



HAL
open science

Chromosome-level genome assembly and annotation of the emblematic silver-lipped pearl oyster *Pinctada maxima* Jameson 1901

Laura Benestan, A. Cormier, T. Destanque, Océane Romatif, Jérémie Vidal-Dupiol, David Byron Jones, D. R. Jerry, Chin-Long Ky, Jérémy Le Luyer

► To cite this version:

Laura Benestan, A. Cormier, T. Destanque, Océane Romatif, Jérémie Vidal-Dupiol, et al.. Chromosome-level genome assembly and annotation of the emblematic silver-lipped pearl oyster *Pinctada maxima* Jameson 1901. *Scientific Data*, 2026, <10.1038/s41597-026-06905-0>. <hal-05613895>

HAL Id: hal-05613895

<https://hal.science/hal-05613895v1>

Submitted on 6 May 2026

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY-NC 4.0 - Attribution - Non-commercial use - International License

Chromosome-level genome assembly and annotation of the emblematic silver-lipped pearl oyster *Pinctada maxima* Jameson 1901

Received: 1 April 2025

Accepted: 12 February 2026

Cite this article as: Benestan, L., Cormier, A., Destanque, T. *et al.* Chromosome-level genome assembly and annotation of the emblematic silver-lipped pearl oyster *Pinctada maxima* Jameson 1901. *Sci Data* (2026). <https://doi.org/10.1038/s41597-026-06905-0>

L. Benestan, A. Cormier, T. Destanque, O. Romatif, J. Vidal-Dupiol, D. B. Jones, D. R. Jerry, C.-L. Ky & J. Le Luyer

We are providing an unedited version of this manuscript to give early access to its findings. Before final publication, the manuscript will undergo further editing. Please note there may be errors present which affect the content, and all legal disclaimers apply.

If this paper is publishing under a Transparent Peer Review model then Peer Review reports will publish with the final article.

Chromosome-level genome assembly and annotation of the emblematic silver-lipped pearl oyster *Pinctada maxima* Jameson 1901

Benestan L. ^{†,1,2}, Cormier A. ^{†,3}, Destanque T.¹, Romatif O.⁴, Vidal-Dupiol J.⁴, Jones D.B.⁵, Jerry D.R.^{5,6}, Ky C.-L.^{1,4} & Le Luyer J.^{1,2}

¹ Ifremer, IRD, Institut Louis-Malardé, Université de Polynésie française, EIO, F-98719 Taravao, Tahiti, Polynésie française, France

² Ifremer, PHYTNESS, Centre Bretagne - ZI de la Pointe du Diable - CS 10070 - 29280 Plouzané

³ Ifremer, IRSI, SeBiMER Service de Bioinformatique de l'Ifremer, F-29280 Plouzané, France

⁴ Ifremer, IHPE, Univ. Montpellier, CNRS, Univ. Perpignan Via Dominitia, Montpellier, France

⁵ ARC Research Hub for Supercharging Tropical Aquaculture through Genetic Solutions, James Cook University Townsville, Australia

⁶ Tropical Futures Institute, James Cook University, Singapore

[†] These authors contributed equally to this work

***Authors for correspondence:** laura.benestan@icloud.com

Abstract

The silver-lipped pearl oyster, *Pinctada maxima*, is an iconic bivalve farmed in tropical Australia and South-east Asia to produce high valued South Sea pearls. Here, we present a high-quality, chromosome-level reference genome for *P. maxima*. The final genome assembly has a total size of 1.27 Gb, with a scaffold N50 length of 65.69 Mb. Using high-throughput chromosome conformation capture (Hi-C) technology, we have anchored 80.99% of the assembly to 14 chromosomes. The BUSCO analysis revealed a high level of accuracy and completeness, identifying 87.9% (n = 4,658) of Mollusca core genes, which exceeds the average completeness of 79.2% reported for other mollusc genomes sequenced and annotated to date. The assembly includes 62.79% repeat elements (809 Mb), and gene annotation predicted 25,752 protein-coding genes. This genomic resource offers a powerful tool for unraveling the molecular mechanisms underlying pearl formation and facilitating continued research into the evolutionary biology of pearl oysters worldwide.

Background & Summary

Molluscs are a dominant phylum in the marine environment, comprising 23% of all known marine species. Within this phylum, the class Bivalvia is the second largest group after Gastropoda, encompassing more than 8,000 species that inhabit a diverse range of environments, from intertidal zones to deep-sea habitats. Bivalvia play a pivotal role in tropical aquaculture, notably with the genus *Pinctada* that supports cultured pearl production. These bivalves are widely distributed across shallow and deep waters of the Indo-Pacific and western Atlantic regions ¹. Renowned as the “queen of jewels”, pearls represent an essential economical resource in many tropical regions, often serving as one of the few available sources of income ². Consequently, significant efforts have been dedicated to ensuring pearl production by deciphering the genetic and environmental determinants of pearl quality ³⁻⁵.

The silver-lipped pearl oyster, *Pinctada maxima*, native to the Indo-Australian archipelagos, produces white, silver and gold South Sea pearls, the largest and most valuable cultured pearl among pearl-producing species ⁶. This species alone accounts for over 50% of the global pearl market ⁷. Large scale, genetics-assisted selection programs have been developed to enhance pearl quality in *P. maxima* ⁸⁻¹⁰. However, our understanding of the genes influencing pearl quality remains limited, partly due to the scarcity of genomic resources. Furthermore, wild and farmed stock of *P. maxima* in Australia have been massively impacted by episodic and large-scale mortality events due to a yet unidentified causative agent. These mortality events, termed juvenile pearl oyster mortality syndrome (JOMS and/or oyster oedema disease OOD) have significantly reduced the economic value of the sector and led to strong uncertainty for the future of the industry ¹¹. As a response, large scale genetic assisted selection programs have emerged toward developing more resistant lines ⁹.

