Cephalopod Sex Determination and its Ancient Evolutionary Origin Revealed by Chromosome-level Assembly of the California Two-Spot Octopus

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], Andrew D. Kern[1*]

[1]Department of Biology, University of Oregon,
77 Klamath Hall, 1210 University of Oregon, Eugene, OR, 97403, USA

[*]To whom correspondence should be addressed; E-mail: adkern@uoregon.edu.

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. Here, using a chromosome-level 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 show 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 other chromosomes and that it is present in all coleoid cephalopods that we have examined. Our results suggest that the cephalopod Z chromosome originated between 455 and 248 million years ago and has been conserved to the present, making it the among the oldest conserved animal sex

chromosomes 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 (*1*). 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 (*2, 3*). 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 (*4*).

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, unique origin, dating to between 455-248 million years ago, and has been conserved to the present day in both the octopus and squid lineages.

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 (*5*) and subsequently placed into scaffolds in 2022 (*6*). 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. A comparison of assembly statistics to existing *O. bimaculoides* assemblies is shown in Table S1. Generally speaking, 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 370,699 (*6*) to 583, thus 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 (*6*) 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 S2) 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 (*7, 8*). Indeed ZZ/ZO and XX/XO sex determination systems have been described in other groups including Lepidopterans (*9*) and

plants (*8*).

Whole genome alignment among cephalopods shows the Z chromosome is an outlier

To determine if the Z chromosome was unique to *O. bimaculoides* or more broadly distributed among cephalopods, we next created the first whole-genome, multiple alignment among existing cephalopod genomes using the Progressive Cactus / Comparative Genomics toolkit pipeline (*10*) (see Supplemental Methods for details). This alignment compared three other Octopus genomes (*Hapalochlaena maculosa*, *Octopus minor*, and *Octopus sinensis*), three squid genomes (*Architeuthis dux*, *Euprymna scolopes*, and *Sepia pharaonis*), and the outgroup to coleoid cephalopods, the chambered nautilus (*Nautilus pompilius*) (Figure 2A). 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. 2). Looking at 1 megabase windowed divergence between *O. bimaculoides* and *O. sinensis* revealed that the Z chromosome is evolving significantly slower than autosomes (Mann–Whitney U test, $p = 0.0011$). The same pattern holds true for divergence between *O. bimaculoides* and the bobtail squid *Euprymna scolopes* (Mann–Whitney U test, $p < 0.0001$; Fig. 2B). This result mirrors what is seen in primate autosomes in comparison to the X chromosome (*11, 12*) and is likely due to purging of recessive, mildly deleterious mutations from the hemizygous chromosome.

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 classes of repetitive element: 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, 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 (*5*), 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 this 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 (Fig. 2D) 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* and *E. scolopes* yielded a striking pattern– a single chromosome in each of those assemblies (chr20 and chr43 respectively) 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; $p < 0.0001$ and $p = 0.00023$ for *O. sinensis* and *Euprymna scolopes*, respectively). Upon inspection of the much more distant nautilus genome, no such LINE element outlier chromosome could be identified (Fig. S3). Thus evidence from LINE element enrichment suggests that this is a unique feature of the Z chromosome and that the Z chromosome may have originated before the split of the squid and octopus lineages.

Syntenic relationships of genes on the Z chromosome are conserved

We hypothesized that the chromosomes we identified in *O. sinensis* and *E. scolopes* might be orthologous to the *O. bimaculoides* Z chromosome on the basis of LINE element content. To test this hypothesis, we compared synteny among chromosomes using gene-level annotations as revealed by the GENESPACE software package (*13*), which uses 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 synteny conserved on both chr20 in *O. sinensis* and on chr43 in *E. scalopes*. Dotplots showing finer resolution pairwise alignments are shown in Figure S4.

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 chr20 in *O. sinensis* and to chr43 in *E. scolopes*. While the divergence time between *O. bimaculoides* and *O. sinensis* is relatively modest at approximately 34 million years (*14*), divergence between squids and octopus is much older, with an approximate divergence time between *O.*

bimaculoides and *E. scolopes* of 248 - 339 million years (*15*). It was thus imperative to examine whether the orthologous chromosome in *E. scolopes* was functioning as a Z chromosome in that lineage. While single animal sequence libraries from sexed *E. scolopes* were not available, we examined coverage of a set of Illumina short read experiments derived from single, unsexed embryos from (*16*). These results again confirmed our hypothesis– while coverage among chromosomes was nearly uniform, chr43 stands out as an outlier with multiple distinct coverage classes, suggesting females are hemizygous as observed in *O. bimaculoides* (Figs. 3C, S6, and S7).

