci, however,
were also detected in azooxanthellate anthozoans, such as *Chrysogorgia tricaulis*.

**In Silico Test**

We generated two alignment matrices, one consisting of the exon loci taken from the
transcriptome-enabled taxa and the other one consisting of the UCE loci taken from the genome-
enabled taxa. The alignment matrix generated with the UCE loci, which included the *H.*
magnipapillata* outgroup, had a total of 522 loci, with a trimmed mean locus length of 373 bp
(95% CI: 8.4) and a total alignment length of 138,778 bp. The alignment matrix generated with
the exon loci included 407 loci, with a trimmed mean locus length of 462 bp (95% CI: 5.8) and a
total length of 220,139 bp. The ML phylogenies generated from these alignments were well
supported and recovered monophyletic sub-classes and established ordinal relationships (Fig. 1).
The phylogeny generated with the UCE loci had 100% support at all the nodes (Fig. 1a) whereas
the phylogeny generated with the exon loci had complete support at the majority (86%) of the

This article is protected by copyright. All rights reserved
nodes (Fig. 1b). Trees produced using Bayesian inference were congruent with ML results.

**In Vitro Test**

The total number of reads obtained from Illumina sequencing ranged from 460,724 to 17,283,798 reads per sample (mean: 5,938,769 ± 3,407,199 SD reads) across all bait strengths and Kapa kits tested (Table S2). Quality and adapter trimming lead to the removal of 1.8 to 10.5% reads from each sample, resulting in a mean of 5,486,800 ± 2,092,161 SD trimmed reads per sample (Tables S2, S3). Trimmed reads were assembled into 4,699 to 327,623 contigs per sample (mean: 92,076 ± 65,772 SD contigs) with a mean length of 384 ± 27 bp (range: 224 to 32,406 bp) using Trinity (Tables 2 and S3). Coverage averaged 2.5 to 9.9X per contig. No differences in numbers of contigs or reads were evident between libraries prepared using the two different Kapa kits (Hyper Prep or Hyper Plus) at 1/2 bait strength or between the different bait strengths used (1/8, 1/4, 1/2, full) (Fig. S1, Tables 2 and S3). Using Abyss, trimmed reads were assembled into 43,428 to 763,227 contigs per sample with a mean length of only 179 ± 24 bp. Because contig sizes were much smaller from Abyss than those assembled via Trinity, remaining analyses were conducted on the Trinity-assembled data.

A total of 713 UCE loci and 1,061 exon loci (1,774 total loci out of 1,791 targeted loci) were recovered from the assembled contigs. Mean length of UCE contigs was 598 ± 158 bp (range: 224 to 3,995 bp) and mean length of exon contigs was 593 ± 156 bp (range: 224 to 4,500 bp) (Table S2). No differences in numbers of loci were evident between the two different Kapa kits (Hyper Prep or Hyper Plus) at 1/2 bait strength or between the individuals subjected to the four different bait strengths used (Fig. S1, Tables 2 and S3). The number of loci recovered from each species using a Kapa Hyper prep kit with 1/2 bait strength was highly variable, ranging between 172 to 1034 total loci per sample (mean: 638 ± 222 loci) (Tables 2 and S3), although few loci (172) were recovered from the sample with the fewest contigs (15,433). More loci were recovered from octocorals (mean: 783 ± 138 loci, range: 569-1036 loci) compared to hexacorals (mean: 475 ±187 loci, range: 172-786 loci), even after removing the sample with the fewest loci (498 ±172 loci).

Alignment lengths, locus number and length, and the number of parsimony informative sites varied depending upon percent (25 or 50%) of taxon occupancy per locus and type of taxonomic dataset (Anthozoa+genome+outgroup, Anthozoa, Hexacorallia, Octocorallia)
included in the GBLOCKS trimmed alignments (Table 3). The average percentage of parsimony informative sites across all alignments was 39%. For the comparisons within each of three genera (Acropora, Alcyonium, Sinularia), 382 to 426 loci were retained in the 100% alignment matrices (Table 4). Mean % variable sites per locus ranged from 4.7 to 30%, with the most variation found in the Alcyonium dataset and the least found within Acropora. Percent variation per locus ranged from 0 to 55%, with only one non-polymorphic locus found in the Acropora dataset.

Tree topologies were mostly congruent between the 25% and 50% Anthozoa+genome+outgroup data matrices using all loci and the Bayesian and ML analyses (Figs. 2 and S2). Bootstrap support and posterior probabilities were higher overall in the 25% Anthozoa+genome+outgroup ML tree (Fig. 2) compared to the 50% dataset tree (Fig. 2, Fig. S2). By rooting to the outgroup H. magnipapillata, monophyly for the currently established anthozoan subclasses and the hexacoral orders was recovered in all analyses except that the sister relationship of Ceriantharia to the rest of the hexacorals was not supported in the Bayesian analysis of the 25% dataset. Only a few branches shifted between the ML trees produced with either of the data matrices. Acropora digitifera was sister to A. muricata in the 50% dataset, but sister to A. millepora in the 25% dataset. In Octocorallia, both Cornularia pabloi and Erythropodium caribaeorum shifted positions between 25% and 50% datasets. These two species and Tubipora musica also changed positions between Bayesian and ML analyses of the 50% dataset.