Currently, three pearl oyster genomes are available: *P. imbricata*¹², *P. fucata*¹³, and recently *P. radiata*¹⁴. Here, we complement these available resources with a high-quality chromosome-level genome of *P. maxima*, achieved through the integration of long-read sequencing (PacBio), short-read sequencing (Illumina) and high-resolution chromosome conformation capture (Hi-C). The final assembly yielded a genome size of 1.27 Gb, anchored to 14 chromosomes, with a scaffold N50 length of 65.69 Mb. We have annotated protein-coding genes and conducted a comparative synteny analysis between *P. maxima* and *P. fucata* species. This genomic resource will advance continued research into the ecology and evolution of *Pinctada*, while also supporting efforts to improve oyster health and optimize pearl production¹⁵.

Methods

Libraries preparation and sequencing

A single *Pinctada maxima* individual from the Roko Pearls farm (Roko Island Cape York, Queensland, Australia) was collected. The individual was air-shipped alive to James Cook University (July 2019) for dissection. To limit exogenous contamination, 800 mg of tissue was cut from the interior adductor muscle and diced on a petri dish. Tissue was lysed overnight at 50 °C in a blend of 9.5 ml of Qiagen buffer, 19 µL RNase A buffer and 1ml proteinase K. Lysates were centrifuged at 500g for 10 min at 4 °C. Lysates were then loaded onto Qiagen Genomic-tip 100/G DNA extraction columns and washed with the QC buffer as specified in the manufacturer's protocol. Samples were eluted in 5 ml of QF buffer and precipitated by adding 3.5 ml of isopropanol. Samples were then centrifuged at 8,000 g for 30 min at 4 °C. Supernatant was then removed and pellets were washed with 70% ethanol at -20 °C and centrifuged again at 8,000 g for 30 min at 4 °C. Pellets were then air-dried for 10 minutes on ice. Quality and quantity were assessed using a Nanodrop spectrophotometer and a 0.8% agarose gel. Extracted DNA was kept

at 4 °C until it was received by the sequencing facilities. Long-read PacBio libraries were prepared using a PacBio sheared gDNA kit and sequenced on eight runs of PacBio Sequel II at the Genome Quebec sequencing platform (Canada). Short-read Illumina libraries, Chromium 10X and DNA-seq, were prepared using TrueSeq DNA nano kit and sequenced on a full chip of Illumina NovaSeq 6000 150bp paired-end (PE) at the Australian Genome Research Facilities (Melbourne, Australia). Overall, library sequencing resulted in a total of 6M (mean size 7.5 Kb; ranges from 50 to 114 Kb; 45X coverage) PacBio reads and 431M Illumina reads (Table 1). In parallel, gill tissue, from the same individual, was sent on dry ice to the NextGen platform (Plateforme Génomique IHPE Ifremer facilities Montpellier, France) for Dovetail Hi-C library preparation. Libraries were prepared using the Omni-C kit (Dovetail Genomic, DNase enzyme) and were sequenced in two runs on the Illumina 6000 NovaSeq (1/3 full chip).

Genome properties

Nanoplot v1.32.1 and FastQC v0.11.9 were used to keep only high-quality long (PacBio) and short reads (Hi-C, DNA-seq and 10X), respectively. Bias in GC content of short read data was investigated using KAT v2.4.2 with the function *gcp* that creates a matrix of the number of k-mers found given a GC count and a K-mer count ¹⁶. Dovetail Hi-C libraries were trimmed using Cutadapt v4.1 ¹⁷ and seqkit v2.3.1 was applied to compute genome quality metrics (Table 1). Then, k-mers were counted using KMC v3.1.2rc1 and genome properties were investigated using GenomeScope v2.0 to get a picture of the genome size and heterozygosity profile. To optimize the assembly, we assessed the performance of three recommended *de novo* assemblers, namely Flye ¹⁸, Canu ¹⁹ and wtdbg2 ²⁰. The best genome assembly for Pacbio reads was obtained with Flye v2.9 using the *--no-alt-contigs* option. Contigs of the Flye assembly were polished for base errors and indels with 10X reads using Hapog v1.3.2 to ensure accuracy ²¹. Haplotigs and contig overlaps

were removed using `purge_dups v1.2.5`²². The final genome assembly has a size of 1.27 Gb with a 39x coverage, which is comparable in size (1.15 Gb) to the *Pinctada fucata* genome previously published¹³. The average length of contigs was 106.9 Kb (ranging from 61 to 145.6Kb) with a contig N50 size of 209 Kb. The GC content was around 36% (Table 2), an estimate close to what has been reported for *P. fucata* (GC = 34%).

Chromosome-level assembly

To generate a chromosome-level assembly, Hi-C proximity ligation data was used to order and orient the contigs along chromosomes. The scaffolding process was carried out by Dovetail Genomics (Santa Cruz, CA, USA) using the Hi-C library reads to connect and order the input set of contigs. First, paired fastq from Hi-C sequencing data were mapped using BWA v1.1.4 and ligation events were recorded²³. The pair file was then sorted, and PCR duplicates were filtered to produce the final mapping file. A total of 44.5% of reads were retained after applying the filters (i.e. quality, ligation and duplicates). Scaffolding was performed using YAHS v1.1²⁴ and then manually cleaned using Juicebox v2.0²⁵. The final chromosome-level genome assembly of the silver-lipped pearl oyster contains the 14 chromosome groups with 19.3% of the sequences remaining as unplaced scaffolds. Chromosomes have variable lengths ranging from 33.9 (Chr14) to 128.4 Mb (Chr1; Fig. 1).

