The first *de novo* HiFi genome assemblies for three clownfish-hosting sea anemone species (Anthozoa: Actiniaria)

Aurélien De Jode*[1,2], Benjamin M. Titus[1,2]

[1] Department of Biological Sciences, University of Alabama, Tuscaloosa, AL, USA.

[2] Dauphin Island Sea Lab, Dauphin Island, AL, USA.

*Author for Correspondence: Aurélien De Jode, Department of Biological Sciences, University of Alabama, Tuscaloosa, AL, USA, aurelien.dejode@laposte.net

Abstract

The symbiosis between clownfish and giant tropical sea anemones (Order Actiniaria) is one of the most iconic on the planet. Distributed on tropical reefs, 28 species of clownfishes form obligate mutualistic relationships with 10 nominal species of venomous sea anemones. Our understanding of the symbiosis is limited by the fact that most research has been focused on the clownfishes. Chromosome scale reference genomes are available for all clownfish species, yet there are no published reference genomes for the host sea anemones. Recent studies have shown that the clownfish-hosting sea anemones belong to three distinct clades of sea anemones that have evolved symbiosis with clownfishes independently. Here we present the first high quality long read assemblies for three species of clownfish hosting sea anemones belonging to each of these clades: *Entacmaea quadricolor*, *Stichodactyla haddoni*, *Radianthus doreensis*. PacBio HiFi sequencing yielded 1,597,562, 3,101,773, and 1,918,148 million reads for *E. quadricolor*, *S. haddoni*, and *R. doreensis*, respectively. All three assemblies were highly contiguous and complete with N50 values above 4Mb and BUSCO completeness above 95% on the Metazoa dataset. Genome structural annotation with BRAKER3 predicted 20,454, 18,948 and 17,056 protein coding genes in *E. quadricolor*, *S. haddoni* and *R. doreeensis* genome, respectively. These new resources will form the basis of comparative genomic analyses that will allow us to deepen our understanding of this mutualism from the host perspective.

Keywords: Cnidaria, Actinioidea, Symbiosis

Significance

Chromosome-scale genomes are available for all 28 clownfish species yet there are no high-quality reference genomes published for the clownfish-hosting sea anemones. The lack of genomic resources impedes our ability to understand evolution of this iconic symbiosis from the host perspective. The clownfish-hosting sea anemones belong to three clades of sea anemones that have evolved mutualism with clownfish independently. Here we assembled the first high-quality long-read genomes for three species of host sea anemones each belonging to a different host clade: *Entacmaea quadricolor*, *Stichodactyla haddoni*, *Radianthus doreensis.* These resources will enable in depth comparative genomics of clownfish-hosting sea anemones providing a critical perspective for understanding how the symbiosis has evolved. Finally, these reference genomes present a significant increase in the number of high-quality long-read genome assemblies for sea anemones (11 currently published) and double the number of high-quality reference genomes for the sea anemone superfamily Actinoidea.

Introduction

Among all mutualistic symbioses none are more recognizable than the association between clownfishes (also called anemonefishes: Subfamily Amphiprioninae) and their giant host sea anemones (Anthozoa: Actiniaria). Distributed broadly on tropical coral reefs throughout the Indian and Pacific Oceans, 28 species of clownfishes have rapidly radiated to form obligate mutualistic relationships with 10 nominal species of venomous sea anemones (Litsios et al., 2012; Gaboriou et al., 2024; Titus et al. 2024). Famously, clownfishes have evolved to live unharmed in their toxic and venomous hosts benefiting from their association with sea anemones by receiving shelter against predators (Fautin 1991), protection for their externally brooded eggs (Fautin & Allen 1992), removal of external parasites (Lubbock 1981) and potentially even nourishment (Verde et al., 2015). In turn, sea anemones are granted reciprocal protection against predation (Dunn 1981, Godwin & Fautin, 1992, Holbrook & Schmitt, 2005), novel sources of nitrogen and carbon from fish excrement (Roopin et al. 2008, Roopin & Chadwick, 2009, Cleveland et al., 2011), and increased oxygenation and gas transfer as clownfishes move through their tentacles (Szczebak et al. 2013).

The symbiosis has attracted a great deal of popular and scientific attention. The small size of clownfishes and their well-defined hierarchical social groups, reproductive biology, and amenability to aquaculture have made them tractable systems for understanding fundamental biological processes. They are now considered an emerging model organism for a wide range of research (Roux et al., 2020; Laudet & Ravasi, 2022). Chromosome-level reference genomes are now available for all 28 clownfish species (Marcionetti et al., 2019; Marcionetti & Salamin, 2023; Gaboriau et al., 2024) and comparative analyses are revealing the first candidate genes likely to be associated with the ability of clownfishes to remain

protected from toxic sea anemone hosts (Marcionetti et al., 2019; Marcionetti & Salamin, 2023).

Our evolutionary understanding of the symbiosis, however, is limited by the fact that most research and genomic resources have been focused on clownfishes. No reference genomes exist for any of the 10 clownfish-hosting sea anemones. This perspective is expected to be especially critical as recent work has demonstrated that the host sea anemones are driving phenotypic convergence in clownfishes (Gaboriau et al., 2024), and that host sea anemones have diversification times that are broadly concomitant with the clownfish radiation (De Jode et al. 2024). These findings highlight the importance of the host sea anemones for a holistic understanding of the entire symbiosis. High-quality reference genomes for the host sea anemones thus hold the potential to identify genomic regions associated with establishing and maintaining mutualism with clownfishes, and other genomic consequences of a symbiotic lifestyle.

