

Title: Whole genome assembly and annotation of the endangered Caribbean coral *Acropora cervicornis*

Authors: Jason D. Selwyn* and Steven V. Vollmer

Affiliations:

Department of Marine and Environmental Sciences, Northeastern University; 430 Nahant Road, Nahant, MA 01908, USA.

*Corresponding author. Email: j.selwyn@northeastern.edu

Keywords:

Acropora cervicornis, Acroporidae, Scleractinia, *de novo* assembly, comparative genomics, time-calibrated phylogeny

Abstract

Coral species in the genus *Acropora* are key ecological components of coral reefs worldwide and represent the most diverse genus of scleractinian corals. While key species of Indo-Pacific *Acropora* have annotated genomes, no annotated genome has been published for either of the two species of Caribbean *Acropora*. Here we present the first fully annotated genome of the endangered Caribbean staghorn coral, *Acropora cervicornis*. We assembled and annotated this genome using high-fidelity nanopore long-read sequencing with gene annotations validated with mRNA sequencing. The assembled genome size is 318 Mb, with 28,059 validated genes. Comparative genomic analyses with other *Acropora* revealed unique features in *A. cervicornis*, including contractions in immune pathways and expansions in signaling pathways. Phylogenetic analysis confirms previous findings showing that *A. cervicornis* diverged from Indo-Pacific relatives around 41 million years ago, with the closure of the western Tethys Sea, prior to the primary radiation of Indo-Pacific *Acropora*. This new *A. cervicornis* genome enriches our understanding of the speciose *Acropora* and addresses evolutionary inquiries concerning speciation and hybridization in this diverse clade.

Introduction

The *Acropora* are one of the most speciose and important genera of reef-building scleractinian corals globally (Wallace 1999). The genus *Acropora* are divided into multiple speciose Indo-Pacific clades and a single depauperate Caribbean clade (Wallace 1999). The two sister species of Caribbean *Acropora* – the Staghorn coral *A. cervicornis* and Elkhorn coral *A. palmata* – and their hybrid – called *A. prolifera* (Vollmer and Palumbi 2002) are thought to have diverged from the Indo-Pacific *Acropora* during the late Eocene after the closure of the western Tethys Sea prior to the rapid diversification in the Indo-Pacific *Acropora* (Wallace 1999; Wallace and Portell 2022). To date, all 16 published *de novo* assembled and annotated *Acropora* genomes are of Indo-Pacific species (Shinzato *et al.* 2011, 2021; Ying *et al.* 2019; Fuller *et al.* 2020; López-Nandam *et al.* 2023).

Acropora, like all corals, are severely threatened by anthropogenic climate change leading to elevated water temperatures that can cause acute bleaching and subsequently death (Hughes *et al.* 2018). The Caribbean *Acropora* are also experiencing a secondary range-wide pressure in the form of White Band Disease, which has resulted in ~95% population losses of both species Caribbean wide and is the direct cause of their listing on the Endangered Species List (Aronson and Precht 2001; National Marine Fisheries Service 2006). Because these two species are such important foundational species in the Caribbean reef ecosystem these losses have likely had tremendous unknown effects on higher order taxa which depend on *Acropora* dominant reefs for survival.

Here we present the first fully annotated genome for the endangered Caribbean staghorn coral *Acropora cervicornis*, importantly representing the first Caribbean species of this diverse clade. This genome was assembled using a combination of long-read nanopore and short-read shotgun sequences and annotated and validated using mRNA sequencing. This reference genome will accelerate genomic research on this endangered coral and address fundamental evolutionary questions about speciation and hybridization in the speciose Acroporids.

Materials and Methods

Sample Collection and Sequencing

High molecular weight genomic DNA was extracted in June 2021 from adult tissue of the K2 genotype maintained in the Coral Restoration Foundation (CRF) Key Largo, Florida nursery. Three libraries were prepared using Oxford Nanopore Technologies (ONT) kit SQK-LSK112. Two libraries were not size selected while the third included 20+kb PippinPrep size-selection. All ONT prepared libraries were sequenced separately on three Minion flow cells (FLO-MIN112). High-quality base-calling was performed using Guppy v6.1.7 (Oxford Nanopore Technologies).

Four additional Illumina PCR-Free shotgun libraries were constructed using the DISCOVAR protocol to produce libraries with fragments between 400 and 600 bp (Love *et al.* 2016). KAPA PCR-free library kits were leveraged with the addition of a second round of 0.7x Agencourt AmPure XP SPRI bead cleanup post adapter ligation. Libraries were multiplexed and sequenced on a single rapid-run HiSeq 2500 flowcell with 250 bp paired-end sequencing. Additionally, a library of paired 150 bp reads was prepared using the Illumina DNA Prep kit and sequenced as part of a NovaSeq S4 run.

To acquire transcriptome data, mRNA sequence data was obtained for 48 individuals (including the K2 genotype) using NEBs unidirectional mRNA library preparations sequenced on an Illumina NEXTSEQ 550 platform and combined with previously published RNA sequencing data from 38 additional *A. cervicornis* (PRJNA222758; Libro and Vollmer 2016; PRJNA423227; Parkinson *et al.* 2018).

Sequence Quality Control

Nanopore long-reads (DNA) were quality controlled using PORECHOP (v 0.2.3_seqan2.1.1, <https://github.com/rrwick/Porechop>) to remove adapter sequences and then quality trimmed into

longer assembly reads (minimum average quality 3, minimum length 1,000 bp) and shorter polishing reads (minimum average quality 5, minimum length 500 bp) using NANOFILT (De Coster *et al.* 2018). Illumina sequenced short-reads (DNA and RNA) were quality controlled initially using FASTP (Chen *et al.* 2018) to remove adapters and barcodes, filter low quality sequences (PHRED < 30), trim sequences shorter than 140 bp, and PCR artifacts. Contaminants were removed with FASTQ_SCREEN (Wingett and Andrews 2018) by mapping reads against a suite of potential contaminant genomes (e.g. human, viral, and bacterial) as well the 13 available genomes of *Symbiodiniaceae* (Table S1) and removing reads which had hits to any potential contaminant genome.

Genome Assembly

We estimated the genome size using a k-mer counting approach implemented in JELLYFISH based on the quality controlled Illumina short reads (Marçais and Kingsford 2011). An initial genome assembly of the ONT sequenced reads was built using FLYE (Kolmogorov *et al.* 2019) with the nano-hq parameter setting to fit with the chemistry and base-calling method of the sequencing. After initial assembly of the raw nanopore reads, duplicated sequences were removed using purge_dups (Guan *et al.* 2020; Guiguelmoni *et al.* 2021). We polished the genome using two rounds of long read polishing with RACON (Vaser *et al.* 2017) followed by a round of polishing with MEDAKA v1.7.2 (Oxford Nanopore Technologies) and then two rounds of polishing with paired-end Illumina sequences using PILON (Walker *et al.* 2014). To ensure a final genome assembly of the highest quality and contiguity, we corrected misassembly errors with short reads using MEC (Wu *et al.* 2020) and remove the mitochondrial genome and other potential genomic contaminants using BLOOTOOLS (Laetsch and Blaxter 2017). This was followed by misassembly correction using the long read data (Coombe *et al.* 2021), scaffolding (Coombe *et al.* 2021), gap closing (Xu *et al.* 2020), and a final round of long read polishing with RACON and short read polishing with PILON (Walker *et al.* 2014; Vaser *et al.* 2017). Finally, all contigs shorter than 1,000 bp and/or with no gene annotations (see below) were removed as they did not contain any BUSCOs.

