Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

The chromosome-scale genome assembly of the reef-building coral Platygyra daedalea

Craig Michell

craig.michell@kaust.edu.sa

Biological and Environmental Science and Engineering Division, King Abdullah University of Science and Technology, Thuwal, Kingdom of Saudi Arabia https://orcid.org/0000-0003-4706-7256

Patrick Micke

National Marine Science Centre, Faculty of Science and Engineering, Southern Cross University, Coffs Harbour, Australia 2450 https://orcid.org/0009-0001-8160-2147

Shoug Alguthmi

Biological and Environmental Science and Engineering Division, King Abdullah University of Science and Technology, Thuwal, Kingdom of Saudi Arabia https://orcid.org/0009-0006-7219-1995

Chloe-Marie Hawley

Biological and Environmental Science and Engineering Division, King Abdullah University of Science and Technology, Thuwal, Kingdom of Saudi Arabia

Eli Meyer

Holland Elder

James Cook University, Australian Institute of Marine Science, Townsville, Australia 4811. https://orcid.org/0000-0001-8432-4632

Emily J. Howells

National Marine Science Centre, Faculty of Science and Engineering, Southern Cross University, Coffs Harbour, Australia 2450. Lizard Island Research Station, Australian Museum, Sydney, New South Wales, Australia https://orcid.org/0000-0001-7732-2372

Manuel Aranda

Biological and Environmental Science and Engineering Division, King Abdullah University of Science and Technology, Thuwal, Kingdom of Saudi Arabia https://orcid.org/0000-0001-6673-016X

Data Note

Keywords: Platygyra, Chromosome, Genome, Coral, Telomere-2-telomere

Posted Date: April 7th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8967172/v2

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

The chromosome-scale genome assembly of the reef-building coral *Platygyra daedalea*

Michell, C.T[1], Micke, P.[2], Alguthmi, S.[1], Hawley, C-M.[1], Meyer, E., Elder, H,[3]., Howells, E. J. [2,4], Aranda, M.[1]

Affiliations:
1- King Abdullah University of Science and Technology, Biological and Environmental Science and Engineering Division, Thuwal, Kingdom of Saudi Arabia 23955.
2- National Marine Science Centre, Faculty of Science and Engineering, Southern Cross University, Coffs Harbour, Australia 2450.
3- James Cook University, Australian Institute of Marine Science, Townsville, Australia 4811.
4- Lizard Island Research Station, Australian Museum, Sydney, New South Wales, Australia

Keywords:
Platygyra, Chromosome, Genome, Coral, Telomere-2-telomere

Abstract

The genus *Platygyra* represents an essential group of reef-building corals in marine ecosystems. Here, we present the first high-quality genome assembly of the brain coral *Platygyra daedalea*, a widespread species known for its ecological resilience across diverse marine environments. Using long-read sequencing combined with chromosomal-level scaffolding based on a genetic linkage map, we assembled an 878.6 Mb genome with a contig N50 of 34.8 Mb. A total of 99.78% of the assembly was anchored into 16 linkage groups, with only eight contigs remaining unplaced. Annotation identified 27,516 protein-coding genes. This reference genome provides a critical resource for understanding coral biology, aiding conservation efforts, and exploring adaptive potential under climate change.

Introduction

Platygyra daedalea, a reef-building coral in the family Merulinidae, is a vital component of Indo-Pacific reef ecosystems. Known for its distinctive brain-like morphology, characterized by deeply grooved valleys and thick ridges, *P. daedalea* plays a critical role in reef structure and biodiversity. Its wide geographic range spans the Red Sea, Arabian Gulf, and western Pacific, where it inhabits diverse reef environments, from lagoons to reef slopes(Veron et al., 2016). This ecological versatility is complemented by its reproductive strategy as a broadcast spawner(Howells et al., 2014; Mangubhai & Harrison, 2008), resulting in high connectivity between populations(Buitrago-López et al., 2023)

Populations of *P. daedalea* in the Red Sea and Arabian Gulf are particularly notable for their ability to thrive under some of the most extreme marine conditions globally, including elevated sea temperatures (>32°C) and high salinity (>40 ppt)(Smith et al., 2022). These environmental extremes make the species an excellent model for studying coral resilience and adaptation(Elder et al., 2022; Hoey et al., 2016; Howells et al., 2021, 2025; Liew et al., 2020; Smith et al., 2022). In particular, populations from the Arabian Gulf offer valuable insight into the genetic and physiological mechanisms that enable persistence in extreme environments, insights that are increasingly important in the context of climate change.

Advancements in sequencing technologies have facilitated the generation of chromosome-scale genomes for non-model organisms, offering new opportunities to investigate the molecular basis of coral adaptation(Formenti et al., 2022; Guiglielmoni et al., 2024; Mc Cartney et al., 2023). To date, chromosome-scale assemblies have been published for several coral species, predominantly in the genus Acropora(Fiesinger et al., 2024; Salazar et al., 2022; Shinzato et al., 2021; Voolstra et al., 2017), as well as for *Stylophora pistillata(Voolstra et al., 2017)*. However, a high-quality genomic resource for *P. daedalea* remains missing, despite the number of quality datasets available for this species(Elder et al., 2022; Liew et al., 2020; Smith et al., 2022), hindering efforts to study its unique adaptations and evolutionary history. Genomic studies of *P. daedalea* can reveal key insights into genetic connectivity, symbiosis, and the molecular pathways involved in stress tolerance, insights that are crucial for addressing the escalating threats posed by climate change and supporting conservation initiatives(Hoey et al., 2016).

In this study, we present the first chromosome-scale genome assembly of *P. daedalea*. Using PacBio HiFi sequencing and a genetic linkage map constructed from 2b RADseq data, we generated a high-quality genomic resource. This assembly provides a foundation for exploring the molecular mechanisms underlying coral adaptation to extreme environments, facilitating comparative studies across coral species, and informing conservation and restoration efforts. As reefs continue to decline globally, such genomic resources are critical for understanding and protecting the biodiversity and ecological integrity of coral ecosystems.

Methods

Coral Collection, DNA Extraction and Sequencing

The *P. daedalea* adult JUB01 used in this study was collected by SCUBA in Jubail, Saudi Arabia, and transported to KAUST wrapped in wet bubble wrap. It has been maintained in the KAUST aquaria since collection and remains alive to this day.

