Cephalopod Sex Determination and its Ancient Evolutionary Origin

Gabrielle C. Coffing[1,♠], Silas Tittes[1], Scott T. Small[1], Jeremea O. Songco-Casey[1], Denise M. Piscopo[1], Judit R. Pungor[1], Adam C. Miller[1], Cristopher M. Niell[1], and Andrew D. Kern[1,♠]

[1]Department of Biology, University of Oregon
[♠]corresponding author

Abstract

Sex chromosomes are critical elements of sexual reproduction in many animal and plant taxa, however they show incredible diversity and rapid turnover even within clades. Until now, the mechanism of sex determination in cephlaopods has been a mystery. Using a chromosome-level genome assembly generated with long read sequencing, we report the first evidence for genetic sex determination in cephalopods. We have uncovered a sex chromosome in California two-spot octopus (*Octopus bimaculoides*) in which males/females have ZZ/ZO karyotypes respectively. We show that the octopus Z chromosome is an evolutionary outlier with respect to divergence and repetitive element content as compared to autosomes and that it is present in all cephalopods that we have examined including Nautilus, the outgroup to squids and octopuses. Our results suggest that the cephalopod Z chromosome system originated before the split of all extant cephalopod lineages, over 480 million years ago and has been conserved to the present, making it the among the oldest conserved animal sex chromosome systems known.

Introduction

Octopuses, squids, and cuttlefishes – the coleoid cephalopods – are a remarkable branch in the tree of life whose members exhibit a repertoire of sophisticated behaviors (Hanlon and Messenger, 2018). As a clade, coleoid cephalopods harbor an incredible variety of novel traits including the largest and most complex nervous system in the invertebrate world, independently derived camera-type eyes, and rapid adaptive camouflage abilities (Young, 1971; Hanlon, 2007). The burst of evolutionary novelty that distinguishes cephalopods is even more striking when put into a phylogenetic context; cephalopods are a deeply diverged lineage that last share a common ancestor with other extant molluscan lineages in the Cambrian period roughly 550 million years ago (Ponder and Lindberg, 2008; Huang et al., 2022).

Here, using PacBio long-read sequencing of genomic DNA and IsoSeq full-length mRNA sequencing, we provide a novel chromosome-scale reference genome and annotation for a female California two-spot octopus (*O. bimaculoides*). Strikingly, our assembly reveals evidence for a hemizygous chromosome, the first evidence of genetic sex determination in cephalopods. We use our assembly and annotation in combination with existing genomic information from other cephalopods to create the first whole genome alignments from this group and demonstrate that the sex chromosome is of an ancient origin, before the radiation of extant cephalopod lineages approximately 480 million years ago (Huang et al., 2022), and has been conserved to the present day in all cephalopods.

A chromosome-level assembly reveals a hemizygous Z chromosome in female octopus

The California two-spot octopus was the first cephalopod genome to be sequenced in 2015 (Albertin et al., 2015) and subsequently placed into scaffolds in 2022 (Albertin et al., 2022b). Although these resources have been tremendously valuable for cephalopod research, the assembly still contains numerous gaps and many gene annotations remain fragmented due to the highly-repetitive nature of the genome. To sequence through the long, repetitive stretches of the *O. bimaculoides* genome, we re-sequenced a single female individual with PacBio's long, high-fidelity (HiFi) sequencing and used chromosomal conformation capture (Hi-C) to place scaffolds into chromosomes. After scaffolding, the total genome assembly consists of 2.3Gb with 30 chromosomal scaffolds representing the expected $N = 30$ karyotype of octopus (Wang and Zheng, 2018). Additionally, we used IsoSeq sequencing to aid with annotating full-length genes. A comparison of assembly statistics to existing *O. bimaculoides* assemblies is shown in Table S1 and genome annotation statistics are in Table S2. Generally, our new assembly is more complete than previous assemblies, with a contig N50 of 0.86 Mb and a scaffold N50 of 101.05 Mb. The new assembly reduced the number of scaffolds in the assembly from 145,326 (Albertin et al., 2022b) to 583, increasing the average scaffold length by roughly three orders of magnitude.

Strikingly, our Hi-C contact map showed evidence for one chromosome, chromosome 17, having reduced coverage in comparison to other chromosomes in our assembly from a female individual (Figs. 1A & S1). As the original reference assembly from (Albertin et al., 2022b) was from a male individual we were able to compare coverage in that assembly, which showed no difference in coverage between chromosome 17 and any other chromosome. Using short read Illumina data that we generated from unrelated female and male *O. bimaculoides* individuals ($N = 2$ of each sex; Table S3) we confirmed that females of this species are hemizygous for chromosome 17 whereas males are diploid, so hereafter we refer to chromosome 17 as Z (Figs. 1B & S2).

Since the *O. bimaculoides* male genotype is clearly ZZ and the female is hemizygous at Z, we next turned our attention towards identification of a potential W chromosome limited to females. To do so we looked for scaffolds that were only present in the female-derived sequence libraries and absent from the male-derived libraries and assemblies. We found no such candidates. This suggests that females are ZO and males ZZ, perhaps indicating the

evolutionary loss of the W chromosome after substantial degradation (Charlesworth et al., 2005; Charlesworth, 2013). While we note the absence of evidence is not evidence of absence, the abundance of sequence data that we do have presently from this species does not clearly identify a W chromosome. It is possible that a small dot W exists that cannot be identified with our sequencing data. However, ZZ/ZO and XX/XO sex determination systems have been described in other groups including Lepidopterans (Johnson and Lachance, 2012) and plants (Charlesworth, 2013).

Genomic comparisons among cephalopods show the Z chromosome is an outlier

To determine if the Z chromosome was unique to *O. bimaculoides* or more broadly distributed among cephalopods, we inferred a species tree of cephalopods using protein sequences with the STAG algorithm in OrthoFinder Emms and Kelly (2019). This alignment compared three other octopus genomes (*Hapalochlaena maculosa*, *Octopus minor*, and *Octopus sinensis*), two squid genomes (*Architeuthis dux*, and *Euprymna scolopes*), a cuttlefish genome (*Sepia esculenta*), and the outgroup to coleoid cephalopods, the chambered nautilus (*Nautilus pompilius*) (Fig. 2A). Next, we created the first whole-genome, multiple alignment among existing cephalopod genomes using the Progressive Cactus / Comparative Genomics toolkit pipeline (Armstrong et al., 2020) (see Supplemental Methods for details). Our alignment enables a host of analyses, including studies of sequence divergence and synteny, and we were particularly interested to compare patterns of evolution on autosomes versus the Z chromosome.

The Z chromosome is a clear outlier in these comparisons (Fig. 2B). Looking at 1 megabase windowed divergence (calculated using the phast package under the GTR model Siepel et al. (2005)) between *O. bimaculoides* and *O. sinensis* revealed that the Z chromosome is evolving significantly slower than autosomes (Mann–Whitney U test, $p = 0.0011$; Fig. 2B). The same pattern holds true for Z chromosome divergence between *O. bimaculoides* and the Hawaiian bobtail squid *Euprymna scolopes* (Mann–Whitney U test, $p < 0.0001$; Fig. 2B). We also analysed patterns of coding substitutions between *O. bimaculoides* and *O. sinensis*. While we found that the rate of nonsynonymous substitution was roughly equivalent between Z-linked and autosomal genes (Mann–Whitney U test, $p = 0.86$), Z-linked genes have decreased levels of synonymous substitution (mean Z-linked dS= 0.092, mean autosomal dS= 0.123; Mann–Whitney U test, $p < 0.00001$; Fig. S7). Reduced divergence on the Z mirrors what is observed in the primate X chromosome relative to autosomes (Hobolth et al., 2007; Scally et al., 2012), and may be caused by a variety of reasons including purging of recessive, mildly deleterious mutations from the hemizygous chromosome and sex-biased mutation rate differences between males and females. Using the ratio of the substitution rate at Z-linked versus autosomal loci allows us to estimate the degree of sex-biased mutation (Miyata et al., 1987), generally termed α. We find that for synonymous divergence $\alpha = 0.1396$ (95% C.I.= $0.029 - 0.300$) meaning that for these positions we estimate the female mutation rate to be ~ 7.2 times that of the male mutation rate. While synonymous substitutions are likely a good proxy for neutral divergence, we can get a bulk estimate of

α from the genome by simply considering divergence in large genomic windows. Using 1Mb windowed divergence yields an estimate of $\alpha = 0.539$ (95% C.I.$= 0.406 - 0.703$) indicating that the female mutation rate is ~ 1.82 times that of the male mutation rate genome-wide. This observation of female-biased mutation is notable and has been seen in fish lineages (Bergeron et al., 2023) but contrasts to the commonly-observed male bias in amniote lineages (Wilson Sayres and Makova, 2011; Bergeron et al., 2023).

Repetitive element characteristics of the Z chromosome are unique

Our whole-genome alignment allows us to examine repetitive element evolution in a phylogenetic framework. Figures 2A & 2C show a phylogenetic tree representing the relationship among these cephalopods along with the percent of the genome occupied by each of four repeat element classes: DNA transposons, LINEs, LTRs, and SINEs. It is clear from this comparison that the lineage leading to octopuses underwent a dramatic increase in the number of SINE elements, with SINEs taking up approximately 4.64% of the genome sequence in *O. bimaculoides*. The SINE expansion in octopus genomes has been noted previously with sparser comparisons (Albertin et al., 2015), however when seen in a phylogenetic light (Fig. 2) this pattern is abundantly clear.

While a historical expansion of SINE elements is a striking feature of the octopus genome, we also found an abundance of LINE elements, which composed a higher proportion of the genome than SINE elements, at 12.73%. Two clades of LINE elements, R2/R4/NeSL and RTE/Bov-B, made up the bulk of this with 3.10% and 4.59% of the genomic sequence respectively (Fig. 2C).

Our chromosome-level assembly allows us to explore if there is genomic heterogeneity in the accumulation of transposable elements (TEs). Visualization of the TE contributions to individual chromosomes (Figs. 2D, S3, S4, S5, S6) suggests that while most chromosomes have little variation in their proportions of TEs, the Z chromosome is again a notable outlier with respect to LINE elements, harboring LINEs at approximately twice the density as other chromosome arms (Mann-Whitney test; $p < 0.0001$; Fig. 2D).

LINE elements are a signature of cephalopod sex chromosomes

We hypothesized that this huge abundance of LINEs could be a clear signature of the cephalopod Z chromosome in other genomic assemblies. Looking at the landscape of TEs in the high quality, chromosome-level assemblies of *O. sinensis*, *E. scolopes*, and *S. esculenta* yielded a striking pattern: a single chromosome in each of those assemblies (chrZ in *S. esculenta*, formerly chr20 in *O. sinensis* and chr43 in *E. scolopes*) are outliers for LINE elements, perhaps suggesting the presence of an orthologous Z chromosome (Fig. 2D). As in *O. bimaculoides*, each of these chromosomes harbors a significantly greater density of LINEs in comparison to other chromosomes/putative autosomes in the genome (Mann-Whitney tests; $p < 0.0001$,

$p = 0.00023$ and $p < 0.0001$, for *O. sinensis*, *Euprymna scolopes*, and *Sepia esculenta* respectively; Table S5). As a result of these findings, we renamed the outlying chromosomes to Z in all figures for clarity. Upon inspection of the much more distant *Nautilus* genome, no such LINE element outlier chromosome could be identified. Thus evidence from LINE element enrichment suggests that this is a unique feature of the coleoid Z chromosome that may have originated before the split of the squid and octopus lineages, but after divergence from *Nautilus*.

Syntenic relationships of genes on the Z chromosome are conserved

On the basis of LINE element content, we hypothesized that the chromosomes we identified in *O. sinensis*, *E. scolopes*, and *S. esculenta* might be orthologous to the *O. bimaculoides* Z chromosome. To test this hypothesis, we compared synteny among chromosomes using gene-level annotations as revealed using the GENESPACE and MCScanX software packages (Lovell et al., 2022; Wang et al., 2012, 2024), which use orthologous gene identification to define conserved synteny blocks. As can be seen in Figures 3A and 3B, gene-based synteny confirms our hypothesis – the *O. bimaculoides* Z chromosome has a large block of conserved synteny on the putative Z chromosomes in *O. sinensis*, *E. scolopes*, and *S. esculenta*. Dotplots showing finer resolution pairwise alignments are shown in Figure S8.

Due to the huge evolutionary distances between *Nautilus pompilius* and the other cephalopod taxa studied here, coupled with the extensive genome expansion along the coleoid lineage, calling syntentic regions on the basis of colinear blocks of one-to-one orthologs across multiple taxa is challenging. Instead we used MCscanX, a method not constrained to single-copy genes, to compare chromosome Z of *O. bimaculoides* to the *N. pompilius* genome in a more targeted fashion. This revealed 1,057 collinear genes (8.31% of all genes) contained on 135 alignment blocks between *O. bimaculoides* Z-linked proteins and *N. pompilius*. Syntenic blocks are found across almost all of the *N. pompilius* chromosomes, indicative of the extensive genome and gene duplication since coleoid cephlapods diverged from *N. pompilius* (Fig. S9). Of the 300 single copy, unique Z-linked genes in *O. bimaculoides*, more than half were contained on one of two chromosomes in *N. pompilius*: 87 were contained in collinear blocks on chromosome 7 and 77 were found on *N. pompilius* chromosome 4. We note that neither of these chromosomes is unusual for LINE density (see Fig. 2D).