Having established strong support for homology of the Z chromosome among squid and octopus, we were lastly interested in examining what genes might be shared in orthologous regions of the Z chromosome among lineages. We found 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 S3). Using publicly available summaries from GeneCards (*17*) 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 here is the protein obimac0008950 which shows strong orthology to the human Sperm Associated Antigen 9 (SPAG9; e-val=0.0). Thus we believe 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 coleoid 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. Further, our results suggest that the cephalopod Z chromosome evolved once in the lineage leading to the common ancestor of squids and octopuses and has thus been conserved for between 455 and

248 million years of evolution. This is an astoundingly long time for a sex chromosome to be preserved (*7*). A few other ancient sex chromosomes have been described previously from liverworts ($\sim$ 430Mya; (*18*)) and mosses ($\sim$ 300Mya; (*19*)), and there is some evidence the insect X chromosome is quite ancient ($\sim$ 450Mya; (*20*)), although it is not well conserved and shows rapid evolution in some clades. Thus in context, the cephalopod Z chromosome may be the oldest conserved animal sex chromosome 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, 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.

References and Notes

1. R. T. Hanlon, J. B. Messenger, *Cephalopod Behaviour* (Cambridge University Press, 2018), second edn.

2. J. Z. Young, *The anatomy of the nervous system of Octopus vulgaris* (Clarendon Press, 1971).

3. R. Hanlon, *Current biology* **17**, R400 (2007).

4. W. Ponder, D. R. Lindberg, *Phylogeny and Evolution of the Mollusca* (Univ of California Press, 2008).

5. C. B. Albertin, *et al.*, *Nature* **524**, 220 (2015).

6. C. B. Albertin, *et al.*, *Nature communications* **13**, 2427 (2022).

7. D. Charlesworth, B. Charlesworth, G. Marais, *Heredity* **95**, 118 (2005).

8. D. Charlesworth, *Journal of experimental botany* **64**, 405 (2013).

9. N. A. Johnson, J. Lachance, *Annals of the New York Academy of Sciences* **1256**, E1 (2012).

10. J. Armstrong, *et al.*, *Nature* **587**, 246 (2020).

11. A. Hobolth, O. F. Christensen, T. Mailund, M. H. Schierup, *PLoS genetics* **3**, e7 (2007).

12. A. Scally, *et al.*, *Nature* **483**, 169 (2012).

13. J. T. Lovell, *et al.*, *Elife* **11**, e78526 (2022).

14. D. Jiang, *et al.*, *BMC biology* **20**, 1 (2022).

15. A. R. Tanner, *et al.*, *Proceedings of the Royal Society B: Biological Sciences* **284**, 20162818 (2017).

16. H. Schmidbaur, *et al.*, *Nature communications* **13**, 2172 (2022).

17. G. Stelzer, *et al.*, *Current protocols in bioinformatics* **54**, 1 (2016).

18. M. Iwasaki, *et al.*, *Current Biology* **31**, 5522 (2021).

19. S. B. Carey, *et al.*, *Science Advances* **7**, eabh2488 (2021).

20. M. A. Toups, B. Vicoso, *Evolution* **77**, 2504 (2023).

21. J. O. Songco-Casey, *et al.*, *Current Biology* **32**, 5031 (2022).

22. G. Marçais, C. Kingsford, *Bioinformatics* **27**, 764 (2011).

23. T. R. Ranallo-Benavidez, K. S. Jaron, M. C. Schatz, *Nature communications* **11**, 1432 (2020).

24. H. Cheng, G. T. Concepcion, X. Feng, H. Zhang, H. Li, *Nature methods* **18**, 170 (2021).

25. D. Guan, *et al.*, *Bioinformatics* **36**, 2896 (2020).

26. N. Durand, *et al.*, Juicer provides a one-click system for analyzing loop-resolution hi-c experiments. cell syst 3: 95–98 (2016).

27. O. Dudchenko, *et al.*, *Science* **356**, 92 (2017).

28. J. M. Flynn, *et al.*, *Proceedings of the National Academy of Sciences* **117**, 9451 (2020).

29. N. Chen, *Current protocols in bioinformatics* **5**, 4 (2004).

30. T. Bruna, K. J. Hoff, A. Lomsadze, M. Stanke, M. Borodovsky, *NAR genomics and bioinformatics* **3**, lqaa108 (2021).