Approximately 50 mg of fresh *P. daedalea* tissue, collected from the adult colony JUB01, was used for DNA extraction. High-molecular-weight DNA was extracted using a Phenol:Chloroform method. The tissue was lysed in an extraction buffer containing Tris-HCl pH 8.0 (100 mM), EDTA pH 8.0 (50 mM), PVP-40 (1%), Sodium metabisulfite (1%), NaCl (0.5 M), and SDS (2%). The mixture was incubated at 55°C for 60 minutes on a ThermoMixer at 450 rpm. A 1:1 volume of phenol:chloroform:isoamyl alcohol (25:24:1) was added to the lysate, and the

tube was shaken by hand for approximately 30 seconds. Phase separation was achieved by centrifugation at 16,000 × g for 5 minutes at room temperature, and the aqueous phase was carefully removed. An equal volume of chloroform:isoamyl alcohol (24:1) was then added to the aqueous phase, shaken for 30 seconds, and centrifuged under the same conditions. The final aqueous phase was collected for downstream purification. DNA was further purified using the Tube-O-dialyzer from G-Biosciences, following the manufacturer's recommendations. The sample was dialyzed in TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8.0) for 12 hours, with buffer changes performed twice during this period. Purified DNA was quantified using the Qubit dsDNA Broad Range kit, and purity was assessed using a Nanodrop spectrophotometer, with 260/280 and 230/280 ratios confirming high-quality DNA. DNA integrity and fragment size were evaluated using the Femto Pulse system, and only samples with fragment sizes greater than 100 kb were used for genomic library preparation. Extracted DNA was stored in 0.1X TE buffer at 4°C to preserve quality. High-quality genomic DNA from *P. daedalea* (JUB01) was sequenced on the PacBio Sequel IIe platform at the KAUST Bioscience Core Lab to generate HiFi long reads (Table 1).

RNA Extraction and Sequencing

RNA was extracted from the same *P. daedalea* individual (JUB01) used for DNA extraction, using approximately 50 mg of fresh tissue. To minimize RNase contamination, pipettes and the workspace were cleaned thoroughly with RNAseZAP, and all steps were conducted with RNase-free reagents and on ice. Extraction was performed immediately on the fresh tissue to preserve RNA integrity. The procedure followed the Trizol Plus RNA Purification Kit (ThermoFisher) protocol, including on-column DNase digestion to remove residual genomic DNA contamination. RNA quantity was measured using the Qubit RNA Broad Range assay, and integrity was evaluated using the Agilent Tapestation RNA ScreenTape system. A concentration threshold of 50 ng/µL was applied, as 1 µg of RNA was required as the input for library preparation. Extracts with a RIN value below 8 or concentrations below the threshold were excluded from downstream applications. Extracted RNA was stored at -80°C in nuclease-free water to preserve integrity until library preparation.

RNA sequencing libraries were prepared using the NEBNext Poly(A) mRNA Magnetic Isolation Module in conjunction with the NEBNext Ultra II Directional RNA Library Prep Kit for Illumina. All steps were performed as directed by the protocol, with modifications as described in the protocol appendices for a 350 bp RNA fragment. The RNA library was sequenced as part of a pool on the Illumina NovaSeq 6000 platform using an SP flowcell, generating paired-end reads of 150 bp (Table 1). In addition to short Illumina reads, a direct RNA sequencing library was prepared from the same RNA extraction at the KAUST Bioscience Core Lab for direct RNA sequencing on the PromethION platform using a R9.4.1 chemistry (Table 1).

Linkage Map Generation

A genetic linkage map for *P. daedalea* was constructed using data from 144 individuals, with sequencing libraries prepared following the 2bRADseq protocol(Wang et al., 2012) and sequenced on the Illumina HiSeq 2000 platform. The raw sequencing data are publicly available under NCBI project accessions PRJNA302771 and PRJNA302770. Linkage analysis followed the Snelling et al. (2017) protocol,(Snelling et al., 2017) employing the R package Onemap (Margarido et al., 2007), which estimates linkage and linkage phases using maximum likelihood, enabling the integration of both partially and fully informative markers into a single unified map. Before analysis, we applied stringent filtering to SNP genotypes, excluding samples with excessive missing data, poorly genotyped markers, highly variable sequence tags, and loci with extreme segregation distortions, ensuring a high-quality dataset. Linkage groups were identified based on two-point recombination frequencies (LOD ≥ 8, $r \leq 0.5$), followed by an iterative mapping procedure to refine marker positions within each group.

Genome Assembly

The *P. daedalea* genome was assembled using PacBio HiFi reads generated on the Sequel IIe platform. Raw HiFi reads were first filtered to remove adapter sequences and low-quality reads using HiFiAdapterFilt v3.0.0 (Sim et al., 2022). The filtered reads were then assembled using HiFiasm v0.24.0 (Cheng et al., 2021) with the --primary option. HiFiasm is optimized for highly accurate long-read data and produces a primary assembly that represents a single haplotype. To further refine the assembly and ensure non-redundant representation of the genome, purge_haplotigs v1.1.2 (Roach et al., 2018) was applied. This step removed duplicate or alternative haplotypes, resulting in a haploid assembly suitable for downstream analysis. We then used the long HiFi reads to scaffold and fill in gaps between scaffolded contigs using NtLINK v1.3.11 (Coombe et al., 2023) following the recommended settings with 5 iterations. Finally, the integration of the genetic linkage map was performed by mapping reads to the haploid primary assembly using Bowtie(Langmead et al., 2009). Chromonomer v1.15 (Catchen et al., 2020) was then used to integrate the linkage map into the assembly, resulting in the final chromosome-scale assembly.

Mitochondrial Genome Assembly and Annotation

The mitochondrial genome was assembled from PacBio HiFi reads using MitoHiFi v3.2.2. A previously published *P. daedalea* mitochondrial genome (GenBank accession: PP856480.1) was provided as a reference seed in FASTA and GenBank format to guide the identification of mitochondrial reads and annotation. MitoHiFi uses sequence similarity to recruit mitochondrial reads, which are subsequently assembled de novo into a circular contig. Gene annotation was performed based on homology to the reference mitogenome. Default parameters were used for all steps. Circularization was confirmed based on overlapping terminal regions and uniform read coverage across the contig.