A single, ancient origin of the cephalopod Z chromosome

Our evidence, when taken together, demonstrates that the *O. bimaculoides* Z chromosome is a genomic outlier with clear homology at the gene level and with respect to repeat content to the putative Z chromosomes in *O. sinensis*, *E. scolopes*, and *S. esculenta*. While the divergence time between *O. bimaculoides* and *O. sinensis* is relatively modest, between 13 - 34 million years (Huang et al., 2022), divergence between Decapodiformes (squid and cuttlefish) and Octopodiformes (octopus) is much older, with an approximate divergence time between *O. bimaculoides* and *E. scolopes* of 285 - 340 million years (Huang et al.,

2022). Thus, it was imperative to examine whether the orthologous chromosomes in other cephalopod species are functioning as a Z chromosome. Figure 4 presents sequencing coverage data from four additional species, including short read libraries that we generated from sexed *Octopus bimaculatus* individuals (see methods), the sister taxa to *O. bimaculoides*, a female cuttlefish *S. esculenta* recently released by the Darwin Tree of Life project (NCBI, GCA_964036315.1), two unsexed individuals of *N. pompilius* (Zhang et al., 2021; Huang et al., 2022) and a set of eight Illumina short read libraries derived from single, unsexed embryos of *E. scolopes* from Schmidbaur et al. (2022). These results again confirmed our hypothesis in each case. In species where we have sexed individuals (*O. bimaculatus* and *S. esculenta*), the putative Z chromosome appears at approximately half coverage in the female in comparison to representative autosomes (chromosomes 1 and 10 shown as automsomal comparisons in Figs. 4A, 4B; all chromosomes shown in S10, and S11). For *Euprymna scolopes* we did not have sexed individuals, however in the eight embryos coverage among chromosomes was nearly uniform across individuals except at the putative Z (chr43). The putative Z is an outlier in comparison to the other chromosomes with multiple distinct coverage classes among individuals, suggesting females are hemizygous as observed in *O. bimaculoides* (Figs. 4C; all chromosomes shown in S12, & S13). Finally, in comparing sequencing coverage among the two available unsexed *Nautilus* genomes, we again observed a single chromosome, chr4, at reduced coverage in one of the two libraries (Fig. 4D). Chr4 is one of two chromosomes in *Nautilus* that shows synteny to the *O. bimaculoides* Z as determined by our MCScanX analysis. Thus, while we were not able to detect a Z chromosome based on LINE element density in *Nautilus*, sequencing coverage among individuals and synteny strongly suggests that a sex chromosome is present and that it shows orthology to the Z in coleoid cephalopods.

Having established strong support for homology of the Z chromosome among cephalopods, we were lastly interested in examining what genes might be shared in orthologous regions of the Z chromosome among lineages. We first compared annotations from the shared syntenic chromosome Z blocks defined by GENESPACE across coleoid cephalopods, using annotation from *O. bimaculoides*, *O. sinensis*, and *E. scolopes*. This revealed 19 unique protein coding loci shared among these three taxa that are housed in this region of the genome. We did blastp homology searches of these genes to human proteins and found 16 of 19 had strong hits (Table S6). Using publicly available summaries from GeneCards (Stelzer et al., 2016) we report that all 16 show mRNA expression in human reproductive tissues, and 15 of 16 of these show protein expression in human reproductive tissues. A particularly leading hit is the protein obimac0008950, which shows strong homology to the human Sperm Associated Antigen 9 (SPAG9; e-val=0.0). We next asked if the 19 proteins we identified above also occur in the colinear blocks identified between the Z in *O. bimaculoides* and *N. pompilius* and found that 9/19 were present, and 6 of these 9 could be found in humans. All 6 show evidence of mRNA expression and 5 of these 6 show protein expression in human reproductive tissues. Thus we speculate that the genes retained on the Z chromosome may represent an ancient set of proteins essential to animal reproduction and/or gametogenesis.

Our results provide the first glimpse of sex determination in cephalopods, a phenomenon which until now has remained a mystery. The clear presence of a hemizygous Z chromosome in females suggests a ZZ/ZO sex determination system that had previously been missed in cytological comparisons, likely as a result of the lack of heterogamy. Taken together, our evidence suggests that the cephalopod Z chromosome evolved once in the lineage leading

to the common ancestor of all extant lineages. Recent molecular estimates of the time to the common ancestor of *Nautilus* and coleoids place it at $\sim$ 482 million years ago (range 398-520 Mya; Huang et al. (2022)) thus the Z chromosome would had to have originated at least this long ago. This is an astoundingly long time for a sex chromosome to be preserved (Charlesworth et al., 2005). A few other ancient sex chromosomes have been described previously from liverworts ($\sim$ 430Mya; Iwasaki et al. (2021)) and mosses ($\sim$ 300Mya; Carey et al. (2021)), and there is evidence that the insect X chromosome is quite ancient ($\sim$ 450Mya; Toups and Vicoso (2023)), although it shows rapid evolution in some clades. Thus in context, the cephalopod Z chromosome is among the very oldest sex chromosomes yet described.

Acknowledgements

We thank Caroline Albertin and Matthew Birk for providing samples, Mara Lawniczak for her input on the project and comments on the manuscript, and John Postlethwait, Peter Ralph, Melissa Toups, Graham Coop, Clara Rehmann, and members of the Kern Ralph co-lab for their input and comments on this manuscript. Sequencing and sample preparation was performed by the University of Oregon Genomics & Cell Characterization Core Facility. GCC was funded by NSF GRFP 1842486. ADK was funded in part by NIH awards R35148253 and R01HG010774. CMN was funded in part by NIH award R01NS118466. CMN and ACM were funded in part by a University of Oregon Renee James Seed Grant.

Author Contributions

G.C.C. and A.D.K designed the study. J.O.S.C., D.M.P., and J.R.P. prepared samples for sequencing. G.C.C. assembled the genome, G.C.C and S.T. annotated the genome, S.T. conducted the repeat analysis and generated the multiple species alignment, G.C.C. and S.T.S. conducted the synteny analyses. A.D.K did homology searches. G.C.C, S.T., S.T.S, and A.D.K wrote the manuscript. A.C.M, C.M.N, and A.D.K. helped supervise the project. All authors read and approved the final manuscript.

Data availability

The genome assembly and associated sequencing reads have been deposited to BioProject ID PRJNA1076263. Additional data have been deposited to a Zenodo repository at 10.5281/zenodo.13278524.

References

Caroline B Albertin, Oleg Simakov, Therese Mitros, Z Yan Wang, Judit R Pungor, Eric Edsinger-Gonzales, Sydney Brenner, Clifton W Ragsdale, and Daniel S Rokhsar. The oc-

topus genome and the evolution of cephalopod neural and morphological novelties. *Nature*, 524(7564):220–224, 2015.

Caroline B Albertin, Sofia Medina-Ruiz, Therese Mitros, Hannah Schmidbaur, Gustavo Sanchez, Z Yan Wang, Jane Grimwood, Joshua JC Rosenthal, Clifton W Ragsdale, Oleg Simakov, et al. Genome and transcriptome mechanisms driving cephalopod evolution. *Nature communications*, 13(1):2427, 2022a.

Caroline B Albertin, Sofia Medina-Ruiz, Therese Mitros, Hannah Schmidbaur, Gustavo Sanchez, Z Yan Wang, Jane Grimwood, Joshua JC Rosenthal, Clifton W Ragsdale, Oleg Simakov, et al. Genome and transcriptome mechanisms driving cephalopod evolution. *Nature communications*, 13(1):2427, 2022b.

Joel Armstrong, Glenn Hickey, Mark Diekhans, Ian T Fiddes, Adam M Novak, Alden Deran, Qi Fang, Duo Xie, Shaohong Feng, Josefin Stiller, et al. Progressive cactus is a multiple-genome aligner for the thousand-genome era. *Nature*, 587(7833):246–251, 2020.

Lucie A Bergeron, Søren Besenbacher, Jiao Zheng, Panyi Li, Mads Frost Bertelsen, Benoit Quintard, Joseph I Hoffman, Zhipeng Li, Judy St. Leger, Changwei Shao, et al. Evolution of the germline mutation rate across vertebrates. *Nature*, 615(7951):285–291, 2023.

Tomas Bruna, Katharina J Hoff, Alexandre Lomsadze, Mario Stanke, and Mark Borodovsky. Braker2: automatic eukaryotic genome annotation with genemark-ep+ and augustus supported by a protein database. *NAR genomics and bioinformatics*, 3(1):lqaa108, 2021.

Sarah B Carey, Jerry Jenkins, John T Lovell, Florian Maumus, Avinash Sreedasyam, Adam C Payton, Shengqiang Shu, George P Tiley, Noe Fernandez-Pozo, Adam Healey, et al. Gene-rich uv sex chromosomes harbor conserved regulators of sexual development. *Science Advances*, 7(27):eabh2488, 2021.

D Charlesworth, B Charlesworth, and Gabriel Marais. Steps in the evolution of heteromorphic sex chromosomes. *Heredity*, 95(2):118–128, 2005.

Deborah Charlesworth. Plant sex chromosome evolution. *Journal of experimental botany*, 64(2):405–420, 2013.

Nansheng Chen. Using repeat masker to identify repetitive elements in genomic sequences. *Current protocols in bioinformatics*, 5(1):4–10, 2004.

Haoyu Cheng, Gregory T Concepcion, Xiaowen Feng, Haowen Zhang, and Heng Li. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nature methods*, 18(2):170–175, 2021.

Peter JA Cock, Tiago Antao, Jeffrey T Chang, Brad A Chapman, Cymon J Cox, Andrew Dalke, Iddo Friedberg, Thomas Hamelryck, Frank Kauff, Bartek Wilczynski, et al. Biopython: freely available python tools for computational molecular biology and bioinformatics. *Bioinformatics*, 25(11):1422, 2009.

Rute R Da Fonseca, Alvarina Couto, Andre M Machado, Brona Brejova, Carolin B Albertin, Filipe Silva, Paul Gardner, Tobias Baril, Alex Hayward, Alexandre Campos, et al. A draft genome sequence of the elusive giant squid, architeuthis dux. *GigaScience*, 9(1):giz152, 2020.

J Dainat. Agat: Another gff analysis toolkit to handle annotations in any gtf/gff format. *Zenodo*, 2024.

Olga Dudchenko, Sanjit S Batra, Arina D Omer, Sarah K Nyquist, Marie Hoeger, Neva C Durand, Muhammad S Shamim, Ido Machol, Eric S Lander, Aviva Presser Aiden, et al. De novo assembly of the aedes aegypti genome using hi-c yields chromosome-length scaffolds. *Science*, 356(6333):92–95, 2017.

Neva C Durand, Muhammad S Shamim, Ido Machol, Suhas SP Rao, Miriam H Huntley, Eric S Lander, and Erez Lieberman Aiden. Juicer provides a one-click system for analyzing loop-resolution hi-c experiments. *Cell systems*, 3(1):95–98, 2016.

David M. Emms and Steven Kelly. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biology*, 20(1):238, 2019. ISSN 1474-760X. doi: 10.1186/s13059-019-1832-y. URL `https://doi.org/10.1186/s13059-019-1832-y`.

Jullien M Flynn, Robert Hubley, Clément Goubert, Jeb Rosen, Andrew G Clark, Cédric Feschotte, and Arian F Smit. Repeatmodeler2 for automated genomic discovery of transposable element families. *Proceedings of the National Academy of Sciences*, 117(17):9451–9457, 2020.

Dengfeng Guan, Shane A McCarthy, Jonathan Wood, Kerstin Howe, Yadong Wang, and Richard Durbin. Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics*, 36(9):2896–2898, 2020.

Roger Hanlon. Cephalopod dynamic camouflage. *Current biology*, 17(11):R400–R404, 2007.

Roger T. Hanlon and John B. Messenger. *Cephalopod Behaviour*. Cambridge University Press, 2 edition, 2018. doi: 10.1017/9780511843600.

Asger Hobolth, Ole F Christensen, Thomas Mailund, and Mikkel H Schierup. Genomic relationships and speciation times of human, chimpanzee, and gorilla inferred from a coalescent hidden markov model. *PLoS genetics*, 3(2):e7, 2007.

Da Wei Huang, Brad T. Sherman, and Richard A. Lempicki. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. 4(1):44–57. ISSN 1750-2799. doi: 10.1038/nprot.2008.211. URL `https://www.nature.com/articles/nprot.2008.211`. Number: 1 Publisher: Nature Publishing Group.

Zekun Huang, Wanlong Huang, Xiaolin Liu, Zhaofang Han, Guangjian Liu, Grace Afumwaa Boamah, Yi Wang, Feng Yu, Yang Gan, Qizhen Xiao, et al. Genomic insights into the adaptation and evolution of the nautilus, an ancient but evolving "living fossil". *Molecular ecology resources*, 22(1):15–27, 2022.

Melissa J Hubisz, Katherine S Pollard, and Adam Siepel. Phast and rphast: phylogenetic analysis with space/time models. *Briefings in bioinformatics*, 12(1):41–51, 2011.

Miyuki Iwasaki, Tomoaki Kajiwara, Yukiko Yasui, Yoshihiro Yoshitake, Motoki Miyazaki, Shogo Kawamura, Noriyuki Suetsugu, Ryuichi Nishihama, Shohei Yamaoka, Dierk Wanke, et al. Identification of the sex-determining factor in the liverwort marchantia polymorpha reveals unique evolution of sex chromosomes in a haploid system. *Current Biology*, 31(24): 5522–5532, 2021.

Norman A Johnson and Joseph Lachance. The genetics of sex chromosomes: evolution and implications for hybrid incompatibility. *Annals of the New York Academy of Sciences*, 1256(1):E1–E22, 2012.

Bo-Mi Kim, Seunghyun Kang, Do-Hwan Ahn, Seung-Hyun Jung, Hwanseok Rhee, Jong Su Yoo, Jong-Eun Lee, SeungJae Lee, Yong-Hee Han, Kyoung-Bin Ryu, et al. The genome of common long-arm octopus octopus minor. *Gigascience*, 7(11):giy119, 2018.

Fenghui Li, Li Bian, Jianlong Ge, Fengming Han, Zhihong Liu, Xuming Li, Yongsheng Liu, Zhishu Lin, Huilai Shi, Changlin Liu, et al. Chromosome-level genome assembly of the east asian common octopus (octopus sinensis) using pacbio sequencing and hi-c technology. *Molecular ecology resources*, 20(6):1572–1582, 2020.

Heng Li. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*, 34(18): 3094–3100, 2018.

