

# Chromosome-level genomes of hard clams *Meretrix lamarckii* (Deshayes, 1853) and *Meretrix meretrix* (Linnaeus, 1758)

Received: 18 September 2025

Accepted: 19 March 2026

Cite this article as: Law, S.T., Nong, W., Au, M.F. *et al.* Chromosome-level genomes of hard clams *Meretrix lamarckii* (Deshayes, 1853) and *Meretrix meretrix* (Linnaeus, 1758). *Sci Data* (2026). <https://doi.org/10.1038/s41597-026-07119-0>

Sean Tsz Sum Law, Wenyan Nong, Ming Fung Franco Au, Leni Hiu Tung Cheung, Cheryl Wood Yee Shum, Shing Yip Lee, Siu Gin Cheung & Jerome Ho Lam Hui

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 genomes of hard clams *Meretrix lamarckii* (Deshayes, 1853) and *Meretrix meretrix* (Linnaeus, 1758)**

Sean Tsz Sum Law<sup>1,^</sup>, Wenyan Nong<sup>1,^</sup>, Ming Fung Franco Au<sup>1</sup>, Leni Hiu Tung Cheung<sup>1</sup>, Cheryl Wood Yee Shum<sup>1</sup>, Shing Yip Lee<sup>2,\*</sup>, Siu Gin Cheung<sup>3,\*</sup>, Jerome Ho Lam Hui<sup>1,\*</sup>

1. School of Life Sciences, Simon F.S. Li Marine Science Laboratory, Institute of Environment, Energy and Sustainability, State Key Laboratory of Agrobiotechnology, The Chinese University of Hong Kong, Hong Kong SAR, China

2. Simon F.S. Li Marine Science Laboratory, School of Life Sciences, The Chinese University of Hong Kong, Hong Kong SAR, China; Australian Rivers Institute, Griffith University Gold Coast campus, Southport, Qld 4222, Australia

3. Department of Chemistry, State Key Laboratory of Marine Pollution, City University of Hong Kong, Hong Kong SAR, China

<sup>^</sup> = contributed equally

<sup>\*</sup> = correspondence (joesylee@cuhk.edu.hk; bhsgche@cityu.edu.hk; jeromehui@cuhk.edu.hk)

Journal: Scientific Data

## Abstract

The genus *Meretrix* contains marine bivalve molluscs commonly known as the Venus clams or hard clams, which can be found in the estuarine and marine habitats in Asia. Given their edibility, they have been exploited in clam digging activities and also been farmed in some places. Here, we provide two new high-quality genomes of species *M. meretrix* and *M. lamarckii*. Utilising a combination of PacBio HiFi and Omni-C sequencing technologies, genome assemblies of *M. meretrix* and *M. lamarckii* are obtained with sizes 835.1 Mb (scaffold N50 = 46Mb) and 890.5 Mb (scaffold N50 = 46Mb), respectively. More than 99% of sequences were anchored to 19 pseudochromosomes, and high completeness was also obtained estimated by BUSCO scores (~99.5%, mollusca\_odb12). The two new genomic resources provided in this study will be useful for further understanding biology, ecology, and evolution of edible clams.

## Background & Summary

The family Veneridae contains more than 700 species of estuarine and marine bivalves worldwide. Species in the genus *Meretrix* are widely distributed in the coastal waters of the Indo-Pacific region<sup>1</sup>, and some of them are economically significant in aquaculture industry in the East and Southeast Asia<sup>1-3</sup>. Natural populations of *Meretrix* spp. are also susceptible to anthropogenic activities such as over-harvesting at industrial scale, recreational clam-digging and habitat degradation<sup>2,4,5</sup>.

Among the 17 described *Meretrix* species<sup>6</sup>, some of them have high morphological similarities, and molecular analyses are required for identification of distinct populations in different geographical locations<sup>3,7,8</sup>. For instance, a distinct clade (and later become a new species) of *M. taiwanica* has been detected from COI barcoding sequences<sup>7,9</sup>, and utilization of microsatellite DNA markers have revealed subpopulations of *M. meretrix*<sup>3</sup>. Genomic resources of *Meretrix* are thus important to understand and reveal the diversity between species and populations. As of to date, a total of four *Meretrix* genomes are available including *M. petechialis*<sup>10</sup>, *M. lamarckii*<sup>7</sup>, and two distinct *Meretrix* sp. MF1 and *Meretrix* sp. MT1<sup>7</sup>. Here, we present two new chromosomal-level genomes of *M. meretrix* and *M. lamarckii* for the further investigation of genetic diversity and of evolutionary of *Meretrix* clams.

## Methods

### *Sample collection, DNA extraction and Species identification*

Clams were collected in Tung Wan of Lantau Island in Hong Kong on 13 August 2022. Genomic DNA was extracted from the adductor muscle using Purelink Genomic Mini Kit (Invitrogen) following the manufacturer's instructions. The quality and quantity of the isolated DNA was assessed by gel electrophoresis and NanoDrop spectrophotometer (Thermo Scientific). The species identity of the specimen was determined by the amplification of mitochondrial DNA cytochrome oxidase subunit I (COI) gene in polymerase chain reaction (PCR) using the universal COI primers LCO1490 and HCO2198<sup>11</sup>. The PCR was performed on a T100<sup>TM</sup> thermocycler (Bio-Rad) with the following parameters: 95 °C for 3 minutes, followed by 39 cycles of 30 seconds at 95 °C, 30 seconds at 55 °C, and 45 seconds at 72 °C, and a final extension step at 72 °C for 5 minutes. The fragment size of amplified product was inspected on gel electrophoresis and was sent to BGI Hong Kong for Sanger sequencing. The chromatograms of the sequences were checked on Chromas v2.6.6 (<https://technelysium.com.au/wp/chromas/>) for quality control. Sequences were searched for homology in NCBI nr database using BLASTn. Subsequently, the sequences were aligned with reference COI sequences from previous studies<sup>12-14</sup> using MAFFT<sup>15</sup>. A maximum likelihood phylogenetic tree was constructed using FastTree<sup>16</sup>.

### *High molecular weight DNA extraction and long read sequencing*

~300 mg of gill and adductor muscle tissues from *M. meretrix* and *M. lamarckii* respectively were ground into powder under liquid nitrogen. High molecular weight DNA was then isolated from the powder using the NucleoBond HMW DNA kit (Macherey-Nagel) following the manufacturer's instructions. The resulting DNA was subjected to quality and quantity measurement using NanoDrop spectrophotometer (Thermo Scientific), Qubit® Fluorometer and pulse-field gel electrophoresis. The qualified DNA samples were sent to Macrogen for PacBio high-fidelity (HiFi) sequencing on the PacBio Revo system. 62.87 Gb and 57.89 Gb of HiFi reads were generated for *M. meretrix* and *M. lamarckii*, respectively (Table 1).

### *Omni-C library preparation and sequencing*

Omni-C libraries of *M. meretrix* and *M. lamarckii* were constructed from ~100 mg of gill tissues using the Dovetail® Omni-C® Library Preparation Kit (Cantata Bio) following the manufacturer's instructions. The fragment size distribution and concentration of the resulting libraries were assessed with Agilent D5000 ScreenTape Assay and Qubit® Fluorometer. The qualified libraries were sent to Novogene for sequencing on an Illumina PE150 platform to obtain 3.55 Gb and 3.36 Gb Omni-C data for *M. meretrix* and *M. lamarckii*, respectively (Table 1).

### *Genome assemblies*

*De novo* genome assemblies of *M. meretrix* and *M. lamarckii* were performed with Hifiasm<sup>17</sup>. Haplotypic duplications were further removed using *purge\_dups*<sup>18</sup>. The contigs of the draft assemblies were searched against the NT database using BLASTn to filter out potential contaminations with BlobTools (v1.1.1)<sup>19</sup>. Subsequently, the assemblies were scaffolded with the Omni-C data using YaHS<sup>20</sup> and manually checked using Juicebox (v1.11.08)<sup>21</sup>. Finally, the assembled genome sequences were assessed with NCBI Foreign Contamination Screen (FCS)<sup>22</sup> to remove contaminant sequences. The final assemblies of *M. meretrix* and *M. lamarckii* were 853.1 Mb and 890.5 Mb in size, and 99.04% and 99.43% of the sequences were anchored to 19 chromosomal scaffolds, respectively (Figure 1A-B; Table 2-3). Both *Meretrix* genomes were of high sequence continuity with scaffold N50 of ~46.2 Mb (Table 2).

### *Genome annotation*

Gene model prediction was performed using Braker (v3.0.8)<sup>23</sup> with default parameters. Protein hints with 389,399 Mollusca reference protein sequences downloaded from NCBI were used to train GeneMark-EP+ without RNA-Seq data, whereas genes with high extrinsic evidence support were subsequently used to train AUGUSTUS<sup>24</sup>. A total of 28,699 and 33,346 gene models were predicted from the *M. meretrix* and *M. lamarckii* genomes, respectively (Table 2).

### *Repeat annotation*

Transposable elements (TEs) of the two genomes were annotated using Earl Grey (v1.2)<sup>25</sup> with the option “-r eukarya” specified for the initial mask of known TEs within the Earl Grey automated pipeline. In total, 357.2 Mb and 397.5 Mb of repetitive elements were identified in the genomes of *M. meretrix* and *M. lamarckii*, which account for 42.77% and 44.64% of respective genome (Figure 1C; Table 4).

