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

Summary

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, coleoids harbor an incredible variety of novel traits including the most complex nervous system among invertebrates, 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 in a phylogenetic context; cephalopods are a deeply diverged lineage that last share a common ancestor with other extant molluscs in the Cambrian period, roughly 550 million years ago (Ponder and Lindberg, 2008; Huang et al., 2022). With recent advances in genome sequencing technologies, we have the capability to explore the genomic foundations of cephalopod novelties. 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*). Our assembly reveals that the female octopus has just one sex chromosome, consistent with a ZO karyotype, while the male has two (ZZ), providing 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 cephalopods approximately 480 million years ago (Huang et al., 2022), and has been conserved to the present day in all cephalopod genomes available.

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 size is 2.3Gb with 30 chromosomal scaffolds representing a $N = 30$ karyotype, which matches the findings of cytogenetic studies in other octopus species (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 searched 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, suggesting that females are ZO and males are ZZ. This perhaps indicates the evolutionary loss of the W chromosome after substantial degradation (Charlesworth et al., 2005; Charlesworth, 2013). It is possible that a small dot W exists that cannot be identified with our sequencing data. However, while ZZ/ZO sex determination systems are relatively rare among species found in the literature (Jonika et al., 2022), they have been described in several groups including Lepidopterans (Johnson and Lachance, 2012), plants (Charlesworth, 2013), amphibians (Perkins et al., 2019), and more (Bachtrog et al., 2014; Tree of Sex Consortium, 2014).

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). Of 237,704 proteins input to OrthoFinder, 166,258 (69.9%) were assigned to 18,791 orthogroups. 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.

We calculated divergence in 1 megabase windows between *O. bimaculoides* and *O. sinensis* using the phast package (GTR model) (Siepel et al., 2005), which revealed that the Z chromosome evolves 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 note that by using *O. bimaculoides* as a reference, the observed differences in divergence rates between Z and autosomes is likely conservative, as any loci which have moved moved between Z and autosomal chromosomes over evolutionary time should only obscure a difference, if present.

We also analyzed patterns of coding substitutions between *O. bimaculoides* and *O. sinensis*, restricting calculations to Z-linked genes in both species. While we found that the rate of nonsynonymous substitution was roughly equivalent between Z-linked and autosomal genes (Mann–Whitney U test, $p = 0.38$), Z-linked genes have decreased levels of synonymous substitution (mean Z-linked dS= 0.091, 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 $\alpha = 0.127$ (95% C.I.$= 0.022 - 0.268$) meaning that for these positions we estimate the female mutation rate to be ~ 7.8 times that of the male mutation rate. While synonymous substitutions are a decent 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). Although female-biased mutaiton rate is quite rare, *O. bimaculoides* females undergo synchronous ovulation (Ibáñez et al., 2021), a process that has been hypothesized to contribute to higher mutation rate in females (Bergeron et al., 2023).

Repetitive element characteristics of the Z chromosome are unique

Our whole-genome alignment allowed 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 allowed 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 *O. bimaculoides* Z chromosome has an elevated density of LINE elements, harboring LINEs at approximately twice the density as the genome-wide background (Mann-Whitney test; $p < 0.0001$; Fig. 2D).

LINE elements are a signature of coleoid sex chromosomes

We hypothesized that the abundance of LINE elements could be a clear signature of Z chromosomes in other cephalopod genome 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 large chromosome in each of those assemblies (chrZ in *S. esculenta*, formerly chr20 in *O. sinensis* and formerly chr43 in *E. scolopes*) contained a high density of 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 the rest of their genomes (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 supporting our hypothesis, we renamed

the putative Z chromosomes to Z in all figures for clarity. Upon inspection of the much more distant *Nautilus* genome, no chromosomes with elevated LINE element content could be identified. Thus evidence from LINE element enrichment suggests that this could be 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