Data records

The raw sequencing data and genome assembly of *Pinctada maxima* have been deposited in the ENA under accession number PRJEB92378 (<https://identifiers.org/ena.embl:PRJEB92378>).³⁶

Technical validation

The final assembly was compared to a conserved set of 5,295 Benchmarking Universal Single-Copy Orthologs (BUSCOs) from the *mollusca* ODB10 dataset to assess for completeness (Fig. 2). A total of 4,658 genes out of the curated list of 5,295 highly conserved genes (87.9%) were found in the assembly, of which 4,567 were single copy (86.2%). We reported only 91 duplicated (1.7%), 156 fragmented (2.9%) and 481 missing genes (9.1%), indicating high quality of the genome assembly (Table 3). Our BUSCO completeness of 87.9% reported here for *Pinctada maxima* is equivalent to *Pinctada fucata* (89.3%; Table S1; Fig. 2), as closely related species tend to share conserved single copy orthologs. In comparison, *P. imbricata* was one of the 10 least complete genomes with a weak completeness of 76.2% and then chromosome synteny was not assessed as the difference in the quality of genome assembly could have affected the results observed²⁶.

Evolution of bivalves

We established a phylogeny across the 32 bivalve species represented in the mollusc_odb10 lineage dataset (Table S1). We used IQ-TREE v1.6.12 (-b 1000 -altr 100) to build a fast and accurate phylogenetic tree based on strict orthologous genes. The tree was visualized with FigTree v1.4.4 setting up the time scale option (Fig. 3)²⁷. The reference-calibrated time point for the divergence between Cephalopoda and Bivalvia was set to 543 MYA [480.0 - 559.4 MYA] based on TimeTree database (<http://timetree.org/>)²⁸.

Genome annotation

Prediction of protein coding genes were completed using BRAKER v.3.0.3²⁹ based on the protein extracted from the Uniprot/Swissprot database³⁰, *Pinctada fucata* and *Crassostera gigas* genomes, available in Marine genomics (<https://marinegenomics.oist.jp>) and EnsemblMetazoa (<https://metazoa.ensembl.org>), respectively. A total of 25,752 genes and 32,338 mRNA were

found; genes had mean length of 16.91 Kb (ranging from 78 Kb to 385 Kb; Table 2). The mean number of exons, introns and CDS per mRNA was 8.2, 7.2 and 1.0 CDS per gene. Number of genes per chromosome varies from 631 (Chr14) to 2,790 (Chr02; Fig. 4). Functional annotation of the genome was performed using eggNOG-mapper v2³¹ for orthology-based inference, and BLASTp searches against the UniProt-SwissProt database³⁰ for high-confidence protein annotations (threshold e-value < 10e-5).

Repeat elements

We performed *de novo* identification of repeat sequences using RepeatModeler2 v2.0.1³². Repetitive regions represent a major component of the *P. maxima* genome with a total of 62.79% repetitive elements comprising 45.97% of unclassified elements. Notably, this percent of repetitive elements in *P. maxima* is higher than in *P. fucata* (53.30%). Similarly to *P. fucata*, the most abundant classes of transposable elements for *P. maxima* are LINES followed by the LTR elements and other elements (Table 2).

Chromosome synteny and orthologs detection

Gene orthology was inferred using Orthofinder v2.5.2³³ across the genomes of *P. maxima* and *P. fucata*. A total of 16,604 orthogroups (22,306 genes) were shared between *P. maxima* and *P. fucata* while 1,257 genes were specific to *P. maxima*. We then investigated interspecific chromosomal evolution using Genespace v2.4³⁴. The Chr10 of *P. maxima* shared the strongest synteny (98.0%) with Chr5a of *P. fucata* and the maximum syntenic block length of 37 Mb was reported between the Chr2 of *P. maxima* and the Chr1a of *P. fucata*. All the other chromosomes matched their corresponding chromosomes with moderate synteny (on average 78.1%), suggesting frequent interspecific chromosome rearrangement in *Pinctada* genus. Chr12 and Chr8 of the *P. maxima*

assembly lacked homology with only 26.1 and 59.5% of syntenic regions with Chr9a and Chr7a, respectively, of the *P. fucata* assembly (Fig. 5). This mismatch could result from significant divergent time between the two species as *P. fucata* belongs to the “smaller toothed” oyster clade while *P. maxima* belong to the larger-toothless clade that was estimated to diverge at 13.68 [7.33, 20.09] Mya³⁵. A more detailed study of the divergence at the scale of the *Pinctada* genus will be necessary to shed light on the different eco-evolutionary processes shaping the genome and generating the particularities specific to each species in terms of biomineralization.

Code availability

The scripts and the dataset can also be found on Zenodo (<https://doi.org/10.5281/zenodo.17379201>). All analysis scripts are accessible at <https://bioinfo.gitlab-pages.ifremer.fr/bioanalysis/internal/pmaxima>.