John T Lovell, Avinash Sreedasyam, M Eric Schranz, Melissa Wilson, Joseph W Carlson, Alex Harkess, David Emms, David M Goodstein, and Jeremy Schmutz. Genespace tracks regions of interest and gene copy number variation across multiple genomes. *Elife*, 11: e78526, 2022.

Guillaume Marçais and Carl Kingsford. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics*, 27(6):764–770, 2011.

Anand Mayakonda, De-Chen Lin, Yassen Assenov, Christoph Plass, and H Phillip Koeffler. Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome research*, 28(11):1747–1756, 2018.

T Miyata, H Hayashida, K Kuma, K Mitsuyasu, and T Yasunaga. Male-driven molecular evolution: a model and nucleotide sequence analysis. In *Cold Spring Harbor symposia on quantitative biology*, volume 52, pages 863–867. Cold Spring Harbor Laboratory Press, 1987.

Emmanuel Paradis and Klaus Schliep. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in r. *Bioinformatics*, 35(3):526–528, 2019.

Brent S Pedersen and Aaron R Quinlan. Mosdepth: quick coverage calculation for genomes and exomes. *Bioinformatics*, 34(5):867–868, 2018.

Winston Ponder and David R Lindberg. *Phylogeny and Evolution of the Mollusca*. Univ of California Press, 2008.

Aaron R Quinlan and Ira M Hall. Bedtools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6):841–842, 2010.

T Rhyker Ranallo-Benavidez, Kamil S Jaron, and Michael C Schatz. Genomescope 2.0 and smudgeplot for reference-free profiling of polyploid genomes. *Nature communications*, 11 (1):1432, 2020.

Vincent Ranwez, Nathalie Chantret, and Frédéric Delsuc. Aligning protein-coding nucleotide sequences with macse. *Multiple Sequence Alignment: Methods and Protocols*, 2231:51–70, 2021.

Aylwyn Scally, Julien Y Dutheil, LaDeana W Hillier, Gregory E Jordan, Ian Goodhead, Javier Herrero, Asger Hobolth, Tuuli Lappalainen, Thomas Mailund, Tomas Marques-Bonet, et al. Insights into hominid evolution from the gorilla genome sequence. *Nature*, 483(7388):169–175, 2012.

Hannah Schmidbaur, Akane Kawaguchi, Tereza Clarence, Xiao Fu, Oi Pui Hoang, Bob Zimmermann, Elena A Ritschard, Anton Weissenbacher, Jamie S Foster, Spencer V Nyholm, et al. Emergence of novel cephalopod gene regulation and expression through large-scale genome reorganization. *Nature communications*, 13(1):2172, 2022.

Alaina Shumate and Steven L Salzberg. Liftoff: accurate mapping of gene annotations. *Bioinformatics*, 37(12):1639–1643, 2021.

Adam Siepel, Gill Bejerano, Jakob S Pedersen, Angie S Hinrichs, Minmei Hou, Kate Rosenbloom, Hiram Clawson, John Spieth, LaDeana W Hillier, Stephen Richards, et al. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome research*, 15(8):1034–1050, 2005.

Weiwei Song, Ronghua Li, Yun Zhao, Herve Migaud, Chunlin Wang, and Michaël Bekaert. Pharaoh cuttlefish, sepia pharaonis, genome reveals unique reflectin camouflage gene set. *Frontiers in Marine Science*, 8:639670, 2021.

Jeremea O Songco-Casey, Gabrielle C Coffing, Denise M Piscopo, Judit R Pungor, Andrew D Kern, Adam C Miller, and Cristopher M Niell. Cell types and molecular architecture of the octopus bimaculoides visual system. *Current Biology*, 32(23):5031–5044, 2022.

Gil Stelzer, Naomi Rosen, Inbar Plaschkes, Shahar Zimmerman, Michal Twik, Simon Fishilevich, Tsippi Iny Stein, Ron Nudel, Iris Lieder, Yaron Mazor, et al. The genecards suite: from gene data mining to disease genome sequence analyses. *Current protocols in bioinformatics*, 54(1):1–30, 2016.

Manuel Tardaguila, Lorena De La Fuente, Cristina Marti, Cécile Pereira, Francisco Jose Pardo-Palacios, Hector Del Risco, Marc Ferrell, Maravillas Mellado, Marissa Macchietto,

Kenneth Verheggen, et al. Sqanti: extensive characterization of long-read transcript sequences for quality control in full-length transcriptome identification and quantification. *Genome research*, 28(3):396–411, 2018.

Melissa A Toups and Beatriz Vicoso. The x chromosome of insects likely predates the origin of class insecta. *Evolution*, 77(11):2504–2511, 2023.

Md Vasimuddin, Sanchit Misra, Heng Li, and Srinivas Aluru. Efficient architecture-aware acceleration of bwa-mem for multicore systems. In *2019 IEEE international parallel and distributed processing symposium (IPDPS)*, pages 314–324. IEEE, 2019.

Jin-hai Wang and Xiao-dong Zheng. Cytogenetic studies in three octopods, octopusminor, amphioctopusfangsiao, and cistopuschinensis from the coast of china. *Comparative cytogenetics*, 12(3):373, 2018.

Yupeng Wang, Haibao Tang, Jeremy D DeBarry, Xu Tan, Jingping Li, Xiyin Wang, Taeho Lee, Huizhe Jin, Barry Marler, Hui Guo, et al. Mcscanx: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic acids research*, 40(7): e49–e49, 2012.

Yupeng Wang, Haibao Tang, Xiyin Wang, Ying Sun, Paule V Joseph, and Andrew H Paterson. Detection of colinear blocks and synteny and evolutionary analyses based on utilization of mcscanx. *Nature Protocols*, (19):1–24, 2024.

Brooke L Whitelaw, Ira R Cooke, Julian Finn, Rute R da Fonseca, Elena A Ritschard, MTP Gilbert, Oleg Simakov, and Jan M Strugnell. Adaptive venom evolution and toxicity in octopods is driven by extensive novel gene formation, expansion, and loss. *GigaScience*, 9 (11):giaa120, 2020.

Melissa A Wilson Sayres and Kateryna D Makova. Genome analyses substantiate male mutation bias in many species. *Bioessays*, 33(12):938–945, 2011.

John Zachary Young. *The anatomy of the nervous system of Octopus vulgaris*. Clarendon Press, 1971.

Yang Zhang, Fan Mao, Huawei Mu, Minwei Huang, Yongbo Bao, Lili Wang, Nai-Kei Wong, Shu Xiao, He Dai, Zhiming Xiang, et al. The genome of nautilus pompilius illuminates eye evolution and biomineralization. *Nature Ecology & Evolution*, 5(7):927–938, 2021.

Main Figures

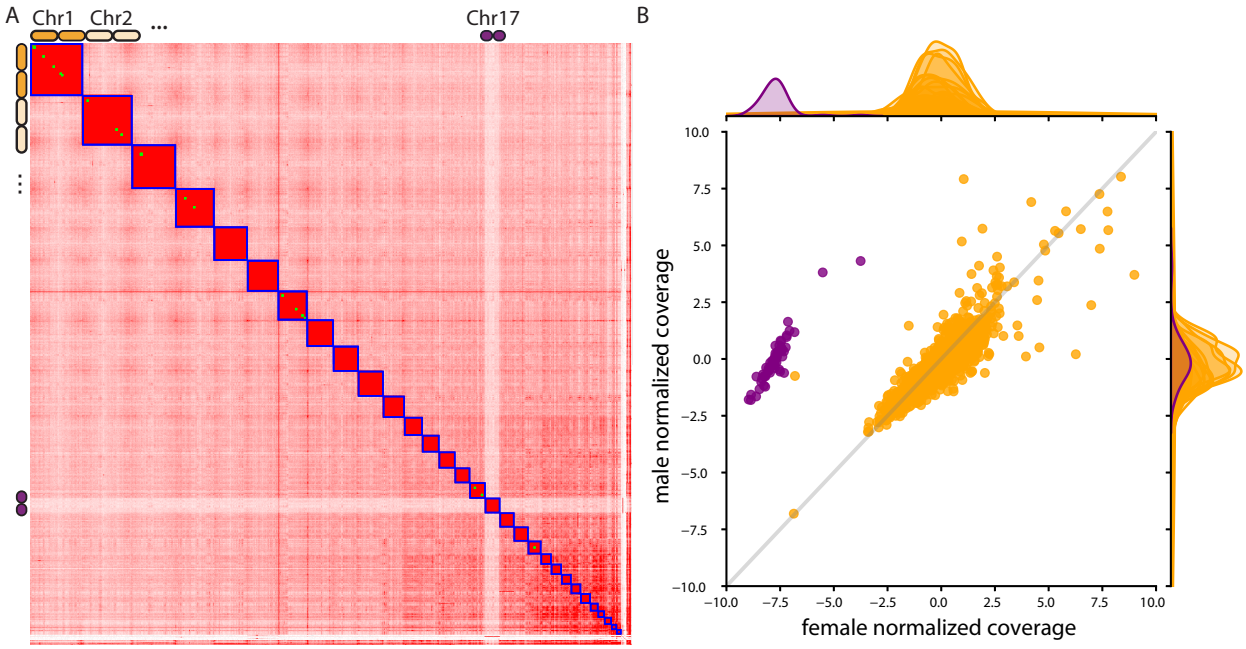

Figure 1: Sequencing data showing half coverage at chromosome 17 of *O. bimaculoides*. **A.** Hi-C contact map from our assembly of a female *O. bimaculoides*. Chromosome 17, our putative Z chromosome, is highlighted as a clear outlier in coverage. **B.** Normalized read depth of male and female whole-genome Illumina short read data. Purple points are from chromosome 17, whereas every other scaffold is shaded in orange. Scatterplot showing male vs. female normalized coverage, while the density estimates on the margins are from female (top) and male (side) separately.

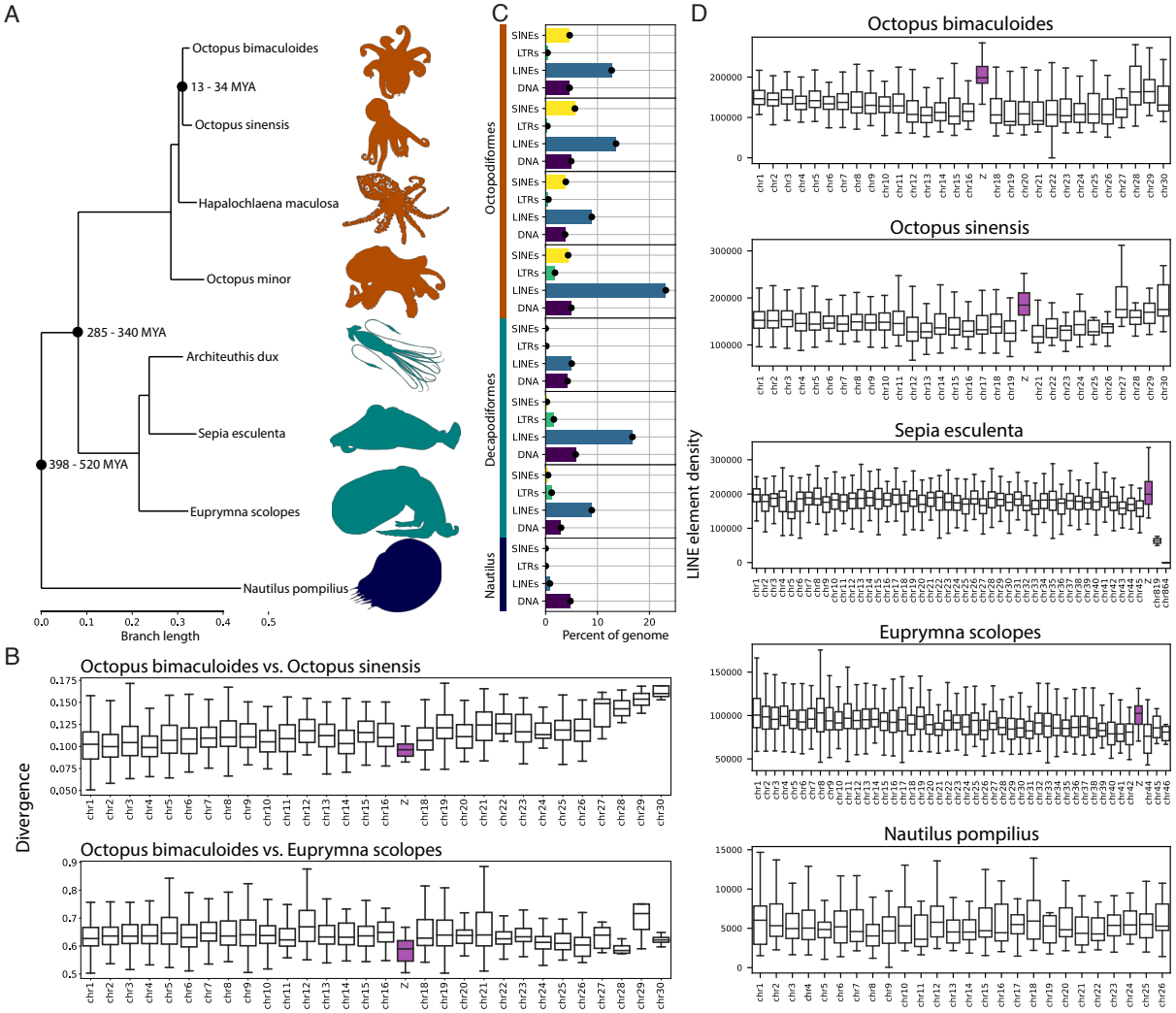

Figure 2: **A.** Cephalopod phylogeny used in this study. Phylogeny inferred using one-to-one protein orthologs identified by OrthoFinder Emms and Kelly (2019). Branch lengths in units of amino acid substitutions per site. Divergence dates are from (Huang et al., 2022). **B.** Windowed pairwise divergence calculated between *O. bimaculoides* and *O. sinesis* (top panel) and *O. bimaculoides* and *E. scolopes* (bottom panel). For these analysis 1Mb non-overlapping windows were used using our a multiple alignment extracted with *O. bimaculoides* as the reference. Chromosome 17, our presumptive Z chromosome is highlighted in purple. **C.** Genomic repeats associated with each leaf taxon. Repeats show percent of genome composed of DNA transposons, LINE, LTR, and SINE elements for each species. **D.** LINE content across genomes of five cephalopod species. In each of the coleoid species, a single outlying chromosome with respect to LINE density is observed. These chromosomes are highlighted in purple and renamed to Z. In *N. pompilius*, the presumptive Z chromosome does not have an outlying LINE density.