We hypothesized that the chromosomes with elevated LINE densities 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 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 hemizygous. 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), Illumina short read libraries of 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 autosomal 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. Taken together, these blast results let us speculate that the genes retained on the Z chromosome may represent an ancient set of proteins essential to animal reproduction and/or gametogenesis. Last, we searched for known sex determination genes from other mollusc species on the Octopus Z. Four are described: *Forkhead box protein L2* (*FOXL2*) in *M. yessoensis* and *C. farreri*, *Zinc finger protein 226-like* (*ZNK226l*) in *P. magellanicus*, and *Cytochrom P450 3A24-like* (*CYP3A24l*) in *A. japonicum* have been found to be female-biased (Han et al., 2022). *Doublesex and mab-3 related transcript factor 1-like* (*DMRT1L*) is mollusc-specific and male-biased in *M. edulis* (Evensen et al., 2022). We downloaded representative protein sequences of these genes from NCBI (*FOXL2* and *DMRT1L* from *A. irradians*, *ZNK226l* and

CYP3A24l from *M. yessoensis*) and ran BlastP searches to a database composed of proteins we annotated on the *O. bimaculoides* Z chromosome. This revealed Z-linked hits for *FOXL2* (obimac_0009054.1, e-val=3.23e-33) and *ZNK226l* (obimac_0008819.2, e-val=2.66e-27), thus the *O. bimaculoides* Z harbors known mollusc sex determination genes.

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. Since ovaries contain fewer cells in meiosis than testes, it is difficult to identify female heterogameity via karyotyping (Jonika et al., 2022). Therefore, it is standard to use testicular tissue (Bonnaud et al., 2004; Vitturi et al., 1982) or unsexed embryos (GAO and Natsukari, 1990) for karyotyping. Although we have not found evidence for a W chromosome in *O. bimaculoides* or any other cephalopod species, assembly of degenerate chromosomes is particularly challenging and so evidence for a W might be gathered in the future. 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.

Acknowledgments

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 colab 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. Silhouettes of cephalopod taxa were obtained from PhyloPic (https://www.phylopic.org/).

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 octopus 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.

Doris Bachtrog, Judith E Mank, Catherine L Peichel, Mark Kirkpatrick, Sarah P Otto, Tia-Lynn Ashman, Matthew W Hahn, Jun Kitano, Itay Mayrose, Ray Ming, et al. Sex determination: why so many ways of doing it? *PLoS biology*, 12(7):e1001899, 2014.

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.

Laure Bonnaud, Catherine Ozouf-Costaz, and Renata Boucher-Rodoni. A molecular and karyological approach to the taxonomy of nautilus. *Comptes Rendus Biologies*, 327(2): 133–138, 2004.

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`.

K Garrett Evensen, William E Robinson, Keegan Krick, Harry M Murray, and Helen C Poynton. Comparative phylotranscriptomics reveals putative sex differentiating genes across eight diverse bivalve species. *Comparative Biochemistry and Physiology Part D: Genomics and Proteomics*, 41:100952, 2022.

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.

Yue Mian GAO and Yutaka Natsukari. Karyological studies on seven cephalopods. *Venus (Japanese Journal of Malacology)*, 49(2):126–145, 1990.

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.

Wentao Han, Liangjie Liu, Jing Wang, Huilan Wei, Yuli Li, Lijing Zhang, Zhenyi Guo, Yajuan Li, Tian Liu, Qifan Zeng, et al. Ancient homomorphy of molluscan sex chromosomes sustained by reversible sex-biased genes and sex determiner translocation. *Nature Ecology & Evolution*, 6(12):1891–1906, 2022.

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. *Nature Protocols*, 4(1): 44–57, 2009. 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.

Christian M Ibáñez, Mariana Díaz-Santana-Iturrios, David A López-Córdova, Sergio A Carrasco, M Cecilia Pardo-Gandarillas, Francisco Rocha, and Erica AG Vidal. A phylogenetic approach to understand the evolution of reproduction in coleoid cephalopods. *Molecular phylogenetics and evolution*, 155:106972, 2021.

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.

Michelle M Jonika, James M Alfieri, Terrence Sylvester, Andrew Riley Buhrow, and Heath Blackmon. Why not y naught. *Heredity*, 129(2):75–78, 2022.

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.