Data Availability

The raw sequencing data and genome assembly of *Pinctada maxima* have been deposited in the ENA under accession number PRJEB92378 (<https://identifiers.org/ena.embl:PRJEB92378>).³⁶

Funding

This project called *Genomax* was funded by an Inter-center Ifremer grant awarded to CLK and DBJ.

Author contributions

CLK, JVD and JLL conceived the project. DBJ and DRJ sourced the *Pinctada maxima* individual and conducted dissection. OR and JVD conducted Hi-C libraries preparation and AC performed HiC scaffolding, respectively. JLL annotated the genome. AC conducted the assembly and the

structural annotation. LB and AC performed the statistical analyses and wrote the draft of the paper.

Competing interests

The authors declare that they have no known competing financial or non-financial interests that could have appeared to influence the work reported in this paper.

References

1. van der Schatte Olivier, A. *et al.* A global review of the ecosystem services provided by bivalve aquaculture. *Reviews in Aquaculture*, **12**, 3-25 (2020).
2. Zhu, C., Southgate, P. C. & Li, T. Production of pearls. *Goods and Services of Marine Bivalves*, 73-93 (2019).
3. Le Luyer, J. *et al.* Whole transcriptome sequencing and biomineralization gene architecture associated with cultured pearl quality traits in the pearl oyster, *Pinctada margaritifera*. *BMC Genomics* **20**, 111 (2019).
4. Jerry, D. R. *et al.* Donor-oyster derived heritability estimates and the effect of genotype×environment interaction on the production of pearl quality traits in the silver-lip pearl oyster, *Pinctada maxima*. *Aquaculture* **338**, 66-71 (2012).
5. Stenger, P. L. *et al.* Environmentally driven color variation in the pearl oyster *Pinctada margaritifera* var. *cumingii* (Linnaeus, 1758) is associated with differential methylation of CpGs in pigment- and biomineralization-related genes. *Front Genet* **12**, 630290 (2021).
6. Taylor, J. & Strack, E. *Pearl Production*. The pearl oyster (pp. 273-302) (2008).
7. Tisdell, C. & Poirine, B. Economics of pearl farming. in *The pearl oyster*, **473**, 495 (2008).

8. Jones, D. B., Jerry, D. R., Khatkar, M. S., Raadsma, H. W. & Zenger, K. R. A high-density SNP genetic linkage map for the silver-lipped pearl oyster, *Pinctada maxima*: A valuable resource for gene localisation and marker-assisted selection. *BMC Genomics* **14**, 810 (2013).
9. Zenger, K. R. *et al.* Genomic selection in aquaculture: Application, limitations and opportunities with special reference to marine shrimp and pearl oysters. *Frontiers in Genetics*, **9**, 693 (2019).
10. Jones, D. B. *et al.* Determining genetic contributions to host oyster shell growth: Quantitative trait loci and genetic association analysis for the silver-lipped pearl oyster, *Pinctada maxima*. *Aquaculture* **434**, 376-375 (2014).
11. Jones, J. B., Crockford, M., Creeper, J. & Stephens, F. Histopathology of oedema in pearl oysters *Pinctada maxima*. *Dis Aquat Organ* **91**, 67-73 (2010).
12. Du, X. *et al.* The pearl oyster *Pinctada fucata martensii* genome and multi-omic analyses provide insights into biomineralization. *Gigascience* **6**, gix059 (2017).
13. Takeuchi, T. *et al.* A high-quality, haplotype-phased genome reconstruction reveals unexpected haplotype diversity in a pearl oyster. *DNA Research* **29**, dsac035 (2022).
14. Vasileiadou, K. *et al.* ERGA-BGE genome of *Pinctada radiata* (Leach, 1814): one of the first Lessepsian migrants. bioXiv (2025) doi: 10.1101/2025.01.17.632165.
15. Rexroad, C. *et al.* Genome to phenome: Improving animal health, production, and well-being - A new USDA blueprint for animal genome research 2018-2027. *Frontiers in Genetics* **10**, 327 (2019).

16. Mapleson, D., Accinelli, G. G., Kettleborough, G., Wright, J. & Clavijo, B. J. KAT: A K-mer analysis toolkit to quality control NGS datasets and genome assemblies. *Bioinformatics* **33**, 574-576 (2017).
17. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J* **17**, 10-12 (2011).
18. Kolmogorov, M., Yuan, J., Lin, Y. & Pevzner, P. A. Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol* **37**, 540-546 (2019).
19. Sergey *et al.* Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res* **27**, 722-736 (2017).
20. Ruan, J. & Li, H. Fast and accurate long-read assembly with wtdbg2. *Nat Methods* **17**, 155- 158 (2020).
21. Aury, J. M. & Istace, B. Hapo-G, haplotype-Aware polishing of genome assemblies with accurate reads. *NAR Genom Bioinform* **3**, Iqab034 (2021).
22. Guan, D. *et al.* Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics* **36**, 2896-2898 (2020).
23. Li, H. Compares BWA to other long read aligners like CUSHAW2] Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv* (2013) <https://doi.org/10.48550/arXiv.1303.3997>
24. Zhou, C., McCarthy, S. A. & Durbin, R. YaHS: yet another Hi-C scaffolding tool. *Bioinformatics* **39**, btac808 (2023).