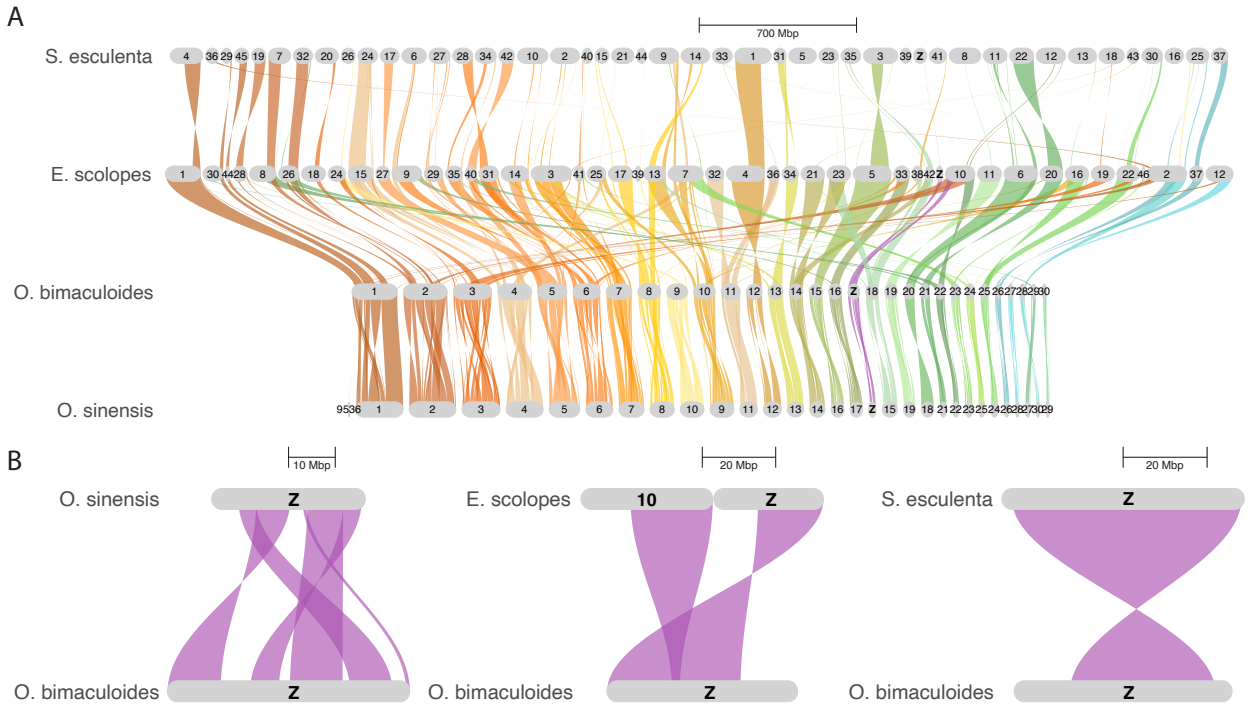

Figure 3: Syntenic relationships between four cephalopod species. **A.** Riparian synteny plot generated using orthogroups between the chromosomes of representative octopus, squid, and cuttlefish species. Ribbons connected to the *O.bimaculoides* Z chromosome are highlighted in purple. **B.** Paired riparian synteny plots between the Z chromosomes of *O. bimaculoides* and three other species, *O. sinensis*, *E. scolopes*, and *S. esculenta*, reveal conservation of the Z chromosome.

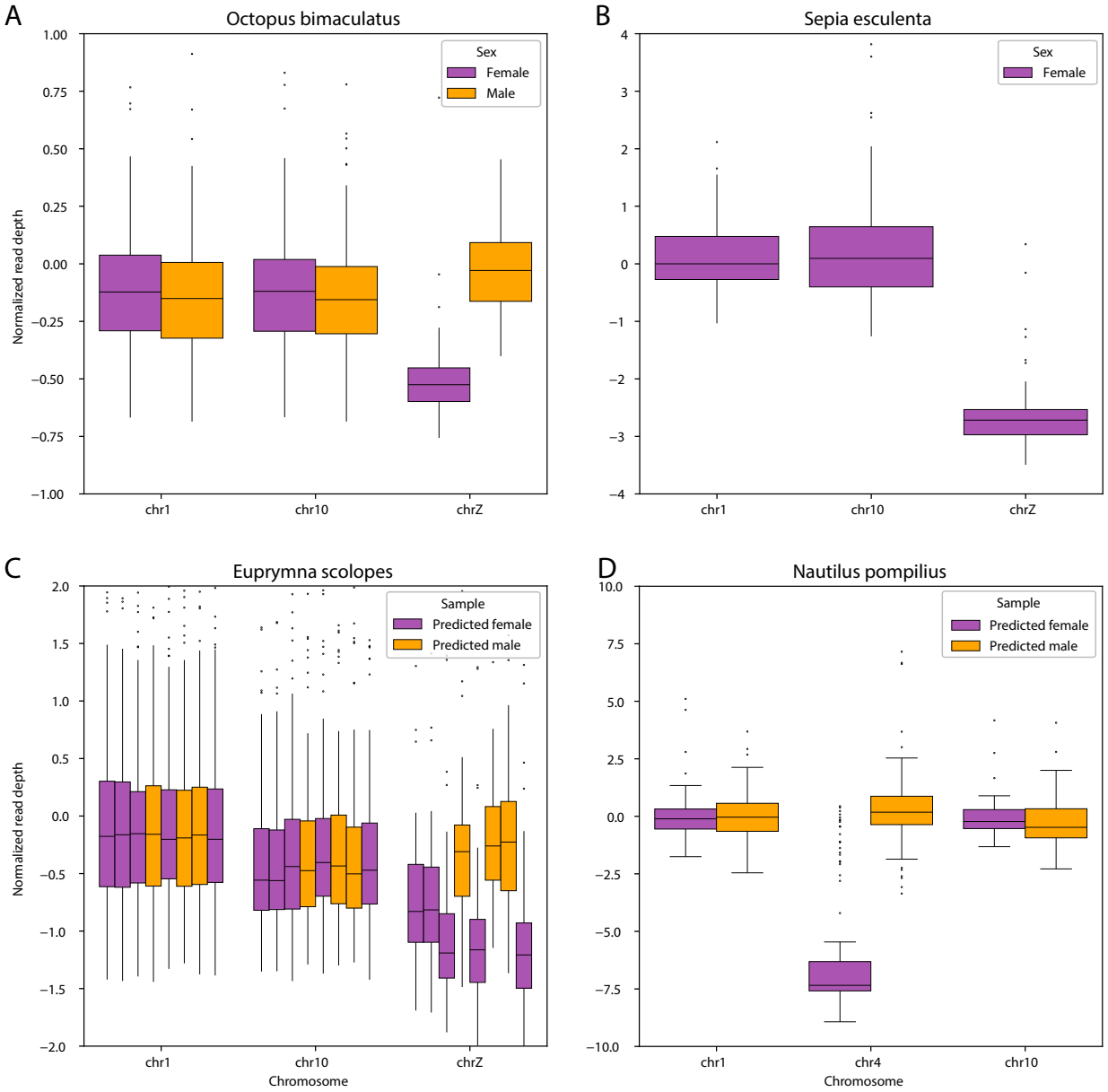

Figure 4: Normalized coverage of sequencing data mapped to representative chromosomes in *O. bimaculatus*, *S. esculenta*, *N. pompillius*, and *E. scolopes*. **A.** For *O. bimaculatus*, we generated Illumina short reads from two male and two female individuals (See Methods for more details). **B.** Female *S. esculenta* short reads were generated by Darwin Tree of Life Project and downloaded from NCBI (GCA_964036315.1). **C.** Sequencing data for unsexed embryos of*E. scolopes* were generated by Schmidbaur et al. (2022). The putative Z chromosome is the only chromosome with variation in coverage across individuals (see Figs S12 & S13). Predicted female individuals (purple) have significantly lower coverage at the presumed Z chromosome compared to the average coverage of all other chromosomes ($p < 0.0005$ in all cases; Mann Whitney U tests) and predicted males (orange) do not (N.S.; Mann-Whitney U tests;). **D.** In *N. pompilius*, sequencing data from the predicted female individual were generated by Zhang et al. (2021) and sequencing data from the predicted male individual were generated by Huang et al. (2022).

Supplemental Materials

Materials and Methods

Sample collection

A description of the genome sample collection and sequencing was published previously in Songco-Casey et al. (2022). Optic lobe tissue was collected from an adult female *O. bimaculoides* that was obtained from the southern coast of California. Optic lobe, central brain, retina, and arm tissue were dissected from related 6-week-old juvenile octopuses to be used for construction of an Iso-Seq library. Additional arm tissue was dissected to construct an Illumina library to be used in Hi-C sequencing. Four additional unrelated individuals were collected from the southern coast of California for sex chromosome coverage analysis. Optic lobe tissue was extracted from two individuals, gill tissue from one individual, and testes from one individual (Table S3).

Library construction and sequencing

All tissue samples were sent to the University of Oregon Genomics & Cell Characterization Core Facility (GC3F) for DNA extraction, sample preparation, and sequencing. High molecular weight genomic DNA was extracted with a Nanobind Tissue Big DNA kit (Circuloimics) from the adult optic lobe tissue. A SMRTbell Express Template Prep Kit 2.0 was used to construct a Pacific Bioscience standard HiFi library. Megaruptor 2 instrument was used to shear genomic DNA at 20kb target size and BluePippin size selection (Sage Science) omitted the smallest fragments (<10-14kb). Two HiFi genomic circular consensus sequencing (CCS) SMRTbell libraries were prepared as input for five HiFi SMRT cells. For Iso-Seq sequencing, a RNeasy Plus Mini Kit (QIAGEN) was used to conduct RNA extractions. GC3F prepared a multiplexed SMRTbell Iso-Seq library according to PacBio's protocol to be used as input for one SMRT cell. Single molecule sequencing of HiFi and Iso-Seq libraries was conducted with a Pacific Biosciences Sequel II system (Table S3). For scaffolding, we constructed one Illumina library with a Phase Hi-C kit to produce two NovaSeq 6000 sequencing Hi-C runs (Table S2). For the male vs. female coverage analysis, genomic DNA from each individual was barcoded with GC3F's TruSeq-style adaptors. The barcoded DNA was pooled and sequenced on two Illumina NovaSeq 6000 S4 lanes.

Post sequencing, the genomic HiFi reads were processed with SMRT Link to generate 5.8 million HiFi reads. The IsoSeq subreads were processed with the IsoSeq3 v.3.0.8 pipeline. Briefly, CCS generated HiFi reads. Primers were removed and barcodes were identified with lima. Reads were further refined by trimming poly(A) tails and removing concatemers. Finally, reads were clustered and collapsed into unique isoforms with parameters –min-aln-coverage 0.75 –min-aln-identity 0.75.

Genome size estimation and assembly

The *O. bimaculoides* genomic HiFi reads were used to estimate genome size. We generated a k-mer frequency distribution with JellyFish v.2.3.0 (Marçais and Kingsford, 2011), which

was input into GenomeScope v.2.0 (Ranallo-Benavidez et al., 2020) to generate genome size estimation plots. Estimation with 21-mers estimated the genome to be 2.0Gb with 0.838% heterozygosity. Estimation with 31-mers estimated the genome to be 2.03Gb with 0.653% heterozygosity (Fig. S14).

Initial genome assembly was conducted using HiFiasm v.0.15.5-r352 (Cheng et al., 2021) with HiFi reads and standard parameters. Hifiasm did not sufficiently remove duplicates, so purge_dups v.1.2.5 (Guan et al., 2020) was used to purge haplotigs and overlaps. We placed the resulting scaffolds into a chromosome-level assembly using the data generated from the Hi-C library. The SLURM-compatible scripts of Juicer v.1.6 (Durand et al., 2016) were used to identify Hi-C contact points on the genome. 3D de novo assembly (3D-DNA) pipeline v.190716 (Dudchenko et al., 2017) was used to correct genome mis-assemblies, anchor order of chromosomes, and orient genomic fragments.

Genome annotation

Transposable elements were identified using Repeat Modeler (Flynn et al., 2020) and were annotated with Repeat Masker v.4.1.5 (Chen, 2004). Protein-coding genes were annotated using both existing RNA-seq data and our newly generated Iso-seq data. Using existing RNA-seq databases, we ran Braker v.2.1.6 (Bruna et al., 2021) to annotate. Separately, we generated annotations with the Iso-seq reads using SQANTI3 v.5.1 (Tardaguila et al., 2018). The full-length isoforms generated by IsoSeq3 were filtered with SQANTI3 filter using default parameters. After generating the RNA-seq and Iso-seq based annotation files, they were merged using TAMA merge with parameters -e longest_ends -d merge_dup. Genome annotation statistics that were calculated with AGAT's agat_sq_stat_basic.pl script (Dainat, 2024) are presented in Table S2.

Orthology inference

OrthoFinder v.2.5.4 was used with default parameters to cluster the protein sequences of *O. bimaculoides* into ortholog groups using sequences from seven additional cephalopod species including *Hapalochlaena maculosa*, *Octopus minor*, *Octopus sinensis*, *Architeuthis dux*, *Euprymna scolopes*, *Sepia pharaonis*, and *Nautilus pompilius*. The rooted species tree was used as input for the multiple species genome alignment.