Transcriptome Assembly

A preliminary transcriptome was assembled using a genome-guided assembly in TRINITY (Grabherr *et al.* 2011). mRNA sequencing reads were splice-aware mapped to the genome using GSNAP (Wu *et al.* 2016). TRINITY was used to perform a genome guided assembly of the transcriptome with an assumed max intron length of 100,000 bp.

Genome Annotation

Genome annotation was performed using MAKER (Cantarel *et al.* 2008; Holt and Yandell 2011). Repetitive elements were identified and masked using REPEATMODELER (Flynn *et al.* 2020) and REPEATMASKER (Smit *et al.* 2015). Evidence-based gene annotation was performed using as the assembled transcriptome and proteins identified in either all Acroporids in the UniProt database (The UniProt Consortium 2023) or the closest reference proteomes in UniRef from *Stylophora pistillata* (Voolstra *et al.* 2017), *Pocillopora damicornis* (Cunning *et al.* 2018), *Actinia tenebrosa* (Surm *et al.* 2019) and *Nematostella vectensis* (Putnam *et al.* 2007). This

initial round of annotation was used to train the *ab initio* gene identification models of Augustus (Stanke *et al.* 2006), Snap (Korf 2004), and Genemark-ES (Lomsadze *et al.* 2005). After the initial round of annotation based solely on protein and RNA evidence, we performed four subsequent rounds of annotation with the results of the previous round being used to train the *ab initio* gene predictors run in the subsequent round of annotation. Genes were functionally annotated using EnTAP (Hart *et al.* 2020) and Interproscan (Zdobnov and Apweiler 2001), and formatted the annotations for NCBI using GAG and ANNIE (Tate *et al.* 2014; Geib *et al.* 2018).

Mitochondrial assembly

The mitochondrial genome was assembled using the quality controlled Illumina short-reads using MITOZ (Meng *et al.* 2019). Briefly, this was done by first assembling a subset of reads into initial contigs which are then identified as mitogenome sequences using a profile Hidden Markov Model (Wheeler and Eddy 2013; Xie *et al.* 2014; Nurk *et al.* 2017). Contigs were then annotated to find the 13 protein coding mitochondrial genes along with tRNAs and rRNAs with any contigs not containing any annotations removed (Birney *et al.* 2004; Gertz *et al.* 2006; Li and Durbin 2009; Jühling *et al.* 2012; Nawrocki and Eddy 2013). Finally, retained contigs were assembled and circularized into the complete mitochondrial genome and visualized (Gertz *et al.* 2006; Krzywinski *et al.* 2009; Meng *et al.* 2019).

Phylogenetic Analysis

Sixteen published *Acropora* genomes with structural genome annotations along with two *Montipora* species, *Montipora capitata* and *Montipora efflorescens*, two Pocilloporids, *Pocillopora damicornis* and *Stylophora pistillata*, and three anemones, *Nematostella vectensis*, *Actina tenebrosa*, and *Exaiptasia diaphana* were downloaded from NCBI (Table 1). Protein sequences for all annotated genes were extracted and clustered into orthogroups derived from a single gene in the last common ancestor of the group using ORTHOFINDER (Emms and Kelly 2019). Orthogroups were used to infer rooted gene and species trees to develop a phylogenetic hypothesis for the group (Emms and Kelly 2017, 2018). Specifically, we used the STAG (Emms and Kelly 2018) algorithm to infer the species tree from 7,110 multi-copy gene trees which had all species present, each created using DendroBLAST (Kelly and Maini 2013), this species tree was then rooted using the STRIDE algorithm (Emms and Kelly 2017). To infer divergence times, we time-calibrated the species tree using least-squares dating and 1,000 bootstraps to estimate divergence time confidence intervals (To *et al.* 2016). Ancestral dates were gathered from the Fossilworks database (Table S2; Behrensmeier and Turner 2013).

Comparative Genomics

Structural gene annotations for all species included in the phylogenetic hypothesis were functionally annotated using BLAST (Altschul *et al.* 1990; Camacho *et al.* 2009) against the Swiss-Prot curated portion of the UniProt database (The UniProt Consortium 2023). To identify KEGG orthologs for all orthogroups, we matched KEGG gene annotations to orthogroups across species and found the consensus KEGG ortholog across all species in which the orthogroup was found. KEGG ortholog membership within KEGG pathways was identified using KEGGREST (Tenenbaum and Volkening 2023).

We used a logistic regression model to test if there are differences in the percentage of Orthogroups with KEGG annotations across taxa. Posthoc tests were used to compare the rate of annotation in *A. cervicornis* to other *Acropora sp.* and make pairwise comparisons between genera. To identify systematic differences in the distribution of genes in pathways within the *Acropora*, we used χ^2 tests followed by a post-hoc analysis with FDR correction to identify the species and KEGG pathways with significantly more or less gene copies (i.e. orthogroups) than would be expected compared to the other *Acropora* (Benjamini and Hochberg 1995; Beasley and Schumacker 1995).

The time-calibrated phylogeny along with the number of genes found within orthogroups for each species was used to estimate the rate of gene family evolution across the phylogeny using CAFE5 (Hahn *et al.* 2005; Mendes *et al.* 2020) and identify KEGG orthologs which exhibit significant expansions/contractions at each node of the phylogeny. In the CAFE5 analysis we assumed a single rate of evolution across all gene families using the error rate estimating model and Poisson prior distribution (De Bie *et al.* 2006; Han *et al.* 2013). We then performed an overrepresentation analysis using Fisher's exact test to determine if any pathways in *A. cervicornis* showed significant expansions/contractions. All statistical analyses were performed using R v4.2.1 (R Core Team 2022).

Results and Discussion

Genome/Transcriptome assembly and annotation

Nanopore sequencing resulted in ~6.6 million reads containing 15.5 Gb with an N50 of 5,072 bp with 3.3 million high-quality reads after filtering used in the initial long-read assembly containing 13.3 Gb of DNA with an N50 of 6,078 bp and a polishing set of 4.5 million reads containing 13.3 Gb of DNA with an N50 of 5,269 bp. 92 million paired-end Illumina reads (totally 40 Gb) were retained after filtering and decontamination for x polishing. mRNA sequencing from 86 corals totaled 1.4 billion single-end reads and 174 Gb of mRNA sequencing data for gene annotation.