Interestingly, unlike the 28 described species of clownfishes which form a monophyletic clade, the 10 clownfish hosting sea anemones belong to three distinct lineages within the sea anemone superfamily Actinioidea that have evolved symbiosis with clownfishes (Titus et al. 2019; 2024; De Jode et al. 2024). These have been coined *Entacmaea*, Stichodactylina, and Heteractina. The clade *Entacmaea* contains only the bubble-tip sea anemone, *Entacmaea quadricolor* (Fig. 1A), which appears to be a species complex (Titus et al. 2019; De Jode et al. 2024). Clade Stichodactylina contains five clownfish hosting sea anemones species: *Cryptodendrum adhaesivum*, *Radianthus magnifica*, *Stichodactyla gigantea*, *S. haddoni* (Fig. 1B), and *S. mertensii*, as well as other non-host species (e.g. De Jode et al. 2024). Clade Heteractina includes *Heteractis aurora*, *Radianthus crispa*, *R. doreensis* (Fig. 1C), and *R. malu* (Titus et al. 2019; De Jode et al. 2024). The three

independent clades of host sea anemones provide a natural starting point for sequencing and assembling high-quality reference genomes.

Here we present the first high quality long reads assemblies for three species of clownfish hosting sea anemones: *Entacmaea quadricolor, Stichodactyla haddoni*, and *Radianthus doreensis* (Fig. 1). These genomes will form the basis for in depth comparative genomics between clownfish hosting and non-hosting sea anemones and deepen our understanding of this iconic symbiosis.

Results and Discussion

A total of 1,597,562 (~ 20 Gbp), 3,101,773 (~ 39 Gbp) and 1,918,148 (~26 Gbp) HiFi reads were sequenced for *E. quadricolor*, *S. haddoni* and *R. doreensis*, respectively. For each species, mitochondrial genomes were assembled, circularized, and annotated. The total lengths of the newly assembled mitochondrial genomes were 20,761 bp, 18,921 bp, and 19,768 bp for *E. quadricolor, S. haddoni and R. doreensis* (Fig. 1), respectively. These are similar to the size of the mitogenomes used as references to identify mitochondrial reads. A characteristic feature of Hexacorallian mitogenomes is the presence of a self-splicing intron in the NAD5 gene (Feng et al., 2023). Interestingly, MitoFinder (the default annotation tool of MitoHiFi) did not retrieve that feature in our mitogenomes, but we were able identify and retrieve the NAD5 intron using MITOS in all species. Feng et al. (2023) identified a conserved mitochondrial gene order in the Order Actiniaria. Using the annotation conducted in GeSeq we confirmed that our three mitogenomes shared that same order (Act1GO sensu Feng et al. 2023).

Nuclear genome size estimates from GenomeScope were comparable between *S. haddoni* (339 Mbp) and *R. doreensis* (330 Mbp). The assembled genomes of *S. haddoni* and *R. doreensis* spanned 447 Mb and 418 Mb, respectively (supplementary table S1), both

exceeding GenomeScope estimates by approximately 100 Mb. *Entacmaea quadricolor* had the largest estimated genome size at 578 Mbp, which was close to our final total assembly length of 599 Mb (supplementary table S1). Our genome size estimate for *E. quadricolor* was smaller than the 863 Mbp estimated using flow cytometry (Adachi et al., 2017), but higher than the 428 Mb assembly available on GenBank (GCA_024752375.1). The discrepancy between our PacBio assembly and the publicly available assembly is likely due to the latter genome being assembled using Illumina short reads.

Heterozygosity rates were comparable between *R. doreensis* (het: 0.767%) and *S. haddoni* (het: 0.855%) but were higher in *E. quadricolor* (het: 1.49 %). Merqury k-mer spectra plots revealed the presence of haplotypic duplications in each primary assembly (supplementary fig. 1). For each species, the assembly we obtained after using the purge_dups algorithm with manual cutoffs and contigs identified as repeats added back into the assembly, showed the best compromise between under and over purging and was kept for further analysis (supplementary fig. 1). A total of 17, 8 and 8 contigs were identified as contaminants and removed from the *E. quadricolor*, *S. haddoni* and *M. doreensis* assemblies, respectively (supplementary fig. 2). One mitochondrial contig was also identified and removed from each assembly.

Final genome assemblies were highly contiguous and comprised of 558, 648 and 230 contigs for *E. quadricolor*, *S. haddoni* and *R. doreensis*, respectively. N50 values were roughly 4 Mb (*E. quadricolor*), 5.3 Mb (*S. haddoni*), and 8 Mb (*R. doreensis*) (supplementary table 1). BUSCO completeness scores for the Metazoa dataset were 95.6 % for *E. quadricolor* and *S. haddoni* and 95.4 % for *R. doreensis* (supplementary table 1). A total of 20,454, 18,948 and 17,056 protein coding genes were predicted by the BRAKER3 annotation pipeline for *E. quadricolor*, *S. haddoni* and *R. doreensis* respectively (Fig. 2). For each species the

total number of predicted proteins were comparable to the number of predicted proteins for other members of Order Actiniaria (Fig. 2).