25. Durand, N. C. *et al.* Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cell Syst* **3**, 95-98 (2016).
26. Liu, D., Hunt, M. & Tsai, I. J. Inferring synteny between genome assemblies: A systematic evaluation. *BMC Bioinformatics* **19**, 26 (2018).
27. Cummings, M. P. FigTree . in *Dictionary of Bioinformatics and Computational Biology* (2004). doi:10.1002/9780471650126.dob0904.
28. Hedges, S. B., Dudley, J. & Kumar, S. TimeTree: A public knowledge-base of divergence times among organisms. *Bioinformatics* **22**, 2971-2972 (2006).
29. Gabriel, L. *et al.* BRAKER3: Fully automated genome annotation using RNA-Seq and protein evidence with GeneMark-ETP, AUGUSTUS and TSEBRA. *Genome research*, **34**, 769-777 (2024).
30. Bateman, A. *et al.* UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Res* **51**, D523-D531 (2023).
31. Cantalapiedra, C. P., Hernandez-Plaza, A., Letunic, I., Bork, P. & Huerta-Cepas, J. eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Mol Biol Evol* **38**, 5825-5829 (2021).
32. Flynn, J. M. *et al.* RepeatModeler2 for automated genomic discovery of transposable element families. *Proc Natl Acad Sci U S A* **117**, 9451-9457 (2020).
33. Emms, D. M. & Kelly, S. OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biol* **20**, 238 (2019).

34. Lovell, J. T. *et al.* GENESPACE tracks regions of interest and gene copy number variation across multiple genomes. *Elife* **11**, e75526 (2022).
35. Cunha, R. L., Blanc, F., Bonhomme, F. & Arnaud-Haond, S. Evolutionary Patterns in Pearl Oysters of the Genus *Pinctada* (Bivalvia: Pteriidae). *Marine Biotechnology* **13**, 181-192 (2011).
36. ENA European Nucleotide Archive <https://identifiers.org/ena.embl:PRJEB92378> (2024).

ARTICLE IN PRESS

Figures legends.

Figure 1. The Hi-C contact map of the assembled *Pinctada maxima* genome. The blue squares represented the 14 chromosomes of the silver-lipped pearl oyster. Chromosomes are ordered per length from longer (left) to smaller (right) lengths.

Figure 2. Genome assembly quality a mollusc species.

Figure 3. Phylogeny of molluscs species

Figure 4. Circos plot of *Pinctada maxima* genome assembly. From outer to inner layers were (a) chromosomes, (b) repetitive elements, (c) gene densities and (d) GC contents respectively.

Figure 5. Chromosomal synteny between *Pinctada maxima* and *Pinctada fucata*

Table 1. Statistics of the *Pinctada maxima* sequencing data used for genome assembly.

Libraries	Average length (bp)	Sequenced reads	Sequenced base pairs (bp)	Sequencing coverage
Hi-C	150	1,333,587,248	401,231,843,781	315X
Illumina	150	431,409,994	129,422,998,200	101X
PacBio	7,584	6,163,824	46,749,537,286	37X
10X	151	439,508,128	132,731,454,656	104X

Table 2. Genome assembly and annotation statistics of *Pinctada maxima*.**Genome assembly**

GC content (%)	36.16
Total size (bp)	1,272,675,302
Contigs	
No.	5,140
N50 length (bp)	65,683,950
Maximum scaffold length (bp)	128,397,519
Minimum scaffold length (bp)	61
Complete BUSCO score - Mollusc	4,658

Genome annotationNo. transposable elements

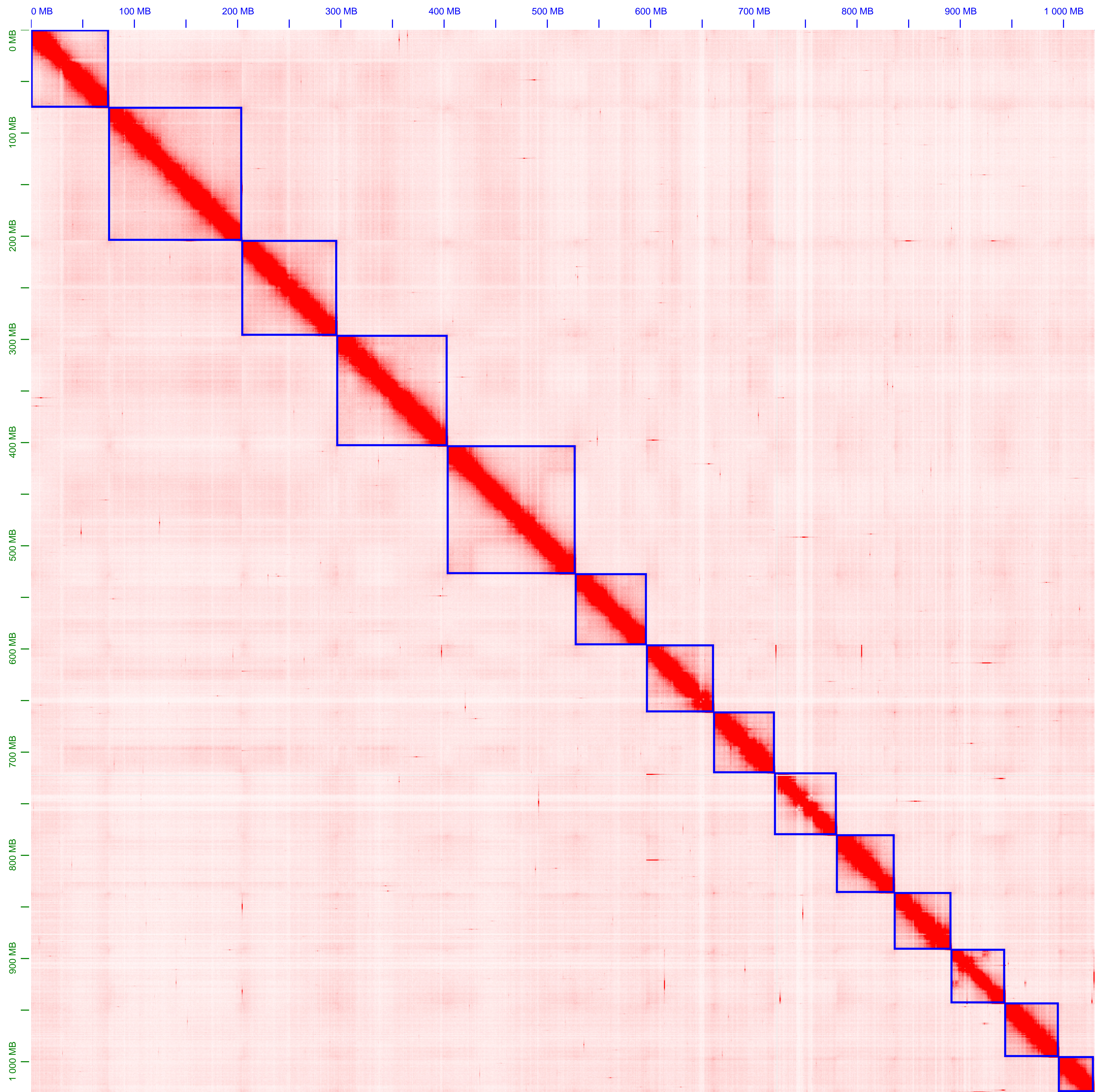
SINES	790
LTR	249,071
LINEs	258,626
DNA elements	168,826
Small RNA	12,531
Unclassified	2,039,206