### *Phylogenomic tree analysis*

To elucidate the phylogenetic relationships between the two genomes generated in this study and other *Meretrix* genomes, gene orthology analysis was carried out from the longest transcript of proteome together with 8 other bivalve genomes using OrthoFinder v. 2.5.2<sup>26</sup>. These genomes include *M. petechialis* (GCA\_046203225.1)<sup>10</sup>, *Meretrix* sp. JML1 (GCA\_049244375.1), *Meretrix* sp. MF1 (GCA\_049244355.1) and *Meretrix* sp. MT1 (GCA\_049244365.1)<sup>7</sup>, *Cyclina sinensis* (GCA\_012932295.1)<sup>27</sup>, *Mercenaria mercenaria* (GCA\_014805675.2)<sup>28</sup>, *Dreissena polymorpha* (GCF\_020536995.1)<sup>29</sup>, *Tridacna crocea* (GCA\_032873355.1)<sup>30</sup>, *Conchocele bisecta* (GCA\_029237695.1)<sup>31</sup>, *Crassostrea nippona* (GCA\_033439105.1)<sup>32</sup>, *Mytilus edulis* (GCF\_036588685.1)<sup>33</sup> and *Ctenoides ales* (GCA\_042257175.1)<sup>34</sup>. 1,210 single-copy orthologs were further used for the construction of a phylogenomic tree in the automated OrthoFinder pipeline, which employs the STAG and STRIDE algorithms for species tree inference and tree rooting, respectively<sup>26</sup>. The resulting tree was used as input to estimate the divergence time using r8s<sup>35</sup>, with a prior divergence time estimate of 494 million years ago between *Conchocele bisecta* and *Tridacna crocea* adapted from the TimeTree database<sup>36</sup>. The resulting phylogenomic tree revealed that *M. meretrix* is sister to other sequenced *Meretrix* genomes (Figure 2).

## **Technical validation**

### *Sequence quality*

Prior to genome assembly scaffolding, the contigs in the draft assemblies were screened for potential contaminations with BlobTools<sup>19</sup>. All 35 and 121 contigs in the draft assemblies of

*M. meretrix* and *M. lamarckii* were affiliated with taxonomic class “Mollusca” or unknown taxonomic class (“no-hit”) and were retained for scaffolding, respectively (Figure 3). After scaffolding with Omni-C data, NCBI Foreign Contamination Screen (FCS)<sup>22</sup> was used for detecting potential contaminated scaffolds in the genome assemblies (Supplementary Table S1). Out of 37 scaffolds in the *M. meretrix* genome assembly, 6 scaffolds were affiliated as prokaryotes, accounting for 214,586 bp in total. On the other hand, of the 102 scaffolds in the *M. lamarckii* genome assembly, 5 scaffolds and 1 scaffold were identified as prokaryotic and mitochondrial sequences, accounting for 239,857 bp and 23,565 bp, respectively. Subsequently, the identified contaminated scaffolds were searched against the respective genome assemblies using BLASTn. While no blast hits of contaminated scaffolds were found in the remaining scaffolds of the *M. meretrix* assembly, 8 scaffolds and 1 scaffold in the *M. lamarckii* assembly were detected for blast hits against two of the prokaryotic contaminated scaffolds and the mitochondrial scaffold, respectively (Supplementary Table S1). Manual BLASTn searches on NCBI core\_nt database were performed for the potential contaminated scaffolds, from which 0%-2% query coverages resulted for prokaryotic contaminated scaffolds and a 100% query coverage was matched with the *M. lamarckii* mitochondrial genome (accession NC\_016174.1) for the mitochondrial scaffold. Therefore, the prokaryotic contaminated scaffolds identified by FCS were removed from the final assemblies and the remaining potential contaminated scaffolds were documented in Supplementary Table S1.

#### *Evaluation of genome assembly and annotation completeness*

The completeness of the genome assemblies and annotations of *M. meretrix* and *M. lamarckii* were assessed with BUSCO (v6.0.0)<sup>37</sup> using databases metazoa\_odb12 and mollusca\_odb12, respectively, and were compared to other available *Meretrix* genomes (Figure 4). Out of 672 single-copy orthologs in the metazoa\_odb12 database, 659 (98.1%) and 661 (98.4%) single-copy orthologs were identified in the genome assemblies of *M. meretrix* and *M. lamarckii*, respectively (Figure 4A). In addition, almost complete sets of single-copy orthologues from the mollusca\_odb12 database could also be identified in genomes of *M. meretrix* (n=4401 (99.5%)) and *M. lamarckii* (n=4402 (99.6%)), respectively. The BUSCO scores of *M. meretrix* and *M. lamarckii* genomes are comparable to the highest BUSCO score from the *M. petechialis* genome

(n=4403 (99.6%) and higher than other available *Meretrix* genomes (n ranging from 4263-4395 (96.4%-99.4%)). For genome annotations, BUSCO scores of 93.0% and 95.2% were attained from *M. meretrix* and *M. lamarckii* with the metazoa\_odb12 database, and 94.3% and 96.3% with the mollusca\_odb12 database, respectively (Figure 4B). The genome annotation BUSCO scores of *M. meretrix* and *M. lamarckii* from the mollusca\_odb12 database are comparable to that of *M. petechialis* (96.6%) and are higher than that of *Meretrix* sp. JML1 (73.8%), *Meretrix* sp. MT1 (84.8%) and *Meretrix* sp. MF1 (90.8%). Therefore, the high BUSCO scores indicate high completeness of both genome assemblies and annotations of both *M. meretrix* and *M. lamarckii* among the available *Meretrix* genomes.

#### *Macrosynteny analysis*

To validate the accuracy of the genome assemblies, the syntenic relationships of *M. meretrix* and *M. lamarckii* genomes were compared to other *Meretrix* genomes (*M. petechialis* (GCA\_046203225.1)<sup>10</sup>, *Meretrix* sp. JML1 (GCA\_049244375.1), *Meretrix* sp. MF1 (GCA\_049244355.1) and *Meretrix* sp. MT1 (GCA\_049244365.1)<sup>7</sup> and *Cyclina sinensis* (GCA\_012932295.1)<sup>27</sup>. The longest transcripts of gene sets of each genome were extracted for macrosynteny analysis using the GENESPACE pipeline<sup>38</sup> with default parameters. Briefly, gene collinearity was detected by a minimum block size of 5 collinear genes using MCScanX<sup>39</sup> and gene orthology inferred by OrthoFinder<sup>26</sup> was used to define collinearity anchors. Macrosynteny analysis revealed a highly conserved chromosome synteny between the *Meretrix* species (Figure 5A). While the *M. lamarckii* genome shared a one-to-one homologous relationship in all 19 chromosomes with other *Meretrix* species, chromosome fusion/fission events were also identified in *M. meretrix*, where chr1 corresponds to chr15 and chr16 of *M. lamarckii* and chr19 to a partial region of chr15 of *M. lamarckii*. For further validation, contigs assembled by Hifiasm were searched against the final assembly of *M. meretrix* using BLASTn. Chr1 was covered by two contigs with 100% identity match, namely “ptg0000211\_1” and “ptg0000051\_1”, where the latter encompasses the genomic region of chromosome fusion/fission (Figure 5B). In addition, raw PacBio HiFi reads were mapped to the *M. meretrix* genome using minimap2<sup>40</sup> and visualised in IGV (v2.19.5)<sup>41</sup>, and the region of chromosome fusion/fission were well supported by aligned

PacBio HiFi reads with even and adequate coverage (>30×) (Figure 5C), ruling out the possibility of a misassembled region.

### Data Records

The genome assemblies of *M. meretrix* and *M. lamarcki* are available at NCBI/ENA GenBank accessions JBTXPG000000000<sup>42,43</sup> and JBTORC000000000<sup>44,45</sup>, DDBJ accessions PRJDB40269<sup>46</sup> and PRJDB40270<sup>47</sup>, NGDC accessions GWHHOJO000000000<sup>48</sup> and GWHHOOY000000000<sup>49</sup>, and CUHK Research Data Repository<sup>60</sup>, respectively. The raw sequencing data of *M. meretrix* and *M. lamarcki* are deposited in the NCBI/ENA database under the SRA accession numbers SRP617717<sup>50,51</sup> and SRP618173<sup>52,53</sup>, DDBJ database under the DRA accession numbers DRR911404<sup>54</sup>-DRR911405<sup>55</sup> and DRR911157<sup>56</sup>- DRR911158<sup>57</sup>, NGDC database under the GSA accession numbers CRA039134<sup>58</sup> and CRA039136<sup>59</sup>, and CUHK Research Data Repository<sup>60</sup>, respectively. The genome annotation files of *M. meretrix* and *M. lamarckii* can be found at CUHK Research Data Repository<sup>60</sup>.