Repetitive Element Annotation

Transposable and repetitive elements in the *P. daedalea* genome were annotated using a combination of EDTA v2.1.3(Su et al., 2021), RepeatModeler v2.0.4(Flynn et al., 2020), CD-Hit-Est v4.8.1(Fu et al., 2012), and RepeatMasker v4.1.7(Smit et al., 2015). The input genome assembly, a haploid assembly in FASTA format, was generated using HiFiasm v0.24.0 and had a total size of 0.87 Gb. EDTA was executed with the options --species others --step all --sensitive 0 --anno 1 to perform comprehensive de novo transposable element detection and annotation. RepeatModeler was used with the -LTRStruct option to enhance the identification of LTR retrotransposons. The repeat libraries generated by EDTA (3,381 sequences) and RepeatModeler (4,848 sequences) were concatenated and processed using CD-Hit to reduce redundancy. CD-Hit-Est was run with the options -d 0 -aS 0.8 -c 0.8 -G 0 -g 1 -b 500, following the 80-80-80 rule to cluster sequences with ≥80% identity across 80% of their length. Finally, the reduced repeat library was classified using RepeatClassifier part of RepeatModeler2, and sequences classified as unknown were further classified using DeepTE v22/10/2022 (Yan et al., 2020) (Metazoans Model). Using this method, only 187 sequences out of 5,908 repetitive elements were not classified to a transposable element family. Telomeric ends of the chromosomes were annotated using TIDK v0.2.31(Brown et al., 2025) with the find --clade Actiniaria and –window 50000 options, identifying and characterizing telomeric repeat sequences specific to the clade. The non-redundant repeat library generated through this process was used as input for RepeatMasker (Smit et al., 2015) to mask repetitive regions in the genome. RepeatMasker was run with the options -gff -nolow -xsmall -a -lib pdae.nonred.TElib.fa.

Gene Annotation

Genome annotation of the masked *P. daedalea* genome was performed using a combination of ab initio, transcriptome, and protein evidence-based approaches. Ab initio gene prediction was initially conducted using Helixer v0.3.4 (Stiehler et al., 2021) without external evidence. To incorporate transcriptomic evidence, RNA-seq data was first trimmed for quality using FastP v0.23.2 (Chen et al., 2018) and subsequently mapped to the genome assembly using HISAT2 v2.2.1 (Kim et al., 2019). The mapped RNA-seq reads were then used as evidence with the BRAKER3 v3.0.8 annotation pipeline(Gabriel et al., 2024). The direct RNA sequencing was incorporated into a gene annotation using StringTie v2.2.1 (Pertea et al., 2015) along with the short Illumina reads via the –mix option. The outputs from the BRAKER3 annotation pipelines were integrated using TSEBRA v1.1.2.5 (Gabriel et al., 2021), which was run with the recommended setting and the default configuration file. This integration produced a high-confidence and comprehensive gene set. Finally, genes that had extrinsic support that overlapped Helixer ab initio predictions were merged to produce the final gene set. Functional annotation of the final predicted gene set was performed using Pannzer2 (Törönen & Holm, 2022), which assigned functional descriptions and gene ontology (GO) terms to the predicted proteins (Table 4).

Genome assembly quality assessment

The quality of the *Platygyra daedalea* genome assembly was evaluated using multiple complementary approaches. Genome completeness was assessed using BUSCO v5.7.1 (Manni et al., 2021) in genome mode (–mode geno -f) with the Metazoa_odb10 database (2021-02-17), providing estimates of complete, duplicated, fragmented, and missing conserved orthologs. Assembly contiguity statistics, including contig and scaffold N50 values, total assembly size, and contig counts, were calculated using QUAST v5.2.0 (Mikheenko et al., 2018) with the parameters -m 5000 -t 64 -s -e -b –large.

K-mer-based analyses were performed to assess genome characteristics and assembly quality. K-mers (k = 27) were counted from PacBio HiFi reads using Jellyfish v2.3.0, and the resulting k-mer frequency distribution was analysed using GenomeScope2 to estimate genome size, heterozygosity, and repeat content. The same k-mer dataset was used for assembly validation using Meryl v1.4.1 (Rhie et al., 2020) and Merqury v1.3 (Rhie et al., 2020). The optimal k-mer size (k = 27) was determined using the Merqury best_k.sh script, and Merqury was run with default parameters to estimate consensus quality (QV) and completeness by comparing the assembly to the k-mer spectrum derived from the sequencing reads.

The predicted gene set was evaluated using BUSCO in transcriptome mode with the Metazoa_odb10 database (2021-02-17). In addition, gene completeness and consistency were assessed using OMArk v0.4.1 (Nevers et al., 2024) with the Cnidaria ancestral clade and the OMAmer database release *All.Jul2024* (OMAmer v2.1.2). To avoid inflation of duplication estimates due to alternative splice isoforms, OMArk analysis was performed on the longest isoform per gene. Taxonomic consistency of the predicted proteome was assessed using AAI-profiler (Medlar et al., 2018), which calculates amino acid identity profiles against reference databases. Results were visualized using Krona plots to confirm lineage assignment and assess potential contamination. In addition, phylogenetic placement of the predicted proteins was evaluated using OrthoFinder2 v2.5.5 (Emms & Kelly, 2019) with default parameters, including comparisons with 17 scleractinian coral species to confirm expected evolutionary relationships.

Results

Nucleic acid extraction and sequencing

High-quality genomic DNA from P. daedalea (JUB01) was sequenced and deposited onto the Sequence Read Archive of NCBI, a summary of the raw data can be found in Table 1 along with the SRA accession numbers.

Table 1 | Summary of raw sequencing data for the genome assembly of *Platygyra daedalea* from the Arabian Gulf (BioProject PRJNA1234051).

Data Type	Platform	Total Raw Reads	Average Read Length	Quality Metrics	SRA Accession
DNA	PacBio Sequel IIe	3,267,008	17 kb	Fragment size > 100 kb	SRR32629517 SRR32629516
RNA	Illumina NovaSeq 6000	80,556,366	150 bp	RIN ≥ 8	SRR32629519
Direct RNA	PromethION	2,175,559	750 bp	RIN ≥ 8	SRR32629518

Genome Assembly

The *P. daedalea* genome was assembled using PacBio HiFi reads, generating a haploid primary assembly of 964,867,186 bp across 1,403 contigs. Following haplotype purging, the assembly was reduced to 890,283,338 bp and 252 contigs. Subsequent scaffolding and gap filling with NtLINK further improved contiguity, resulting in 44 contigs. Integration of a genetic linkage map produced a chromosome-scale assembly with a total size of 878,612,309 bp, a contig N50 of 35.9 Mb, and a scaffold N50 of 59.47 Mb. The final assembly comprises 16 chromosomes (Figure 1) and eight unplaced contigs (not shown in Figure 1), with 99.78% of the genome anchored to chromosomes. A complete summary of assembly statistics is provided in Table 2. A complete summary of assembly statistics is provided in Table 2.