Riddhi D Perkins, Julio Rincones Gamboa, Michelle M Jonika, Johnathan Lo, Amy Shum, Richard H Adams, and Heath Blackmon. A database of amphibian karyotypes. *Chromosome Research*, 27:313–319, 2019.

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.

Tree of Sex Consortium. Tree of sex: a database of sexual systems. *Scientific Data*, 1, 2014. doi: 10.1038/sdata.2014.15.

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.

Roberto Vitturi, Maria Berica Rasotto, and Nunzia Farinella-Ferruzza. The chromosomes of 16 molluscan species. *Italian Journal of Zoology*, 49(1-2):61–71, 1982.

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*, pages 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

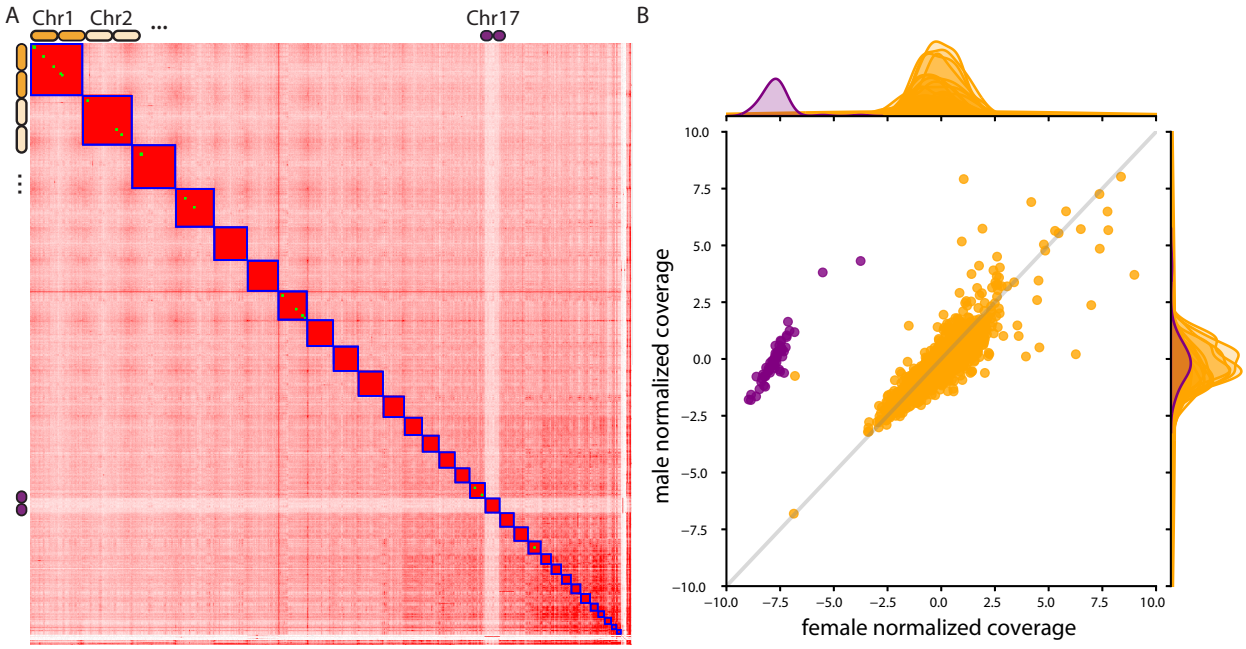

Figure 1: Sequencing data showing half coverage at chromosome 17 of *O. bimaculoides*. **A.** Hi-C contact map from our assembly of a female *O. bimaculoides*. Chromosome 17, our putative Z chromosome, is highlighted as a clear outlier in coverage. **B.** Normalized read depth of male and female whole-genome 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.