The OMArk completeness percentages, based on Hierarchical Orthologous Groups (HOGs) of genes in Eumetazoa, were 92.6%, 92.35 % and 85.53 % for *E. quadricolor*, *S. haddoni* and *R. doreensis*, respectively (Fig. 2, supplementary text). The OMArk consistency assessment was very similar among our three species and other members of Order Actiniaria (Fig. 2, supplementary text). OMArk consistency (the gene family is known to exist in the Eumetazoa lineage) scores 70.16 %, 72.27 % and 71. 56 % for *E. quadricolor*, *S. haddoni* and *R. doreensis*, respectively. No protein sequence was identified as a contaminant in any of our species. The lower amount of protein coding genes predicted in *S. haddoni* and *R. doreensis* and the lower level of completeness in *R. doreensis* are likely due to the lower amount of RNA seq data available for those species (supplementary text). The ratio of mono- to multi-exonic transcripts were 0.34, 0.28 and 0.34 for *E. quadricolor*, *S. haddoni* and *R. doreensis* respectively.

Conclusion

Prior to this study, no genomes had been published for the clownfish-hosting sea anemones, although chromosome level genomes are available for all 28 species of clownfishes (Gaboriau et al. 2024). The three high-quality genomes we've assembled here using PacBio HiFi reads belong to species from the three clades that have independently evolved mutualism with clownfishes (Titus et al., 2019; De Jode et al., 2024). These will thus become critical new resources for comparative genomics and our understanding of clownfish-hosting sea anemones diversification and of the evolution of the iconic mutualism. Finally, only 11 long-read genomes have been assembled and published for Order Actiniaria. Our genomes

represent a significant increase in the number of genomic resources available for understanding sea anemone biology and evolution broadly.

Material and Methods

Live sea anemones were purchased via the ornamental aquarium trade through Quality Marine Inc. to ensure the geographic origin of each animal. We acquired three species: the bubble tip sea anemone *Entacmaea quadricolor* (Fiji), Haddon's carpet sea anemone *Stichodactyla haddoni* (Vietnam), and the long-tentacled sea anemone *Radianthus doreensis* (Vietnam). Fresh or flash frozen pedal disc tissue was thinly sliced from each anemone to avoid endosymbiotic dinoflagellate contamination and used for DNA extractions. High Molecular Weight (HMW) DNA was extracted using PacBio Nanobind® tissue kits following manufacturer instructions with the exception that the quantity of proteinase K was increased to 40µL. DNA samples were purified with Monarch Nucleic Acid Purification Kit (New England BioLabs, Ipswich, Massachusetts) prior to shearing on a Megaruptor 2 (Diagenode, Denville, NJ), at 20kb setting.

Libraries were constructed and sequenced at Maryland Genomics at the Institute for Genome Sciences, University of Maryland School of Medicine. Samples were prepared using SMRTbell Prep Kit 3.0 (Pacific Biosciences, Menlo Park, CA) with barcoded adaptors, and subsequently size-selected on BluePippin instrument (Sage Science, Beverly, MA) to remove DNA fragments less than 10kb in size, followed by an equimolar pooling. The pooled libraries were bound to Revio polymerase, then sequenced with a Revio Sequencing Plate and flow cell on a PacBio Revio platform (Pacific Biosciences, Menlo Park, CA). Mitochondrial genomes were assembled from raw HiFi reads using the MitoHiFi v2. (Uliano-Silva et al., 2023) pipeline with default parameters except for the –max-read-len option that was set to 2. Mitochondrial reads were identified for *E. quadricolor*, *S. haddoni,* and *R. doreensis* using the

mitogenomes of *E. quadricolor* (Genbank NC_049066.1); *Heteractis aurora* (Genbank NC_047219.1) and *S. haddoni* (MW760873.1, Johansen et al., 2021), respectively. Mitochondrial genome annotation was performed using both MitoFinder (Allio et al., 2020) and MITOS (Bernt et al., 2013). A final annotation was conducted using GeSeq (Tillich et al., 2017) (see explanations in Results & Discussion section). Mitogenomes were visualized using Geneious version 2024.0 (https://www.geneious.com) created by Biomatters.

Genome sizes and heterozygosities were estimated using GenomeScope 2.0 (Ranallo-Benavidez et al., 2020) based of k-mer histogram frequencies generated by Jellyfish v2.3.0 (Marçais & Kingsford, 2011) for a k-mer size of 21bp (Vulture et al. 2017). Nuclear genomes were assembled using Hifiasm v0.19.8-r603 (Cheng et al., 2021) and the first set of partially phased contigs was used for the following analysis.

Haplotypic duplications, were removed from the assemblies using the purge_dups v1.2.5 (Guan et al., 2020) algorithm with the following customs cutoffs: (3, 18, 18, 19, 19, 180), (20, 54, 54, 55, 55, 240) and (7, 39, 39, 40, 40, 200) for *E. quadricolor*, *S. haddoni* and *M. doreensis,* respectively. Contigs classified as repeats by purge_dups were added back to the assemblies. The Blobtoolkit (Laetsch & Blaxter, 2017, Challis et al. 2020) v3.4.3 pipeline was then used to assess the presence of contaminant contigs in each assembly. Mitochondrial contigs were removed from decontaminated nuclear contigs using MitoHiFi v2. (Uliano-Silva et al., 2023) with the mitochondrial genome assembled from the reads as a reference.