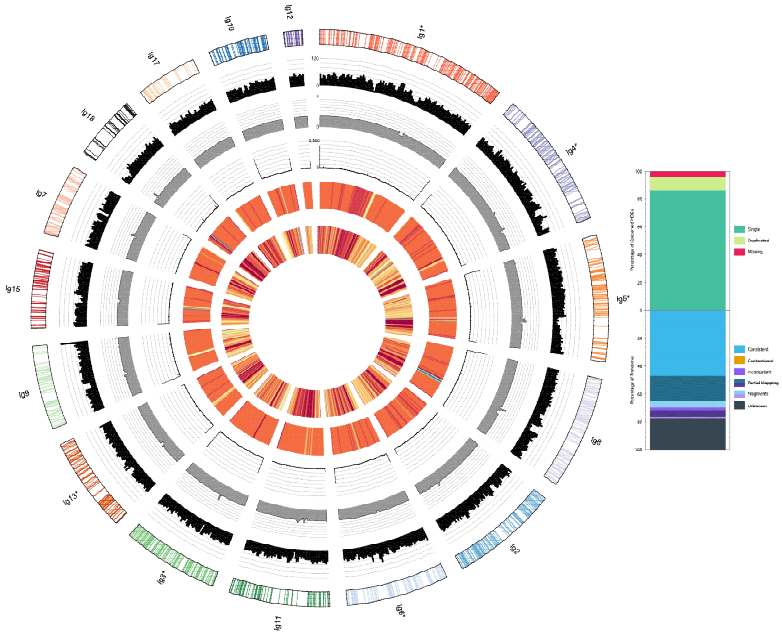


Figure 1 | Chromosome-scale karyotype of *Platygyra daedalea* and genome feature tracks. Shown are chromosomes >1 Mb. Colored bands on the ideograms indicate genetic markers defined by the linkage map, chromosome labels with "*" are telomere-to-telomere. Tracks (outer to inner) are: (1) mean HiFi read coverage per 1 Mb window (genome-wide average 47.51×; reads with mapping quality <Q20 excluded), (2) GC content per 1 Mb window, (3) telomeric repeat density per 50 kb window, (4) repeat density per 1 Mb window, and (5) gene density per 1 Mb window. An inset stacked bar plot summarizes OMark results for the *P. daedalea* genome: 4,117 Hierarchical Orthologous Groups (HOGs), 95.65% complete (86.5 % single-copy, 9.16% duplicated), 4.35% missing; lineage placement—Cnidaria consistent 69.27%, inconsistent 7.84%, contaminants 0.00%, unknown 22.9%.

Table 2 | Summary metrics of the Platygyra daedalea genome assembly.

Metric	**Value**
Total Assembly Size (bp)	878,612,309
GC %	39.53
Contig N50 (bp)	34,798,031
Contig N90 (bp)	13,092,603
Largest Contig (bp)	73,577,125
Scaffold N50 (bp)	59,477,340
Largest Scaffold (bp)	110,899,877
Scaffold L50	6
Scaffold L90	13
Number of Contigs	44
Number of Scaffolds	24
Number of Gaps	20
Linkage group markers anchored in 16 chromosomes	99.78%
Merquery QV score	66.28
BUSCO* (Metazoa odb10)	C:96.9% [S: 94%; D:2.8%]; F: 0.7%; M: 2.4%

*Complete:Single:Duplicate:Fragmented:Missing

Mitochondrial Genome Assembly

The mitochondrial genome of *P. daedalea* was assembled as a single circular contig of 16,412 bp (Figure 2). The assembly exhibited high and uniform read coverage, with a mean depth of 1,361.5× (± 240.95 SD), supporting the accuracy and completeness of the reconstructed mitogenome. Annotation identified a total of 26 genes, including the expected complement of protein-coding genes, ribosomal RNAs, and transfer RNAs typical of scleractinian corals. Gene content and organization were consistent with previously reported *P. daedalea* mitochondrial genomes.

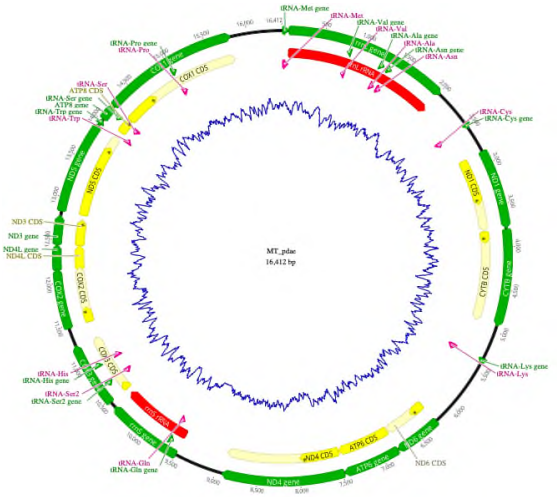


Figure 2 | Circular representation of the mitochondrial genome of *Platygyra daedalea*. The mitochondrial genome (16,412 bp) is shown with annotated protein-coding genes (yellow), transfer RNAs (green writing, CDS in pink), and ribosomal RNAs (red). Gene orientation is indicated by the direction of arrows, with genes located on the forward and reverse strands displayed on the outer and inner tracks, respectively. The inner blue line plot represents the GC content (%) across the genome. Tick marks indicate nucleotide positions along the circular genome.

Repetitive Element Annotation

Repetitive elements accounted for 61.66% of the genome (Table 3). Among these, retrotransposons comprised 16.75% of the genome, with long terminal repeat (LTR) elements representing the dominant fraction (12.54%), primarily driven by Gypsy/DIRS1 elements (10.49%). DNA transposons contributed 25.78% of the genome, representing the largest class of interspersed repeats. A substantial proportion of repeats (17.97%) remained unclassified, reflecting lineage-specific or highly diverged elements. The overall repeat landscape (Figure 3) shows a predominance of LTR retrotransposons and DNA transposons, consistent with patterns observed in other scleractinian coral genomes.