### Data availability

The genome assemblies of *M. meretrix* and *M. lamarcki* are available at NCBI/ENA GenBank under accessions JBTXPG000000000 and JBTORC000000000, DDBJ accessions PRJDB40269 and PRJDB40270, NGDC accessions GWHHOJO000000000 and GWHHOOY000000000, and CUHK Research Data Repository ([doi.org/10.48668/DOVLYS](https://doi.org/10.48668/DOVLYS)), respectively. The raw sequencing data of *M. meretrix* and *M. lamarcki* are deposited in the NCBI/ENA database under the SRA accession numbers SRP617717 and SRP618173, DDBJ database under the DRA accession numbers DRR911404-DRR911405 and DRR911157-DRR911158, NGDC database under the GSA accession numbers CRA039134 and CRA039136, and CUHK Research Data Repository ([doi.org/10.48668/DOVLYS](https://doi.org/10.48668/DOVLYS)), respectively. The genome annotation files of *M. meretrix* and *M. lamarckii* can be found at CUHK Research Data Repository ([doi.org/10.48668/DOVLYS](https://doi.org/10.48668/DOVLYS)).

### Code availability

No specific script was used in this work.

## References

1. Srinun, N., Vidthayanon, C. & Klangnurak, W. Morphological and genetic variations of *Meretrix* spp. (Bivalvia: Veneridae) recently distributed in the coastal waters of the Gulf of Thailand. *Estuar. Coast. Shelf Sci.* **307**, 108889 (2024).
2. Soon, T. K., Sing, O. F., Denil, D. J. & Ransangan, J. Distribution and fishing pressure of hard clam, *Meretrix meretrix* in Marudu Bay, Sabah. *Int. J. Oceans Oceanogr.* **11**, (2017).
3. Ye, Y. *et al.* Genetic Population Structure of the Hard Clam *Meretrix meretrix* Along the Chinese Coastlines Revealed by Microsatellite DNA Markers. *Front. Mar. Sci.* **7**, (2020).
4. Chen, Y. *et al.* The biological resources survey of mudflat clams and other major creatures in Panjin Geligang and Xiao He. *Hebei Fish* **1**, 46–49 (2012).
5. So, K. J. Y., Cheang, C. C., Hui, T. Y. & Chan, J. K. Y. Understanding the behavioural gap between perceived and actual environmental behaviour: Investigating the clam-harvesting pattern in Hong Kong SAR, China. *J. Clean. Prod.* **316**, 128259 (2021).
6. WoRMS. World Register of Marine Species. <https://www.marinespecies.org> (2025)  
doi:10.14284/170.
7. Chen, C.-C., Hsu, T.-H., Lu, H.-Y., Tang, S.-L. & Ho, Y.-N. High-quality chromosome-level genome of three *Meretrix* species using Nanopore and Hi-C technologies. *Sci. Data* **12**, 1141 (2025).
8. Morton, B., Leung, S. F. & Leung, K. F. The biology and functional morphology of *Meretrix* cf. *meretrix* (Bivalvia: Veneridae: Meretricinae) at Tong Fuk Miu Wan, Shui Hau, Lantau Island, Hong Kong. *Reg. Stud. Mar. Sci.* **45**, 101842 (2021).

9. Hsiao, S.-T. & Chuang, S.-C. *Meretrix taiwanica* (Bivalvia: Veneridae), a previously misidentified new species in Taiwan. *Molluscan Res.* **43**, 12–21 (2023).
10. Law, S. T. S. *et al.* Genomes of two indigenous clams *Anomalocardia flexuosa* (Linnaeus, 1767) and *Meretrix petechialis* (Lamarck, 1818). *Sci. Data* **12**, 409 (2025).
11. Folmer, O., Black, M., Hoeh, W., Lutz, R. & Vrijenhoek, R. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Mol. Mar. Biol. Biotechnol.* **3**, 294–299 (1994).
12. Chen, J., Li, Q., Kong, L. & Zheng, X. Molecular phylogeny of venus clams (Mollusca, Bivalvia, Veneridae) with emphasis on the systematic position of taxa along the coast of mainland China. *Zool. Scr.* **40**, 260–271 (2011).
13. Wang, X., Kong, L., Chen, J., Matsukuma, A. & Li, Q. Phylogeography of bivalve *Meretrix petechialis* in the Northwestern Pacific indicated by mitochondrial and nuclear DNA data. *PLOS ONE* **12**, e0183221 (2017).
14. Zhang, S. P., Wang, H. X. & Xu, F. S. Taxonomic study on *Meretrix* (Bivalvia, Veneridae) from China seas. *Acta Zootaxonomica Sin.* **37**, (2012).
15. Katoh, K. & Standley, D. M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
16. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. *PLOS ONE* **5**, e9490 (2010).
17. Cheng, H., Concepcion, G. T., Feng, X., Zhang, H. & Li, H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat. Methods* **18**, 170–175 (2021).
18. Guan, D. *et al.* Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics* **36**, 2896–2898 (2020).

19. Laetsch, D. R. & Blaxter, M. L. BlobTools: Interrogation of genome assemblies. Preprint at <https://doi.org/10.12688/f1000research.12232.1> (2017).
20. Zhou, C., McCarthy, S. A. & Durbin, R. YaHS: yet another Hi-C scaffolding tool. *Bioinformatics* **39**, btac808 (2023).
21. Durand, N. C. *et al.* Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom. *Cell Syst.* **3**, 99–101 (2016).
22. Astashyn, A. *et al.* Rapid and sensitive detection of genome contamination at scale with FCS-GX. *Genome Biol.* **25**, 60 (2024).
23. Hoff, K. J., Lomsadze, A., Borodovsky, M. & Stanke, M. Whole-Genome Annotation with BRAKER. in *Gene Prediction: Methods and Protocols* (ed. Kollmar, M.) 65–95 (Springer, New York, NY, 2019). doi:10.1007/978-1-4939-9173-0\_5.
24. Stanke, M., Diekhans, M., Baertsch, R. & Haussler, D. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics* **24**, 637–644 (2008).
25. Baril, T., Galbraith, J. & Hayward, A. Earl Grey: A Fully Automated User-Friendly Transposable Element Annotation and Analysis Pipeline. *Mol. Biol. Evol.* **41**, msae068 (2024).
26. Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**, 238 (2019).
27. Wei, M. *et al.* Chromosome-Level Clam Genome Helps Elucidate the Molecular Basis of Adaptation to a Buried Lifestyle. *iScience* **23**, 101148 (2020).
28. Song, H. *et al.* The hard clam genome reveals massive expansion and diversification of inhibitors of apoptosis in Bivalvia. *BMC Biol.* **19**, 15 (2021).

29. McCartney, M. A. *et al.* The genome of the zebra mussel, *Dreissena polymorpha*: a resource for comparative genomics, invasion genetics, and biocontrol. *G3 GenesGenomesGenetics* **12**, jkab423 (2022).
30. Li, J. *et al.* Chromosome-level genome assembly and annotation of rare and endangered tropical bivalve, *Tridacna crocea*. *Sci. Data* **11**, 186 (2024).
31. Guo, Y. *et al.* Hologenome analysis reveals independent evolution to chemosymbiosis by deep-sea bivalves. *BMC Biol.* **21**, 51 (2023).
32. Bai, Y. *et al.* Multi-omic insights into the formation and evolution of a novel shell microstructure in oysters. *BMC Biol.* **21**, 204 (2023).
33. Regan, T., Hori, T. S. & Bean, T. P. A chromosome-scale *Mytilus edulis* genome assembly for aquaculture, marine ecology, and evolution. *G3 GenesGenomesGenetics* **14**, jkae138 (2024).
34. McElroy, K. E., Masonbrink, R., Chudalayandi, S., Severin, A. J. & Serb, J. M. A chromosome-level genome assembly of the disco clam, *Ctenoides ales*. *G3 GenesGenomesGenetics* **14**, jkae115 (2024).
35. Sanderson, M. J. r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics* **19**, 301–302 (2003).
36. Hedges, S. B., Dudley, J. & Kumar, S. TimeTree: a public knowledge-base of divergence times among organisms. *Bioinformatics* **22**, 2971–2972 (2006).
37. Manni, M., Berkeley, M. R., Seppey, M. & Zdobnov, E. M. BUSCO: Assessing Genomic Data Quality and Beyond. *Curr. Protoc.* **1**, e323 (2021).
38. Lovell, J. T. *et al.* GENESPACE tracks regions of interest and gene copy number variation across multiple genomes. *eLife* **11**, e78526 (2022).

39. Wang, Y. *et al.* MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* **40**, e49 (2012).
40. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).
41. Robinson, J. T., Thorvaldsdottir, H., Turner, D. & Mesirov, J. P. igv.js: an embeddable JavaScript implementation of the Integrative Genomics Viewer (IGV). *Bioinformatics* **39**, btac830 (2023).
42. *NCBI GenBank* <https://identifiers.org/ncbi/insdc:JBTXPG0000000000> (2026).
43. *European Nucleotide Archive* [https://www.ebi.ac.uk/ena/browser/view/GCA\\_055551385.1](https://www.ebi.ac.uk/ena/browser/view/GCA_055551385.1) (2026).
44. *NCBI GenBank* <https://identifiers.org/ncbi/insdc:JBTORC0000000000> (2026).
45. *European Nucleotide Archive* [https://www.ebi.ac.uk/ena/browser/view/GCA\\_055550865.1](https://www.ebi.ac.uk/ena/browser/view/GCA_055550865.1) (2026).
46. *DNA Data Bank of Japan* <https://ddbj.nig.ac.jp/search/entry/bioproject/PRJDB40269> (2026).
47. *DNA Data Bank of Japan* <https://ddbj.nig.ac.jp/search/entry/bioproject/PRJDB40270> (2026).
48. *National Genomics Data Center* <https://ngdc.cncb.ac.cn/gwh/Assembly/111533/show> (2026).
49. *National Genomics Data Center* <https://ngdc.cncb.ac.cn/gwh/Assembly/111604/show> (2026).
50. *NCBI Sequence Read Archive* <https://identifiers.org/ncbi/insdc.sra:SRP617717> (2026).