Multiple species alignment

We constructed whole genome alignments of eight cephalapod species with available reference genomes (Table S5) using progressive cactus (v2.3.0 via Docker container) (Armstrong et al., 2020). We used the whole genome alignments and phastCons v.1.5 (Hubisz et al., 2011) to identify highly conserved elements on each scaffold. This alignment, as well as our new reference genome and annotation are available as a UCSC genome assembly hub from http: `http://poppy.uoregon.edu/~ssmall/ucsc_genome_hubs/v2.1.0/Octo.v2.1.0/hub.txt`

Coverage calculations

To estimate read coverage of each species, we aligned short reads with bwa-mem2 (Vasimuddin et al., 2019) and long reads with minimap2 (Li, 2018) to the respective chromosome-scale genomes. Sequencing depth was estimated using mosdepth (Pedersen and Quinlan, 2018) in windows of 500,000bp. Plotted coverages of all species were normalized by the median coverage of the first chromosome.

Annotation of *Sepia esculenta*

At the time of these analyses, a genome annotation of *Sepia esculenta* was not yet publicly available. In order to include *Sepia esculenta* into the synteny analysis we lifted over the genome annotation from *Sepia pharaonis* (Song et al., 2021) onto *Sepia esculenta* using the software package Liftoff with default settings (Shumate and Salzberg, 2021).

Chromosome renaming

To simplify our analyses and plots we renamed all full-length chromosomes in all of the genomes that we included in this manuscript. We renamed chromosomes from longest to shortest with the naming scheme chr1, chr2, chr3, etc. with the exception of putative sex chromosomes which were renamed as chrZ.

Assessing sequence divergence among scaffolds

We converted the hierarchical alignment (hal) generated by progressive cactus to maf with hal2maf. We filtered the full, repeat-masked alignment into 1Mb windows along the genome using the phast utility msa_view, where windows were made with bedtools makewindows v2.31.0 (Quinlan and Hall, 2010). We furthered filtered alignments with multiple sequences per species using maftools mafDuplicateFilter v.0.1 (Mayakonda et al., 2018). We used phyloFit to construct phylogenies on each resulting 1Mb alignment region, only using alignments that had less than 40% of positions with gaps using the GTR mutation model. We calculated divergence as the cophenetic distance between *O. sinensis* and *O. bimaculoides* using the cophenetic.phylo function from the R package ape v.5.7-1 (Paradis and Schliep, 2019). The ratio of male to female mutation rates, α, was estimated according to (Miyata et al., 1987). Let R represent the ratio of Z-linked divergence to autosomal divergence. Then $\alpha = \frac{3R-2}{4-3R}$. α was estimated separately for mean dS at the invidual gene level as well as for mean divergence in 1Mb genomic windows. Confidence intervals on α were computed using bootstrap resampling of the observed data.

Synteny analyses

Synteny analyses between all chromosomes of *O. bimaculoides*, *O. sinensis*, *E. scolopes* and *S. esculenta* were conducted using the R package GENESPACE v.1.2.3 (Lovell et al., 2022). *O. bimaculoides* was used as the reference species for all synteny maps. In all riparian plots

and dotplots, chromosomes are scaled by physical position (useOrder = FALSE parameter in GENESPACE).

To identify the macrosynteny relationships between *N pompilius* and other cephalopod species, we performed a macrosynteny analysis using MCScanX (Wang et al., 2012). Input files were prepared following protocols described in (Wang et al., 2024) and using protein sequences assigned to the Z chromosome in *O. bimaculoides* and 15,086 *N pompilius* proteins assigned to chromosomes (Huang et al., 2022). MCScanX was then run with default parameters and results visualized using scripts included with MCScanX program (www.github.com/wyp1125/MCScanX).

dN and dS estimations

To estimate dN and dS, we used protein alignments of *O. bimaculoides* and *O. sinensis*. Proteins of 8,245 single copy orthologs previously identified using OrthoFinder v2.5.4 (Emms and Kelly, 2019), were re-aligned using MACSE v2.07 (Ranwez et al., 2021), with -prog alignSequences and default options. The module cal_dn_ds from BioPython (Cock et al., 2009) was used to estimate the non-synonymous and synonymous substitution in the alignment with options 'method=NG86' and estimating the transition/transversion rate ratio using custom code (see data availability).

GO term enrichment analysis

We conducted a GO term enrichment analysis of *O. bimaculoides* Z genes with the entire genome as the background with DAVID (Huang et al.). No functional clusters were found to be significantly enriched.

Accession numbers for data used in this study

Accession numbers, sources, and citations for all data used and generated in this study are given in Table S7.

Supplemental Figures

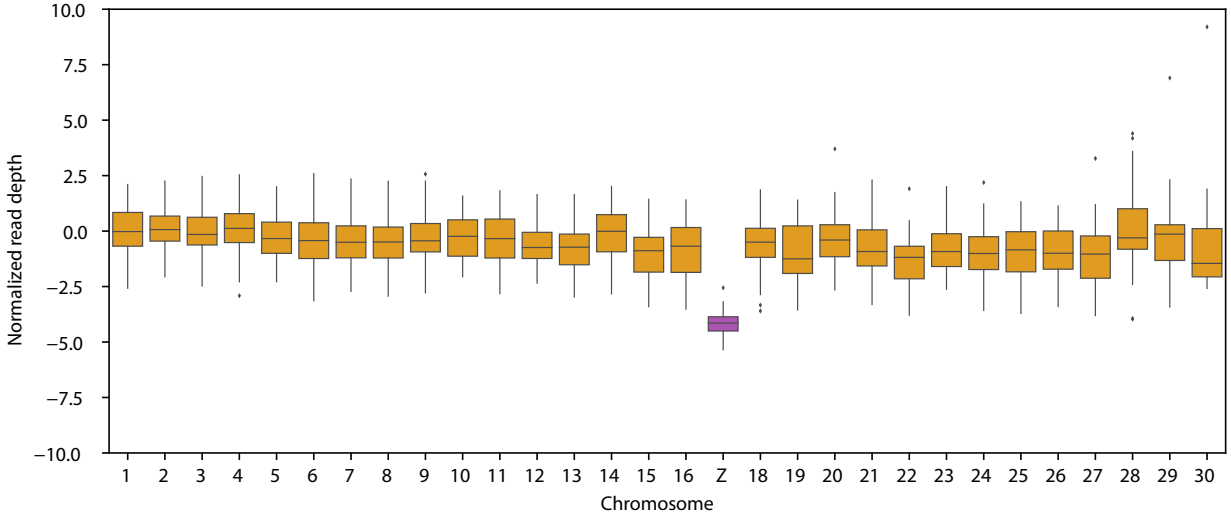

Figure S1: Normalized read depth of PacBio HiFi sequencing data across chromosomes from a female *O. bimaculoides.*

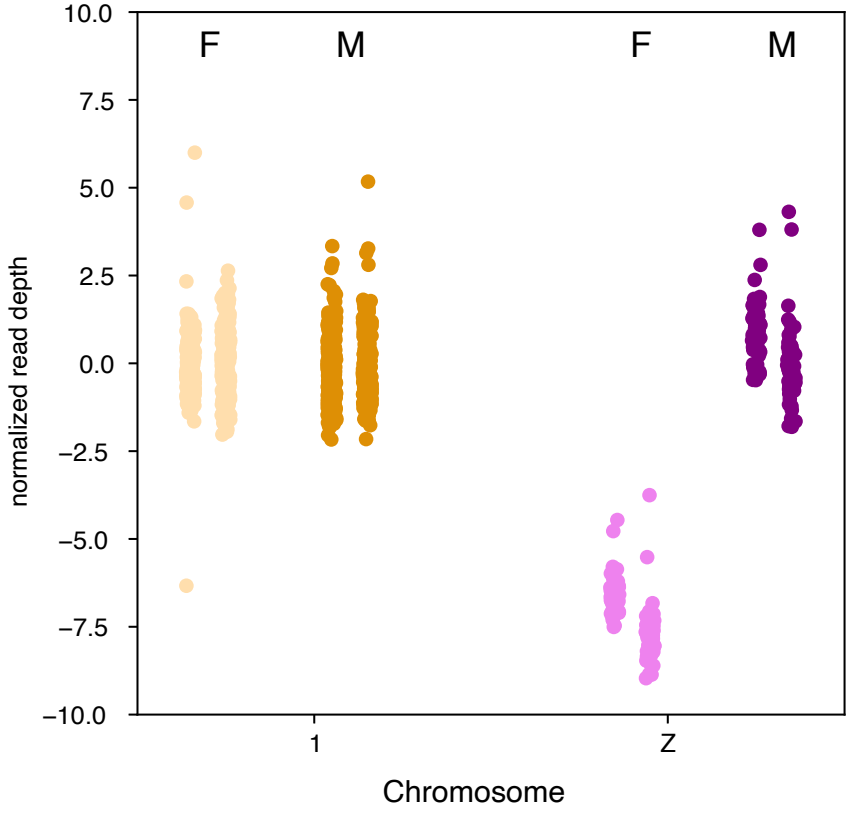

Figure S2: Normalized coverage of known sex individuals of *O. bimaculoides*. Includes two male (darker shades; labelled 'M') and two female (lighter shades; labelled 'F') with short-read sequencing. *O. bimaculoides* individuals at chromosomes 1 and Z are colored to reflect designation in main text. Both female individuals show hemizygous coverage at chromosome Z, whereas both male individuals show the same dosage at chromosome Z as at chromosome 1. See Table S3 for sequencing library information.

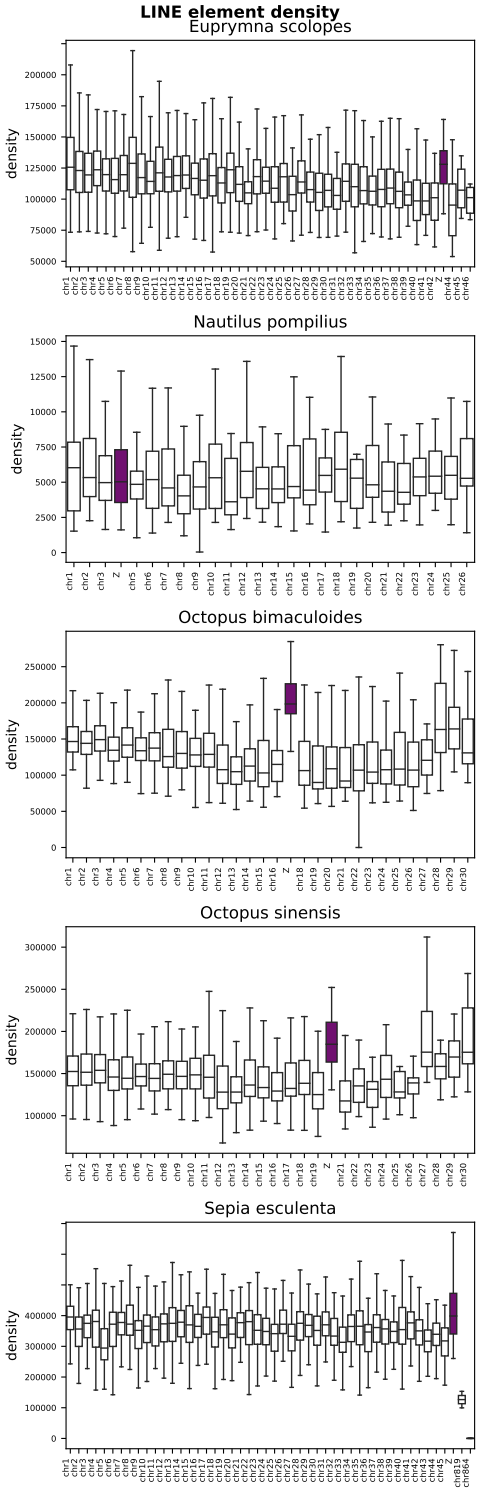

Figure S3: Density of LINE elements content on chromosomes arms across cephalopods. Data as in Figure 2 but with inclusion of *Nautilus*.

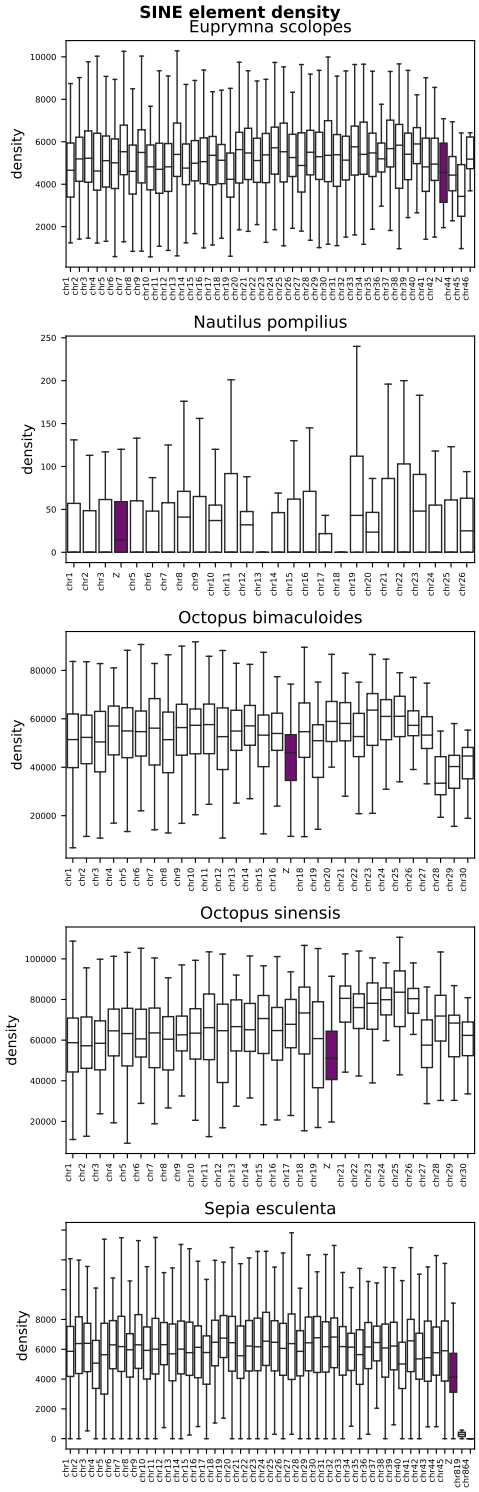

Figure S4: Density of SINE elements content on chromosomes arms across cephalopods.