Telomeric repeat sequences specific to the clade Actiniaria were identified and annotated using TIDK. A total of eight chromosomes contained telomeric repeats at both ends, while a further seven chromosomes exhibited a single telomeric peak, suggesting that these contigs likely represent single chromosomal arms (Figure 1). Additionally, three of the eight unplaced contigs showed high densities of telomeric repeats (>150 copies per 50 kbp window), indicating that these sequences likely correspond to highly repetitive regions that could not be confidently integrated into the chromosomal assembly.

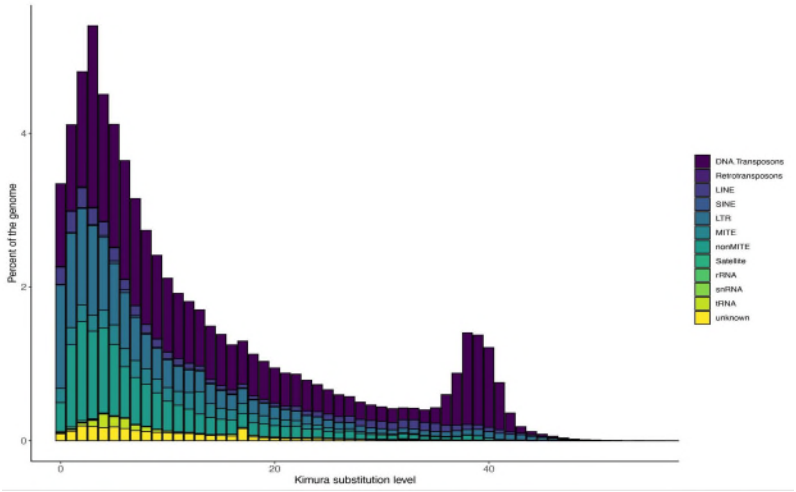


Figure 3 | Repeat landscape of the *Platygyra daedalea* genome. Proportions of major repeat classes are shown as a function of Kimura divergence, providing an estimate of repeat age and historical activity. Total masked repetitive content accounts for 61.66% of the genome.

Table 3 | Summary of repeat content in the *Platygyra daedalea* genome assembly. Repeat classes were identified and quantified across the genome. Percentages indicate the proportion of the total assembly length occupied by each repeat class.

Repeat Class	Count	Genome (%)
Class I Retrotransposons	443,874	16.75
Class II DNA Transposons	1,065,433	25.78
Unclassified repeats	632,293	17.97
Other repeats (small RNAs, Satellites, Simple repeats)	29,047	1.15

Gene Annotation

A total of 27,516 genes were predicted using a combination of ab initio, transcriptomic, and protein evidence-based approaches (Table 4.). Gene completeness was assessed using BUSCO in protein mode, reporting 93.6% completeness. The ratio of mono-exonic to multi-exonic genes was 0.42 which is in line with other model organism genomes (Brůna et al., 2023).

Table 4 | Summary of the final gene annotation for the *Platygyra daedalea* genome. The table summarises gene model statistics for the chromosome-scale assembly, including numbers of predicted genes and transcripts and annotation quality metrics.

Total Predicted Genes	27,516
Total predicted transcripts	32,295
Ratio of mono-exonic to multi-exonic genes	0.42
Average Gene Length (CDS length)	8,205 bp (1,373 bp)
Average number of exons per gene	6.1

Average exon length	223 bp
Average number of introns per gene	5.1
Average intron length	1,329 bp
BUSCO* (Metazoa odb10) longest isoform	C:93.6%[S:91.9%,D:1.7%],F:2.1%,M:4.3%,n:954
Genes with GO annotations (Longest isoform only)	18,422 (17,241)

* Complete:Single:Duplicate:Fragmented:Missing

Genome Assembly Quality Assessment

The quality of the *Platygyra daedalea* genome assembly was evaluated using multiple complementary approaches. Genome completeness assessed using BUSCO in genome mode with the Metazoa_odb10 database indicated a high level of completeness, with 96.9% of conserved single-copy orthologs recovered, including 94.0% as single-copy and 2.8% as duplicated, while 0.7% were fragmented and 2.4% were missing (Table 2). Assembly contiguity metrics further supported the high quality of the assembly, with a contig N50 of 34.8 Mb and a scaffold N50 of 59.47 Mb.

K-mer-based analyses were used to assess genome characteristics and validate assembly quality. GenomeScope2 analysis of 27-mers estimated a haploid genome size of approximately 731–733 Mb, with a heterozygosity rate of ~1.74% and a repeat content of ~45.7% (Figure 4). These estimates are consistent with the overall assembly size and repeat composition inferred from repeat annotation. Independent k-mer-based validation using Merqury yielded a consensus quality value (QV) of 66.28, indicating a highly accurate assembly with low base-level error rates. K-mer completeness analysis showed that 98.28% of k-mers derived from the sequencing reads were present in the combined assembly, indicating near-complete representation of the genome. Lower completeness values observed for the primary (75.30%) and alternate (69.71%) assemblies reflect the expected partitioning of haplotypes in a diploid genome following haplotig purging.

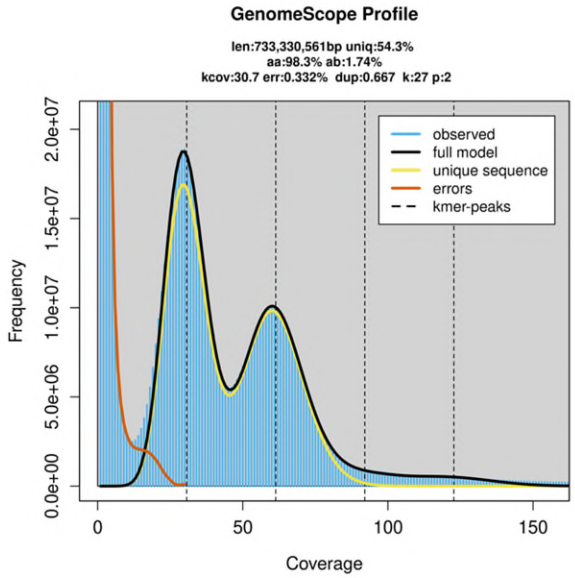


Figure 4 | K-mer spectrum analysis of the Platygyra daedalea genome. Genome characteristics were inferred from 27-mer frequency distributions generated from PacBio HiFi reads and analysed using GenomeScope2. The observed k-mer profile (blue) and fitted model (black) indicate a predominantly homozygous genome with low heterozygosity (~1.7%). The estimated

haploid genome size is approximately 731–733 Mb, with a substantial proportion of repetitive sequence (~334–335 Mb) and unique sequence (~396–398 Mb). The distribution shows a dominant homozygous peak and a smaller heterozygous peak, consistent with a diploid genome with moderate heterozygosity. Error k-mers are represented at low frequencies.