Genomes were annotated using the BRAKER3 pipeline v.3.0.8 (Gabriel et al. 2024) specifying a BUSCO lineage (--busco_lineage=metazoa_odb10). The pipeline uses several sub-pipelines and third-party tools (Stanke et al., 2006, Gotoh, 2008, Stanke et al., 2008, Iwata & Gotoh, 2012, Simão et al., 2015, Buchfink et al., 2015, Hoff et al., 2016, Hoff et al., 2019, Kovaka et al., 2019, Pertea & Pertea, 2020, Brůna et al., 2021, Gabriel et al., 2021, Huang & Li, 2023, Li, 2023, Brůna et al., 2024) to predict highly reliable genes in reference

genome using protein and RNA-seq data. RNA-seq data (see SI for list of accession numbers) were downloaded from Genbank using the fastq-dump function of the SRA-toolkit v.3.0.7 (SRA Toolkit Development Team) and reads were mapped to the genome using HISAT2 v.2.2.1 (Kim et al. 2019) with the --dta option. The Metazoa partition of the OrthoDB v11 protein database was used as evidence for BRAKER 3 (Kuznetsov et al., 2023). To compute the ratio of mono- to multi-exon genes we used the analyze_exons.py script from the GALBA github repository (https://github.com/Gaius-Augustus/GALBA). Annotation quality was assessed with use OMArk v.0.3.0 (Nevers et al., 2022) on the OMArk public webserver (https://omark.omabrowser.org/home/). Genome quality was assessed for the final assembly and at several intermediate steps during the assembly process. Contiguity and completeness metrics were computed using gfastats v1.3.6 (Formenti et al., 2022) and BUSCO v5.6.1 (Manni et al. 2021). Completeness, quality and levels of haplotypic duplication were assessed using Merqury (Rhie et al. 2020) based on HiFi reads.

Acknowledgments

We thank members of the Titus Lab for help with sample processing and Rebecca Varney for advice on high molecular weight DNA extractions. We thank Maryland Genomics at the Institute for Genome Sciences, University of Maryland School of Medicine for library preparation and sequencing. We also thank Giulio Formenti and Jonathan M. Wood for helpful discussion regarding the purging of haplotypic duplications. Funding for this work was provided by the University of Alabama and National Science Foundation Grant DEB-1934274 to BMT.

Data Availability

The HiFi PacBio sequencing reads, are deposited in NCBI under the BioProject accession numbers PRJNA1076568. The final genome assembly and genome annotation will be deposited under the same BioProject upon publication.

Literature Cited

Adachi, K., Miyake, H., Kuramochi, T., Mizusawa, K., & Okumura, S. 2017. Genome size distribution in phylum Cnidaria. *Fisheries Science*, *83*(1), 107–112. https://doi.org/10.1007/s12562-016-1050-4

Allio, R., Schomaker-Bastos, A., Romiguier, J., Prosdocimi, F., Nabholz, B., & Delsuc, F. 2020. MitoFinder: Efficient automated large-scale extraction of mitogenomic data in target enrichment phylogenomics. *Molecular Ecology Resources*, 20(4), 892–905. https://doi.org/10.1111/1755-0998.13160

Bernt, M., Donath, A., Jühling, F., Externbrink, F., Florentz, C., Fritzsch, G., Pütz, J., Middendorf, M., & Stadler, P. F. 2013. MITOS: Improved *de novo* metazoan mitochondrial genome annotation. *Molecular Phylogenetics and Evolution*, *69*(2), 313–319. https://doi.org/10.1016/j.ympev.2012.08.023

Brůna T, Hoff KJ, Lomsadze A, Stanke M, Borodovsky M. 2021. BRAKER2: Automatic Eukaryotic Genome Annotation with GeneMark-EP+ and AUGUSTUS Supported by a Protein Database. *NAR Genomics and Bioinformatics*. 3(1), lqaa108.

Brůna, T., Lomsadze, A., & Borodovsky, M. 2024. GeneMark-ETP significantly improves the accuracy of automatic annotation of large eukaryotic genomes. *Genome Research*, *34*(5), 757–768. https://doi.org/10.1101/gr.278373.123

Buchfink B, Xie C, Huson DH. 2015. Fast and sensitive protein alignment using DIAMOND. *Nature Methods*. 12(1), 59-60.

Challis R, Richards E, Rajan J, Cochrane G, Blaxter M. 2020. BlobToolKit – Interactive Quality Assessment of Genome Assemblies. *G3* (Bethesda) 10:1361–1374. doi: 10.1534/g3.119.400908.

Cleveland, A., Verde, E.A. and Lee, R.W., 2011. Nutritional exchange in a tropical tripartite symbiosis: direct evidence for the transfer of nutrients from anemonefish to host anemone and zooxanthellae. *Marine Biology, 158,* pp.589-602.

De Jode, A., Quattrini, A. M., Chiodo, T., Daly, M., McFadden, C. S., Berumen, M. L., Meyer, C. P., Mills, S., Beldade, R., Bartholomew, A., Scott, A., Reimer, J. D., Yanagi, K., Fuji, T., Rodríguez, E., & Titus, B. M. 2024. Phylogenomics reveals coincident divergence between giant host sea anemones and the clownfish adaptive radiation. *bioRxiv*, 2024.01.24.576469. https://doi.org/10.1101/2024.01.24.576469

Dunn, D.F., 1981. The clownfish sea anemones: Stichodactylidae (Coelenterata: Actiniaria) and other sea anemones symbiotic with pomacentrid fishes. *Transactions of the American Philosophical Society*, *71*(1), pp.3-115.

Fautin, D.G., 1991. Review article The Anemonefish Symbiosis: What is Known and What is Not. *Symbiosis*.

Feng, H., Lv, S., Li, R., Shi, J., Wang, J., & Cao, P. 2023. Mitochondrial genome comparison reveals the evolution of cnidarians. *Ecology and Evolution*, *13*(6), e10157. https://doi.org/10.1002/ece3.10157

Formenti G, et al. 2022. Merfin: improved variant filtering, assembly evaluation and polishing via k-mer validation. *Nature Methods*.19:696–704. doi: 10.1038/s41592-022-01445-y.