Using the filtered paired-end short-read sequences, we estimated the genome size of *A. cervicornis* to be ~318 Mb, somewhat smaller than the Pacific Acroporids (384 – 475 Mb; Shinzato *et al.* 2011, 2021; Ying *et al.* 2019; Fuller *et al.* 2020), and obtained ~42x genomic coverage of the long-read nanopore sequences for the genomic assembly. The initial genome assembly measured 328 Mb with 3,014 contigs (N50 = 0.94 Mb; L50 = 99) with the longest contig being 4.9 Mb and a BUSCO completeness of 92.8% (1.2% duplicated, 3.7% fragmented). After genomic post-processing, polishing, and scaffolding, the final genome assembly measured 307 Mb (96.5% the estimated genome size, Table 3) with 398 scaffolded contigs (N50 = 2.8 Mb; L50 = 35) with the longest scaffold measuring 8.3 Mb and a BUSCO completeness of 92.4% (0.4% duplicated, 3.6% fragmented, Figure 1). The pooled transcriptome contained 374,749 transcripts with a total N50 of 3,659 and longest isoform N50 of 1,483 with a BUSCO completeness of 93.4% (78% duplicated, 3.5% fragmented). The high degree of transcriptome duplication in the BUSCO value is likely due to the pooling of individuals to create the transcriptome. Similar to other *Acropora* genomes (Shinzato *et al.* 2021), *A. cervicornis*

contained 39% interspersed repeats (Table S3) with the most common identifiable class of repeat being short interspersed nuclear elements (SINEs), though the plurality (17%) of repeats in the genome were unable to be classified and may be taxon specific (Shinzato *et al.* 2021). *A. cervicornis* had 28,059 validated genes (Table 4), intermediate among other Acroporids (25,278 - 41,860; Shinzato *et al.* 2011, 2021; Ying *et al.* 2019; Fuller *et al.* 2020), with 53% having a SwissProt annotation at an e-value $< 10^{-6}$ and a BUSCO completeness of 81.6% (2.8% duplicated, 10% fragmented).

Mitochondrial Genome

The *A. cervicornis* mitochondrial genome was assembled into a single circular structure containing 18,259 bp, 13 protein coding genes two tRNA genes (tRNA-Met and tRNA-Trp), and two rRNA genes (16s and 12s; Figure 2, GenBank accession number: OQ772303). Like all *Acropora*, the *A. cervicornis* mitochondrial genome has a particular dearth of tRNA coding sequences when compared to other metazoans, likely due to this diversification occurring after the split between cnidarians and metazoans (van Oppen *et al.* 1999, 2002a; Liu *et al.* 2015; Tian and Niu 2017; Colin *et al.* 2021). The gene order found in *A. cervicornis* is the same as that found in other Acroporids, including ND5 being split into two portions with a large intron containing all genes except the two tRNA coding genes, the large ribosomal subunit rRNA, ATP8, and COI (Figure Y; van Oppen *et al.* 2002a; Liu *et al.* 2015; Zhang *et al.* 2016; Colin *et al.* 2021).

Comparative Genomics

The 28,059 genes found in *A. cervicornis* belong to 15,191 distinct orthogroups of which 54.4% had a KEGG annotation. The percent of orthogroups with KEGG annotations varied significantly by species ($\chi^2_{(23)} = 595.8$, $p < 0.0001$) due primarily to differences between genera (*Acropora* vs *Montipora*, $Z = 12.9$, $p < 0.0001$) and higher order taxonomic comparisons outside of the *Acropora* (*Acropora* vs outgroups, $Z = -17.6$, $p < 0.0001$) rather than within *Acropora* which all show a similar degree of annotation (*A. cervicornis* vs other *Acropora*, $Z = -0.87$, $p = 0.38$). The number of orthogroups per KEGG pathway were similar for all *Acropora* sp. ($\chi^2_{(4,976)} = 478.1$, $p = 1$) indicating that no *Acropora* genomes had broad gains or losses of genes. Amongst five major KEGG categories, 25.7% of the orthogroups were found in pathways involved in organismal systems and 24.8% were involved in environmental information processing. The remaining orthogroups were split between cellular processes (18.8%), metabolism related processes (16.7%), and genetic information processing (14.1%). *Acropora cervicornis* possessed 135 unique orthogroups, more than the other *Acropora* species (17 – 70), possibly as a result of the separate evolutionary history in the Caribbean. An alternative hypothesis for this discrepancy could be that the *ab initio* gene models for most other *Acropora* were trained on *A. digitifera* rather than being trained for each species independently (Shinzato *et al.* 2021).

Six gene pathways were significantly overrepresented in *Acropora cervicornis* among the significantly expanded and contracted KEGG orthologs (Table S4). Two immune system pathways - NOD-like receptor signaling (map04621, OR = 19.7, $p < 0.001$, $p_{adj} < 0.001$) and Neutrophil extracellular trap formation (map04613, OR = 16.9, $p < 0.001$, $p_{adj} = 0.018$) – and the

related cellular growth and death pathway Necroptosis (map04217, OR = 17.6, $p < 0.001$, $p_{adj} = 0.018$) were significantly overrepresented due to one expansion and three contractions in NACHT, LRR and PYD domain-containing proteins, three contractions in histone proteins (H2A, H3, and H4) and one expansion in a cation channel protein (TRPM7). In corals, NOD-like receptor signaling pathway has been found to be an integral part of the innate immune system (Hamada *et al.* 2013) and histone phosphorylation has been shown to be a key response to nutrient stress and regulating the dinoflagellate symbionts (Rodriguez-Casariago *et al.* 2018).

The second major group of overrepresented pathways are signaling pathways, particularly the calcium signaling pathway (map04020, OR = 9.93, $p < 0.001$, $p_{adj} = 0.025$) and the neuroactive ligand-receptor interaction (map04080, OR = 39.8, $p < 0.001$, $p_{adj} < 0.001$). While the immune related pathways were predominantly gene contractions in these pathways all KEGG orthologs showed expansions (Table S4). The calcium signaling and neuroactive ligand-receptor pathways have been found to be important in regulation reproduction (Hilton *et al.* 2012; Rosenberg *et al.* 2017), nematocyst regulation (Russell and Watson 1995), and biomineralization (Reyes-Bermudez *et al.* 2009). Further these pathways appear to be differentially expressed between the growth tips and bases in the Caribbean *Acropora* (Hemond *et al.* 2014).

The final KEGG pathway overrepresented among the rapidly evolving KEGG orthologs is the Taurine and hypotaurine metabolism pathway (map00430, OR = 91.6, $p < 0.001$, $p_{adj} = 0.04$). This pathway was driven by the loss of a single KEGG orthogroup, cysteine dioxygenase (-2, K00456). Unlike other scleractinian corals, Indo-Pacific *Acropora* lack the cystathionine beta-synthase gene involved in cysteine biosynthesis (Shinzato *et al.* 2011) suggesting that *Acroporid* corals must rely on their algal symbionts to supply this vital amino acid (Shinzato *et al.* 2014). However, an alternative cysteine biosynthesis pathway has recently been discovered in *Acropora loripes* suggesting *Acropora* corals can natively synthesize cysteine through this alternate pathway (Salazar *et al.* 2022). Similar to the other studied *Acropora*, *A. cervicornis* lacks genes coding for cystathionine beta-synthase but does possess the genes required for the alternate cysteine biosynthesis pathway (Salazar *et al.* 2022). The loss of cysteine dioxygenase in *A. cervicornis* may suggest that the alternative pathway does not fully compensate the biosynthesis of cysteine resulting in a lack of excess cysteine needing to be metabolized.

Phylogeny

Acropora cervicornis is estimated to have diverged from the other *Acropora* 41 mya (35 – 47 95% CI) during the Paleogene, approximately coinciding with the initial closure of the Tethys Sea (van Oppen *et al.* 2001; Wallace and Portell 2022). The addition of *Acropora cervicornis* as a Caribbean *Acropora* to the published *Acropora* phylogenomic tree (Shinzato *et al.* 2021) suggests that the *Acropora* radiation in the Indo-Pacific occurred in at least two stages with an initial split occurring between 58 – 68 million years ago resulting in clades I and II prior to the divergence of the Caribbean and Indo-Pacific *Acropora*, followed by a second set of radiations of the Indo-Pacific *Acropora* 35 – 47 mya and resulting in the more speciose clades III and IV (Figure 3). Low bipartition support values within *Acropora* could result from the relatively rapid radiation within *Acropora*, incomplete lineage sorting and/or introgressive hybridization (van Oppen *et al.* 2002b; Vollmer and Palumbi 2002; Minh *et al.* 2020). The Acroporids (*Montipora*

and *Acropora*) likely diverged from the Pocilloporids (*S. pistillata* and *P. damicornis*) during the Triassic period 220 mya (188 – 225, 95% CI) with *Acropora* and *Montipora* diverging during the early Cretaceous 136 mya (116 – 136, 95% CI).