Assessment of the predicted gene set demonstrated high completeness and consistency. BUSCO analysis in protein mode recovered 93.6% of conserved orthologs, consistent with a high-quality annotation. OMArk analysis of the longest isoform per gene, using the Cnidaria ancestral clade, identified 4,117 conserved hierarchical orthologous groups (HOGs), of which 95.65% were complete. Of these, 86.5% were single-copy and 9.16% were duplicated, with only 4.35% missing. Whole-proteome assessment indicated that 69.27% of proteins had consistent lineage placement, with no detectable contamination, although 22.9% of proteins remained unclassified.

Taxonomic and phylogenetic validation further confirmed the quality of the assembly and annotation. Amino acid identity profiling showed that the top matches corresponded exclusively to cnidarian sequences, supporting correct taxonomic assignment. Phylogenetic reconstruction using orthologous protein sequences placed *P. daedalea* within the genus Platygyra and correctly resolved major scleractinian coral clades, consistent with established evolutionary relationships (Figure 4).

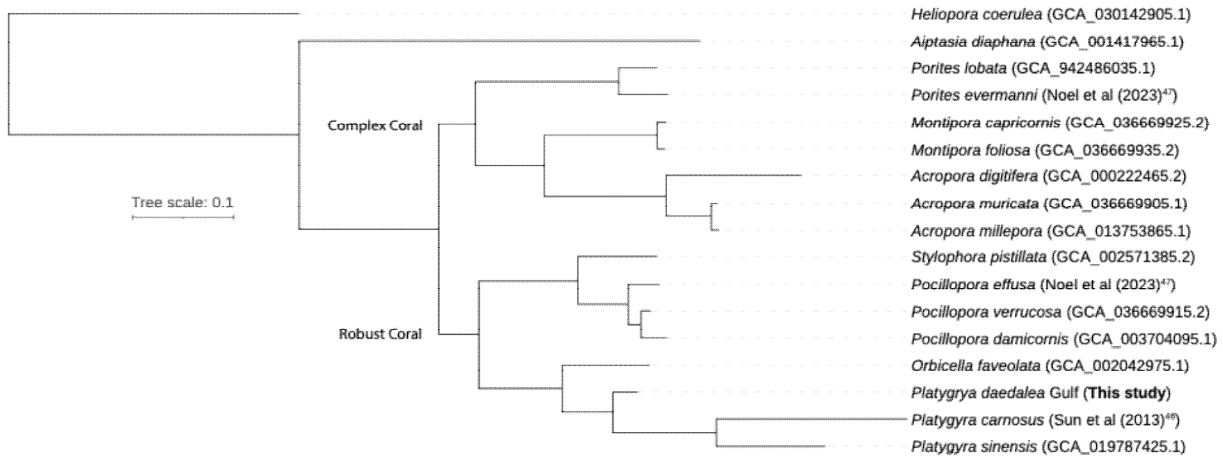


Figure 4 | Species tree of representative reef-building corals (order Scleractinia). The phylogeny is inferred from orthogroups of protein-coding sequences. GenBank accession numbers and corresponding citations for all taxa are provided in the figure.

Discussion

This study presents the first chromosome-scale genome assembly for *Platygyra daedalea*, providing a highly contiguous and accurate genomic resource for a key reef-building coral species. The integration of PacBio HiFi sequencing with a genetic linkage map enabled the recovery of 16 chromosome-length scaffolds, with 99.78% of the assembly anchored and minimal fragmentation. Assembly quality metrics, including high BUSCO completeness, strong k-mer concordance, and a Merqury QV exceeding 60, indicate a near-complete and highly accurate representation of the genome. The consistency of gene content, repeat composition, and phylogenetic placement further supports the reliability of the assembly and annotation. Compared to previously published coral genomes, this assembly represents a substantial improvement in

contiguity and completeness for the genus Platygyra, filling a critical gap in available genomic resources for corals.

The genomic features of *P. daedalea* provide important insights into coral genome evolution and resilience. The high proportion of repetitive elements (~61.7%) is notably greater than that reported for many other scleractinian corals, which typically exhibit repeat contents closer to ~40%, suggesting lineage-specific expansion or retention of transposable elements. This repeat landscape is dominated by LTR retrotransposons and DNA transposons, consistent with patterns observed across anthozoan genomes, but at an elevated scale. The identification of telomeric repeats across most chromosomes, including evidence for near telomere-to-telomere assemblies, highlights the structural completeness of the genome and provides a foundation for studying chromosome evolution in corals. In addition, this genome provides a valuable reference for resolving phylogenetic relationships within Scleractinia, where genome-scale data are increasingly needed to disentangle complex evolutionary histories and lineage diversification. Importantly, this resource enables future investigations into the genetic basis of thermal tolerance and environmental adaptation in *P. daedalea*, a species known for its persistence in extreme environments such as the Arabian Gulf. The availability of a high-quality reference genome will facilitate population genomics, genotype–environment association studies, and conservation genomics approaches aimed at enhancing coral resilience under climate change.

Data Records

All data generated in this study are deposited in publicly accessible repositories. Sequencing data accession numbers, together with descriptions of the associated datasets, are provided in Table 1. The genome assembly is available through GenBank under BioProject accession number **PRJNA1234051**. Genome annotation (GFF), repeat annotation (GFF), mitochondrial genome assembly and annotation files are archived on Drayd (https://doi.org/10.5061/dryad.3j9kd520f). All bioinformatic tools used in this study are described in the Methods section; default parameters were applied unless otherwise stated.

Acknowledgments

This work was funded by the KAUST Office of Research Administration (ORA) award no. ORA-2021-URF/1/4697-01-01.