Formenti G, et al. 2022. Gfastats: conversion, evaluation and manipulation of genome sequences using assembly graphs. *Bioinformatics* 38:4214–4216. doi: 10.1093/bioinformatics/btac460.

Gaboriau, T., Marcionetti, A., Garcia Jimenez, A., Schmid, S., Fitzgerald, L.M., Micheli, B., Titus, B. and Salamin, N. 2024. Host-use drives convergent evolution in clownfish and disentangles the mystery of an iconic adaptive radiation. *bioRxiv*, pp.2024-07.

Gabriel, L., Brůna, T., Hoff, K. J., Ebel, M., Lomsadze, A., Borodovsky, M., & Stanke, M. 2024. BRAKER3: Fully automated genome annotation using RNA-seq and protein evidence with GeneMark-ETP, AUGUSTUS, and TSEBRA. *Genome Research*. https://doi.org/10.1101/gr.278090.123

Gotoh, O. 2008. A space-efficient and accurate method for mapping and aligning cDNA sequences onto genomic sequence. Nucleic Acids Res. 36(8), 2630-2638.

Godwin, J. and Fautin, D.G., 1992. Defense of host actinians by anemonefishes. *Copeia. 1992(3)* pp.902-908

Guan D, et al. 2020. Identifying and removing haplotypic duplication in primary genome assemblies. Bioinformatics 36:2896–2898. doi: 10.1093/bioinformatics/btaa025.

Hoff KJ, Lange S, Lomsadze A, Borodovsky M, Stanke M. 2016. BRAKER1: unsupervised RNA-Seq-based genome annotation with GeneMark-ET and AUGUSTUS. Bioinformatics 32(5), 767-769.

Hoff KJ, Lomsadze A, Borodovsky M, Stanke M. 2019. Whole-genome annotation with BRAKER. Methods Mol Biol. 1962:65-95.

Holbrook, S.J. and Schmitt, R.J., 2005. Growth, reproduction and survival of a tropical sea anemone (Actiniaria): benefits of hosting anemonefish. *Coral Reefs*, *24*, pp.67-73.

Huang, N., & Li, H. 2023. compleasm: a faster and more accurate reimplementation of BUSCO. Bioinformatics 39(10):btad595.

Iwata H, Gotoh O. 2012. Benchmarking spliced alignment programs including Spaln2, an extended version of Spaln that incorporates additional species-specific features. Nucleic Acids Res. 40(20), e161-e161.

Johansen, S. D., Chi, S. I., Dubin, A., & Jørgensen, T. E. 2021. The Mitochondrial Genome of the Sea Anemone Stichodactyla haddoni Reveals Catalytic Introns, Insertion-Like Element, and Unexpected Phylogeny. *Life (Basel, Switzerland)*, *11*(5), 402. https://doi.org/10.3390/life11050402

Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. 2019. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol. 37(8):907-915.

Kovaka S, et al. 2019. Transcriptome assembly from long-read RNA-seq alignments with StringTie2. *Genome Biology*. 20(1):1-13.

Kuznetsov D, et al. 2023. OrthoDB v11: annotation of orthologs in the widest sampling of organismal diversity. *Nucleic Acids Research*. 51(D1), pp.D445-D451.

Laetsch DR, Blaxter ML. 2017. BlobTools: Interrogation of genome assemblies. *F1000 Research*. 6:1287. doi: 10.12688/f1000research.12232.1.

Li, H. 2023. Protein-to-genome alignment with miniprot. *Bioinformatics*, *39*(1), btad014. https://doi.org/10.1093/bioinformatics/btad014

Litsios, G., Sims, C.A., Wüest, R.O., Pearman, P.B., Zimmermann, N.E. and Salamin, N., 2012. Mutualism with sea anemones triggered the adaptive radiation of clownfishes. *BMC Evolutionary Biology*, *12*, pp.1-15.

Lubbock, R., 1981. The clownfish/anemone symbiosis: a problem of cellular recognition. *Parasitology*, *82*(1), pp.159-173.

Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. 2021. BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for

Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. *Molecular Biology & Evolution.* 38:4647–4654. doi: 10.1093/molbev/msab199.

Marçais, G., & Kingsford, C. 2011. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics*, *27*(6), 764–770. https://doi.org/10.1093/bioinformatics/btr011

Marcionetti, A. and Salamin, N., 2023. Insights into the genomics of clownfish adaptive radiation: the genomic substrate of the diversification. *Genome Biology and Evolution*, *15*(7), p.evad088.

Marcionetti, A., Rossier, V., Roux, N., Salis, P., Laudet, V. and Salamin, N., 2019. Insights into the genomics of clownfish adaptive radiation: genetic basis of the mutualism with sea anemones. *Genome Biology and Evolution*, *11*(3), pp.869-882.

Nevers, Y., Rossier, V., Train, C. M., Altenhoff, A., Dessimoz, C., & Glover, N. 2022. *Multifaceted quality assessment of gene repertoire annotation with OMArk* (p. 2022.11.25.517970). *bioRxiv*. https://doi.org/10.1101/2022.11.25.517970

Pertea, G., & Pertea, M. 2020. GFF utilities: GffRead and GffCompare. F1000Research, 9.

Ranallo-Benavidez, T. R., Jaron, K. S., & Schatz, M. C. 2020. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nature Communications*, *11*(1), 1432. https://doi.org/10.1038/s41467-020-14998-3

Rhie A, Walenz BP, Koren S, Phillippy AM. 2020. Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. Genome Biol. 21:245. doi: 10.1186/s13059-020-02134-9.