Protein coding genes

No.	25,752
Spliced transcript	32,338
Exons	265,936

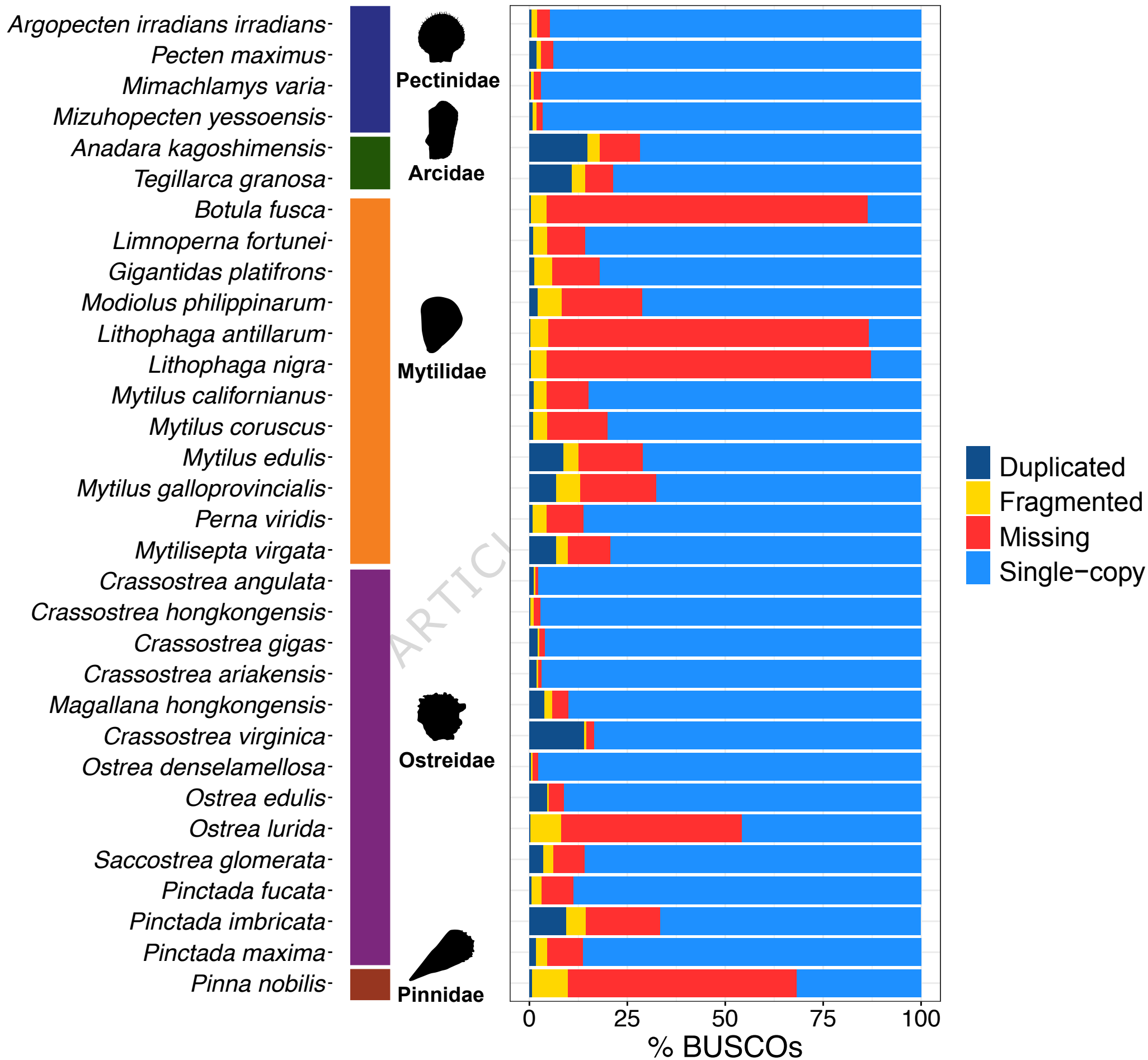
Table 3. Summary of BUSCO completeness scores for all species