Lower bootstrap support was found in ML trees created with only the exon loci (Fig. S3C, S3D) or the UCE loci (Fig. S3A, S3B), but tree topologies were congruent with the few exceptions noted above (Fig. S3). Cerianthids were also found to be sister to all other anthozoans in both 25% and 50% exon-locus datasets, but sister to hexacorals in the UCE-locus datasets. Zoanthus cf. pulchellus was sister to the actiniarians in the 25% exon-locus dataset, but sister to a clade containing Actiniaria, Antipatharia, Corallimorpharia, and Scleractinia in all other datasets (Fig S3).

Discussion

Our results demonstrate the utility of the target-capture enrichment approach for inferring phylogenomic relationships in the class Anthozoa. To date, a few studies based on transcriptomic
data have recovered well-supported phylogenomic relationships within Anthozoa, but these studies were based on only a handful (≤15) of taxa (Zapata et al. 2015; Lin et al. 2016, Pratlong et al. 2017) and were limited in scope. In general, phylogenomic studies based on transcriptomic data have provided well-supported and well-resolved phylogenies based on 100s to 1000s of orthologs (Dunn et al. 2008; Kocot et al. 2011; Zapata et al. 2015). However, obtaining these types of sequencing data can be relatively expensive and requires high-quality RNA, two limitations that hinder the transcriptomic-approach for large datasets. In addition, it is often not feasible to obtain RNA from rare taxa or taxa that have not been properly preserved for transcriptomics, such as museum specimens. In our study, we show that a sequence-capture approach for both UCEs and exons can be used to capture genome-wide data in anthozoans. To date, this approach has not been applied to anthozoans or to marine invertebrates more generally (except Hugall et al. 2016). We successfully designed a novel bait set based on existing transcriptomes and genomes, and captured 1,774 loci from a diversity of anthozoans spanning >500 million years of divergence (Peterson et al. 2004). This target-enrichment approach has the capability to resolve evolutionary relationships at a wide range of divergence levels, from deep (orders, sub-orders) to shallow levels (species). This novel genomic resource can help to advance studies of systematics, divergence-time estimation, and character evolution in the species-rich class Anthozoa.

**In Vitro Test Results**

The newly designed bait set successfully enriched 713 UCE loci and 1,061 exon loci across a diversity of anthozoans. These loci had an average of 39% parsimony informative sites, comparable to the arachnid (30% PI sites, Starrett et al. 2016) UCE dataset, which targeted ~1,000 loci. The large range of loci recovered per anthozoan species (172 to 1036 loci) was also similar to the arachnid results (170 to 722 loci). We note that the number of loci recovered from octocorals was much higher than what was recovered from hexacorals. This result is perhaps because we added more octocoral-specific baits to the final bait set. And as we added more octocoral-specific baits, we removed baits that were potential paralogs; the majority of these were designed based on the hexacorals. As was done for the hymenopteran UCE bait set (Branstetter et al. 2017), we need to re-design the baitset and include additional octocoral-specific baits and hexacoral-specific baits to increase the success of locus capture. We will also
design separate octocoral- and hexacoral-specific bait sets so that additional loci specific to each sub-class can be targeted. Nevertheless, this first bait design and *in vitro* results from 33 taxa demonstrate the promising utility of the target-capture method for resolving anthozoan relationships across deep divergence levels.

The number of variable sites found at loci recovered from within three genera demonstrates that this is also a promising approach for species-level phylogenetics. Within all three genera examined, variable sites ranged up to 55% per locus, with a mean variation across all loci of 4.7, 5.5, and 30% in *Sinularia, Acropora,* and *Alcyonium,* respectively. The high variation seen within *Alcyonium* is consistent with unpublished data (C. McFadden, unpubl. data) that suggest the three species are perhaps different genera. For *Sinularia,* average divergence estimates are also higher (~10X) than what has been demonstrated in other studies using mitochondrial barcoding markers (McFadden et al. 2009). In fact, a 0.5% divergence level at an extended mitochondrial barcode (*mtMutS+igrI+COI*) was proposed as a conservative criterion for species delimitation (McFadden et al. 2011, 2014). Similarly, low divergence estimates at mitochondrial barcoding markers have been found among hexacoral congeners (Shearer and Coffroth 2008; Brugler et al. 2013; Gonzalez-Muñoz et al. 2015). Thus, these UCE and exon loci are promising for resolving species boundaries, although the level of intraspecific variation has yet to be determined. Our UCE and exon locus datasets may serve as an alternative resource to RADseq to address species-boundary questions while simultaneously allowing for data to be combined and examined across deeper levels.