Roopin, M., Henry, R.P. and Chadwick, N.E., 2008. Nutrient transfer in a marine mutualism: patterns of ammonia excretion by anemonefish and uptake by giant sea anemones. *Marine Biology*, *154*, pp.547-556.

Roopin, M. and Chadwick, N.E., 2009. Benefits to host sea anemones from ammonia contributions of resident anemonefish. *Journal of Experimental Marine Biology and Ecology*, *370*(1-2), pp.27-34.

Szczebak, J.T., Henry, R.P., Al-Horani, F.A. and Chadwick, N.E., 2013. Anemonefish oxygenate their anemone hosts at night. *Journal of Experimental Biology*, *216*(6), pp.970-976.

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. 2015. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, *31*(19), 3210–3212. https://doi.org/10.1093/bioinformatics/btv351

Smit, AFA and Green, P RepeatMasker at http://www.repeatmasker.org

SRA Toolkit Development Team. 2020. SRA Toolkit. https://trace.ncbi.nlm.nih.gov/Traces/sra/sra.cgi?view=software.

Stanke M, Diekhans M, Baertsch R, Haussler, D. 2008. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics*, 24(5), 637-644.

Stanke, M., Schöffmann, O., Morgenstern, B., Waack, S. 2006. Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC Bioinformatics*, 7(1), 62.

Tillich, M., Lehwark, P., Pellizzer, T., Ulbricht-Jones, E. S., Fischer, A., Bock, R., & Greiner, S. 2017. GeSeq – versatile and accurate annotation of organelle genomes. *Nucleic Acids Research*, *45*(W1), W6–W11. https://doi.org/10.1093/nar/gkx391

Titus, B.M., Benedict, C., Laroche, R., Gusmão, L.C., Van Deusen, V., Chiodo, T., Meyer, C.P., Berumen, M.L., Bartholomew, A., Yanagi, K. and Reimer, J.D., 2019. Phylogenetic relationships among the clownfish-hosting sea anemones. *Molecular Phylogenetics and Evolution*, *139*, p.106526.

Titus, B.M., Bennett-Smith, M.F., Chiodo, T. and Rodríguez, E., 2024. The clownfish-hosting sea anemones (Anthozoa: Actiniaria): updated nomenclature, biogeography, and practical field guide. *Zootaxa*, *5506*(1), pp.1-34.

Verde, E.A., Cleveland, A. and Lee, R.W., 2015. Nutritional exchange in a tropical tripartite symbiosis II: direct evidence for the transfer of nutrients from host anemone and zooxanthellae to anemonefish. *Marine Biology*, *162*, pp.2409-2429.

Vurture GW, et al. 2017. GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics*. 33:2202–2204. doi: 10.1093/bioinformatics/btx153.

Figures

Figure 1. Long-read mitochondrial and nuclear genome assemblies for the clownfish hosting sea anemone species A) *Entacmaea quadricolor,* B) *Stichodactyla haddoni*, and C) *Radianthus doreensis.* Pictures of each species are the individuals sequenced in this study. Circularized and annotated mitogenomes were visualized with Geneious version 2024.0. BlobToolKit snailplots representing nuclear genome assembly statistics including genome size, largest scaffold length, N50 and N90 lengths, and GC composition. Snailplots are divided into 1,000 size-ordered bins around the circumference with each bin representing 0.1% of the assembly length. The distribution of contig lengths is shown in dark gray with the plot radius scaled to the longest contig present in the assembly (the smallest of the three arcs, also shown in red). Next two sectors in size (orange and pale-orange arcs) show the N50 and N90 contig lengths, respectively. The pale gray spiral shows the cumulative scaffold count on a log scale with white scale lines showing successive orders of magnitude. The dark-blue and pale-blue area around the outside of the plot shows the distribution of GC, AT, and N percentages in the same bins as the inner plot.