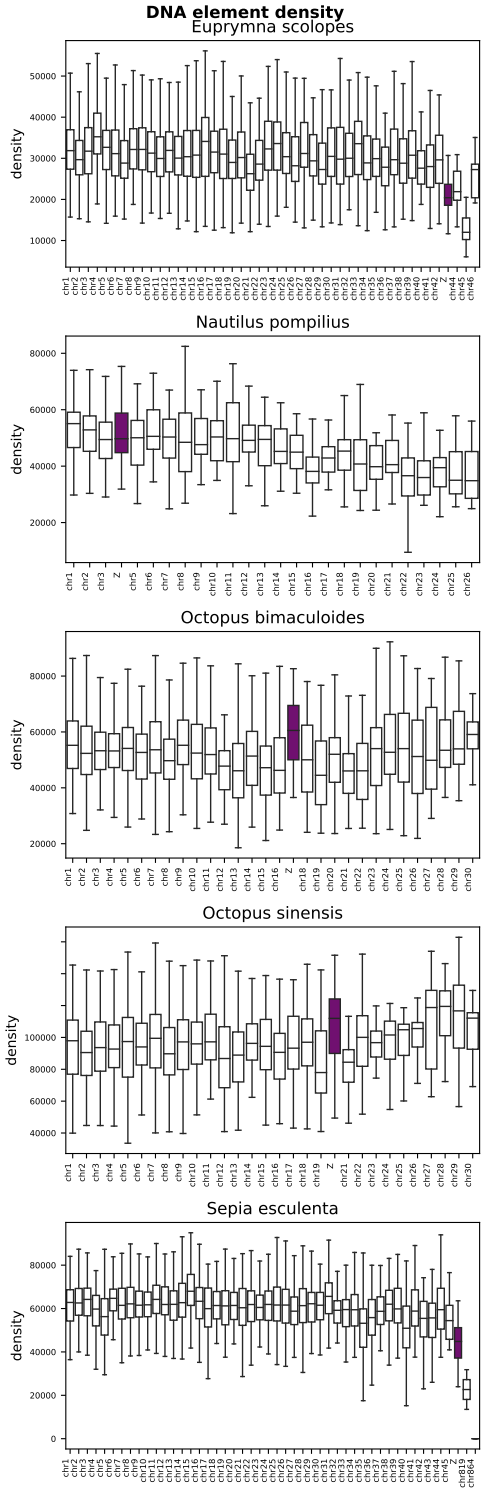

Figure S5: Density of DNA elements content on chromosomes arms across cephalopods.

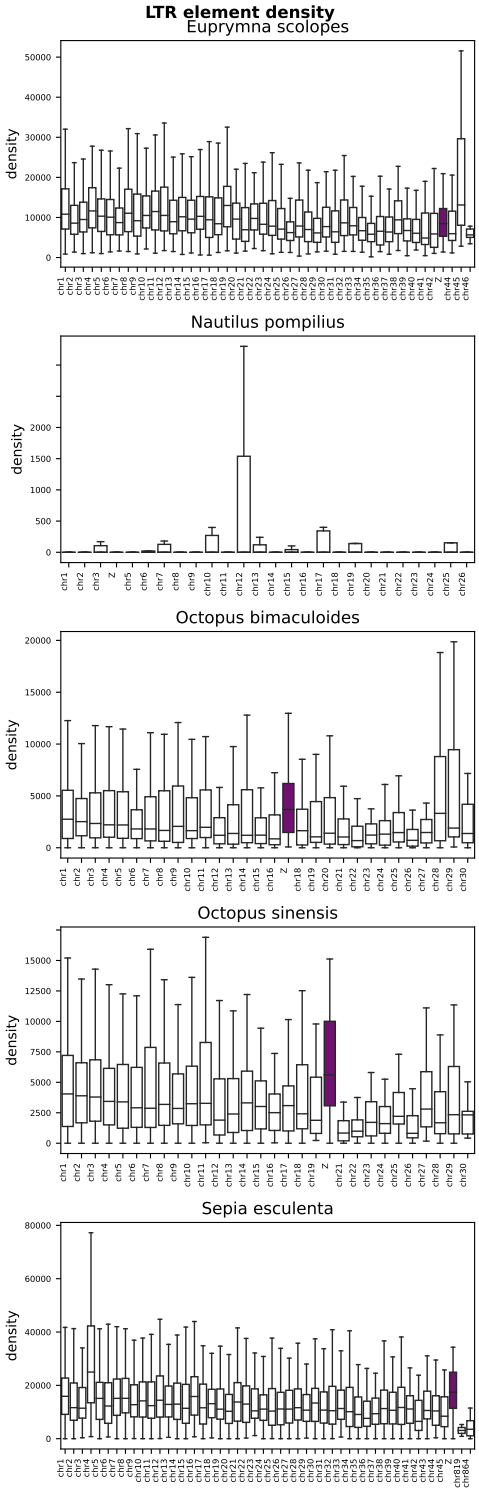

Figure S6: Density of LTR elements content on chromosomes arms across Cephalopods.

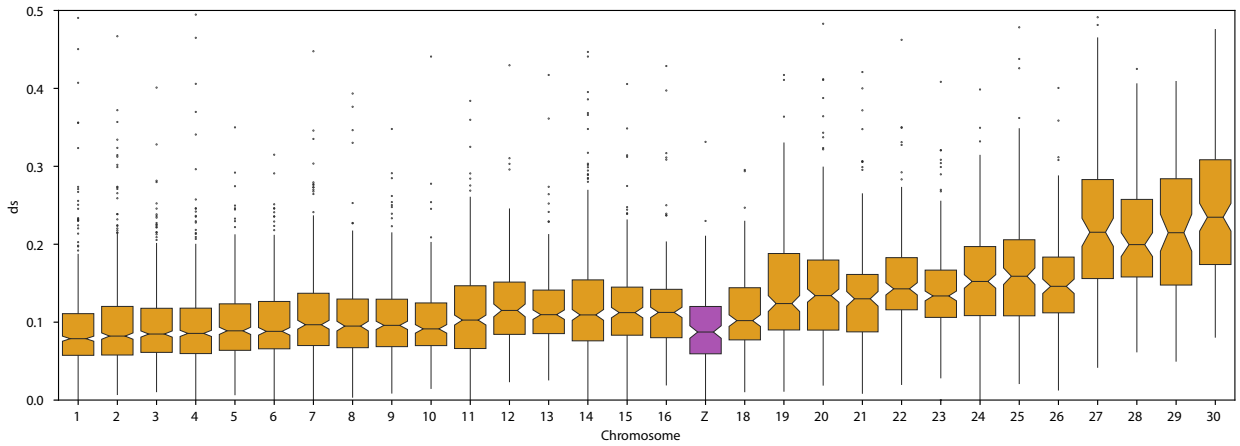

Figure S7: Synonymous substitution rate (dS) between *O. bimaculoides* and *O. sinensis* calculated for each gene per chromosome arm. The Z chromosome is highlighted in purple whereas autosomes are colored orange.

	Albertin et al. 2015	**Albertin et al. 2022**	**This Assembly**
Size	2.34 Gb	2.37 Gb	2.34 Gb
Number of scaffolds	151,674	145,326	583
Average contig/scaffold length	$15,415$ bp	$6,399$ bp	4.02 Mb
Longest scaffold	4.06 Mb	199.87 Mb	202.95 Mb
Contig N50	$5,532$ bp ($n = 97,860$)	$5,420$ bp ($n = 100,821$)	0.86 Mb ($n = 775$)
Scaffold N50	0.475 Mb ($n = 1333$)	96.88 Mb ($n = 9$)	101.05 Mb ($n = 8$)
Percent gaps	15.134%	15.016%	0.118%
GC content	36.04%	36.03%	35.79%
Coverage	$\sim 60x$	$\sim 60x$	$34x$
BUSCO (eukaryote)	C:91.8% [S:90.6%,D:1.2%], F:6.7%, M:1.5%, n : 255	C:92.6% [S:91.4%,D:1.2%], F:6.3%, M:1.1%, n : 255	C:92.2% [S:91.0%,D:1.2%], F:6.7%, M:1.1%, n : 255
BUSCO (metazoa)	C:94.1% [S:93.8%,D:0.3%], F:3.6%, M:2.3%, n : 954	C:94.5% [S:94.1%,D:0.4%], F:3.2%, M:2.3%, n : 954	C:93.9% [S:93.2%,D:0.7%], F:2.6%, M:3.5%, n : 954

Table S1: Comparison of Genome Assembly Statistics. Comparison of the initial genome assembly Albertin et al. (2015), a subsequent reassembly of those same sequencing data Albertin et al. (2022b) and the present assembly we introduce here.

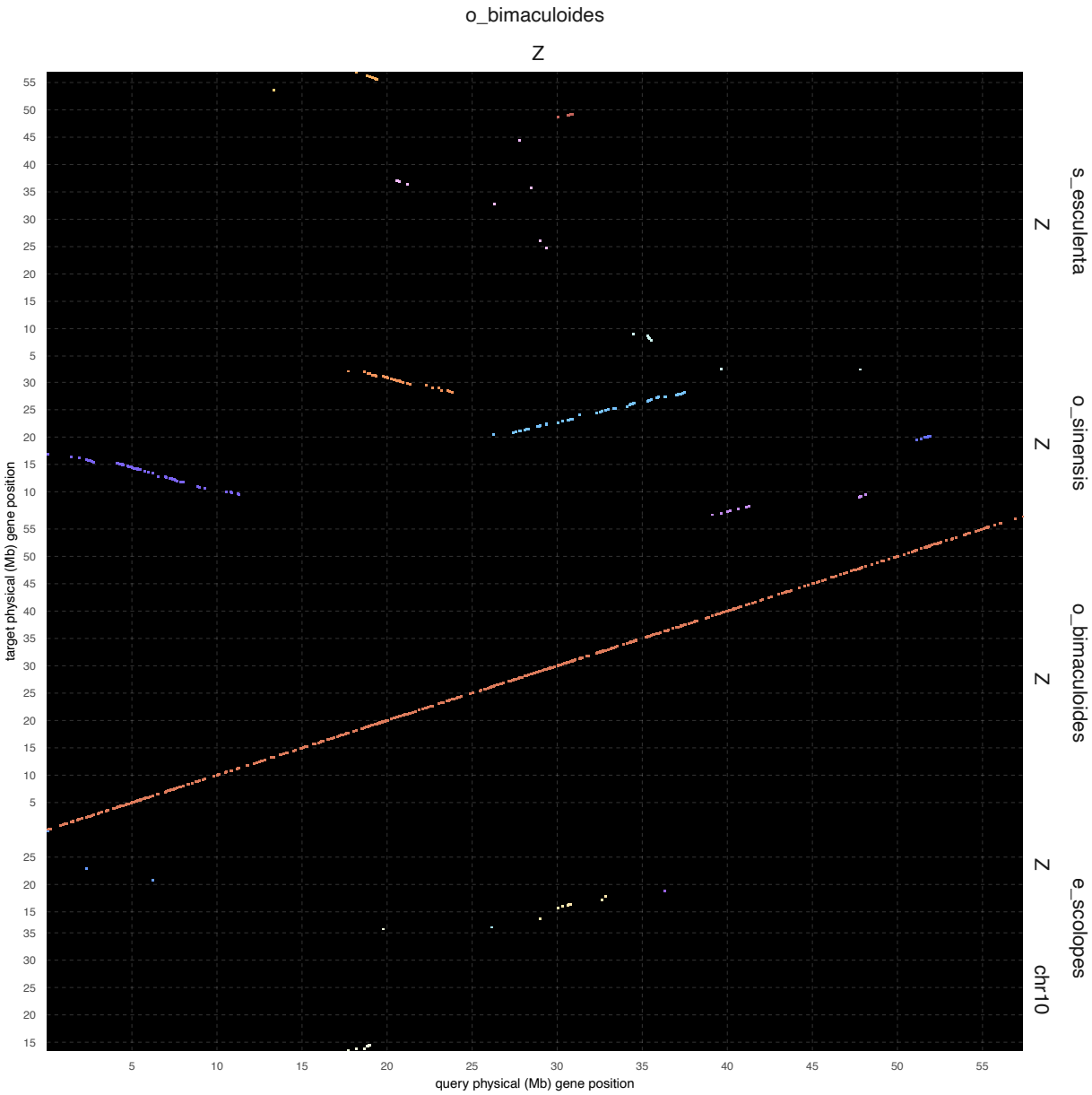

Figure S8: Dotplot of syntenic hits between the Z chromosomes of *O. bimaculoides*, *O. sinensis*, *E. scolopes*, and *S. esculenta*. The dotplot was generated with GENESPACE v.1.2.3 (Lovell et al. (2022)) and uses *O. bimaculoides* as the reference species.