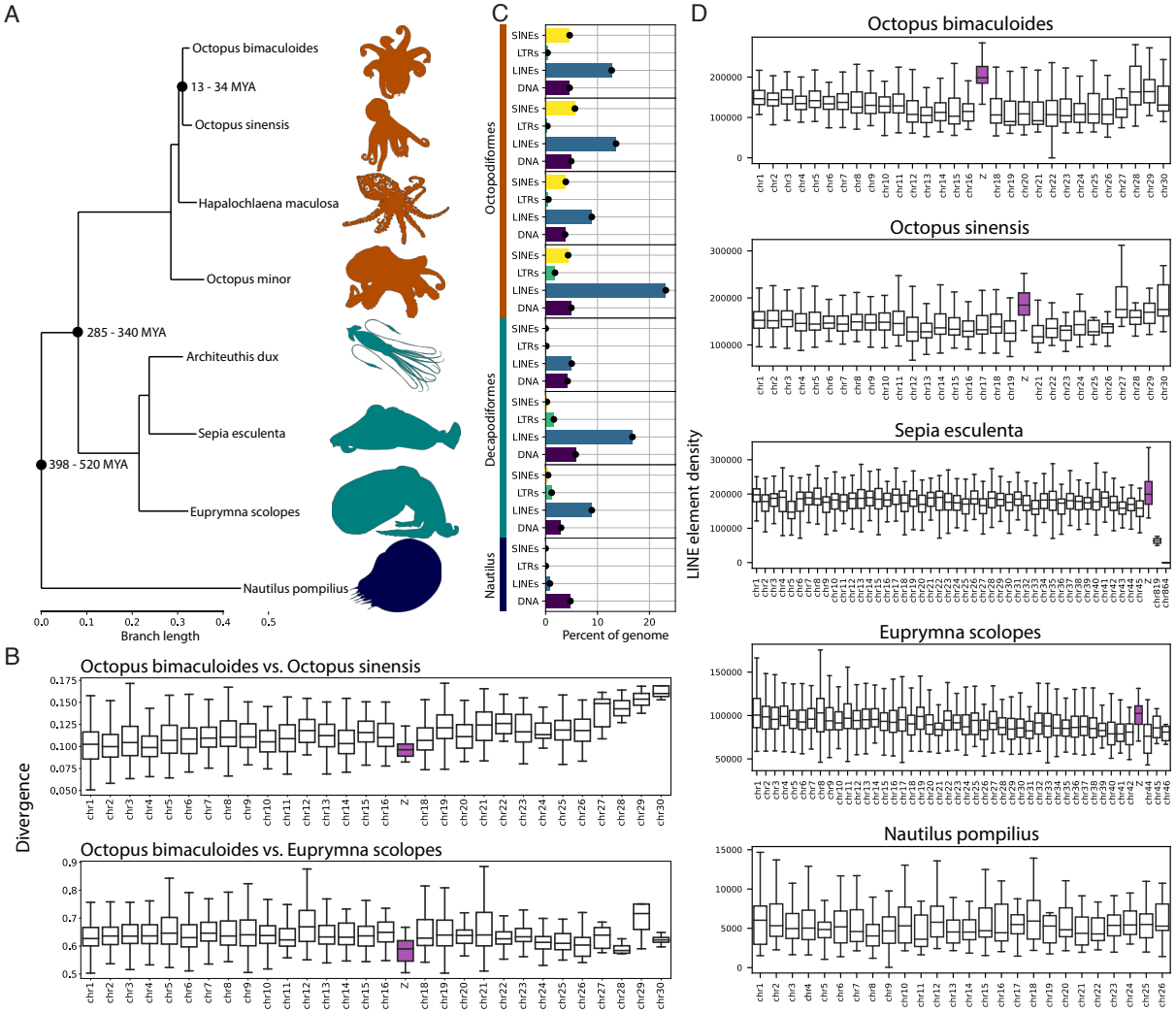

Figure 2: **A.** Cephalopod phylogeny used in this study. Phylogeny inferred using one-to-one protein orthologs identified by OrthoFinder Emms and Kelly (2019). Branch lengths in units of amino acid substitutions per site. Divergence dates are from (Huang et al., 2022). **B.** Windowed pairwise divergence calculated between *O. bimaculoides* and *O. sinesis* (top panel) and *O. bimaculoides* and *E. scolopes* (bottom panel). Divergence was calculated in 1Mb, non-overlapping windows using our multiple genome alignment with *O. bimaculoides* as the reference. The Zchromosome is highlighted in purple. **C.** Genomic repeats associated with each leaf taxon. Repeats show percent of genome composed of SINE elements (yellow), LTR elements (green), LINE elements (blue), and DNA transposons (purple) for each species. **D.** LINE content across genomes of five cephalopod species. The number of LINE base pairs were counted in 1Mb windows. In each of the coleoid species, we found a single chromosome with elevated LINE density. These putative Z chromosomes have been highlighted in purple. In *N. pompilius*, no chromosomes have outlying LINE densities.