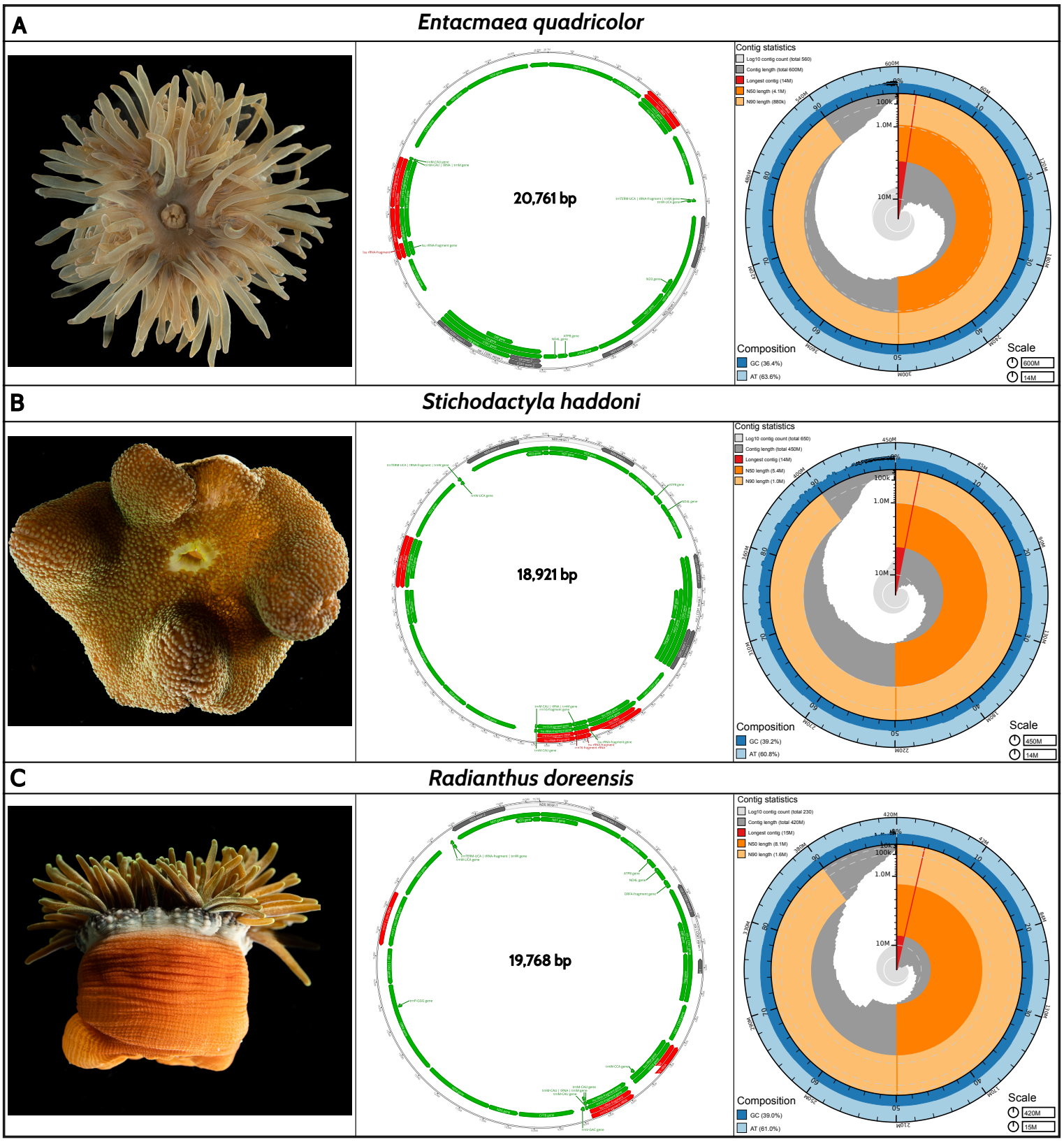
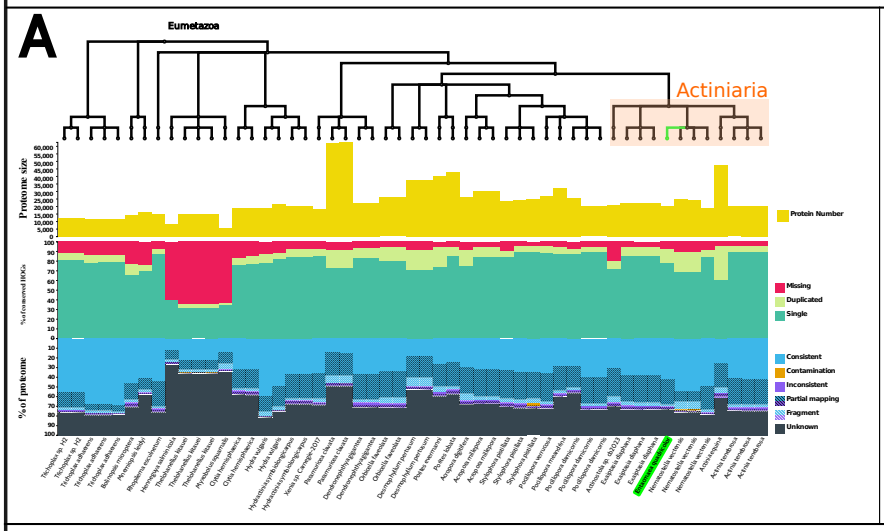
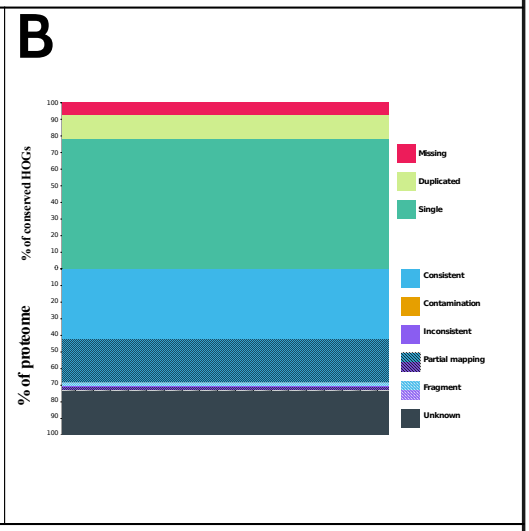

A
Entacmaea quadricolor
20,761 bp
Contig statistics
Log10 contig count (total 560)
Contig length (total 600M)
Longest contig (14M)
N50 length (4.1M)
N90 length (880k)
Composition
GC (36.4%)
AT (63.6%)
Scale
600M
14M
B
Stichodactyla haddoni
18,921 bp
Contig statistics
Log10 contig count (total 650)
Contig length (total 450M)
Longest contig (14M)
N50 length (5.4M)
N90 length (1.0M)
Composition
GC (39.2%)
AT (60.8%)
Scale
450M
14M
C
Radianthus doreensis
19,768 bp
Contig statistics
Log10 contig count (total 230)
Contig length (total 420M)
Longest contig (15M)
N50 length (8.1M)
N90 length (1.6M)
Composition
GC (39.0%)
AT (61.0%)
Scale
420M
15M