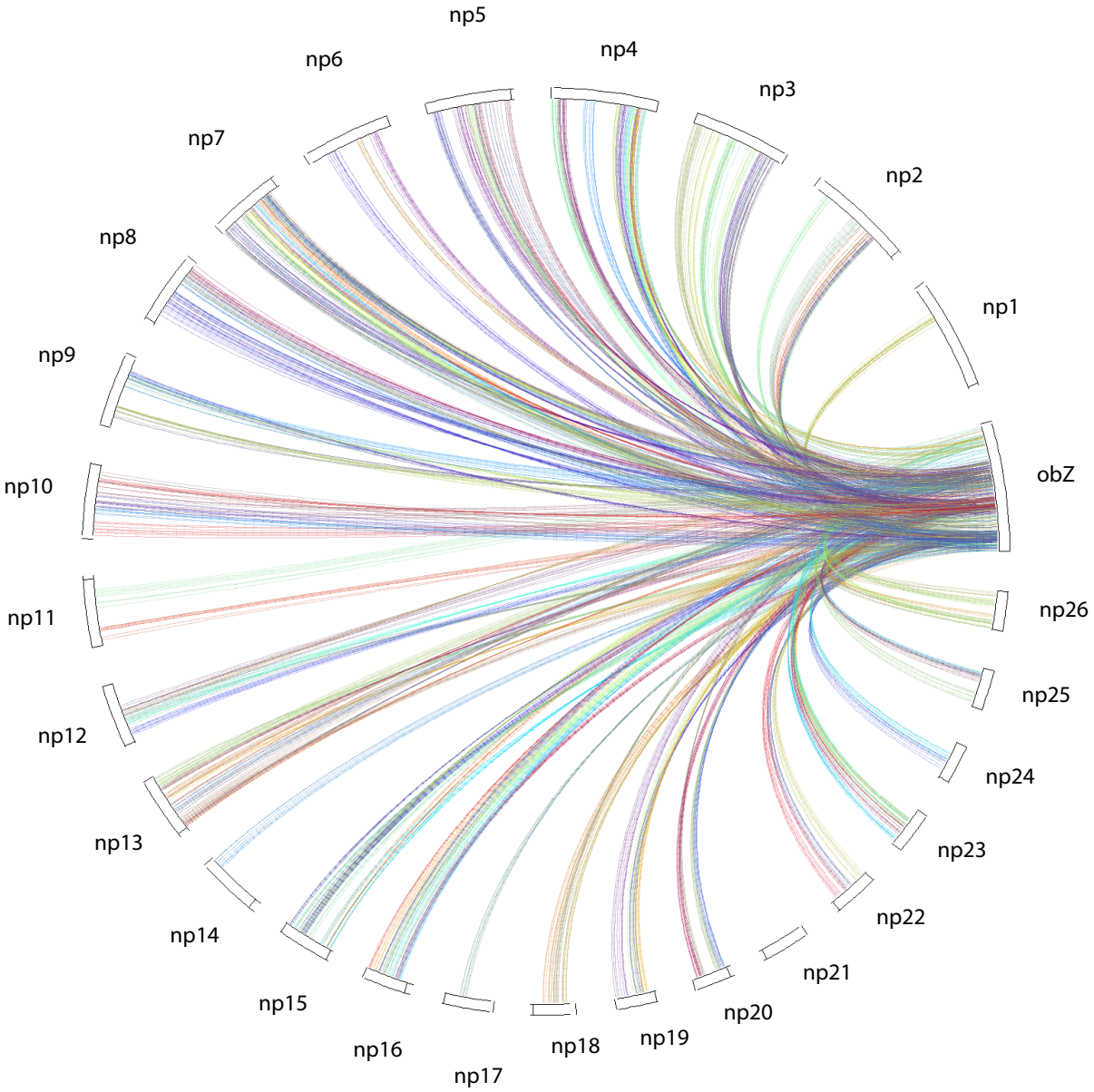

Figure S9: Syntenic blocks between the genes of *O. bimaculoides* Z chromosome and all genes of *N. pompilius*.

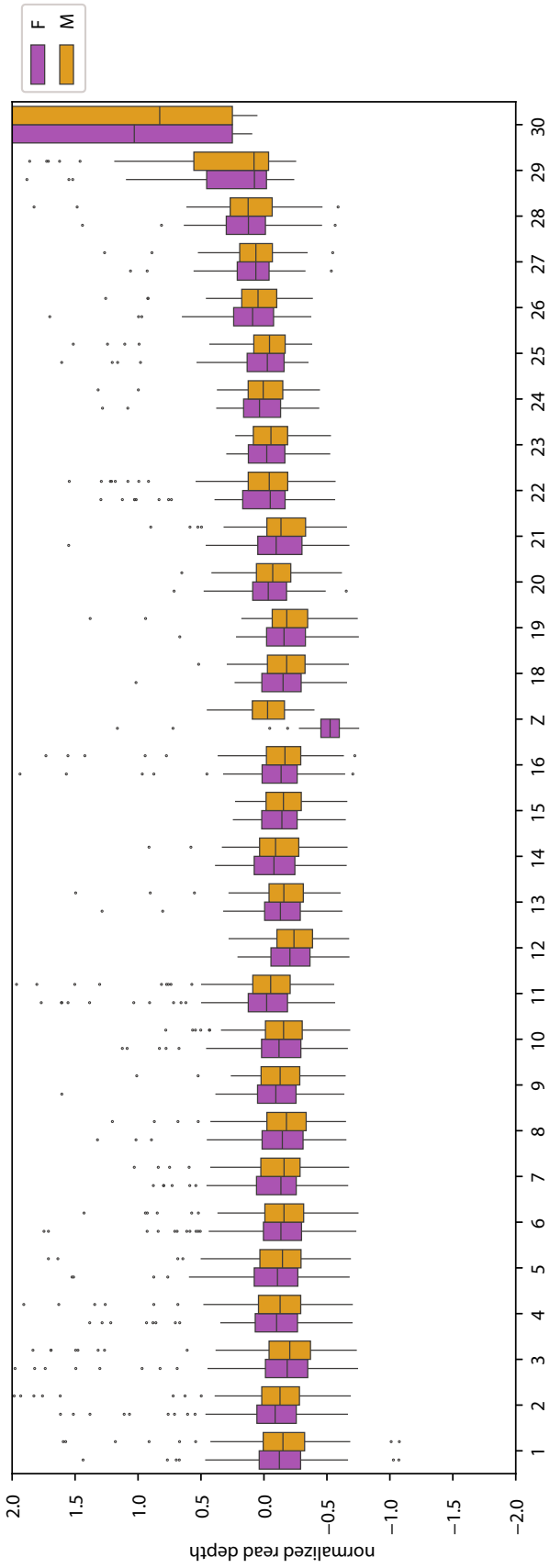

Figure S10: Normalized coverage of *O. bimaculatus* male and female short read libraries mapped to our *O. bimaculoides* genome assembly. Females have lower coverage than males in the Z chromosome.

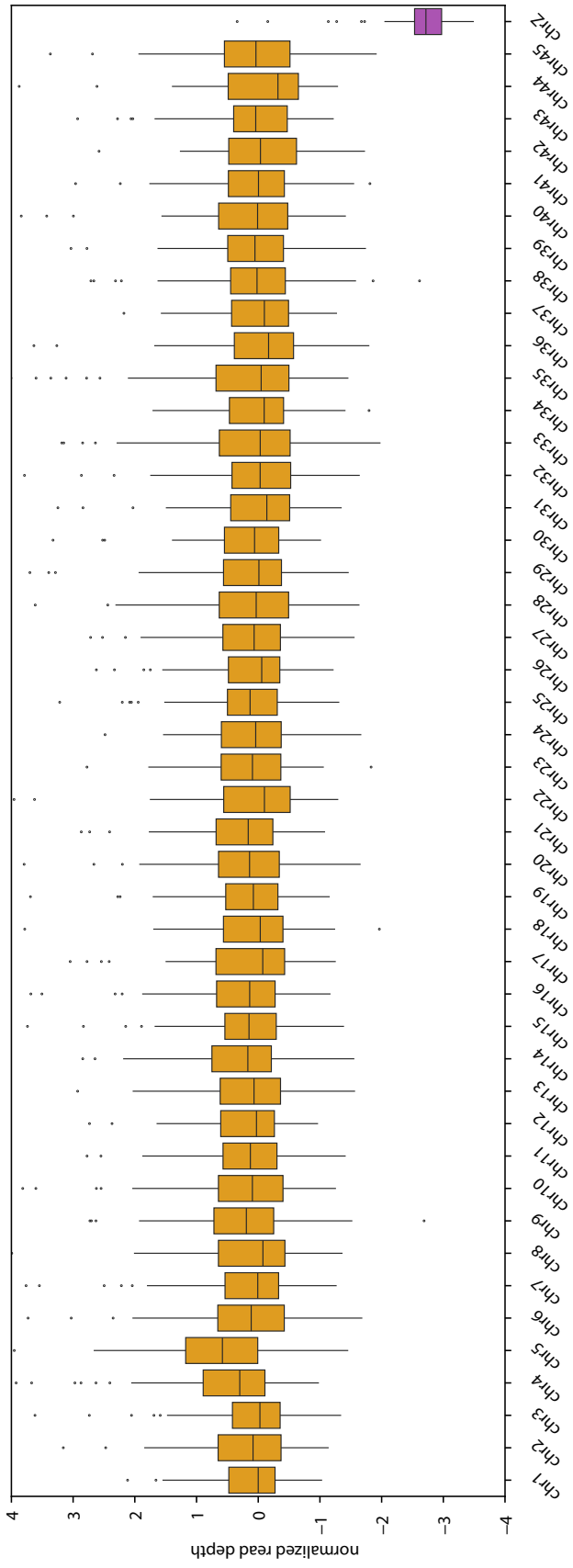

Figure S11: Normalized coverage of *S. esculenta* female short read library mapped to the *S. esculenta* genome assembly. Data and assembly was generated by the Darwin Tree of Life Project.

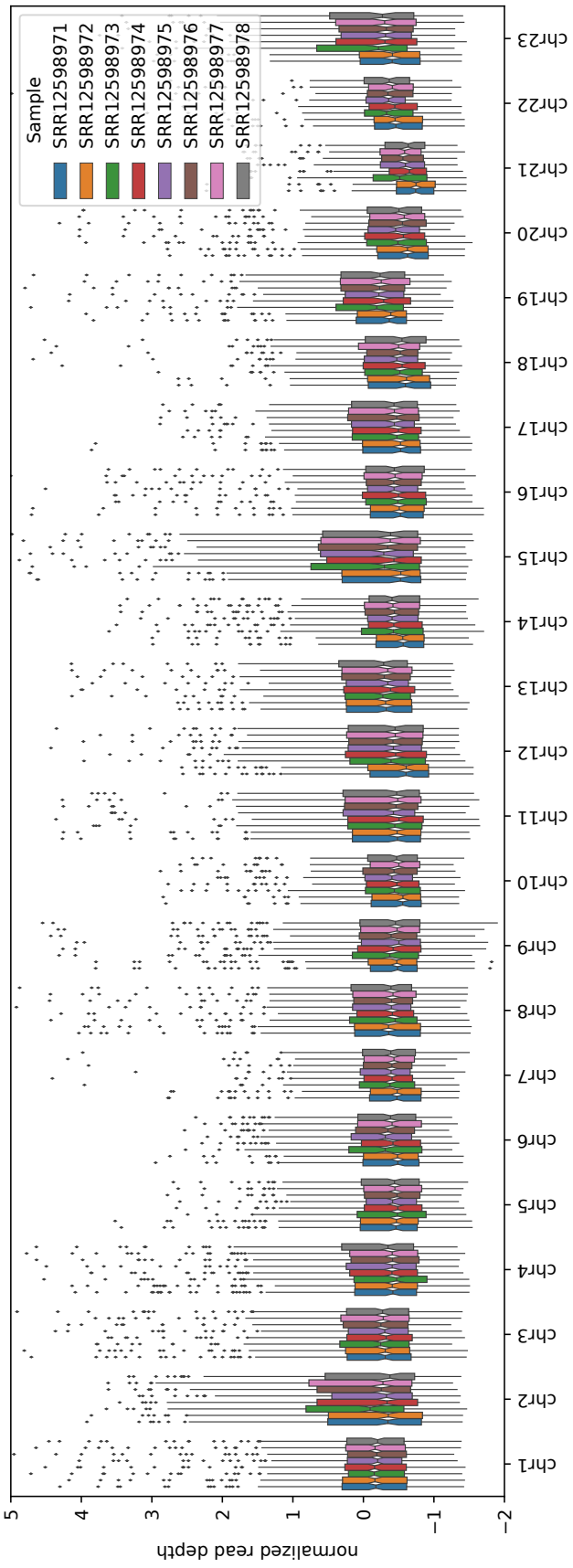

Figure S12: Normalized coverage of unsexed embryo short read libraries mapped to chromosomes 1 through 23 in *E. scolopes*. Chromosome 43 is the only chromosome with distinct variation in coverage. SRA identifiers for each library are shown in the legend. Data was generated by Schmidbaur et al. (2022).

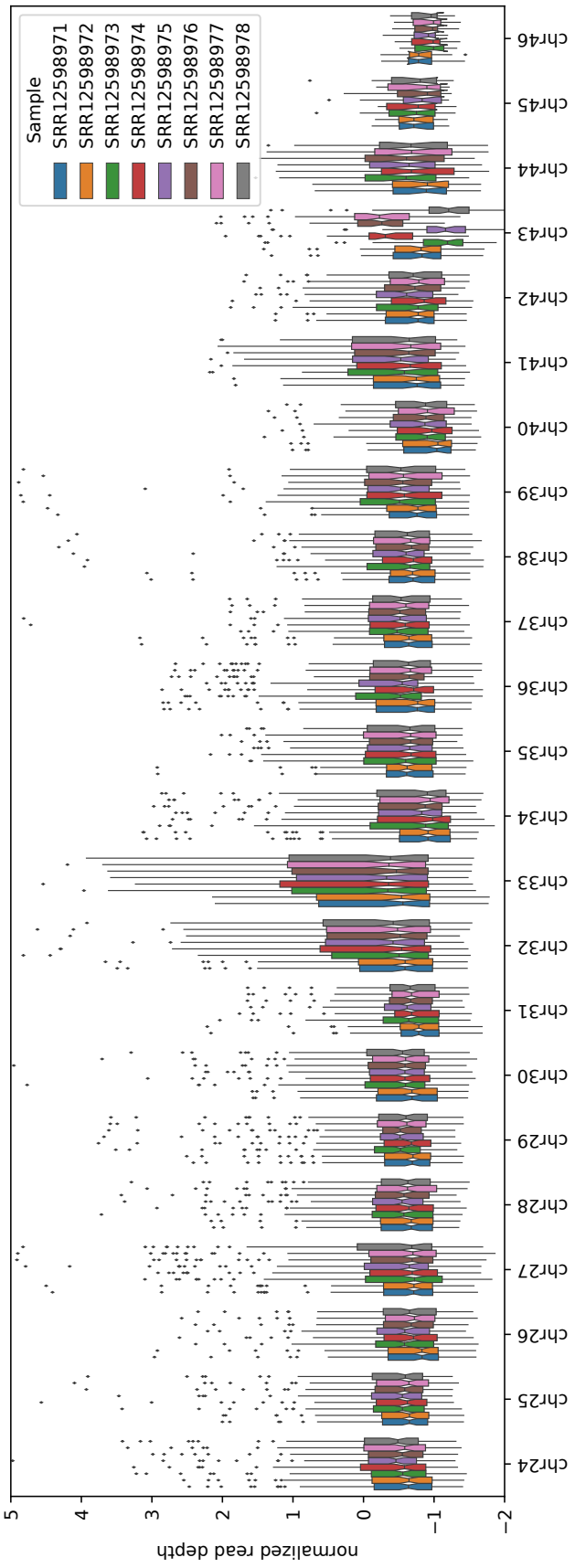

Figure S13: Normalized coverage of unsexed embryo short read libraries mapped to chromosomes 24 through 46 in *E. scolopes*. Chromosome 43 is the only chromosome with distinct variation in coverage. SRA identifiers for each library are shown in the legend. Data was generated by Schmidbaur et al. (2022).