Because this was the first time the target-enrichment UCE approach had been tested on anthozoans, we compared different concentrations of baits and different library preparation kits to determine whether or not particular methods would recover more loci. We found no differences in the number of loci recovered using different concentrations of baits in the hybridization and enrichment protocols. This bait-strength test suggested that the number of hybridizations obtained from one standard reaction could, at least, be doubled. We also found no differences between the two different Kapa kits used. The enzymatic DNA shearing that can be performed with the Kapa Hyper Plus kit may be useful for researchers who do not have access to a sonicator.

Following internal trimming with GBlocks and aligning of conserved loci, the mean locus length was much shorter (~190 bp) compared to the mean length of un-trimmed loci (~600...
Therefore, some of the loci included in the ML analyses were relatively short (< 100 bp), particularly in the Anthozoa+genome+outgroup dataset. In alignments between highly divergent taxa (such as between hexacorals and octocorals), numerous poorly aligned positions and divergent positions were filtered with GBlocks. In contrast, the locus size was considerably higher within genera (~525 bp) because of fewer poorly aligned and divergent positions. Perhaps re-performing the GBlocks internal trimming with less stringent parameters would increase the size of loci in alignments of divergent taxa. Stringent alignment filtering, as done with GBlocks, can not only increase the proportion of unresolved branches, but can also lead to well-supported branches that are in fact incorrect (Tan et al. 2015). Different methods of aligning and filtering data will be explored in future work.

The phylogenies produced from the in vitro data were highly supported despite low overall taxon occupancy (>25 or 50% matrices) and inclusion of short loci. There were a few nodes that had low support and a few branches that shifted between the different taxon occupancy datasets, particularly in the Octocorallia. In addition to stringent filtering as discussed above, sources of incongruence and low bootstrap support could include compositional bias, saturation, violations of model assumptions (Jeffroy et al. 2006) and/or missing data. Missing data, however, are generally not problematic if there are a reasonable number of informative characters (see Streicher et al. 2015). Rather, incongruence and low support at a few nodes is perhaps due to incomplete taxon sampling (Wiens 2005; Wiens and Tiu 2012). Although a diversity of taxa from across the clades was selected for in vitro analyses, several lineages were not represented, particularly in the Octocorallia. Outgroup choice and taxon evenness can also impact topology and clade support in UCE phylogenomics (Branstetter et al. 2017). Future efforts will need to incorporate more thorough taxon sampling.

In general, the inferred phylogenetic relationships corresponded to those found in previous studies (Zapata et al. 2015; Rodríguez et al. 2014), although there were a few exceptions. One exception was the position of the stoloniferan octocoral C. pabloi. In all datasets, this species was nested within the clade containing sea pens (Pennatulacea) and calcaxonians (C. tricaulis, Keratoisidinae sp.), but this species has been previously found to be sister to the rest of the octocorals based on mitochondrial data (McFadden and van Ofwegen 2012). The superfamily Actinostoloidea (Sicyonis sp., Stomphia sp.) was recovered as sister to superfamily Actinioidea (Actinostella sp., Isosicyonis alba) in all datasets. This result differed
from the combined mitochondrial and nuclear rDNA dataset of Rodríguez et al. (2014), which
instead recovered Actinostoloidea as sister to both Actinioidea and Metridioidea (*Lebrunia
danae, E. pallida, Bunodeopsis sp.*). Furthermore, trees in our study were rooted to *H.
magnipapillata*, based on the results of Zapata et al. (2015); however, the unrooted trees
indicated that *H. magnipapillata* was sister to the Octocorallia, a relationship (i.e., a paraphyletic
Anthozoa) that has been noted by mitochondrial data (Park et al. 2012, Kayal et al. 2013), but
not supported by phylogenomic analyses of transcriptomic data (Zapata et al. 2015). Zapata et al.
(2015) also found that the position of the order Ceriantharia was phylogenetically unstable.
Similarly, we found that the placement of Ceriantharia changed between the different exon and
UCE datasets. The topologies resulting from exon data placed the ceriantharians as sister to all
remaining anthozoans, a relationship also suggested by analysis of 16S and 18S ribosomal DNA
(Stampar et al. 2014). Trees from UCE loci had ceriantharians as sister to hexacorals; a
relationship supported by combined mitochondrial and nuclear rDNA data (Rodríguez et al.
2014). Future work must include different outgroup choices (i.e., sponges, bilateria, other
cnidarians), while closely examining the distribution and strength of phylogenetic signal. This
will help clarify the source of incongruence and resolve which loci strongly influence the
resolution of a given ‘contentious’ branch (Shen et al. 2017).

Whether or not scleractinians are monophyletic has been a controversial topic as a result
of different phylogenetic analyses. In 2006, Medina et al. reported that scleractinians were
polyphyletic with corallimorpharians. The “naked coral hypothesis” was thus proposed,
suggesting that corallimorpharians arose from a scleractinian ancestor that had undergone
skeletal loss during paleoclimate conditions when the oceans experienced increased CO₂
concentrations (Medina et al. 2006). Since that study, other studies based on transcriptomic data
(Lin et al. 2016), rDNA (Fukami et al. 2008), and mitochondrial data (Fukami et al. 2008; Park
et al. 2012; Kayal et al. 2013; Kitahara et al. 2014) recovered a monophyletic Scleractinia with
corallimorpharians as the sister clade. Our results also recovered a monophyletic Scleractinia,
thus supporting the conclusions of others that corallimorpharians are not naked corals. However,
increased sampling of robust, complex, and basal scleractinians is necessary to conclusively
address this issue.