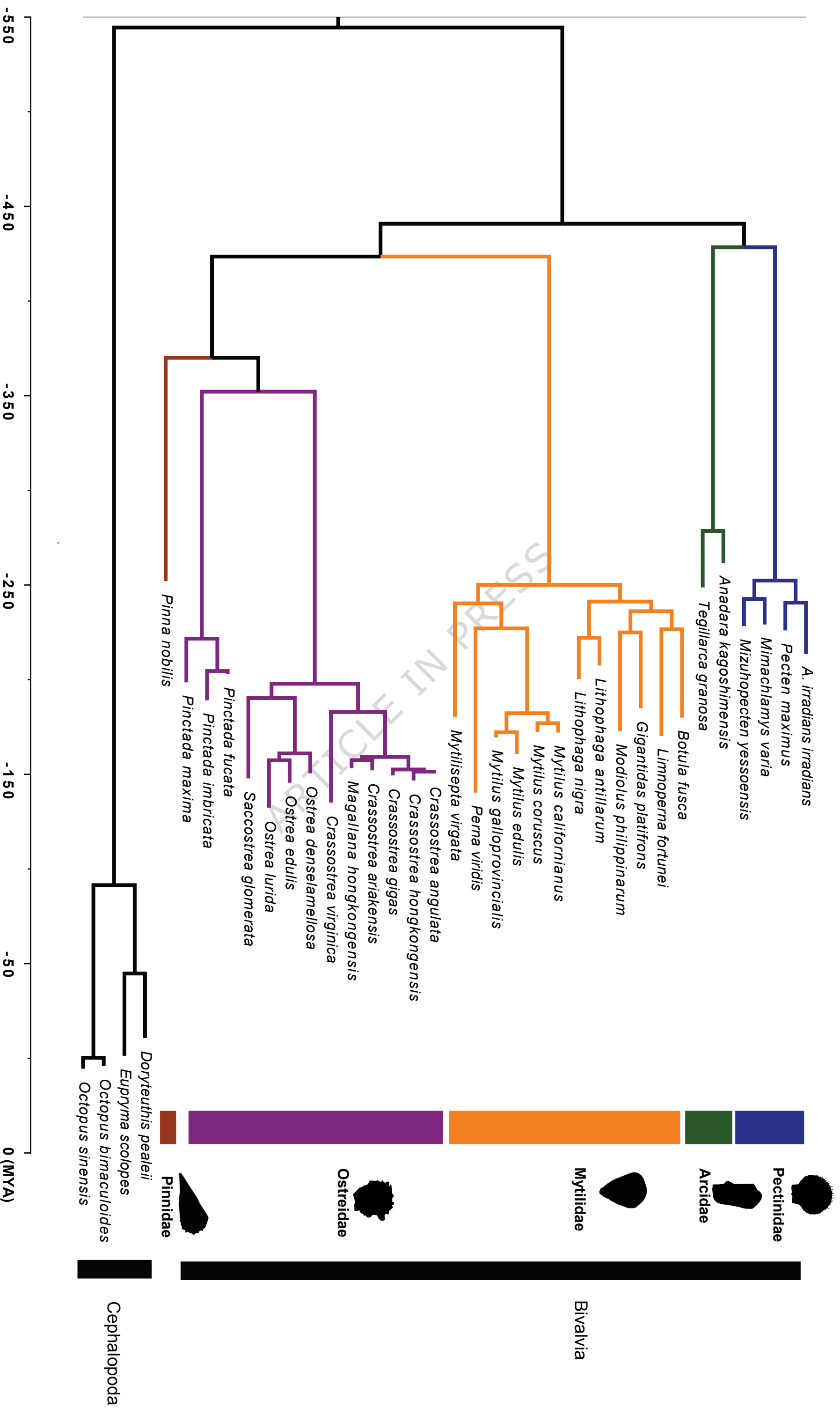
Code	Name	Complete	Single-copy	Duplicated	Fragmented	Missing
AKAG	<i>Anadara kagoshimensis</i>	86.7	71.8	14.9	3.1	10.2
AIRR	<i>Argopecten irradians irradians</i>	95.4	94.8	0.6	1.4	3.2
BUSC	<i>Botula fusca</i>	14.0	13.6	0.4	4.1	81.9
CANG	<i>Crassostrea angulata</i>	98.9	97.8	1.1	0.4	0.7
CARI	<i>Crassostrea ariakensis</i>	98.8	96.9	1.9	0.4	0.8
CGIG	<i>Magallana gigas</i>	98.2	96.1	2.1	0.5	1.3
CHON	<i>Crassostrea hongkongensis</i>	97.5	97.2	0.3	0.9	1.6
CVIR	<i>Crassostrea virginica</i>	97.4	83.5	13.9	0.7	1.9
GPLA	<i>Gigantidas platifrons</i>	83.4	82.1	1.3	4.6	12.0
LFOR	<i>Limnoperna fortunei</i>	86.8	85.8	1.0	3.6	9.6
LANT	<i>Lithophaga antillarum</i>	13.7	13.4	0.3	4.6	81.7
LNIG	<i>Lithophaga nigra</i>	13.2	12.8	0.4	4.0	82.8
MHON	<i>Magallana hongkongensis</i>	94.0	90.1	3.9	1.9	4.1
MVAR	<i>Mimachlamys varia</i>	97.6	97.1	0.5	0.6	1.8
MYES	<i>Mizuhopecten yessoensis</i>	97.6	96.7	0.9	0.9	1.5
MPHI	<i>Modiolus philippinarum</i>	73.3	71.2	2.1	6.2	20.5
MVIR	<i>Mytilisepta virgata</i>	86.1	79.3	6.8	3.1	10.8
MCAL	<i>Mytilus californianus</i>	86.0	84.9	1.1	3.3	10.7
MCOR	<i>Mytilus coruscus</i>	81.0	80.1	0.9	3.7	15.3
MEDU	<i>Mytilus edulis</i>	79.8	71.1	8.7	3.9	16.3
MGAL	<i>Mytilus galloprovincialis</i>	74.6	67.7	6.9	6.1	19.3
ODEN	<i>Ostrea denselamellosa</i>	98.2	97.8	0.4	0.5	1.3
OEDU	<i>Ostrea edulis</i>	95.7	91.2	4.5	0.5	3.8
OLUR	<i>Ostrea lurida</i>	46.1	45.8	0.3	7.9	46.0
PMAU	<i>Pecten maximus</i>	95.8	93.9	1.9	1.1	3.1
PVIR	<i>Perna viridis</i>	87.0	86.2	0.8	3.7	9.3
PFUC	<i>Pinctada fucata</i>	89.4	88.8	0.6	2.5	8.1
PIMB	<i>Pinctada imbricata</i>	76.2	66.7	9.5	4.9	18.9
PMAA	<i>Pinctada maxima</i>	87.9	86.2	1.7	2.9	9.1
PNOB	<i>Pinna nobilis</i>	32.5	31.8	0.7	9.2	58.3
SGLO	<i>Saccostrea glomerata</i>	89.4	85.9	3.5	2.7	7.9
TGRA	<i>Tegillarca granosa</i>	89.5	78.6	10.9	3.4	7.1
PRAD	<i>Pinctada radiata</i>	95.2	94.2	1	0.6	4.2