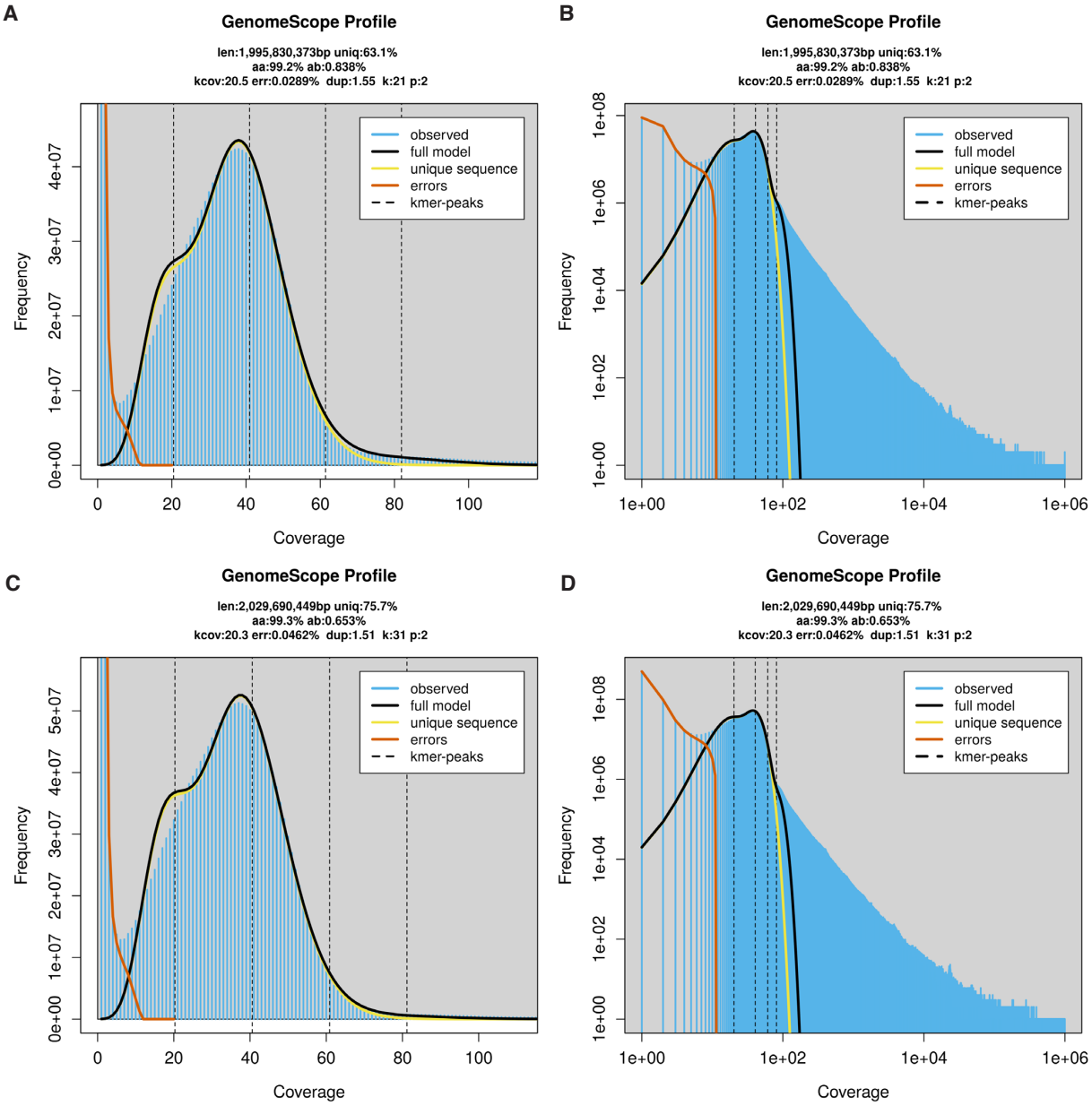

Figure S14: K-mer based genome size prediction of O. bimaculoides. **A** and **B** are 21-mer. **C** and **D** are 31-mer.

Number of genes	30655
Number of transcripts	61652
Number of cds	61639
% of genome is cds	1.3%
Number of exons	504491
Mean transcripts per gene	2.0
Mean exons per transcript	8.2
Total gene length (bp)	758146852
Total transcript length (bp)	2799696948
Total cds length (bp)	77744706
Mean gene length (bp)	24731
Mean transcript length (bp)	45411
Mean cds length (bp)	1261
Mean exon length (bp)	279
Mean intron in cds length (bp)	5154
Mean intron in exon length (bp)	6004

Table S2: *O. bimaculoides* genome annotation statistics.

Individual	**Sex**	**Tissue**	**Data type**	**Total bases**	**Read Number**	**Depth**	**Notes**
Ob1	F	optic lobe	HiFi, PacBio Sequel II 8M	133.4Gb, post-CCS	5.9M	34x	five SMRT cells, two libraries
Ob2	hatchling	optic lobe, central brain, arm crown, retina	IsoSeq, PacBio Sequel II 8M	10.4Gb	1.6M (64,190 post IsoSeq3)	0.04x	one SMRT cell
Ob3	hatchling	arm crown	Hi-C, Illumina NovaSeq 6000	555.8Gb	3.7B	–	one flow cell, two runs
Obmac_F_2	F	optic lobe	Illumina NovaSeq 6000	90.3Gb	1.24B	71x	–
Obmac_M_3	M	optic lobe	Illumina NovaSeq 6000	76.2Gb	1.05B	60x	–
Obmac_F_10	F	gill	Illumina NovaSeq 6000	80.7Gb	1.13B	64x	–
Obmac_M_14	M	testes	Illumina NovaSeq 6000	66.0Gb	924M	52x	–

Table S3: Stats of raw sequencing data collected from *O. bimaculoides* samples.

Individual	**Sex**	**Tissue**	**Data type**	**Total bases**	**Number of reads**	**Depth**
Oblat_F_2	F	optic lobe	Illumina NovaSeq 6000	94.0Gb	669M	35x
Oblat_M_3	M	optic lobe	Illumina NovaSeq 6000	92.6Gb	664M	34x
Oblat_F_10	F	optic lobe	Illumina NovaSeq 6000	111.4Gb	764M	42x
Oblat_M_4	M	optic lobe	Illumina NovaSeq 6000	111.6Gb	779M	42x

Table S4: Stats of raw sequencing data collected from *O. bimaculatus* samples.

Species	Repeat type	p-value
Sepia esculenta	DNA	**4.06E-20**
Euprymna scolopes	DNA	**2.91E-17**
Octopus bimaculoides	DNA	**1.34E-04**
Octopus sinensis	DNA	3.04E-03
Octopus bimaculoides	LINE	**9.90E-25**
Octopus sinensis	LINE	**5.71E-09**
Sepia esculenta	LINE	**6.21E-05**
Euprymna scolopes	LINE	**2.34E-04**
Octopus bimaculoides	LTR	**1.23E-04**
Octopus sinensis	LTR	1.06E-03
Sepia esculenta	LTR	**3.64E-03**
Euprymna scolopes	LTR	3.05E-01
Octopus bimaculoides	SINE	**3.66E-06**
Sepia esculenta	SINE	**7.28E-06**
Octopus sinensis	SINE	1.45E-03
Euprymna scolopes	SINE	1.80E-02

Table S5: Mann–Whitney U tests by repeat element type, testing for a significant difference between the sex chromosome element density against all autosomes. Bold indicates significance after Bonferroni correction.

octo_gene_id	human_gene_id	human_gene_symbol	e-val	function	human reproductive mRNA expression?	human reproductive protein expression?	found on *Nautilus* Z?
obimac_0008617.1	NP_001121055.1	ING4	1.00E-109	tumor suppressor protein that contains a PHD-finger	ovary, uterus, placenta, prostate, testis	ovary (fetal), testis, testis (fetal)	0
obimac_0008634.1	NP_001182257.1	RAB9A	2.00E-92	GDP binding activity; GTP binding activity; and GTPase activity.	ovary, uterus, placenta, prostate, testis	uterus, cervix, ovary, ovary (fetal), testis, testis (fetal)	1
obimac_0008637.2	NP_001007469.1	SMARCB1	0.00E+00	SWI/SNF Related, Matrix Associated, Actin Dependent Regulator Of Chromatin,	ovary, uterus, placenta, prostate, testis	uterus, ovary, ovary (fetal), testis, testis (fetal)	0
obimac_0008638.1	no hit	-	-	-	-	-	0
obimac_0008640.1	NP_006310.1	CDIPT	8.00E-81	Catalyzes the biosynthesis of phosphatidylinositol	ovary, uterus, placenta, prostate, testis	uterus, cervix, ovary, ovary (fetal), testis, testis (fetal)	0
obimac_0008642.1	no hit	-	-	-	-	-	1
obimac_0008643.1	NP_005207.3	DDOST	8.00E-128	Subunit of the oligosaccharyl transferase (OST) complex	ovary, uterus, placenta, prostate, testis	uterus, cervix, ovary, ovary (fetal), testis, testis (fetal), seminal vesicle	0
obimac_0008647.1	NP_003123.2	SRM	2.00E-114	Spermidine Synthase	ovary, uterus, placenta, prostate, testis	uterus, cervix, ovary, ovary (fetal), testis, testis (fetal), seminal vesicle	0
obimac_0008648.1	NP_001011547.2	SLC5A9	0.00E+00	Electrogenic Na(+)-coupled sugar symporter	ovary, uterus, prostate, testis	none	1
obimac_0008657.1	NP_004152.1	RAB1A	9.00E-134	Ras superfamily of GTPases.	ovary, uterus, placenta, prostate, testis	uterus, cervix, ovary, ovary (fetal), testis, testis (fetal), seminal vesicle	1
obimac_0008668.1	NP_001287785.1	FOSL1	2.00E-13	FOS Like 1, AP-1 Transcription Factor Subunit	ovary, uterus, prostate, testis	ovary	0
obimac_0008926.1	AAH14937.1	FHIP1A	3.00E-129	protein localization to perinuclear region of cytoplasm	ovary, uterus, prostate, testis	testis (fetal)	1
obimac_0008950.6	XP_005257828.1	SPAG9	0.00E+00	Sperm Associated Antigen; cancer testis antigen gene family	ovary, uterus, placenta, prostate, testis	uterus, cervix, ovary, ovary (fetal), testis, testis (fetal)	1
obimac_0008956.1	no hit	-	-	-	-	-	1
obimac_0008962.2	NP_002755.1	PRPS1	0.00E+00	catalyzes the phosphoribosylation of ribose 5-phosphate to 5-phosphoribosyl-1-pyrophosphate	ovary, uterus, placenta, prostate, testis	uterus, cervix, ovary, ovary (fetal), testis, testis (fetal)	0
obimac_0008963.1	NP_005330.1	UBE2K	9.00E-89	ubiquitin-conjugating enzyme	ovary, uterus, placenta, prostate, testis	uterus, cervix, ovary, ovary (fetal), testis, testis (fetal), seminal vesicle	0
obimac_0008964.1	NP_001073308.1	OCIAD1	9.00E-22	OCIA Domain Containing 1; Maintains stem cell potency	ovary, uterus, placenta, prostate, testis	uterus, cervix, ovary, ovary (fetal), testis, testis (fetal)	0
obimac_0008989.8	NP_001381782.1	KLC1	0.00E+00	Kinesin Light Chain 1	ovary, uterus, placenta, prostate, testis		1
obimac_0008993.1	NP_001960.2	EIF5	8.00E-147	Eukaryotic translation initiation factor-5	ovary, uterus, placenta, prostate, testis	uterus, cervix, ovary, ovary (fetal), testis, testis (fetal)	1

Table S6: Table of blast results from orthologous region genes of chromosome Z to humans.

Species	**Source**	**Accession ID**	**Citation**
Nautilus pompilius	CNCB	GWHBECW00000000	Zhang et al. (2021)
Nautilus pompilius	www.doi.org/10.6084/m9.figshare.14236208	N/A	Huang et al. (2022)
Euprymna scolopes	NCBI	GCA_024364805.1	Albertin et al. (2022a)
Architeuthis dux	NCBI	GCA_006491835.1	Da Fonseca et al. (2020)
Sepia pharaonis	NCBI	GCA_903632075.3	Song et al. (2021)
Sepia esculenta	Darwin Tree of Life	GCA_964036315.1	N/A
Octopus minor	gigadb	N/A	Kim et al. (2018)
Hapalochlaena maculosa	NCBI	GCA_015501135.1	Whitelaw et al. (2020)
Octopus bimaculoides	NCBI	PRJNA1076263	This Article
Octopus sinensis	NCBI	GCF_006345805.1	Li et al. (2020)
Eldone cirrhosa	Darwin Tree of Life Project	GCA_964016885.1	N/A

Table S7: Source of whole genome assemblies by species and the associated publications.