*Future Research Directions*

This article is protected by copyright. All rights reserved
The in silico and in vitro tests of the novel bait set demonstrate that the target-enrichment approach of UCEs and exons is a promising new genomic resource for inferring phylogenetic relationships among anthozoans. Using this bait set, target-capture enrichment of the UCE and exon loci from at least 192 additional anthozoans is currently underway to further our understanding of character evolution and systematics of the clade. Adding more taxa will likely increase the accuracy of the phylogenetic inference. We also plan to use additional outgroup taxa, including medusozoan cnidarians and sponges, to help address whether or not octocorals are sister to hexacorals or medusozoans and resolve the position of ceriantharians. Finally, we plan to re-design the bait sets to create hexacoral- and octocoral-specific bait sets. We will include additional baits to increase the capture efficiency of loci that were targeted in this study, while adding more loci that are specific to each sub-class. This target-enrichment approach provides a promising genomic resource to resolve phylogenetic relationships at deep to shallow levels of divergence, considerably advancing the current state of knowledge of anthozoan evolution.

Acknowledgements

Funding was provided by NSF-DEB #1457817 to CSM and #1457581 to ER. The 2013 HIMB Pauley program also provided funds to AMQ for preliminary analyses. C. Oliveros and J. Salter helped during lab work and J. Bast helped during preliminary data analyses. F. Zapata and C. Dunn provided transcriptomic datasets. S. Lengyel, E. Cordes, and R. Kulathinal aided DD with the Paramuricea transcriptome assembly, which was funded by a grant from Gulf of Mexico Research to support the “Ecosystem Impacts of Oil and Gas in the Gulf” (ECOGIG) research consortium. E. Bush provided computing support. Specimens from Irish waters were collected during cruise CE14004 on the RV Celtic Explorer led by L. Allcock, NUI Galway, and supported by the Marine Institute's Shiptime Programme. Special thanks to Z. Forsman, R. Toonen, and I. Knapp for organizing the 2013 HIMB Pauley program. Thanks to J. Ryan and three anonymous reviewers for constructive comments and suggestions.

Author Contributions
AMQ, CSM, ER, and BCF conceived and designed this study. AMQ designed the baits, conducted library preparation, target enrichment, and data analyses, and wrote the initial draft of the manuscript with significant contributions from CSM. BCF developed protocols and guided AMQ in laboratory and bioinformatic analyses. LFD helped with preliminary analyses. MB, ER, and CSM extracted DNA. ICB, DMD, SF, SH, SL, DJM, CP, GRB, CRP, and JAS provided genomic or transcriptomic data for analysis. TB provided samples. All authors edited and approved the final version of this manuscript.

Data Accessibility
Tree and alignment files: Data Dryad Entry http://dx.doi.org/10.5061/dryad.36n40
Raw Data: SRA Genbank SUB3122367, BioSample #SAMN07774920-4952
Anthozoan bait set: Data Dryad Entry http://dx.doi.org/10.5061/dryad.36n40

References

This article is protected by copyright. All rights reserved


Crawford, N. G., Faircloth, B. C., McCormack, J. E., Brumfield, R. T., Winker, K., & Glenn, T. C. (2012). More than 1000 ultraconserved elements provide evidence that turtles are the sister group of archosaurs. Biology Letters, 8(5), 783-786.

This article is protected by copyright. All rights reserved


This article is protected by copyright. All rights reserved


This article is protected by copyright. All rights reserved


This article is protected by copyright. All rights reserved


This article is protected by copyright. All rights reserved


This article is protected by copyright. All rights reserved
Wiens, J. J., & Tiu, J. (2012). Highly incomplete taxa can rescue phylogenetic analyses from the negative impacts of limited taxon sampling. PLoS One, 7(8), e42925.

Zapata, F., Goetz, F. E., Smith, S. A., Howison, M., Siebert, S., Church, S. H., ... & Daly, M. (2015). Phylogenomic analyses support traditional relationships within Cnidaria. PLoS One, 10(10), e0139068

Figure Captions

Figure 1. Maximum likelihood phylogenies from in silico analyses. A) Phylogeny constructed with a 138,778 bp concatenated genomic dataset (522 loci) and rooted to Hydra magnipapillata. B) Phylogeny constructed with 220,139 bp concatenated transcriptome dataset (407 loci) with the Hexacorallia rooted to the Octocorallia. Bootstrap support (b.s.) values are followed by posterior probabilities (p.p.) from Bayesian analyses. *=100% b.s. and 1.0 p.p. Branches are color coded by order (green=Ceriantharia, pink=Zoantharia, purple=Scleractinia, blue=Actiniaria, red=Alcyonacea, grey=Pennatulacea)