51. *European Nucleotide Archive* <https://www.ebi.ac.uk/ena/browser/view/PRJNA1312090> (2026).
52. *NCBI Sequence Read Archive* <https://identifiers.org/ncbi/insdc.sra:SRP618173> (2026).
53. *European Nucleotide Archive* <https://www.ebi.ac.uk/ena/browser/view/PRJNA1315236> (2026).
54. *DNA Data Bank of Japan* <https://ddbj.nig.ac.jp/search/entry/sra-run/DRR911404> (2026).
55. *DNA Data Bank of Japan* <https://ddbj.nig.ac.jp/search/entry/sra-run/DRR911405> (2026).
56. *DNA Data Bank of Japan* <https://ddbj.nig.ac.jp/search/entry/sra-run/DRR911157> (2026).
57. *DNA Data Bank of Japan* <https://ddbj.nig.ac.jp/search/entry/sra-run/DRR911158> (2026).
58. *National Genomics Data Center* <https://ngdc.cnbc.ac.cn/gsa/browse/CRA039134> (2026).
59. *National Genomics Data Center* <https://ngdc.cnbc.ac.cn/gsa/browse/CRA039136> (2026).
60. Law, S. T. S. *et al.* Chromosome-level genomes of hard clams *Meretrix lamarckii* (Deshayes, 1853) and *Meretrix meretrix* (Linnaeus, 1758). *CUHK Research Data Repository* <https://doi.org/10.48668/DOVLYS> (2025).

## Acknowledgements

This work was supported by Lantau Conservation Fund (RE-2020-39), Hong Kong Research Grant Council Collaborative Research Fund (C4015-20EF), Innovation Technology Fund of Innovation Technology Commission: Funding Support to State Key Laboratory of Agrobiotechnology, and Direct Grant of The Chinese University of Hong Kong.

## Contributions

S.Y.L., S.G.C. and J.H.L.H. conceived and supervised the study; S.T.S.L., M.F.F.A., L.H.T.C. and C.W.Y.S. carried out sample collection; S.T.S.L. and W.N. performed data curation on the analysis; J.H.L.H. and S.T.S.L. wrote the initial manuscript; all authors revised and contributed to the final version of the manuscript.

## Competing interests

The authors declare no competing interests.

## Tables and Figures

**Table 1.** Genome sequencing data of *M. meretrix* and *M. lamarckii*.

**Table 2.** Genome statistics of *M. meretrix* and *M. lamarckii* and other published *Meretrix* genomes.

**Table 3.** Chromosome information of *M. meretrix* and *M. lamarckii*.

**Table 4.** Repeat annotations of *M. meretrix* and *M. lamarckii*.

**Figure 1.** **A)** Summary of genome assemblies of *M. meretrix* and *M. lamarckii* and other published *Meretrix* genomes. Scaffold N50n refers to the number of scaffolds that contribute to the N50 length; **B)** Omni-C contact maps of the genome assemblies of *M. meretrix* (left) and *M. lamarckii* (right); **C)** Pie charts of annotated transposable elements of *M. meretrix* (left) and *M. lamarckii* (right).

**Figure 2.** Phylogenomic tree of *Meretrix* species and 8 other bivalve species. Mya: million years ago.

**Figure 3.** Blobtools GC-coverage plots for the draft genome assembly control prior to scaffolding for *M. meretrix* (A) and *M. lamarckii* (B). The size of the circles are proportional to the contig length.

**Figure 4.** BUSCO completeness assessment of the A) genome assemblies and B) predicted gene models of *M. meretrix* and *M. lamarckii* and other published *Meretrix* genomes using the metazoa\_odb12 database (left) and mollusca\_odb12 database (right).

**Figure 5.** A) Macrosynteny of *M. meretrix* and *M. lamarckii* and other published genomes; B) Visualization of contigs assembled by Hifiasm aligned to Chr 1 in the final assembly of *M. meretrix*. Black circle denotes the assembly gap and red triangle indicates the genomic region of fusion/fission of Chr 1 in *M. meretrix*; C) Visualization of PacBio HiFi reads aligned to the genomic region of fusion/fission of Chr 1 in *M. meretrix*. Genes have syntenic relationships to *M. lamarckii* Chr 15 and Chr 16 are coloured in red and blue, respectively. Under the expanded view of aligned reads, positive and negative strands of reads are coloured in pink and purple, respectively.

Species	<i>Meretrix meretrix</i>	<i>Meretrix lamarckii</i>	<i>Meretrix petechialis</i>	<i>M. lamarckii</i> JML1	<i>Meretrix</i> <i>sp.</i> MF1	<i>Meretrix</i> <i>sp.</i> MT1
References	This study	This study	Law et al., 2025	Chen et al., 20205	Chen et al., 20205	Chen et al., 20205
Total length	835,139,699	890,513,344	1,045,845,039	883,065,649	883,299,404	944,737,021
No. of scaffold	31	97	114	198	36	74
Mean scaffold length	26,939,990	9,180,550	9,174,079	4,459,928	24,536,095	12,766,717
Longest scaffold length	64,908,192	57,236,489	65,874,028	59,194,310	59,289,571	61,949,085
Gaps	14	33	198	0	0	0
Scaffold N50	46,172,035	46,241,283	53,512,864	46,538,327	46,874,007	48,788,946
Scaffold N50n	9	9	9	9	9	9
No. of gene models	28,699	33,346	25,196	55,887	49,050	54,445
No. of longest-transcripts	26,759	30,829	20,084	51,779	45,263	50,420

Chr number	Chr ID	<i>Meretrix meretrix</i>		<i>Meretrix lamarckii</i>	
		Chr length (bp)	% of the whole genome	Chr length (bp)	% of the whole genome
1	scaffold_1	64,908,192	7.77	57,236,489	6.43
2	scaffold_2	54,483,009	6.52	54,346,355	6.10
3	scaffold_3	51,357,909	6.15	53,604,160	6.02
4	scaffold_4	49,509,254	5.93	53,479,476	6.01
5	scaffold_5	48,075,356	5.76	52,294,165	5.87
6	scaffold_6	47,988,444	5.75	51,023,760	5.73
7	scaffold_7	47,650,749	5.71	49,992,336	5.61
8	scaffold_8	47,076,688	5.64	46,475,248	5.22
9	scaffold_9	46,172,035	5.53	46,241,283	5.19
10	scaffold_10	45,271,054	5.42	46,212,281	5.19
11	scaffold_11	45,158,256	5.41	46,002,274	5.17
12	scaffold_12	43,632,693	5.22	45,873,465	5.15

13	scaffold_1 3	42,715,300	5.11	44,241,166	4.97
14	scaffold_1 4	42,254,144	5.06	43,993,078	4.94
15	scaffold_1 5	39,053,538	4.68	43,461,710	4.88
16	scaffold_1 6	38,933,291	4.66	42,595,028	4.78
17	scaffold_1 7	29,759,200	3.56	41,225,765	4.63
18	scaffold_1 8	27,822,048	3.33	37,866,541	4.25
19	scaffold_1 9	15,296,840	1.83	29,279,834	3.29
	<b>sum=</b>	827,118,000	99.04	885,444,414	99.43

Classification	Coverage (bp)	Count	Proportion (%)	No. of distinct classifications	Coverage (bp)	Count	Proportion (%)	No. of distinct classifications
	<i>Meretrix meretrix</i>				<i>Meretrix lamarckii</i>			
DNA	16,308,484	57,727	1.95	5,814	33,779,114	120,855	3.79	14,768
LINE	24,165,708	78,322	2.89	8,606	43,085,090	138,412	4.84	16,107
LTR	9,914,756	28,900	1.19	5,758	17,627,897	43,593	1.98	9,100
Other (Simple Repeat, Microsatellite, RNA)	65,235,986	301,859	7.81	4,627	70,941,785	331,613	7.97	5,672
Penelope	15,486,324	33,740	1.85	5,542	5,276,681	9,932	0.59	3,996
Rolling Circle	3,717,496	12,133	0.45	3,697	4,236,078	9,895	0.48	3,343
SINE	8,010,153	41,567	0.96	9,559	7,064,880	42,137	0.79	5,763
Unclassified	214,347,055	884,731	25.67	10,742	215,536,269	829,761	24.20	14,463
Sum	357,185,962	1,438,979	42.77	54,345	397,547,794	1,526,198	44.64	73,212