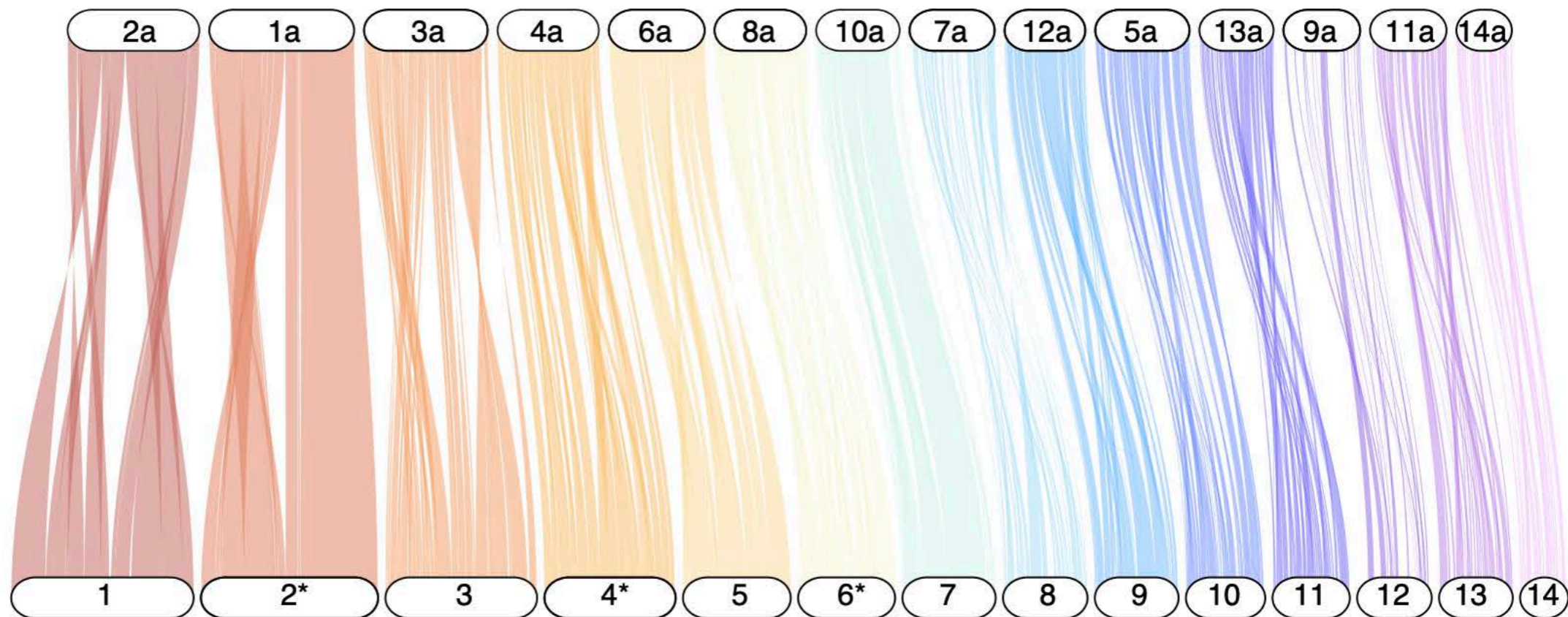


assembly

assembly





Pinctada fucata*Pinctada maxima*

Chromosomes scaled by physical position