Conclusion

In summary, this study presents the first fully annotated genome of the endangered Caribbean staghorn coral, *Acropora cervicornis*. Comparative genomics highlights distinctive genetic traits, including immune pathway contractions and signaling pathway expansions, suggesting further research into the species response to environmental stressors, particularly White Band Disease. Phylogenetic analysis places *A. cervicornis* within the broader *Acropora* genus, dating its divergence to around 41 million years ago, aligning previous morphological findings and supporting the hypothesis of divergence coincident with the closure of the western Tethys Sea. This annotated genome serves as a valuable resource for future research, facilitating conservation and restoration efforts for Caribbean coral reefs and deepening our understanding of speciation and adaptation within *Acropora*.

Data Availability Statement

Genome assembly and associated Nanopore and short-read DNA sequencing data can be accessed from NCBI BioProject: PRJNA948411 with the mitochondrial genome assembly accessible at NCBI GenBank: OQ772303. RNA sequencing data used for annotation can be accessed from NCBI BioProject: PRJNA949884. Genome assembly and annotation pipeline code available from GitHub: https://github.com/VollmerLab/Acerv_Genome.

Acknowledgements

We would like to thank the staff of the Coral Restoration Foundation for providing samples (permit: CRF-2021-001) and logistical support.

Conflict of Interest

Authors declare that they have no competing interests.

Funder Information

Grant funding was provided to SVV by a National Science Foundation Division of Ocean Sciences grant (NSF OCE-1924145).

Literature Cited

- Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman, 1990 Basic local alignment search tool. *Journal of Molecular Biology* 215: 403–410.
- Aronson, R. B., and W. F. Precht, 2001 White-band disease and the changing face of Caribbean coral reefs, pp. 25–38 in *The Ecology and Etiology of Newly Emerging Marine Diseases*, edited by J. W. Porter. Developments in Hydrobiology, Springer Netherlands, Dordrecht.
- Baumgarten, S., O. Simakov, L. Y. Esherrick, Y. J. Liew, E. M. Lehnert *et al.*, 2015 The genome of *Aiptasia*, a sea anemone model for coral symbiosis. *Proc Natl Acad Sci U S A* 112: 11893–11898.
- Beasley, T. M., and R. E. Schumacker, 1995 Multiple Regression Approach to Analyzing Contingency Tables: Post Hoc and Planned Comparison Procedures. *The Journal of Experimental Education* 64: 79–93.
- Behrensmeyer, A. K., and A. Turner, 2013 *Taxonomic occurrences of Suidae recorded in the Paleobiology Database. Fossilworks*.
- Benjamini, Y., and Y. Hochberg, 1995 Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)* 57: 289–300.
- Birney, E., M. Clamp, and R. Durbin, 2004 GeneWise and Genomewise. *Genome Res.* 14: 988–995.
- Camacho, C., G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos *et al.*, 2009 BLAST+: architecture and applications. *BMC Bioinformatics* 10: 421.
- Cantarel, B. L., I. Korf, S. M. C. Robb, G. Parra, E. Ross *et al.*, 2008 MAKER: An easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res.* 18: 188–196.
- Challis, R., 2017 rjchallis/assembly-stats 17.02.
- Chen, S., Y. Zhou, Y. Chen, and J. Gu, 2018 fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34: i884–i890.

- 1 Colin, L., C. Yesson, and C. E. I. Head, 2021 Complete mitochondrial genomes of three reef forming
2 *Acropora* corals (Acroporidae, Scleractinia) from Chagos Archipelago, Indian Ocean. Biodivers
3 Data J 9: e72762.
- 4 Coombe, L., J. X. Li, T. Lo, J. Wong, V. Nikolic *et al.*, 2021 LongStitch: high-quality genome assembly
5 correction and scaffolding using long reads. BMC Bioinformatics 22: 534.
- 6 Cuning, R., R. A. Bay, P. Gillette, A. C. Baker, and N. Traylor-Knowles, 2018 Comparative analysis of the
7 *Pocillopora damicornis* genome highlights role of immune system in coral evolution. Sci Rep 8:
8 16134.
- 9 De Bie, T., N. Cristianini, J. P. Demuth, and M. W. Hahn, 2006 CAFE: a computational tool for the study of
10 gene family evolution. Bioinformatics 22: 1269–1271.
- 11 De Coster, W., S. D’Hert, D. T. Schultz, M. Cruts, and C. Van Broeckhoven, 2018 NanoPack: visualizing and
12 processing long-read sequencing data. Bioinformatics 34: 2666–2669.
- 13 Emms, D. M., and S. Kelly, 2019 OrthoFinder: phylogenetic orthology inference for comparative
14 genomics. Genome Biol 20: 238.
- 15 Emms, D. M., and S. Kelly, 2018 STAG: Species Tree Inference from All Genes. 267914.
- 16 Emms, D. M., and S. Kelly, 2017 STRIDE: Species Tree Root Inference from Gene Duplication Events.
17 Molecular Biology and Evolution 34: 3267–3278.
- 18 Flynn, J. M., R. Hubley, C. Goubert, J. Rosen, A. G. Clark *et al.*, 2020 RepeatModeler2 for automated
19 genomic discovery of transposable element families. Proceedings of the National Academy of
20 Sciences 117: 9451–9457.
- 21 Fuller, Z. L., V. J. L. Mocellin, L. A. Morris, N. Cantin, J. Shepherd *et al.*, 2020 Population genetics of the
22 coral *Acropora millepora*: Toward genomic prediction of bleaching. Science 369: eaba4674.