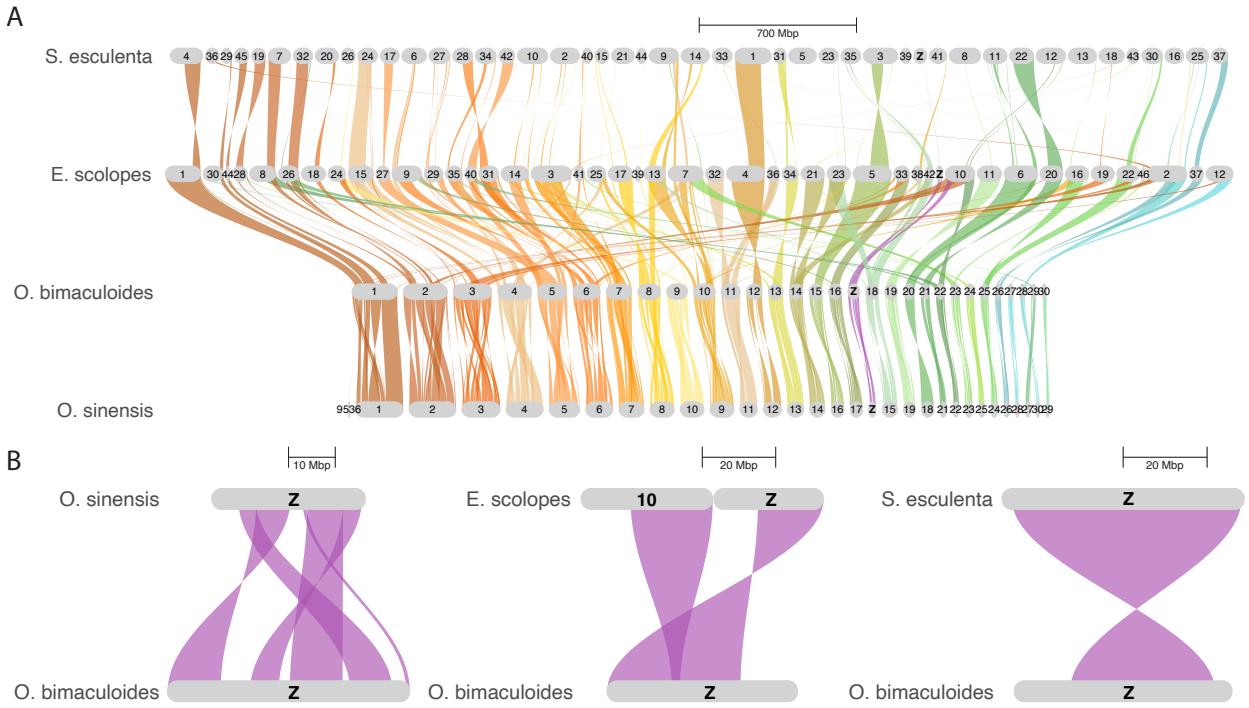

Figure 3: Syntenic relationships between four cephalopod species. **A.** Riparian synteny plot generated using orthogroups between the chromosomes of representative octopus, squid, and cuttlefish species. Ribbons connected to the *O.bimaculoides* Z chromosome are highlighted in purple. **B.** Paired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