Figure 2. Maximum likelihood phylogeny on the Anthozoa+genome+outgroup 25% matrix (257,728 bp, 1378 loci). The tree includes 33 taxa from the in vitro test, 9 genome-enabled taxa, and the outgroup Hydra magnipapillata. Bootstrap support (b.s.) values are followed by posterior probabilities (p.p) from Bayesian analyses *=100% b.s. and 1.0 p.p.; - =not supported by Bayesian analysis. Branches are color coded by order (green=Ceriantharia, pink=Zoantharia, brown=Antipatharia, purple=Scleractinia, lt. blue=Corallimorpharia, blue=Actiniaria, red=Alcyonacea grey=Pennatulacea)

This article is protected by copyright. All rights reserved
Table 1. Number of loci recovered from in silico analyses after initial and final screens for potential paralogs. Also included are the N50 and number of scaffolds for each genome/transcriptome used in analyses.

<table>
<thead>
<tr>
<th>Sub-class</th>
<th>Order</th>
<th>Species</th>
<th>#Scaffolds</th>
<th>N50</th>
<th>Initial Screen</th>
<th>Final Screen</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genomes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hexacoralla</td>
<td>Actiniaria</td>
<td>Exaiptasia pallida</td>
<td>4,312</td>
<td>442,145</td>
<td>518</td>
<td>417</td>
</tr>
<tr>
<td></td>
<td>Actiniaria</td>
<td>Nematostella vectensis</td>
<td>10,804</td>
<td>472,588</td>
<td>496</td>
<td>421</td>
</tr>
<tr>
<td></td>
<td>Actiniaria</td>
<td>Stomphia sp.</td>
<td>479,824</td>
<td>948</td>
<td>44</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>Scleractinia</td>
<td>Acropora digitifera</td>
<td>4,765</td>
<td>191,489</td>
<td>462</td>
<td>395</td>
</tr>
<tr>
<td></td>
<td>Scleractinia</td>
<td>Acropora millepora</td>
<td>12,559</td>
<td>181,771</td>
<td>511</td>
<td>414</td>
</tr>
<tr>
<td>Octocoralla</td>
<td>Alcyonacea</td>
<td>Antillogorgia bipinnata</td>
<td>426,978</td>
<td>3,212</td>
<td>230</td>
<td>134</td>
</tr>
<tr>
<td></td>
<td>Alcyonacea</td>
<td>Pacifigorgia irene</td>
<td>183,211</td>
<td>2,323</td>
<td>547</td>
<td>491</td>
</tr>
<tr>
<td></td>
<td>Alcyonacea</td>
<td>Paragorgia stephencairns</td>
<td>700,190</td>
<td>1,793</td>
<td>453</td>
<td>371</td>
</tr>
<tr>
<td>Octocoralla</td>
<td>Penatulacea</td>
<td>Renilla muelleri</td>
<td>4,114</td>
<td>19,024</td>
<td>599</td>
<td>529</td>
</tr>
<tr>
<td>Hydrozoa</td>
<td>Aplanulata</td>
<td>Hydra magnipapillata</td>
<td>126,667</td>
<td>10,113</td>
<td>449</td>
<td>99</td>
</tr>
<tr>
<td>Transcriptomes</td>
<td>Actiniaria</td>
<td>Anemonia sp.</td>
<td>14,279</td>
<td>703</td>
<td>235</td>
<td>106</td>
</tr>
<tr>
<td></td>
<td>Actiniaria</td>
<td>Anthopleura elegantissima</td>
<td>142,934</td>
<td>1,489</td>
<td>364</td>
<td>207</td>
</tr>
<tr>
<td></td>
<td>Actiniaria</td>
<td>Edwardsia lineata</td>
<td>90,440</td>
<td>1,035</td>
<td>841</td>
<td>623</td>
</tr>
<tr>
<td></td>
<td>Actiniaria</td>
<td>Exaiptasia pallida</td>
<td>60,101</td>
<td>2,159</td>
<td>553</td>
<td>264</td>
</tr>
<tr>
<td>Kingdom</td>
<td>Phylum</td>
<td>Class</td>
<td>Genus</td>
<td>Species</td>
<td>Count</td>
<td>Width</td>
</tr>
<tr>
<td>------------------</td>
<td>-----------------</td>
<td>------------------</td>
<td>----------------------</td>
<td>----------------------</td>
<td>-------</td>
<td>-------</td>
</tr>
<tr>
<td>Hexacorallia</td>
<td>Actiniaria</td>
<td>Metridium sp.</td>
<td></td>
<td></td>
<td>10,885</td>
<td>752</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Nematostella v.</td>
<td></td>
<td></td>
<td>27,273</td>
<td>1,524</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ceriantharia</td>
<td>Cerianthidae sp.</td>
<td></td>
<td>12,074</td>
<td>646</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Scleractinia</td>
<td>Acropora digitifera</td>
<td></td>
<td>36,780</td>
<td>1,575</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Scleractinia</td>
<td>Acropora hyacinthus</td>
<td></td>
<td>67,844</td>
<td>422</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Scleractinia</td>
<td>Fungia scutaria</td>
<td></td>
<td>155,914</td>
<td>1,619</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Scleractinia</td>
<td>Montastraea cavernosa</td>
<td></td>
<td>200,222</td>
<td>2,145</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Scleractinia</td>
<td>Orbicella faveolata</td>
<td></td>
<td>32,463</td>
<td>1,736</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Scleractinia</td>
<td>Pocillopora damicornis</td>
<td></td>
<td>70,786</td>
<td>976</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Scleractinia</td>
<td>Porites astreoides</td>
<td></td>
<td>30,740</td>
<td>661</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Scleractinia</td>
<td>Platygyra daedalea</td>
<td></td>
<td>51,200</td>
<td>684</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Zoantharia</td>
<td>Protopalythoa variabilis</td>
<td></td>
<td>130,118</td>
<td>1,187</td>
</tr>
<tr>
<td>Octocorallia</td>
<td>Alcyonacea</td>
<td>Anthomastus sp.</td>
<td></td>
<td></td>
<td>9,368</td>
<td>610</td>
</tr>
<tr>
<td></td>
<td>Alcyonacea</td>
<td>Corallium rubrum</td>
<td></td>
<td></td>
<td>48,074</td>
<td>2,470</td>
</tr>
<tr>
<td></td>
<td>Alcyonacea</td>
<td>Eunicea flexuosa</td>
<td></td>
<td></td>
<td>165,709</td>
<td>1,095</td>
</tr>
<tr>
<td></td>
<td>Alcyonacea</td>
<td>Gorgonia ventalina</td>
<td></td>
<td></td>
<td>90,230</td>
<td>1,149</td>
</tr>
<tr>
<td></td>
<td>Alcyonacea</td>
<td>Keratoisidinae</td>
<td></td>
<td></td>
<td>12,385</td>
<td>702</td>
</tr>
<tr>
<td></td>
<td>Alcyonacea</td>
<td>Nephthyigorgia sp.</td>
<td></td>
<td></td>
<td>14,677</td>
<td>762</td>
</tr>
<tr>
<td></td>
<td>Alcyonacea</td>
<td>Paramuricea sp.</td>
<td></td>
<td></td>
<td>25,189</td>
<td>2,645</td>
</tr>
<tr>
<td></td>
<td>Alcyonacea</td>
<td>Scleronephthya sp.</td>
<td></td>
<td></td>
<td>8,401</td>
<td>683</td>
</tr>
</tbody>
</table>
Table 2. List of species used in the in vitro test of designed baits with assembly summary statistics. Results are from the Kapa Hyper Prep and Hyper Plus (in bold) library preparation kits with target enrichments performed using 250 ng of baits.