- 1 Geib, S. M., B. Hall, T. Derego, F. T. Bremer, K. Cannoles *et al.*, 2018 Genome Annotation Generator: a
2 simple tool for generating and correcting WGS annotation tables for NCBI submission.
3 GigaScience 7: giy018.
- 4 Gertz, E. M., Y.-K. Yu, R. Agarwala, A. A. Schäffer, and S. F. Altschul, 2006 Composition-based statistics and
5 translated nucleotide searches: Improving the TBLASTN module of BLAST. BMC Biology 4: 41.
- 6 Grabherr, M. G., B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson *et al.*, 2011 Full-length transcriptome
7 assembly from RNA-Seq data without a reference genome. Nat Biotechnol 29: 644–652.
- 8 Guan, D., S. A. McCarthy, J. Wood, K. Howe, Y. Wang *et al.*, 2020 Identifying and removing haplotypic
9 duplication in primary genome assemblies. Bioinformatics 36: 2896–2898.
- 10 Guiguelmoni, N., A. Houtain, A. Derzelle, K. Van Doninck, and J.-F. Flot, 2021 Overcoming uncollapsed
11 haplotypes in long-read assemblies of non-model organisms. BMC Bioinformatics 22: 303.
- 12 Hahn, M. W., T. D. Bie, J. E. Stajich, C. Nguyen, and N. Cristianini, 2005 Estimating the tempo and mode of
13 gene family evolution from comparative genomic data. Genome Res. 15: 1153–1160.
- 14 Hamada, M., E. Shoguchi, C. Shinzato, T. Kawashima, D. J. Miller *et al.*, 2013 The Complex NOD-Like
15 Receptor Repertoire of the Coral *Acropora digitifera* Includes Novel Domain Combinations.
16 Molecular Biology and Evolution 30: 167–176.
- 17 Han, M. V., G. W. C. Thomas, J. Lugo-Martinez, and M. W. Hahn, 2013 Estimating Gene Gain and Loss
18 Rates in the Presence of Error in Genome Assembly and Annotation Using CAFE 3. Molecular
19 Biology and Evolution 30: 1987–1997.
- 20 Hart, A. J., S. Ginzburg, M. (Sam) Xu, C. R. Fisher, N. Rahmatpour *et al.*, 2020 EnTAP: Bringing faster and
21 smarter functional annotation to non-model eukaryotic transcriptomes. Molecular Ecology
22 Resources 20: 591–604.
- 23 Hemond, E. M., S. T. Kaluziak, and S. V. Vollmer, 2014 The genetics of colony form and function in
24 Caribbean *Acropora* corals. BMC Genomics 15: 1133.

- 1 Hilton, J. D., A. K. Brady, S. A. Spaho, and P. D. Vize, 2012 Photoreception and Signal Transduction in
2 Corals: Proteomic and Behavioral Evidence for Cytoplasmic Calcium as a Mediator of Light
3 Responsivity. *The Biological Bulletin* 223: 291–299.
- 4 Holt, C., and M. Yandell, 2011 MAKER2: an annotation pipeline and genome-database management tool
5 for second-generation genome projects. *BMC Bioinformatics* 12: 491.
- 6 Hughes, T. P., K. D. Anderson, S. R. Connolly, S. F. Heron, J. T. Kerry *et al.*, 2018 Spatial and temporal
7 patterns of mass bleaching of corals in the Anthropocene. *Science* 359: 80–83.
- 8 Jühling, F., J. Pütz, M. Bernt, A. Donath, M. Middendorf *et al.*, 2012 Improved systematic tRNA gene
9 annotation allows new insights into the evolution of mitochondrial tRNA structures and into the
10 mechanisms of mitochondrial genome rearrangements. *Nucleic Acids Research* 40: 2833–2845.
- 11 Kelly, S., and P. K. Maini, 2013 DendroBLAST: Approximate Phylogenetic Trees in the Absence of Multiple
12 Sequence Alignments. *PLOS ONE* 8: e58537.
- 13 Kolmogorov, M., J. Yuan, Y. Lin, and P. A. Pevzner, 2019 Assembly of long, error-prone reads using repeat
14 graphs. *Nat Biotechnol* 37: 540–546.
- 15 Korf, I., 2004 Gene finding in novel genomes. *BMC Bioinformatics* 5: 59.
- 16 Krzywinski, M., J. Schein, I. Birol, J. Connors, R. Gascoyne *et al.*, 2009 Circos: An information aesthetic for
17 comparative genomics. *Genome Res.* 19: 1639–1645.
- 18 Laetsch, D. R., and M. L. Blaxter, 2017 BlobTools: Interrogation of genome assemblies.
- 19 Li, H., and R. Durbin, 2009 Fast and accurate short read alignment with Burrows–Wheeler transform.
20 *Bioinformatics* 25: 1754–1760.
- 21 Libro, S., and S. V. Vollmer, 2016 Genetic Signature of Resistance to White Band Disease in the Caribbean
22 Staghorn Coral *Acropora cervicornis*. *PLOS ONE* 11: e0146636.

- 1 Liu, S.-Y. V., C.-L. C. Chan, H. J. Hsieh, S. Fontana, C. C. Wallace *et al.*, 2015 Massively parallel sequencing
2 (MPS) assays for sequencing mitochondrial genomes: the phylogenomic implications for
3 *Acropora* staghorn corals (Scleractinia; Acroporidae). *Mar Biol* 162: 1383–1392.
- 4 Lohse, M., O. Drechsel, S. Kahlau, and R. Bock, 2013 OrganellarGenomeDRAW—a suite of tools for
5 generating physical maps of plastid and mitochondrial genomes and visualizing expression data
6 sets. *Nucleic Acids Research* 41: W575–W581.
- 7 Lomsadze, A., V. Ter-Hovhannisyan, Y. O. Chernoff, and M. Borodovsky, 2005 Gene identification in novel
8 eukaryotic genomes by self-training algorithm. *Nucleic Acids Research* 33: 6494–6506.
- 9 López-Nandam, E. H., R. Albright, E. A. Hanson, E. A. Sheets, and S. R. Palumbi, 2023 Mutations in coral
10 soma and sperm imply lifelong stem cell renewal and cell lineage selection. *Proceedings of the*
11 *Royal Society B*.
- 12 Love, R. R., N. I. Weisenfeld, D. B. Jaffe, N. J. Besansky, and D. E. Neafsey, 2016 Evaluation of DISCOVAR
13 de novo using a mosquito sample for cost-effective short-read genome assembly. *BMC Genomics*
14 17: 187.
- 15 Marçais, G., and C. Kingsford, 2011 A fast, lock-free approach for efficient parallel counting of
16 occurrences of k-mers. *Bioinformatics* 27: 764–770.
- 17 Mendes, F. K., D. Vanderpool, B. Fulton, and M. W. Hahn, 2020 CAFE 5 models variation in evolutionary
18 rates among gene families. *Bioinformatics* 36: 5516–5518.
- 19 Meng, G., Y. Li, C. Yang, and S. Liu, 2019 MitoZ: a toolkit for animal mitochondrial genome assembly,
20 annotation and visualization. *Nucleic Acids Research* 47: e63.
- 21 Minh, B. Q., M. W. Hahn, and R. Lanfear, 2020 New Methods to Calculate Concordance Factors for
22 Phylogenomic Datasets. *Molecular Biology and Evolution* 37: 2727–2733.

- 1 National Marine Fisheries Service, 2006 Endangered and Threatened Species: Final Listing
- 2 Determinations for Elkhorn Coral and Staghorn Coral, 71(FR26852): 26852-26872. Silver Spring,
- 3 MD: National Marine Fisheries Service (NMFS).
- 4 Nawrocki, E. P., and S. R. Eddy, 2013 Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics*
- 5 29: 2933–2935.
- 6 Nurk, S., D. Meleshko, A. Korobeynikov, and P. A. Pevzner, 2017 metaSPAdes: a new versatile
- 7 metagenomic assembler. *Genome Res.* 27: 824–834.
- 8 van Oppen, M. J., J. Catmull, B. J. McDonald, N. R. Hislop, P. J. Hagerman *et al.*, 2002a The Mitochondrial
- 9 Genome of *Acropora tenuis* (Cnidaria; Scleractinia) Contains a Large Group I Intron and a
- 10 Candidate Control Region. *Journal of Molecular Evolution* 55: 1–13.
- 11 van Oppen, M. J., N. R. Hislop, P. J. Hagerman, and D. J. Miller, 1999 Gene content and organization in a
- 12 segment of the mitochondrial genome of the scleractinian coral *Acropora tenuis*: major
- 13 differences in gene order within the anthozoan subclass zoantharia. *Molecular Biology and*
- 14 *Evolution* 16: 1812–1815.
- 15 van Oppen, M. J. H., B. J. McDonald, B. Willis, and D. J. Miller, 2001 The Evolutionary History of the Coral
- 16 Genus *Acropora* (Scleractinia, Cnidaria) Based on a Mitochondrial and a Nuclear Marker:
- 17 Reticulation, Incomplete Lineage Sorting, or Morphological Convergence? *Molecular Biology and*
- 18 *Evolution* 18: 1315–1329.
- 19 van Oppen, M. J. H., B. L. Willis, T. Van Rheede, and D. J. Miller, 2002b Spawning times, reproductive
- 20 compatibilities and genetic structuring in the *Acropora aspera* group: evidence for natural
- 21 hybridization and semi-permeable species boundaries in corals. *Molecular Ecology* 11: 1363–
- 22 1376.