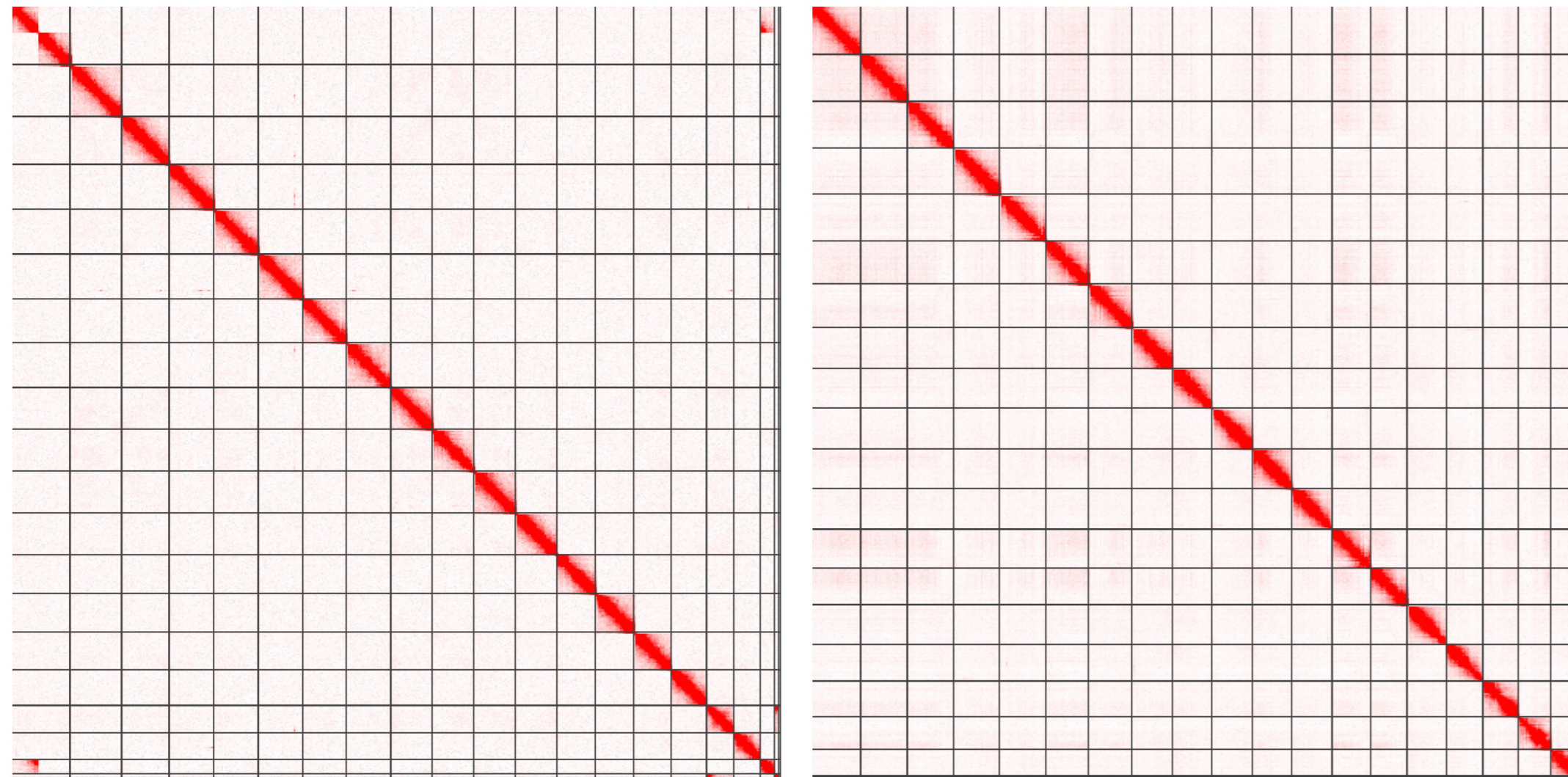
Species	Samples	Reads	Base s	Hifi average length (bp)	NCBI/EN A SRA Accession	DDBJ DRA Accession	NGDC GSA Accession	CUHK Research Data Repository
<i>Meretrix meretrix</i>	Mmer_PacBio	4,376,435	62,875,654,323	14,367	SRR35263659	DRR911404	CRR2744619	doi:10.48668/DOVLYS
<i>Meretrix meretrix</i>	Mmer_OmniC	23,697,482	3,554,622,300	-	SRR35263658	DRR911405	CRR2744620	
<i>Meretrix lamarc kii</i>	Mlam_PacBio	4,402,942	57,891,566,865	13,148	SRR35291242	DRR911157	CRR2744623	
<i>Meretrix lamarc kii</i>	Mlam_OmniC	22,436,556	3,365,483,400	-	SRR35291243	DRR911158	CRR2744624	

A)

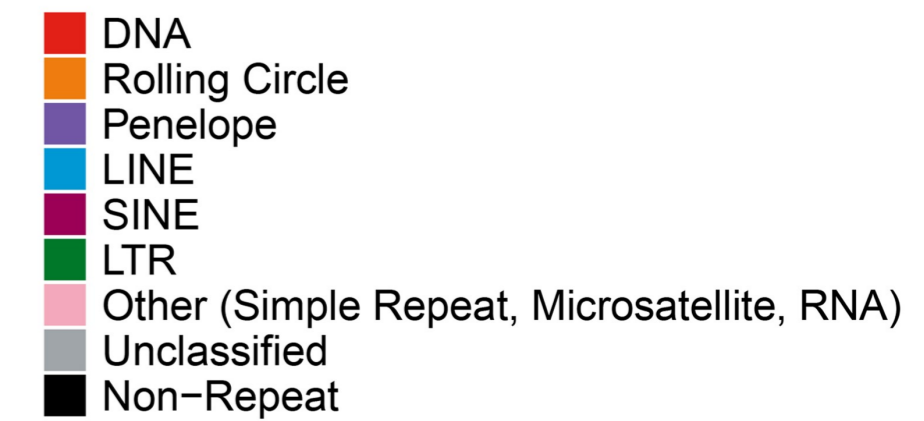
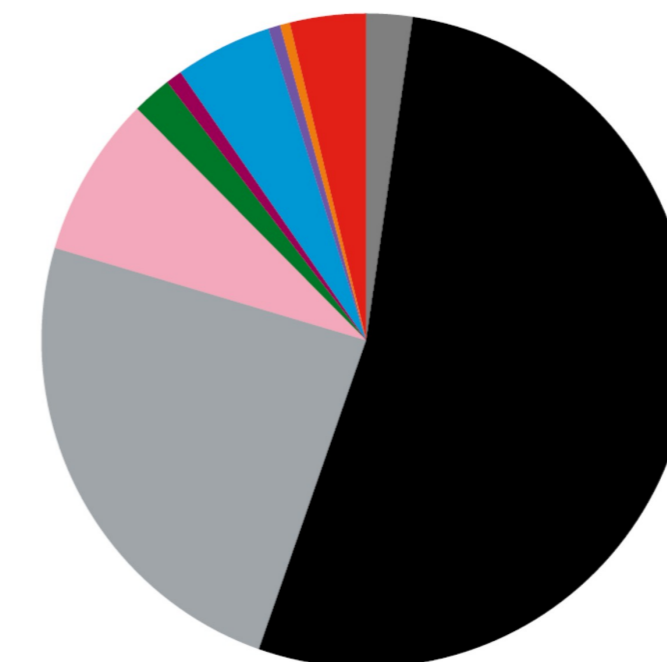
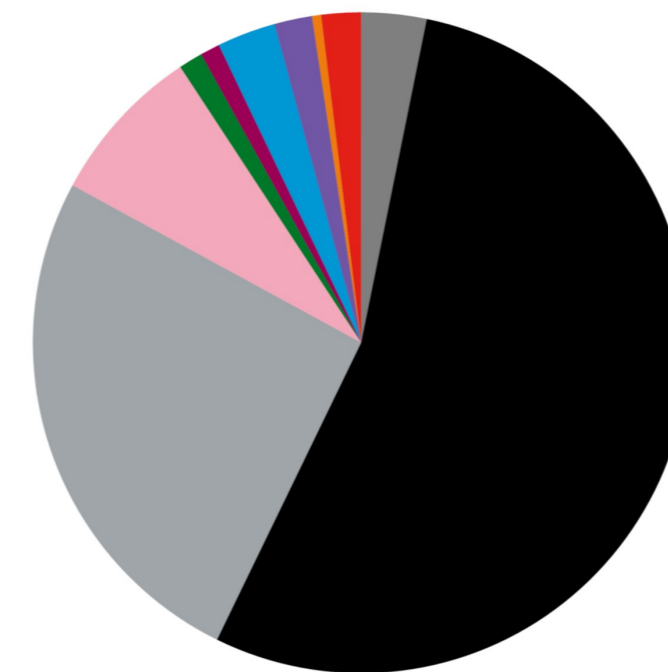