<table>
<thead>
<tr>
<th>Sub-Class</th>
<th>Order</th>
<th>Species</th>
<th># Contigs</th>
<th>Mean Contig Length (bp)</th>
<th># UCEs</th>
<th># Exon Loci</th>
<th>Total # Loci</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hexacorallia</td>
<td>Actiniaria</td>
<td>Actinostella sp.</td>
<td>184,605</td>
<td>440</td>
<td>345</td>
<td>441</td>
<td>786</td>
</tr>
<tr>
<td>Hexacorallia</td>
<td>Actiniaria</td>
<td>Bunodeopsis sp.</td>
<td>82,100</td>
<td>413</td>
<td>285</td>
<td>257</td>
<td>364</td>
</tr>
<tr>
<td>Hexacorallia</td>
<td>Actiniaria</td>
<td>Halcurias pilatus*^</td>
<td>89,449/27,355</td>
<td>379/387</td>
<td>254/158</td>
<td>258/144</td>
<td>512/302</td>
</tr>
<tr>
<td>Hexacorallia</td>
<td>Actiniaria</td>
<td>Isosicyonis alba*^</td>
<td>88,159/37,119</td>
<td>368/360</td>
<td>210/146</td>
<td>184/138</td>
<td>394/284</td>
</tr>
<tr>
<td>Hexacorallia</td>
<td>Actiniaria</td>
<td>Lebrunia danae</td>
<td>187,114</td>
<td>403</td>
<td>340</td>
<td>368</td>
<td>708</td>
</tr>
<tr>
<td>Hexacorallia</td>
<td>Actiniaria</td>
<td>Sicyonis sp.^*</td>
<td>50,490/105,326</td>
<td>402/407</td>
<td>174/238</td>
<td>287/249</td>
<td>461/487</td>
</tr>
<tr>
<td>Hexacorallia</td>
<td>Antipatharia</td>
<td>Antipathes grandis%^</td>
<td>57,950</td>
<td>323</td>
<td>185</td>
<td>197</td>
<td>382</td>
</tr>
<tr>
<td>Hexacorallia</td>
<td>Antipatharia</td>
<td>Myriopathes ulex%^</td>
<td>96,476</td>
<td>356</td>
<td>248</td>
<td>267</td>
<td>515</td>
</tr>
<tr>
<td>Hexacorallia</td>
<td>Ceriantharia</td>
<td>Cerianthus membranaceus^*</td>
<td>146,327/143,221</td>
<td>397/372</td>
<td>206/212</td>
<td>231/227</td>
<td>437/439</td>
</tr>
<tr>
<td>Hexacorallia</td>
<td>Ceriantharia</td>
<td>Pachycerianthus sp.</td>
<td>101,786</td>
<td>426</td>
<td>188</td>
<td>198</td>
<td>386</td>
</tr>
<tr>
<td>Hexacorallia</td>
<td>Corallimorpharia</td>
<td>Corynactis chilensis%^</td>
<td>15,433/44,166</td>
<td>362</td>
<td>95/179</td>
<td>77/187</td>
<td>172/366</td>
</tr>
<tr>
<td>Hexacorallia</td>
<td>Corallimorpharia</td>
<td>Discosoma carlgreni</td>
<td>37,499</td>
<td>353</td>
<td>223</td>
<td>260</td>
<td>483</td>
</tr>
<tr>
<td>Hexacorallia</td>
<td>Scleractinia</td>
<td>Acropora muricata</td>
<td>93,433</td>
<td>378</td>
<td>322</td>
<td>408</td>
<td>730</td>
</tr>
<tr>
<td>Hexacorallia</td>
<td>Scleractinia</td>
<td>Pavona sp.^%</td>
<td>57,223</td>
<td>340</td>
<td>232</td>
<td>251</td>
<td>483</td>
</tr>
<tr>
<td>Hexacorallia</td>
<td>Scleractinia</td>
<td>Pocillipora damicornis</td>
<td>4,699</td>
<td>339</td>
<td>123</td>
<td>105</td>
<td>228</td>
</tr>
</tbody>
</table>