- 1 Parkinson, J. E., E. Bartels, M. K. Devlin-Durante, C. Lustic, K. Nedimyer *et al.*, 2018 Extensive
2 transcriptional variation poses a challenge to thermal stress biomarker development for
3 endangered corals. *Molecular Ecology* 27: 1103–1119.
- 4 Putnam, N. H., M. Srivastava, U. Hellsten, B. Dirks, J. Chapman *et al.*, 2007 Sea Anemone Genome
5 Reveals Ancestral Eumetazoan Gene Repertoire and Genomic Organization. *Science* 317: 86–94.
- 6 R Core Team, 2022 R: A Language and Environment for Statistical Computing.
- 7 Reyes-Bermudez, A., M. K. DeSalvo, C. R. Voolstra, S. Sunagawa, A. M. Szmant *et al.*, 2009 Gene
8 expression microarray analysis encompassing metamorphosis and the onset of calcification in
9 the scleractinian coral *Montastraea faveolata*. *Marine Genomics* 2: 149–159.
- 10 Rodriguez-Casariello, J. A., M. C. Ladd, A. A. Shantz, C. Lopes, M. S. Cheema *et al.*, 2018 Coral epigenetic
11 responses to nutrient stress: Histone H2A.X phosphorylation dynamics and DNA methylation in
12 the staghorn coral *Acropora cervicornis*. *Ecology and Evolution* 8: 12193–12207.
- 13 Rosenberg, Y., T. Doniger, S. Harii, F. Sinniger, and O. Levy, 2017 Canonical and cellular pathways timing
14 gamete release in *Acropora digitifera*, Okinawa, Japan. *Molecular Ecology* 26: 2698–2710.
- 15 Russell, T. J., and G. M. Watson, 1995 Evidence for intracellular stores of calcium ions involved in
16 regulating nematocyst discharge. *Journal of Experimental Zoology* 273: 175–185.
- 17 Salazar, O. R., P. N. Arun, G. Cui, L. K. Bay, M. J. H. van Oppen *et al.*, 2022 The coral *Acropora loripes*
18 genome reveals an alternative pathway for cysteine biosynthesis in animals. *Science Advances* 8:
19 eabq0304.
- 20 Shinzato, C., K. Khalturin, J. Inoue, Y. Zayasu, M. Kanda *et al.*, 2021 Eighteen Coral Genomes Reveal the
21 Evolutionary Origin of *Acropora* Strategies to Accommodate Environmental Changes. *Mol Biol*
22 *Evol* 38: 16–30.

- 1 Shinzato, C., S. Mungpakdee, N. Satoh, and E. Shoguchi, 2014 A genomic approach to coral-dinoflagellate
2 symbiosis: studies of *Acropora digitifera* and *Symbiodinium minutum*. *Frontiers in Microbiology*
3 5:.
- 4 Shinzato, C., E. Shoguchi, T. Kawashima, M. Hamada, K. Hisata *et al.*, 2011 Using the *Acropora digitifera*
5 genome to understand coral responses to environmental change. *Nature* 476: 320–323.
- 6 Smit, A. F. A., R. Hubley, and P. Green, 2015 RepeatMasker Open-4.0. 2013–2015.
- 7 Stanke, M., O. Schöffmann, B. Morgenstern, and S. Waack, 2006 Gene prediction in eukaryotes with a
8 generalized hidden Markov model that uses hints from external sources. *BMC Bioinformatics* 7:
9 62.
- 10 Surm, J. M., Z. K. Stewart, A. Papanicolaou, A. Pavasovic, and P. J. Prentis, 2019 The draft genome of
11 *Actinia tenebrosa* reveals insights into toxin evolution. *Ecol Evol* 9: 11314–11328.
- 12 Tate, R., B. Hall, T. DeRego, and S. Geib, 2014 Annie: the ANnotation Information Extractor (Version 1.0).
13 2014.
- 14 Tenenbaum, D., and J. Volkening, 2023 KEGGREST: Client-side REST access to the Kyoto Encyclopedia of
15 Genes and Genomes (KEGG).
- 16 The UniProt Consortium, 2023 UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids*
17 *Research* 51: D523–D531.
- 18 Tian, P., and W. Niu, 2017 The complete mitochondrial genome of the *Acropora pruinosa*. *Mitochondrial*
19 *DNA B Resour* 2: 652–653.
- 20 To, T.-H., M. Jung, S. Lycett, and O. Gascuel, 2016 Fast Dating Using Least-Squares Criteria and
21 Algorithms. *Systematic Biology* 65: 82–97.
- 22 Vaser, R., I. Sović, N. Nagarajan, and M. Šikić, 2017 Fast and accurate de novo genome assembly from
23 long uncorrected reads. *Genome Res.* 27: 737–746.

- 1 Vollmer, S. V., and S. R. Palumbi, 2002 Hybridization and the evolution of reef coral diversity. *Science* 296:
- 2 2023.
- 3 Voolstra, C. R., Y. Li, Y. J. Liew, S. Baumgarten, D. Zoccola *et al.*, 2017 Comparative analysis of the
- 4 genomes of *Stylophora pistillata* and *Acropora digitifera* provides evidence for extensive
- 5 differences between species of corals. *Sci Rep* 7: 17583.
- 6 Walker, B. J., T. Abeel, T. Shea, M. Priest, A. Abouelliel *et al.*, 2014 Pilon: An Integrated Tool for
- 7 Comprehensive Microbial Variant Detection and Genome Assembly Improvement. *PLOS ONE* 9:
- 8 e112963.
- 9 Wallace, C. C., 1999 *Staghorn Corals of the World: A Revision of the Genus Acropora*. Csiro Publishing.
- 10 Wallace, C. C., and R. W. Portell, 2022 Earliest western Atlantic staghorn corals (*Acropora*) from the lower
- 11 Oligocene Suwannee Limestone of Florida, USA, and their significance for modern coral
- 12 distribution. *Journal of Paleontology* 96: 1390–1399.
- 13 Wheeler, T. J., and S. R. Eddy, 2013 nhmmer: DNA homology search with profile HMMs. *Bioinformatics*
- 14 29: 2487–2489.
- 15 Wingett, S. W., and S. Andrews, 2018 FastQ Screen: A tool for multi-genome mapping and quality
- 16 control. *F1000Res* 7: 1338.
- 17 Wu, B., M. Li, X. Liao, J. Luo, F.-X. Wu *et al.*, 2020 MEC: Misassembly Error Correction in Contigs based on
- 18 Distribution of Paired-End Reads and Statistics of GC-contents. *IEEE/ACM Transactions on*
- 19 *Computational Biology and Bioinformatics* 17: 847–857.
- 20 Wu, T. D., J. Reeder, M. Lawrence, G. Becker, and M. J. Brauer, 2016 GMAP and GSNAP for Genomic
- 21 Sequence Alignment: Enhancements to Speed, Accuracy, and Functionality, pp. 283–334 in
- 22 *Statistical Genomics: Methods and Protocols*, edited by E. Mathé and S. Davis. *Methods in*
- 23 *Molecular Biology*, Springer, New York, NY.