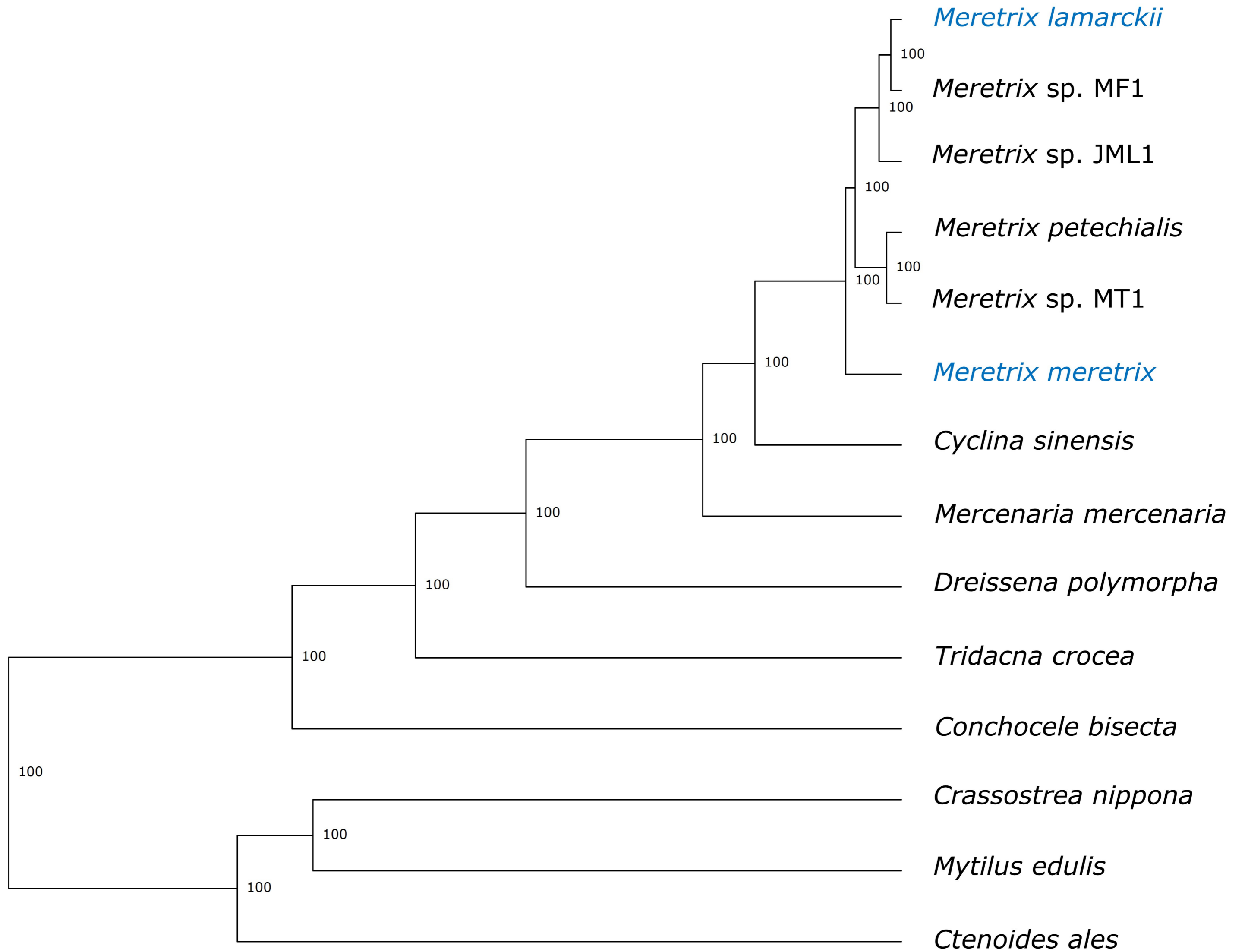
Species	<i>M. meretrix</i>	<i>M. lamarckii</i>	<i>M. petechialis</i>	<i>M. lamarckii</i> JML1	<i>Meretrix</i> sp. MF1	<i>Meretrix</i> sp. MT1
References	This study	This study	Law et al., 2025	Chen et al., 2025	Chen et al., 2025	Chen et al., 2025
Total length (bp)	835,139,699	890,513,344	1,045,845,039	883,065,649	883,299,404	944,737,021
No. of scaffold	31	97	114	198	36	74
Longest scaffold length (bp)	64,908,192	57,236,489	65,874,028	59,194,310	59,289,571	61,949,085
Scaffold N50 (bp)	46,172,035	46,241,283	53,512,864	46,538,327	46,874,007	48,788,946
Scaffold N50n	9	9	9	9	9	9
No. of gene models	28,699	33,346	25,196	55,887	49,050	54,445

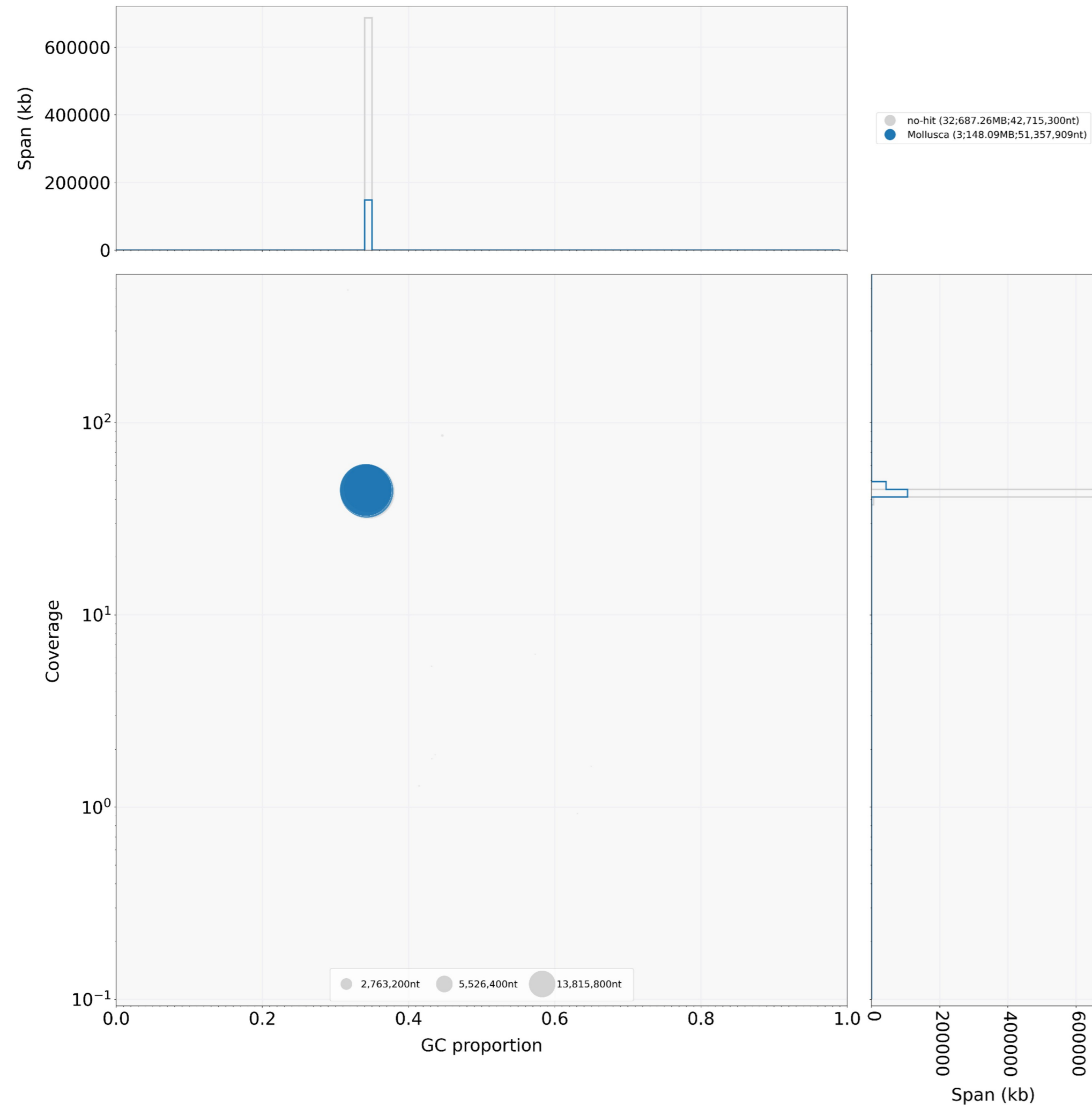
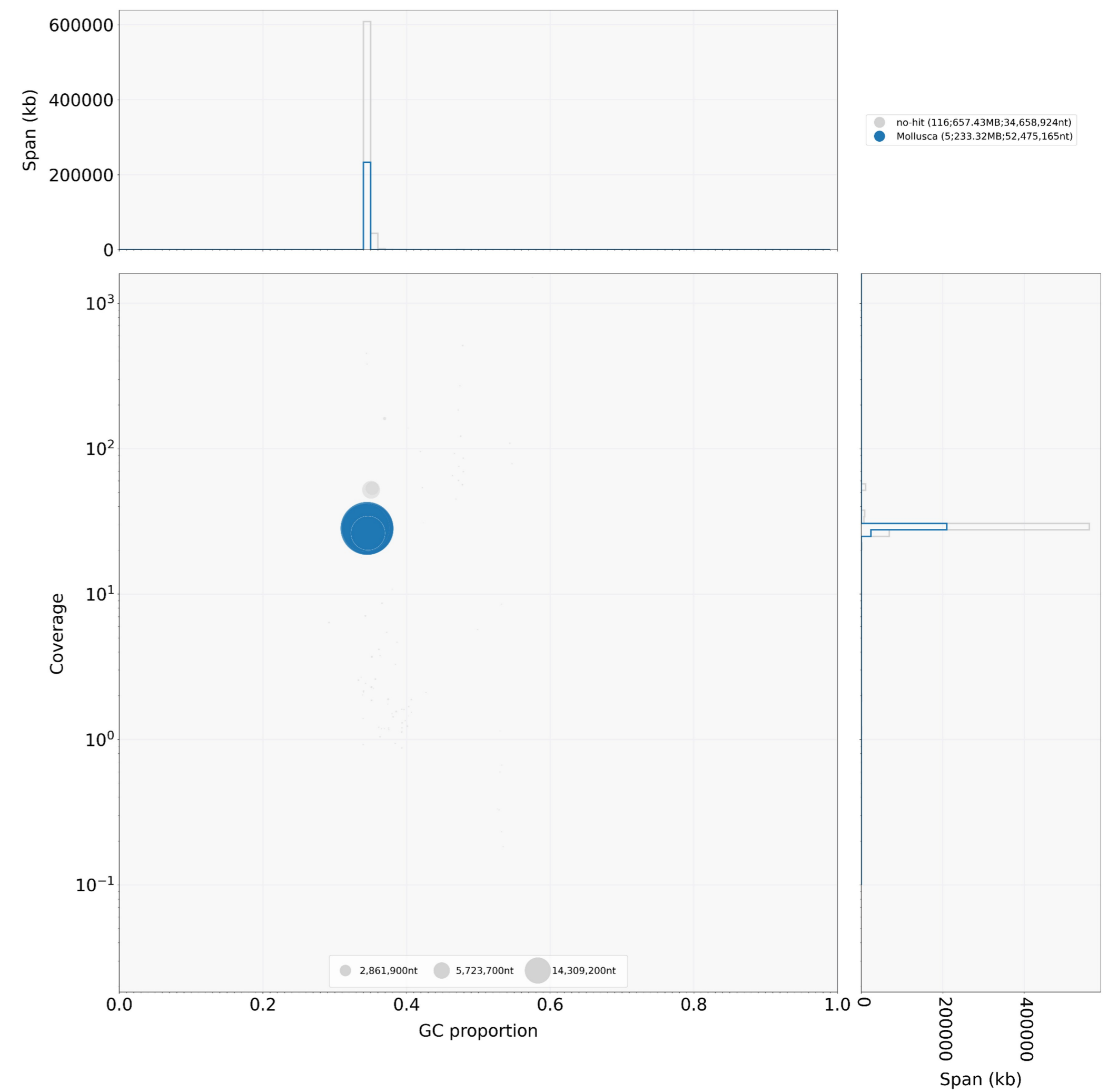
B)