References

Brown, M. R., Gonzalez de La Rosa, P., & Blaxter, M. (2025). tidk: a toolkit to rapidly identify telomeric repeats from genomic datasets. *Bioinformatics*, btaf049. https://doi.org/10.1093/bioinformatics/btaf049

Brůna, T., Li, H., Guhlin, J., Honsel, D., Herbold, S., Stanke, M., Nenasheva, N., Ebel, M., Gabriel, L., & Hoff, K. J. (2023). Galba: genome annotation with miniprot and AUGUSTUS. *BMC Bioinformatics*, *24*(1). https://doi.org/10.1186/s12859-023-05449-z

Buitrago-López, C., Cárdenas, A., Hume, B. C. C., Gosselin, T., Staubach, F., Aranda, M., Barshis, D. J., Sawall, Y., & Voolstra, C. R. (2023). Disparate population and holobiont structure of pocilloporid corals across the Red Sea gradient demonstrate species-specific evolutionary trajectories. *Molecular Ecology*, *32*(9). https://doi.org/10.1111/mec.16871

Catchen, J., Amores, A., & Bassham, S. (2020). Chromonomer: A tool set for repairing and enhancing assembled genomes through integration of genetic maps and conserved synteny. *G3: Genes, Genomes, Genetics*, *10*(11). https://doi.org/10.1534/g3.120.401485

Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). Fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, *34*(17). https://doi.org/10.1093/bioinformatics/bty560

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

Coombe, L., Warren, R. L., Wong, J., Nikolic, V., & Birol, I. (2023). ntLink: A Toolkit for De Novo Genome Assembly Scaffolding and Mapping Using Long Reads. *Current Protocols*, *3*(4). https://doi.org/10.1002/cpz1.733

Elder, H., Weis, V. M., Montalvo-Proano, J., Mocellin, V. J. L., Baird, A. H., Meyer, E., & Bay, L. K. (2022). Genetic Variation in Heat Tolerance of the Coral Platygyra Daedalea Indicates Potential for Adaptation to Ocean Warming. *Frontiers in Marine Science*, *9*. https://doi.org/10.3389/fmars.2022.925845

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

Fiesinger, A., Buitrago-López, C., Sharaf, A., Cárdenas, A., & Voolstra, C. R. (2024). A draft genome assembly of the reef-building coral Acropora hemprichii from the central Red Sea. *Scientific Data* , *11*(1). https://doi.org/10.1038/s41597-024-04080-8

Flynn, J. M., Hubley, R., Goubert, C., Rosen, J., Clark, A. G., Feschotte, C., & Smit, A. F. (2020). RepeatModeler2 for automated genomic discovery of transposable element families. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(17). https://doi.org/10.1073/pnas.1921046117

Formenti, G., Theissinger, K., Fernandes, C., Bista, I., Bombarely, A., Bleidorn, C., Ciofi, C., Crottini, A., Godoy, J. A., Höglund, J., Malukiewicz, J., Mouton, A., Oomen, R. A., Paez, S., Palsbøll, P. J., Pampoulie, C., Ruiz-López, M. J., Svardal, H., Theofanopoulou, C., … Zammit, G. (2022). The era of reference genomes in conservation genomics. In *Trends in Ecology and Evolution* (Vol. 37, Number 3). https://doi.org/10.1016/j.tree.2021.11.008

Fu, L., Niu, B., Zhu, Z., Wu, S., & Li, W. (2012). CD-HIT: Accelerated for clustering the next-generation sequencing data. *Bioinformatics*, *28*(23). https://doi.org/10.1093/bioinformatics/bts565

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*, *34*(5). https://doi.org/10.1101/gr.278090.123

Gabriel, L., Hoff, K. J., Brůna, T., Borodovsky, M., & Stanke, M. (2021). TSEBRA: transcript selector for BRAKER. *BMC Bioinformatics*, *22*(1). https://doi.org/10.1186/s12859-021-04482-0

Guiglielmoni, N., Villegas, L. I., Kirangwa, J., & Schiffer, P. H. (2024). Revisiting genomes of non-model species with long reads yields new insights into their biology and evolution. *Frontiers in Genetics*, *15*. https://doi.org/10.3389/fgene.2024.1308527

Hoey, A. S., Howells, E., Johansen, J. L., Hobbs, J. P. A., Messmer, V., McCowan, D. M., Wilson, S. K., & Pratchett, M. S. (2016). Recent advances in understanding the effects of climate change on coral reefs. In *Diversity* (Vol. 8, Number 2). https://doi.org/10.3390/d8020012

Howells, E. J., Abrego, D., Liew, Y. J., Burt, J. A., Meyer, E., & Aranda, M. (2021). Enhancing the heat tolerance of reef-building corals to future warming. *Science Advances*, *7*(34). https://doi.org/10.1126/sciadv.abg6070

Howells, E. J., Abrego, D., Schmidt-Roach, S., Puill-Stephan, E., Denis, H., Harii, S., Bay, L. K., Burt, J. A., Monro, K., & Aranda, M. (2025). Marine heatwaves select for thermal tolerance in a reef-building coral. *Nature Climate Change*. https://doi.org/10.1038/s41558-025-02381-3

Howells, E. J., Abrego, D., Vaughan, G. O., & Burt, J. A. (2014). Coral spawning in the Gulf of Oman and relationship to latitudinal variation in spawning season in the northwest Indian Ocean. *Scientific Reports*, *4*. https://doi.org/10.1038/srep07484

Kim, D., Paggi, J. M., Park, C., Bennett, C., & Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology*, *37*(8). https://doi.org/10.1038/s41587-019-0201-4

Langmead, B., Trapnell, C., Pop, M., & Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*, *10*(3). https://doi.org/10.1186/gb-2009-10-3-r25

Liew, Y. J., Howells, E. J., Wang, X., Michell, C. T., Burt, J. A., Idaghdour, Y., & Aranda, M. (2020). Intergenerational epigenetic inheritance in reef-building corals. In *Nature Climate Change* (Vol. 10, Number 3). https://doi.org/10.1038/s41558-019-0687-2

Mangubhai, S., & Harrison, P. L. (2008). Gametogenesis, spawning and fecundity of Platygyra daedalea (Scleractinia) on equatorial reefs in Kenya. *Coral Reefs*, *27*(1), 117–122. https://doi.org/10.1007/s00338-007-0297-8

Manni, M., Berkeley, M. R., Seppey, M., Simão, F. A., & Zdobnov, E. M. (2021). BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. *Molecular Biology and Evolution*, *38*(10). https://doi.org/10.1093/molbev/msab199

Margarido, G. R. A., Souza, A. P., & Garcia, A. A. F. (2007). OneMap: Software for genetic mapping in outcrossing species. *Hereditas*, *144*(3). https://doi.org/10.1111/j.2007.0018-0661.02000.x