- 1 Xie, Y., G. Wu, J. Tang, R. Luo, J. Patterson *et al.*, 2014 SOAPdenovo-Trans: de novo transcriptome
2 assembly with short RNA-Seq reads. *Bioinformatics* 30: 1660–1666.
- 3 Xu, M., L. Guo, S. Gu, O. Wang, R. Zhang *et al.*, 2020 TGS-GapCloser: A fast and accurate gap closer for
4 large genomes with low coverage of error-prone long reads. *GigaScience* 9: giaa094.
- 5 Ying, H., D. C. Hayward, I. Cooke, W. Wang, A. Moya *et al.*, 2019 The Whole-Genome Sequence of the
6 Coral *Acropora millepora*. *Genome Biology and Evolution* 11: 1374–1379.
- 7 Zdobnov, E. M., and R. Apweiler, 2001 InterProScan--an integration platform for the signature-
8 recognition methods in InterPro. *Bioinformatics* 17: 847–848.
- 9 Zhang, Y., X. Yu, Z. Zhou, and B. Huang, 2016 The complete mitochondrial genome of *Acropora aculeus*
10 (cnidaria, scleractinia, acroporidae). *Mitochondrial DNA A DNA Mapp Seq Anal* 27: 4276–4277.
- 11
- 12

1 Table 1: National Center for Biotechnology Information accession numbers and citations for
 2 Coral genomes with structural gene annotations used to build phylogeny and for comparative
 3 genomic analysis.

Species	Genome Size (Mb)	Number of Genes	Accession Number	Citation
<i>Acropora acuminata</i>	394.7	26,151	GCA_014633975	(Shinzato <i>et al.</i> 2021)
<i>Acropora awi</i>	428.8	26,801	GCA_014634005	(Shinzato <i>et al.</i> 2021)
<i>Acropora cervicornis</i>	308	28,059		(This Study)
<i>Acropora cytherea</i>	426.3	27,327	GCA_014634045	(Shinzato <i>et al.</i> 2021)
<i>Acropora digitifera</i>	415.8	25,278	GCA_014634065	(Shinzato <i>et al.</i> 2021)
<i>Acropora echinata</i>	401.5	26,170	GCA_014634105	(Shinzato <i>et al.</i> 2021)
<i>Acropora florida</i>	442.8	27,573	GCA_014634605	(Shinzato <i>et al.</i> 2021)
<i>Acropora gemmifera</i>	401	26,269	GCA_014634125	(Shinzato <i>et al.</i> 2021)
<i>Acropora hyacinthus</i>	447.2	27,215	GCA_014634145	(Shinzato <i>et al.</i> 2021)
<i>Acropora intermedia</i>	416.9	26,982	GCA_014634585	(Shinzato <i>et al.</i> 2021)
<i>Acropora microphthalma</i>	383.9	26,384	GCA_014634165	(Shinzato <i>et al.</i> 2021)
<i>Acropora millepora</i>	475.4	41,860	GCA_013753865	(Fuller <i>et al.</i> 2020)
<i>Acropora muricata</i>	420.7	27,409	GCA_014634545	(Shinzato <i>et al.</i> 2021)
<i>Acropora nasuta</i>	416.4	27,379	GCA_014634205	(Shinzato <i>et al.</i> 2021)
<i>Acropora selago</i>	392.9	27,036	GCA_014634525	(Shinzato <i>et al.</i> 2021)
<i>Acropora tenuis</i>	403.1	27,236	GCA_014633955	(Shinzato <i>et al.</i> 2021)
<i>Acropora yongei</i>	438	27,452	GCA_014634225	(Shinzato <i>et al.</i> 2021)
<i>Montipora cactus</i>	652.7	29,158	GCA_014634245	(Shinzato <i>et al.</i> 2021)
<i>Montipora efflorescens</i>	643.3	29,424	GCA_014634505	(Shinzato <i>et al.</i> 2021)
<i>Nematostella vectensis</i>	356.6	34,311	GCF_000209225	(Putnam <i>et al.</i> 2007)
<i>Pocillopora damicornis</i>	234.3	25,183	GCF_003704095	(Cunning <i>et al.</i> 2018)
<i>Stylophora pistillata</i>	397.6	33,252	GCF_002571385	(Voolstra <i>et al.</i> 2017)
<i>Actinia tenebrosa</i>	238.2	27,037	GCF_009602425	(Surm <i>et al.</i> 2019)
<i>Exaiptasia diaphana</i>	256.1	27,753	GCF_001417965	(Baumgarten <i>et al.</i> 2015)

4

5

1 Table 2: *A. cervicornis* assembly statistics

Cumulative scaffold length (bp)	307.4
Number of scaffolds	398
Number of contigs	415
G + C content (%)	38.95
Number of Ns / 100 kbp	37.23
Largest Scaffold (Mb)	8.337
Scaffold N50 (Mb)	2.8
Scaffold L50	35
Largest contig (Mb)	8.337
Contig N50 (Mb)	2.7
Contig L50 (Mb)	36
BUSCO complete single-copy	878
BUSCO complete multi-copy	4
BUSCO fragmented	34
BUSCO missing	38

2

3

1 Table 3: *A. cervicornis* annotation statistics

Number of Genes	28,059
Cumulative length of CDSs (bp)	36,143,406
Median gene length (bp)	4,072
Median CDS length (bp)	119
Median exon length (bp)	124
Median intron length (bp)	569
Number of intronless genes	4,591
Median number of exons per gene	4
Median number of exons per multi-exon gene	4
BUSCO complete single-copy	752
BUSCO complete multi-copy	27
BUSCO fragmented	95
BUSCO missing	80

2

3

Figure Legends

Figure 1: Genome snail plot showing genome contiguity and completeness statistics. For an interactive version of the figure see: <https://jdselwyn.github.io/assembly-stats/>. Created using (Challis 2017).

Figure 2: Mitochondrial genome plot showing the relative locations and orientations of genes on the mitochondria with gene groups color coded by type. Plot made using (Lohse *et al.* 2013).

Figure 3: Species phylogeny based on identified orthogroups. Node color indicates the proportion of single-locus gene trees supporting each bipartition (i.e., gene concordance factor, a more stringent metric of node support than bootstrap support, Emms and Kelly 2018; Minh *et al.* 2020). Red bars at each node indicate the 95% confidence interval of the estimated time of divergence marked in geologic periods on the x-axis. Green/red node numbers indicate the number of KEGG orthologs which significantly expanded/contracted at each divergence point. Clade groupings from Shizato *et al.* (2021).