*Meretrix meretrix**Meretrix lamarckii*

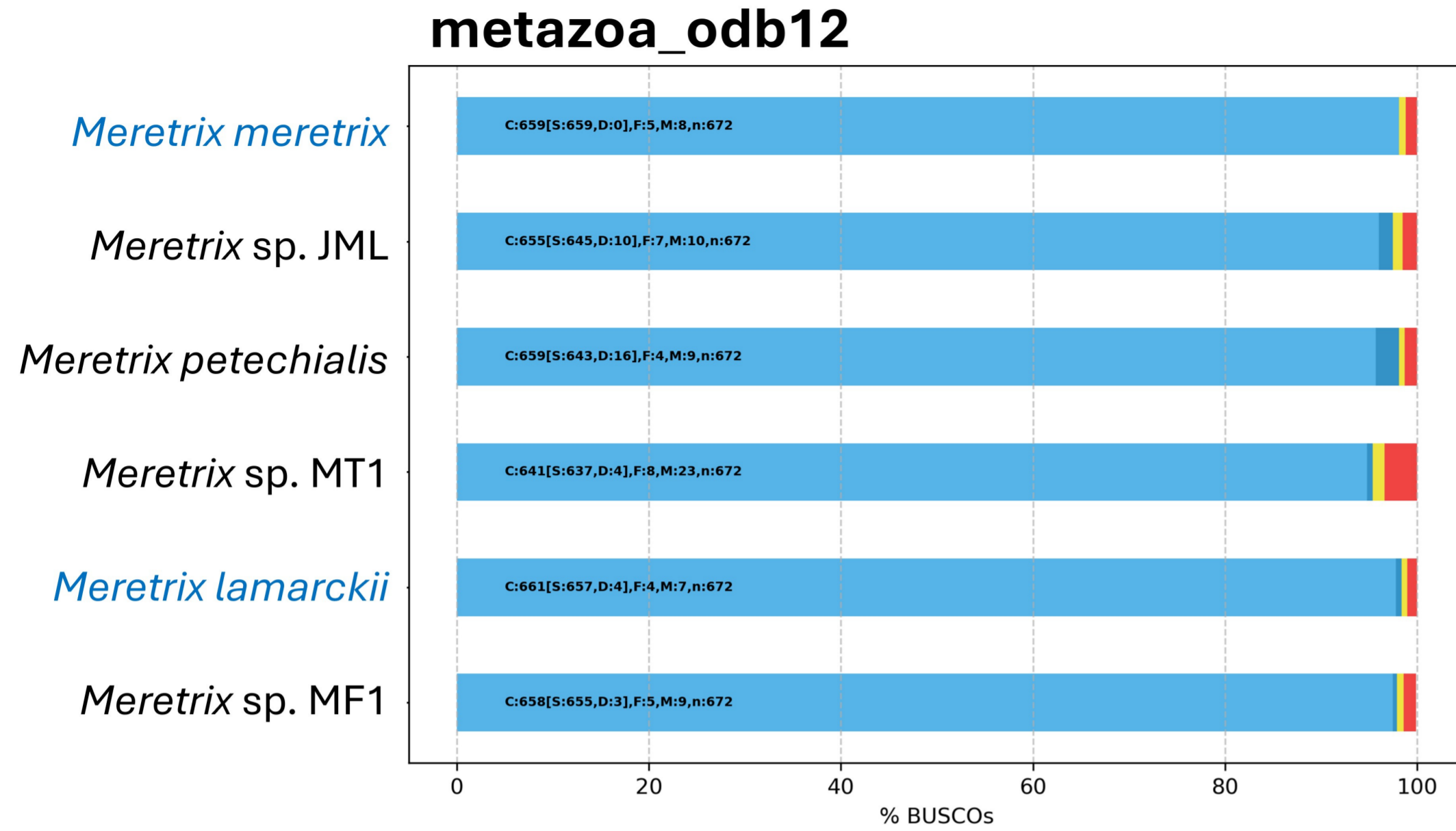
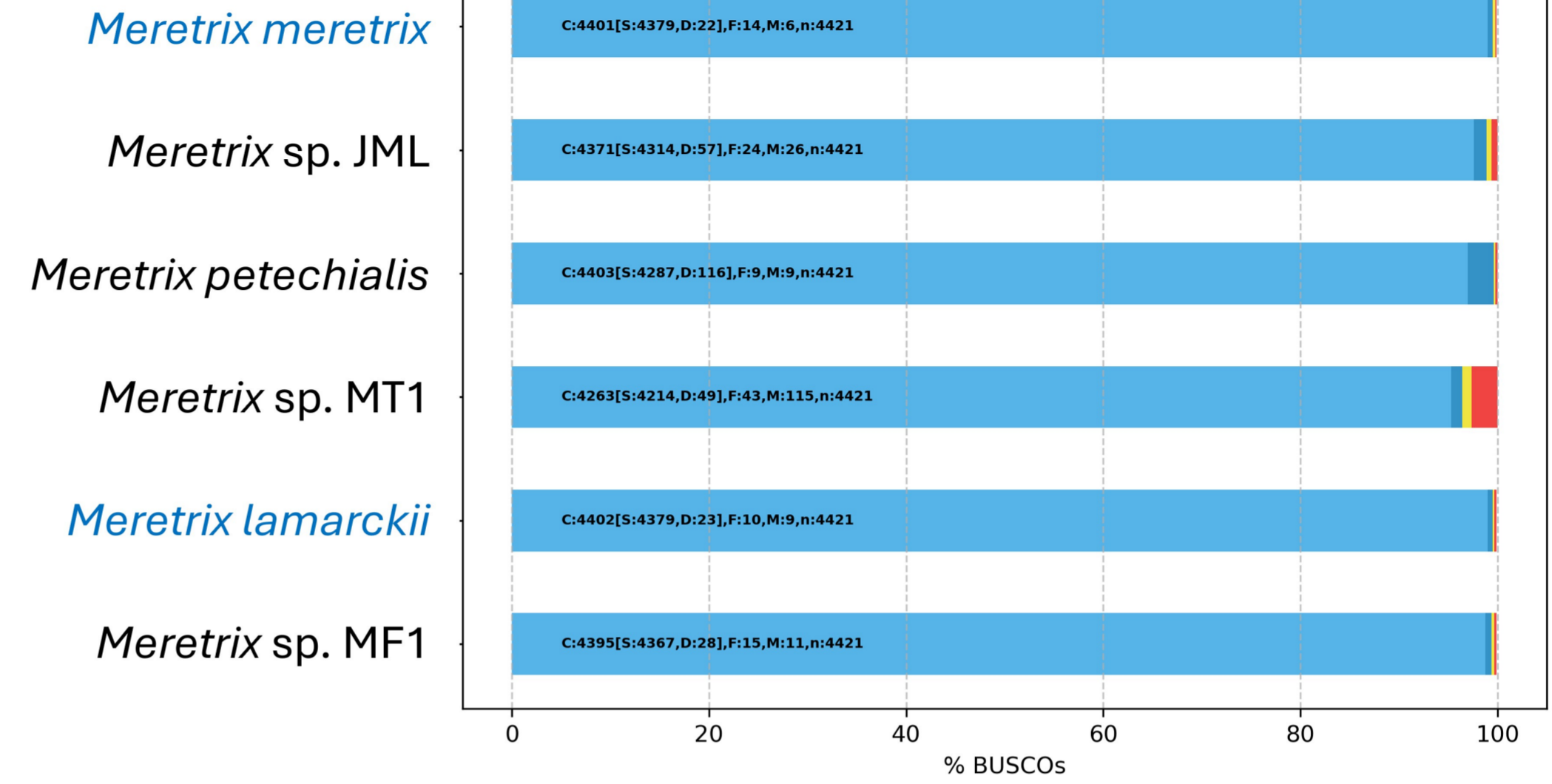
C)