31. M. Tardaguila, *et al.*, *Genome research* **28**, 396 (2018).

32. M. J. Hubisz, K. S. Pollard, A. Siepel, *Briefings in bioinformatics* **12**, 41 (2011).

33. M. Vasimuddin, S. Misra, H. Li, S. Aluru, *2019 IEEE international parallel and distributed processing symposium (IPDPS)* (IEEE, 2019), pp. 314–324.

34. H. Li, *Bioinformatics* **34**, 3094 (2018).

35. A. R. Quinlan, I. M. Hall, *Bioinformatics* **26**, 841 (2010).

36. A. Mayakonda, D.-C. Lin, Y. Assenov, C. Plass, H. P. Koeffler, *Genome research* **28**, 1747 (2018).

37. E. Paradis, K. Schliep, *Bioinformatics* **35**, 526 (2019).

38. Y. Zhang, *et al.*, *Nature Ecology & Evolution* **5**, 927 (2021).

39. C. B. Albertin, *et al.*, *Nature communications* **13**, 2427 (2022).

40. W. Song, *et al.*, *Frontiers in Marine Science* **8**, 639670 (2021).

41. R. R. Da Fonseca, *et al.*, *GigaScience* **9**, giz152 (2020).

42. B. Kim, S. Kang, D. Ahn, *et al.*, *GigaScience Database* **10**, 100503 (2018).

43. B. L. Whitelaw, *et al.*, *GigaScience* **9**, giaa120 (2020).

44. F. Li, *et al.*, *Molecular ecology resources* **20**, 1572 (2020).

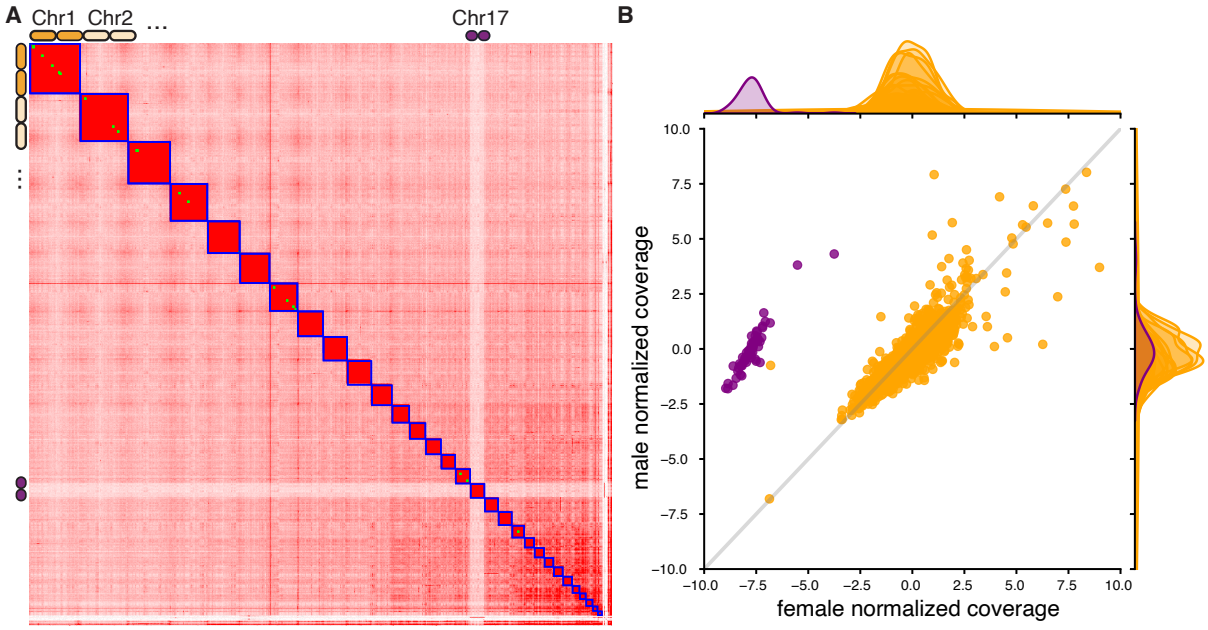

Figure 1: Sequencing data showing half coverage at chromosome 17 of *O. bimaculoides*. **A.** Hi-C contact map from our assembly of a female *O. bimaculoides*. Chromosome 17, our putative Z chromome is highlighted as a clear outlier in coverage **B.** Normalized read depth of male and female whole-genome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