Scaffold statistics

- Log₁₀ scaffold count (total 398)
- Scaffold length (total 307 MB)
- Longest scaffold (8.3 MB)
- N50 length (2.8 MB)
- N90 length (618.2 kB)

BUSCO (n = 954)

- Comp. (92.0%)
- Dup. (0.4%)
- Frag. (3.6%)

Scale

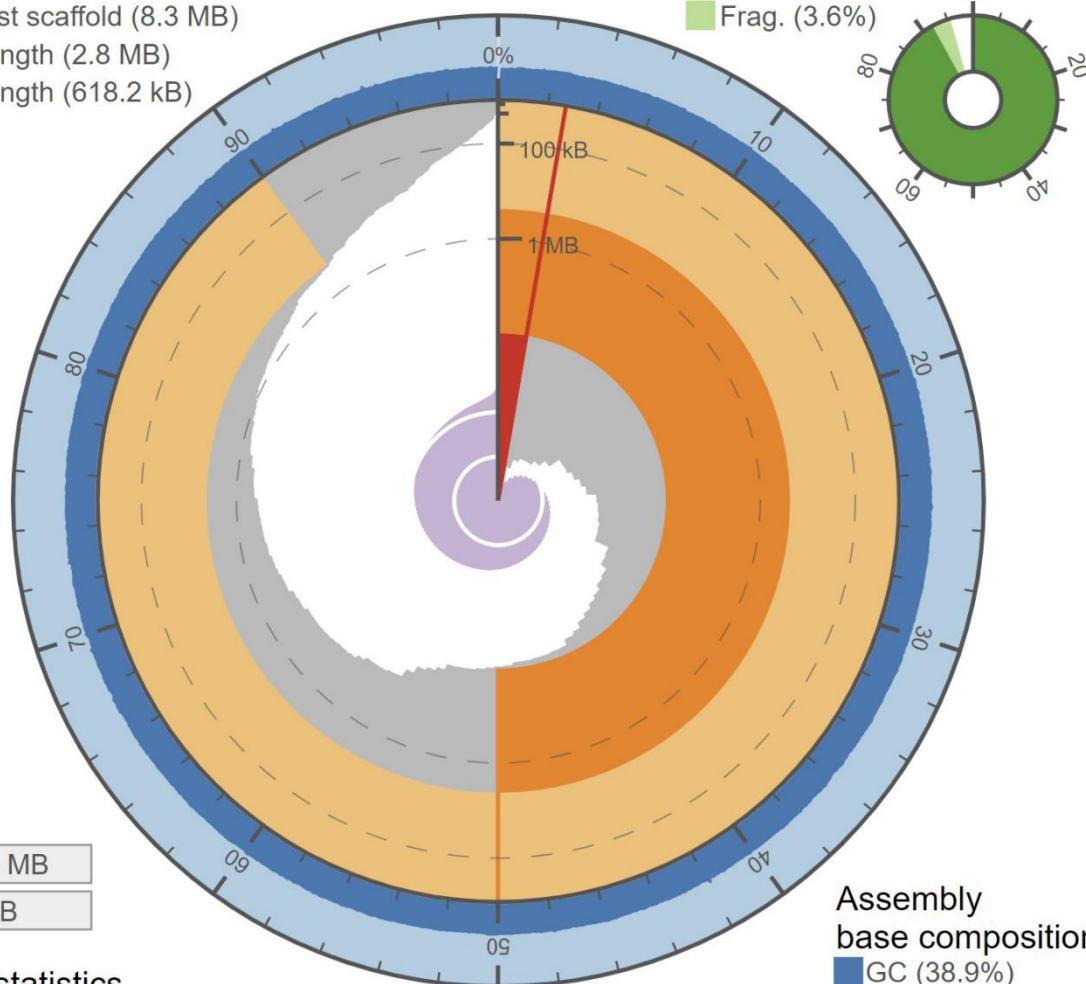
- 307.4 MB
- 8.3 MB

Contig statistics

- Log₁₀ contig count (total 415)
- Contig length (total 307 MB)

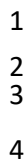
Assembly base composition

- GC (38.9%)
- AT (61.0%)
- N (0.0%)



S

Figure 1
423x416 mm (x DPI)



366x367 mm (x DPI)

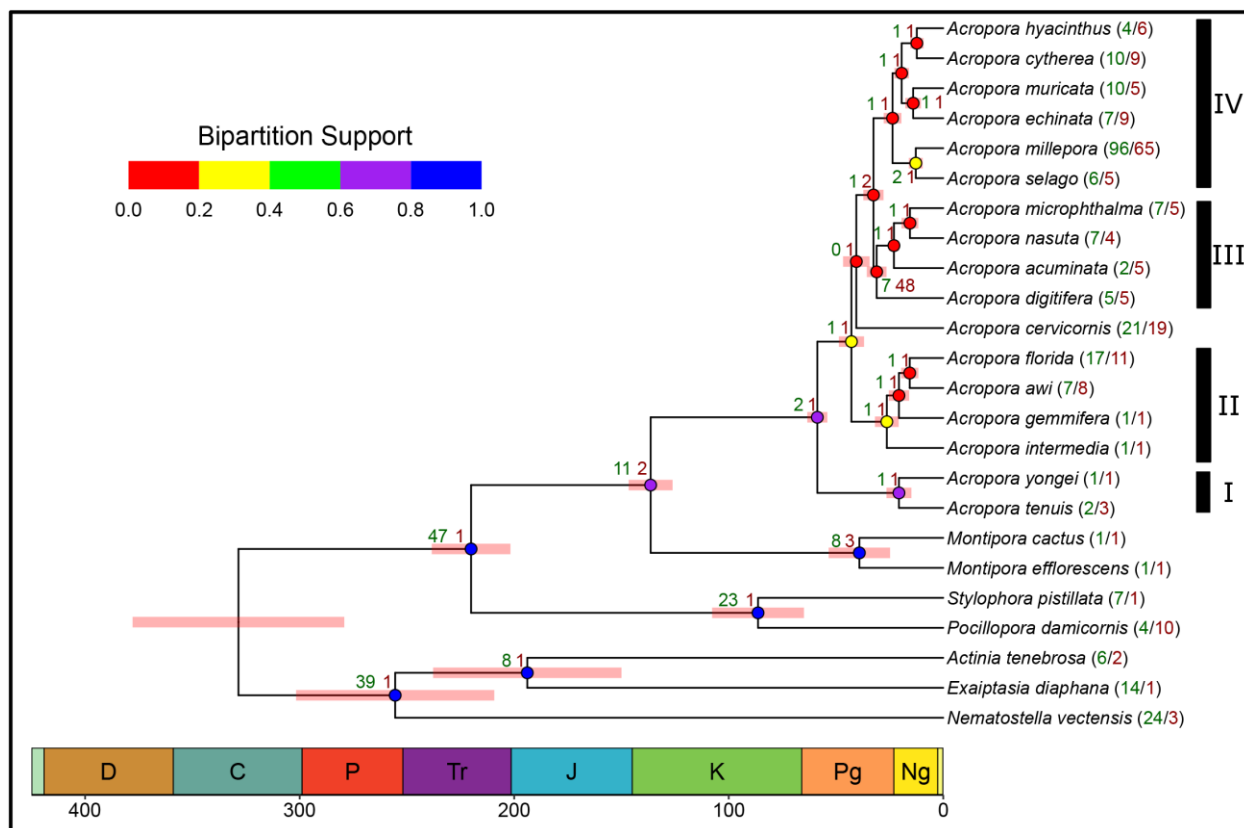


Figure 3
194x128 mm (x DPI)