Mc Cartney, A. M., Formenti, G., Mouton, A., De Panis, D., Marins, L. S., Leitão, H. G., Diedericks, G., Kirangwa, J., Morselli, M., Salces-Ortiz, J., Escudero, N., Iannucci, A., Natali, C., Svardal, H., Fernández, R., De Pooter, T., Joris, G., Strazisar, M., Wood, J., … Mazzoni, C. J. (2023). The European Reference Genome Atlas: piloting a decentralised

approach to equitable biodiversity genomics. In *bioRxiv*. https://doi.org/10.1101/2023.09.25.559365

Medlar, A. J., Törönen, P., & Holm, L. (2018). AAI-profiler: Fast proteome-wide exploratory analysis reveals taxonomic identity, misclassification and contamination. *Nucleic Acids Research*, *46*(W1). https://doi.org/10.1093/nar/gky359

Mikheenko, A., Prjibelski, A., Saveliev, V., Antipov, D., & Gurevich, A. (2018). Versatile genome assembly evaluation with QUAST-LG. *Bioinformatics*, *34*(13). https://doi.org/10.1093/bioinformatics/bty266

Nevers, Y., Warwick Vesztrocy, A., Rossier, V., Train, C. M., Altenhoff, A., Dessimoz, C., & Glover, N. M. (2024). Quality assessment of gene repertoire annotations with OMArk. *Nature Biotechnology*. https://doi.org/10.1038/s41587-024-02147-w

Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T. C., Mendell, J. T., & Salzberg, S. L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature Biotechnology*, *33*(3). https://doi.org/10.1038/nbt.3122

Rhie, A., Walenz, B. P., Koren, S., & Phillippy, A. M. (2020). Merqury: Reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biology*, *21*(1). https://doi.org/10.1186/s13059-020-02134-9

Roach, M. J., Schmidt, S. A., & Borneman, A. R. (2018). Purge Haplotigs: Allelic contig reassignment for third-gen diploid genome assemblies. *BMC Bioinformatics*, *19*(1). https://doi.org/10.1186/s12859-018-2485-7

Salazar, O. R., Prasanna N, A., Cui, G., Bay, L. K., van Oppen, M. J. H., Webster, N. S., & Aranda, M. (2022). The coral Acropora loripes genome reveals an alternative pathway for cysteine biosynthesis in animals. *Science Advances*, *8*(38). https://doi.org/10.1126/sciadv.abq0304

Shinzato, C., Khalturin, K., Inoue, J., Zayasu, Y., Kanda, M., Kawamitsu, M., Yoshioka, Y., Yamashita, H., Suzuki, G., & Satoh, N. (2021). Eighteen Coral Genomes Reveal the Evolutionary Origin of Acropora Strategies to Accommodate Environmental Changes. *Molecular Biology and Evolution*, *38*(1), 16–30. https://doi.org/10.1093/molbev/msaa216

Sim, S. B., Corpuz, R. L., Simmonds, T. J., & Geib, S. M. (2022). HiFiAdapterFilt, a memory efficient read processing pipeline, prevents occurrence of adapter sequence in PacBio HiFi reads and their negative impacts on genome assembly. *BMC Genomics*, *23*(1). https://doi.org/10.1186/s12864-022-08375-1

Smit, A. F. A., Hubley, R., & Grenn, P. (2015). RepeatMasker Open-4.0. In *RepeatMasker Open-4.0.7*.

Smith, E. G., Hazzouri, K. M., Choi, J. Y., Delaney, P., Al-Kharafi, M., Howells, E. J., Aranda, M., & Burt, J. A. (2022). Signatures of selection underpinning rapid coral adaptation to the world's warmest reefs. *Science Advances*, *8*(2). https://doi.org/10.1126/sciadv.abl7287

Snelling, J., Dziedzic, K., Guermond, S., & Meyer, E. (2017). Integrating genomic resources for a threatened Caribbean coral (Orbicella faveolata) using a genetic linkage map developed from individual larval genotypes. *BioRxiv*, 183467. https://doi.org/10.1101/183467

Stiehler, F., Steinborn, M., Scholz, S., Dey, D., Weber, A. P. M., & Denton, A. K. (2021). Helixer: cross-species gene annotation of large eukaryotic genomes using deep learning. *Bioinformatics*, *36*(22–23), 5291–5298. https://doi.org/10.1093/bioinformatics/btaa1044

Su, W., Ou, S., Hufford, M. B., & Peterson, T. (2021). A Tutorial of EDTA: Extensive De Novo TE Annotator. In *Methods in Molecular Biology* (Vol. 2250). https://doi.org/10.1007/978-1-0716-1134-0_4

Törönen, P., & Holm, L. (2022). PANNZER—A practical tool for protein function prediction. *Protein Science*, *31*(1). https://doi.org/10.1002/pro.4193

Veron, J. E. N., Stafford-Smith, M. G., Turak, E., & DeVantier, L. M. (2016). *Corals of the World*. Version 0.01. https://www.coralsoftheworld.org/species_factsheets/species_factsheet_distribution/platygyra-daedalea/

Voolstra, C. R., Li, Y., Liew, Y. J., Baumgarten, S., Zoccola, D., Flot, J. F., Tambutté, S., Allemand, D., & Aranda, M. (2017). Comparative analysis of the genomes of Stylophora pistillata and Acropora digitifera provides evidence for extensive differences between species of corals. *Scientific Reports*, *7*(1). https://doi.org/10.1038/s41598-017-17484-x

Wang, S., Meyer, E., Mckay, J. K., & Matz, M. V. (2012). 2b-RAD: A simple and flexible method for genome-wide genotyping. *Nature Methods*, *9*(8). https://doi.org/10.1038/nmeth.2023

Yan, H., Bombarely, A., & Li, S. (2020). DeepTE: A computational method for de novo classification of transposons with convolutional neural network. *Bioinformatics*, *36*(15). https://doi.org/10.1093/bioinformatics/btaa519

Author Contributions

Michell, C. conducted laboratory work, performed bioinformatics analyses, and wrote the manuscript. Howells, E. contributed to funding acquisition, project design and manuscript revisions. Micke, P. assisted with laboratory work and bioinformatics analyses and revised the manuscript. Alguthmi, S. contributed to laboratory work and manuscript revisions. Hawley, C.-M. assisted with laboratory work and bioinformatics analyses. Meyer, E. contributed to the construction of the linkage map. Aranda, M. designed the project, contributed to writing, and revised the manuscript.

Competing Interests

The authors declare that they have no competing interests.