This article is protected by copyright. All rights reserved
<table>
<thead>
<tr>
<th>Hexacorallia</th>
<th>Scleractinia</th>
<th>Stylophora pistillata</th>
<th>162,597</th>
<th>394</th>
<th>297</th>
<th>311</th>
<th>606</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hexacorallia</td>
<td>Zoantharia</td>
<td>Zoanthus cf. pulchellus</td>
<td>164,870</td>
<td>373</td>
<td>209</td>
<td>195</td>
<td>542</td>
</tr>
<tr>
<td>Octocorallia</td>
<td>Alcyonacea</td>
<td>Alcyonium acaule</td>
<td>93,846</td>
<td>401</td>
<td>363</td>
<td>543</td>
<td>906</td>
</tr>
<tr>
<td>Octocorallia</td>
<td>Alcyonacea</td>
<td>Alcyonium digitatum</td>
<td>43,531</td>
<td>393</td>
<td>343</td>
<td>486</td>
<td>829</td>
</tr>
<tr>
<td>Octocorallia</td>
<td>Alcyonacea</td>
<td>Alcyonium haddoni</td>
<td>66,764</td>
<td>414</td>
<td>348</td>
<td>570</td>
<td>918</td>
</tr>
<tr>
<td>Octocorallia</td>
<td>Alcyonacea</td>
<td>Chrysogorgia tricaulis</td>
<td>111,571</td>
<td>413</td>
<td>235</td>
<td>331</td>
<td>566</td>
</tr>
<tr>
<td>Octocorallia</td>
<td>Alcyonacea</td>
<td>Clavularia inflata</td>
<td>84,673</td>
<td>352</td>
<td>247</td>
<td>325</td>
<td>572</td>
</tr>
<tr>
<td>Octocorallia</td>
<td>Alcyonacea</td>
<td>Coelogorgia palmosa</td>
<td>127,823</td>
<td>437</td>
<td>367</td>
<td>572</td>
<td>939</td>
</tr>
<tr>
<td>Octocorallia</td>
<td>Alcyonacea</td>
<td>Cornularia pabloi</td>
<td>107,331</td>
<td>371</td>
<td>292</td>
<td>359</td>
<td>651</td>
</tr>
<tr>
<td>Octocorallia</td>
<td>Alcyonacea</td>
<td>Erythropodium caribaeorum</td>
<td>119,210</td>
<td>398</td>
<td>316</td>
<td>417</td>
<td>733</td>
</tr>
<tr>
<td>Octocorallia</td>
<td>Alcyonacea</td>
<td>Keratoisidinae sp.</td>
<td>70,544</td>
<td>426</td>
<td>233</td>
<td>344</td>
<td>577</td>
</tr>
<tr>
<td>Octocorallia</td>
<td>Alcyonacea</td>
<td>Parasphaerasclera valdiviae</td>
<td>85,199</td>
<td>404</td>
<td>323</td>
<td>443</td>
<td>766</td>
</tr>
<tr>
<td>Octocorallia</td>
<td>Alcyonacea</td>
<td>Plexaura kuna</td>
<td>105,208</td>
<td>393</td>
<td>423</td>
<td>611</td>
<td>1034</td>
</tr>
<tr>
<td>Octocorallia</td>
<td>Alcyonacea</td>
<td>Sinularia slieringsi</td>
<td>75,970</td>
<td>377</td>
<td>321</td>
<td>516</td>
<td>837</td>
</tr>
<tr>
<td>Octocorallia</td>
<td>Alcyonacea</td>
<td>Sinularia lochmodes</td>
<td>58,759</td>
<td>386</td>
<td>314</td>
<td>514</td>
<td>828</td>
</tr>
<tr>
<td>Octocorallia</td>
<td>Alcyonacea</td>
<td>Sinularia maxima</td>
<td>42,099</td>
<td>366</td>
<td>304</td>
<td>528</td>
<td>832</td>
</tr>
<tr>
<td>Octocorallia</td>
<td>Alcyonacea</td>
<td>Tubipora musica</td>
<td>44,753</td>
<td>369</td>
<td>282</td>
<td>451</td>
<td>733</td>
</tr>
<tr>
<td>Octocorallia</td>
<td>Pennatulacea</td>
<td>Virgularia schultzei</td>
<td>49,954</td>
<td>381</td>
<td>269</td>
<td>509</td>
<td>777</td>
</tr>
</tbody>
</table>