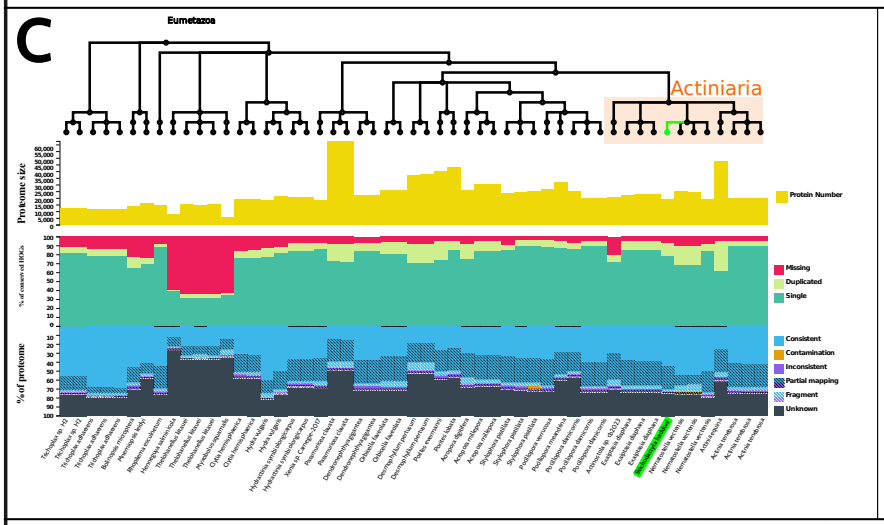
Figure 2. OMArk quality assessment of genome annotations. A, C, &E) OMArk comparison plots showing genome annotation statistics for *Entacmaea quadricolor*, *Stichodactyla. haddoni* and *Radianthus doreensis* in reference to other members of Eumetazoa. Positions of the clownfish-hosting sea anemones sequenced in this study are highlighted in green. List of available proteome accessions used for the comparisons in Supplementary material. B, D, & F) detailed proteome assessment for each species of clownfish-hosting sea anemones sequenced in this study.

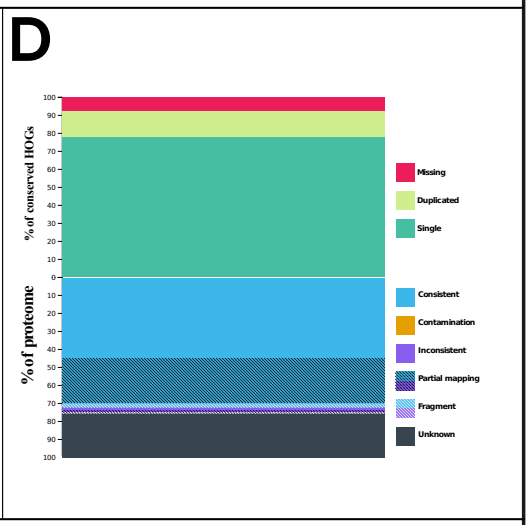
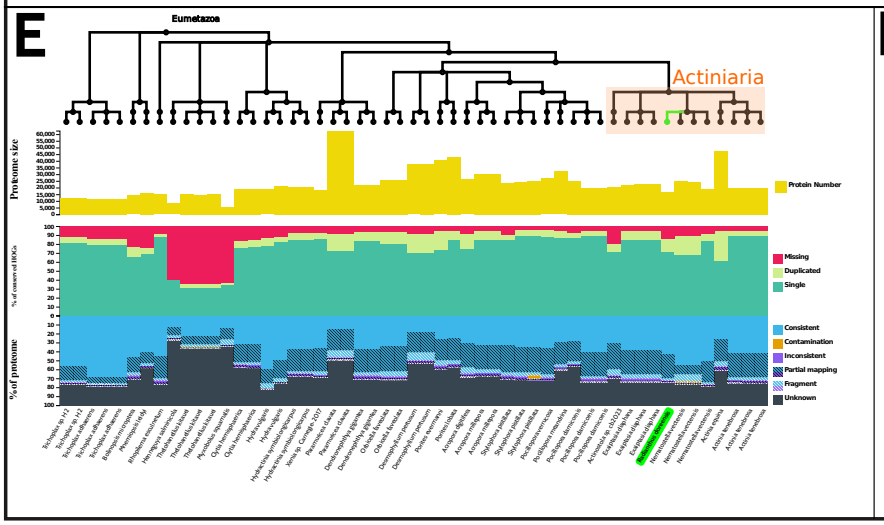
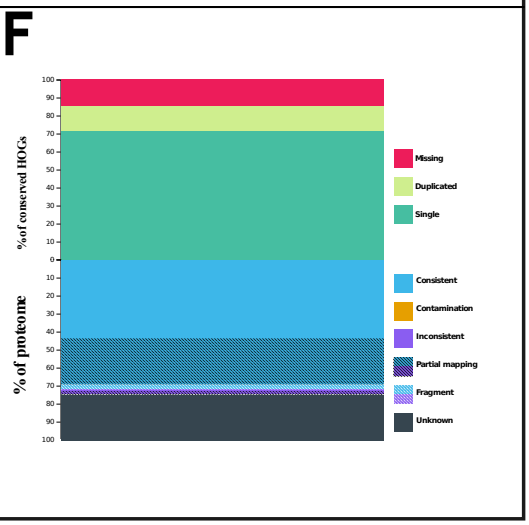
Entacmaea quadricolor





Stichodactyla haddoni





Radianthus doreensis