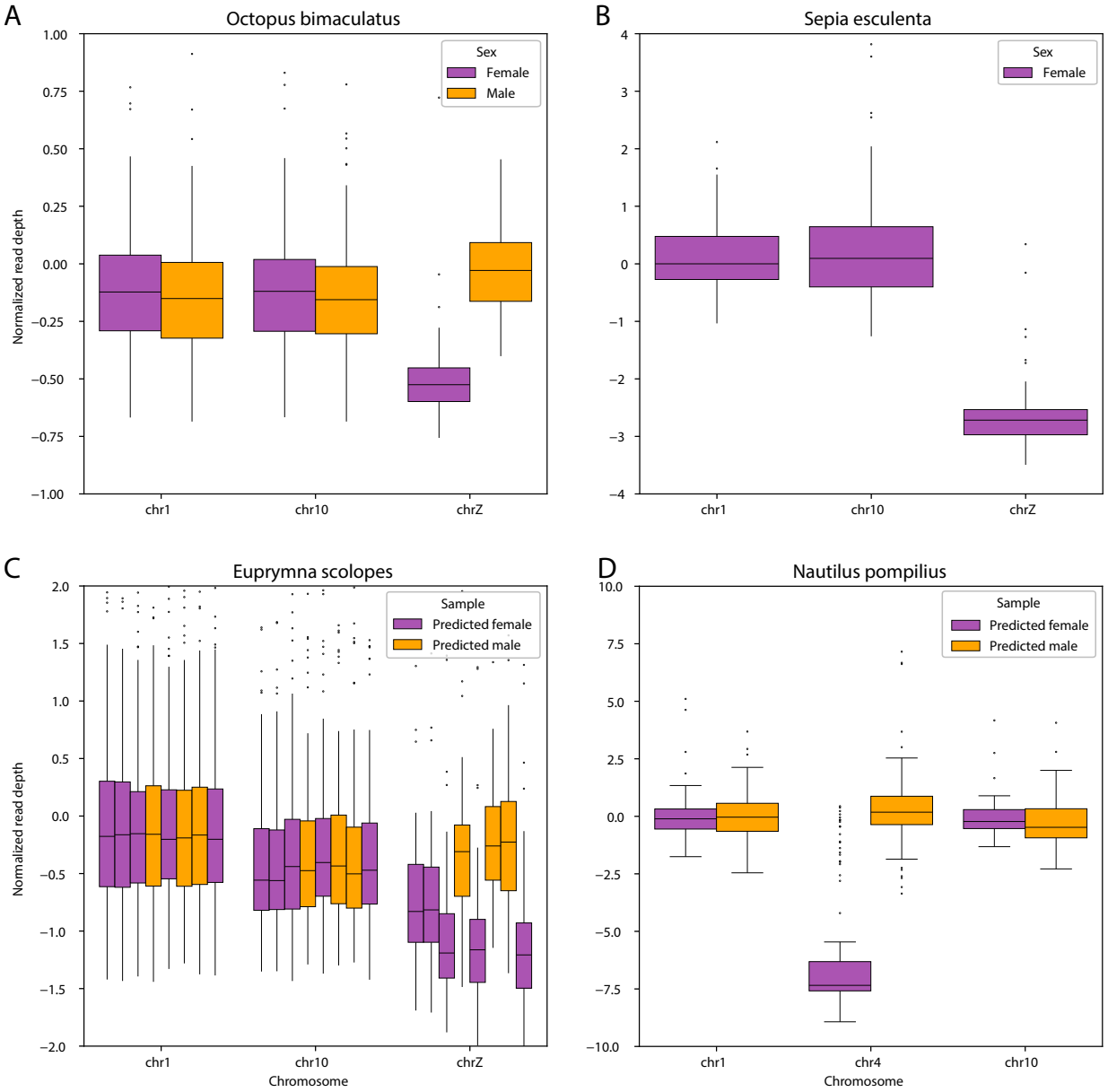

Figure 4: Normalized coverage of sequencing data mapped to representative chromosomes in *O. bimaculatus*, *S. esculenta*, *E. scolopes* and *N. pompillius*. **A.** For *O. bimaculatus*, we generated Illumina short reads from two male and two female individuals (See Methods for more details). **B.** 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*, Illumina short reads from the predicted female individual were generated by Zhang et al. (2021) and Illumina short reads from the predicted male individual were generated by Huang et al. (2022).