*Meretrix meretrix**Meretrix lamarckii*

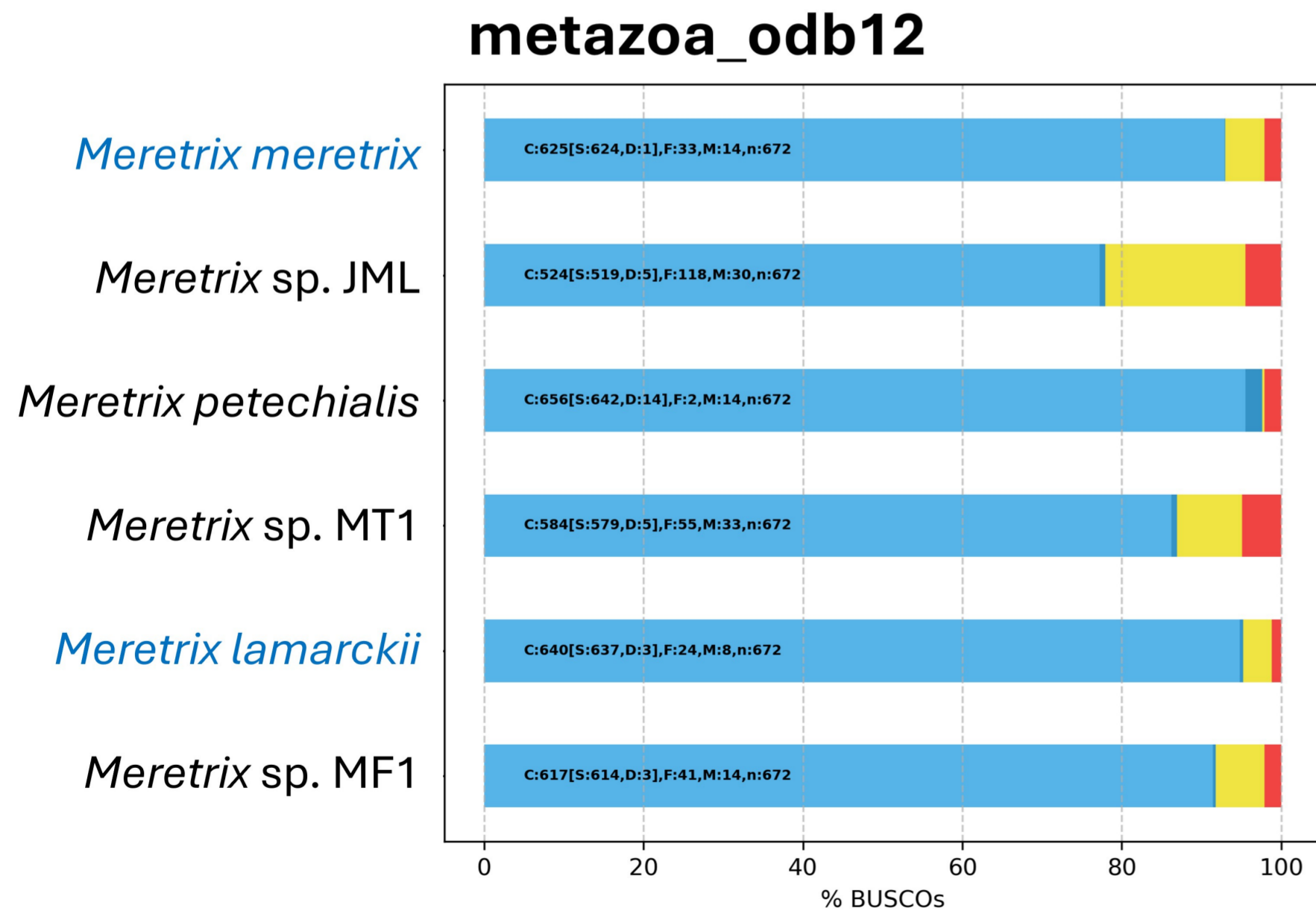
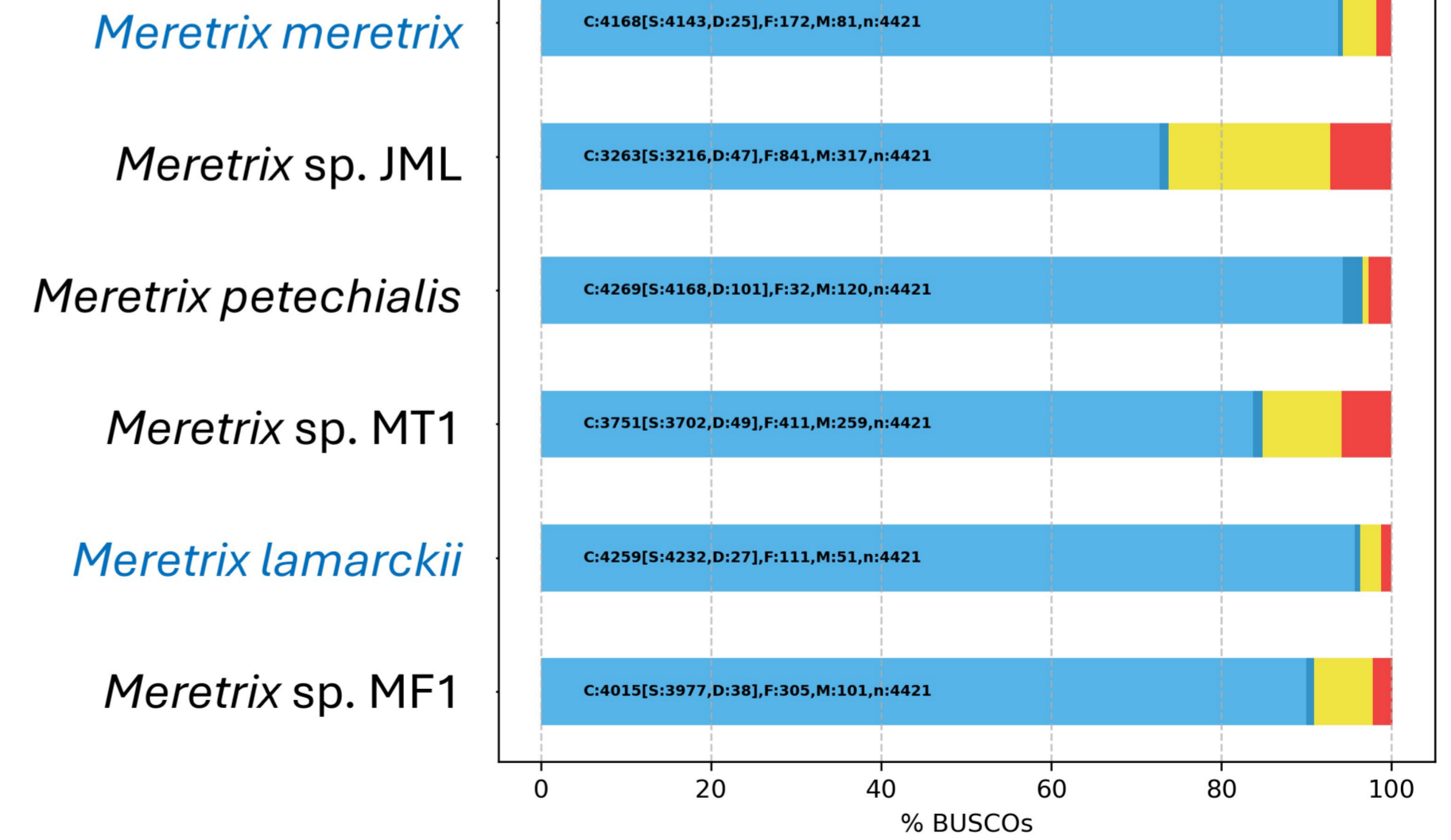


**A)****B)**

A)

**mollusca\_odb12**

B)

**mollusca\_odb12**

Complete (C) and single-copy (S)  
 Complete (C) and duplicated (D)

Fragmented (F)  
 Missing (M)