^ Kapa HyperPlus Kit Trial

% Kapa Hyper Prep Kit Trial Library failed

* Probe Concentration Trials

This article is protected by copyright. All rights reserved
Table 3. Alignment matrix statistics for different taxonomic datasets. Matrix percentage equals the percent occupancy of species per locus. PI=parsimony informative sites.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>%</th>
<th>#</th>
<th># Loci (UCE / exon)</th>
<th>Alignment Length</th>
<th>Mean Locus Length (± SD bp)</th>
<th>Locus Length Range (bp)</th>
<th># PI Sites</th>
<th>%PI Sites</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anthozoa+genome+outgroup*</td>
<td>50</td>
<td>429</td>
<td>228 / 201</td>
<td>81,403</td>
<td>190 ± 89</td>
<td>23-549</td>
<td>40,041</td>
<td>49</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>1375</td>
<td>626 / 749</td>
<td>257,153</td>
<td>187 ± 91</td>
<td>23-601</td>
<td>119,117</td>
<td>46</td>
</tr>
<tr>
<td>Anthozoa</td>
<td>50</td>
<td>464</td>
<td>229 / 235</td>
<td>91,455</td>
<td>197 ± 93</td>
<td>50-667</td>
<td>43,501</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>1330</td>
<td>575 / 755</td>
<td>254,596</td>
<td>191 ± 99</td>
<td>19-823</td>
<td>109,930</td>
<td>43</td>
</tr>
<tr>
<td>Hexacorallia</td>
<td>50</td>
<td>438</td>
<td>223 / 215</td>
<td>89,757</td>
<td>205 ± 93</td>
<td>52-693</td>
<td>34,390</td>
<td>38</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>1052</td>
<td>529 / 523</td>
<td>248,476</td>
<td>236 ± 107</td>
<td>52-1362</td>
<td>63,968</td>
<td>26</td>
</tr>
<tr>
<td>Octocorallia</td>
<td>50</td>
<td>831</td>
<td>334 / 496</td>
<td>208,869</td>
<td>251 ± 127</td>
<td>51-967</td>
<td>70,369</td>
<td>34</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>1366</td>
<td>548 / 818</td>
<td>368,275</td>
<td>270 ± 132</td>
<td>51-1013</td>
<td>96,255</td>
<td>26</td>
</tr>
</tbody>
</table>

* includes 33 taxa used in test run, 9 genome-enabled taxa, and the outgroup Hydra magnipapillata

This article is protected by copyright. All rights reserved
Table 4. Summary statistics for congeneric species alignments. Mean % variation per locus is also included for UCE loci and exon loci, respectively (in parentheses).

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Min</th>
<th># Loci (UCE / Exon)</th>
<th>Alignment Length (± SD bp)</th>
<th>Mean Locus Length</th>
<th>Locus Length Range (bp)</th>
<th># Variable Sites</th>
<th>Range % Variation per Locus</th>
<th>Mean % Variation per Locus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acropora</td>
<td>3</td>
<td>398</td>
<td>206,067</td>
<td>517 ± 73</td>
<td>229-670</td>
<td>9,474</td>
<td>0*-46.0</td>
<td>4.7 (4.3, 5.0)</td>
</tr>
<tr>
<td>Alcyonium</td>
<td>3</td>
<td>382</td>
<td>205,676</td>
<td>538 ± 250</td>
<td>129-1470</td>
<td>60,283</td>
<td>6.0-55.0</td>
<td>30 (28, 31)</td>
</tr>
<tr>
<td>Sinularia</td>
<td>3</td>
<td>426</td>
<td>248,264</td>
<td>583 ± 245</td>
<td>91-1423</td>
<td>14,231</td>
<td>0.3-27.0</td>
<td>5.5 (5.2, 5.6)</td>
</tr>
</tbody>
</table>

* Only one locus was not polymorphic