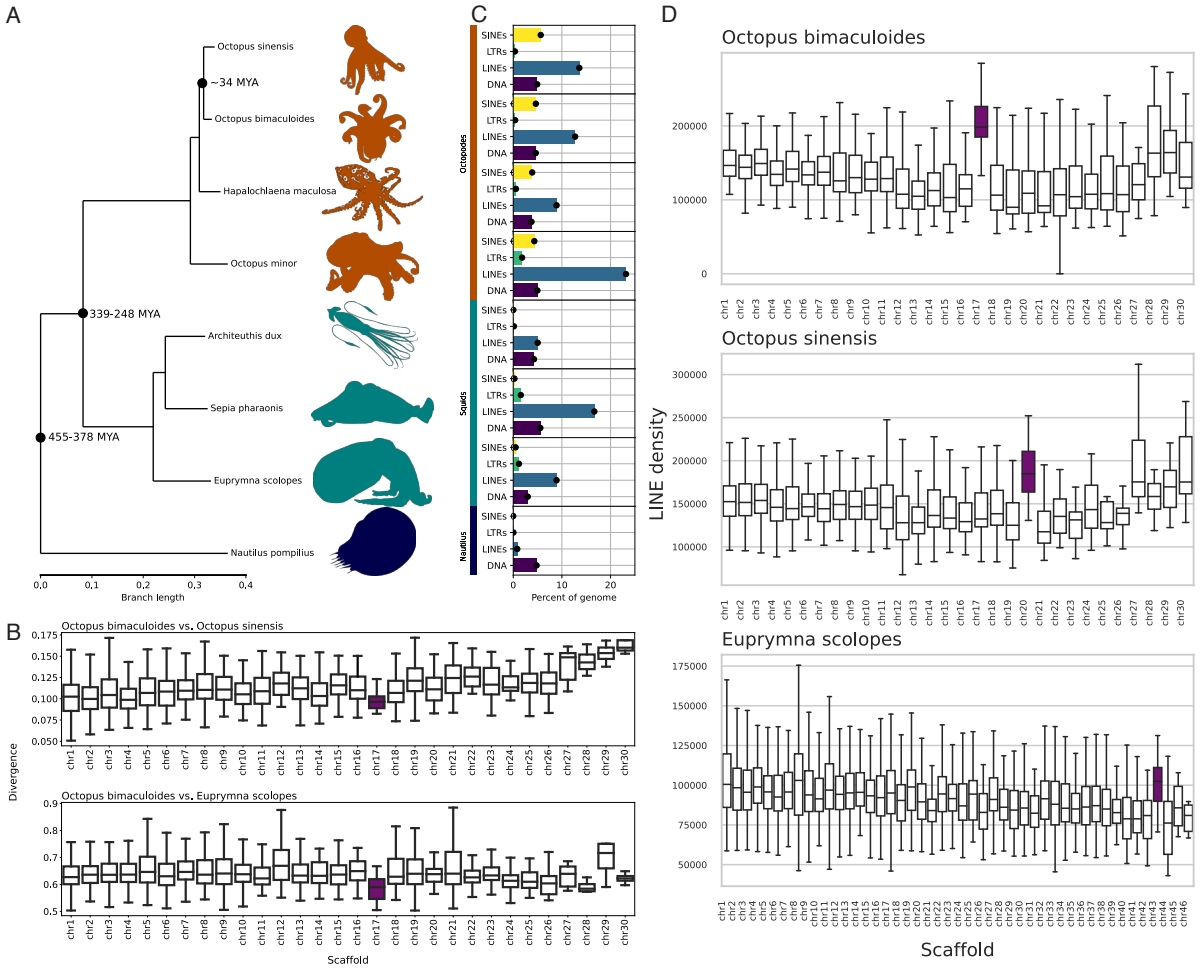

Figure 2: **A.** Cephalopod phylogeny used in this study. Phylogeny inferred using one-to-one protein orthologs identified by Orthofinder. Branch lengths in units of amino acid substitutions per site. Divergence dates are from (*14, 15*) **B.**Windowed pairwise divergence calculated between *O. bimaculoides* 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 three species. A single outlier with respect to LINE density is observed in each species, chromosome 17 in *O. bimaculoides*, chromosome 20 in *O. sinensis*, and chromosome 43 in *E. scolopes*, highlighted in purple.

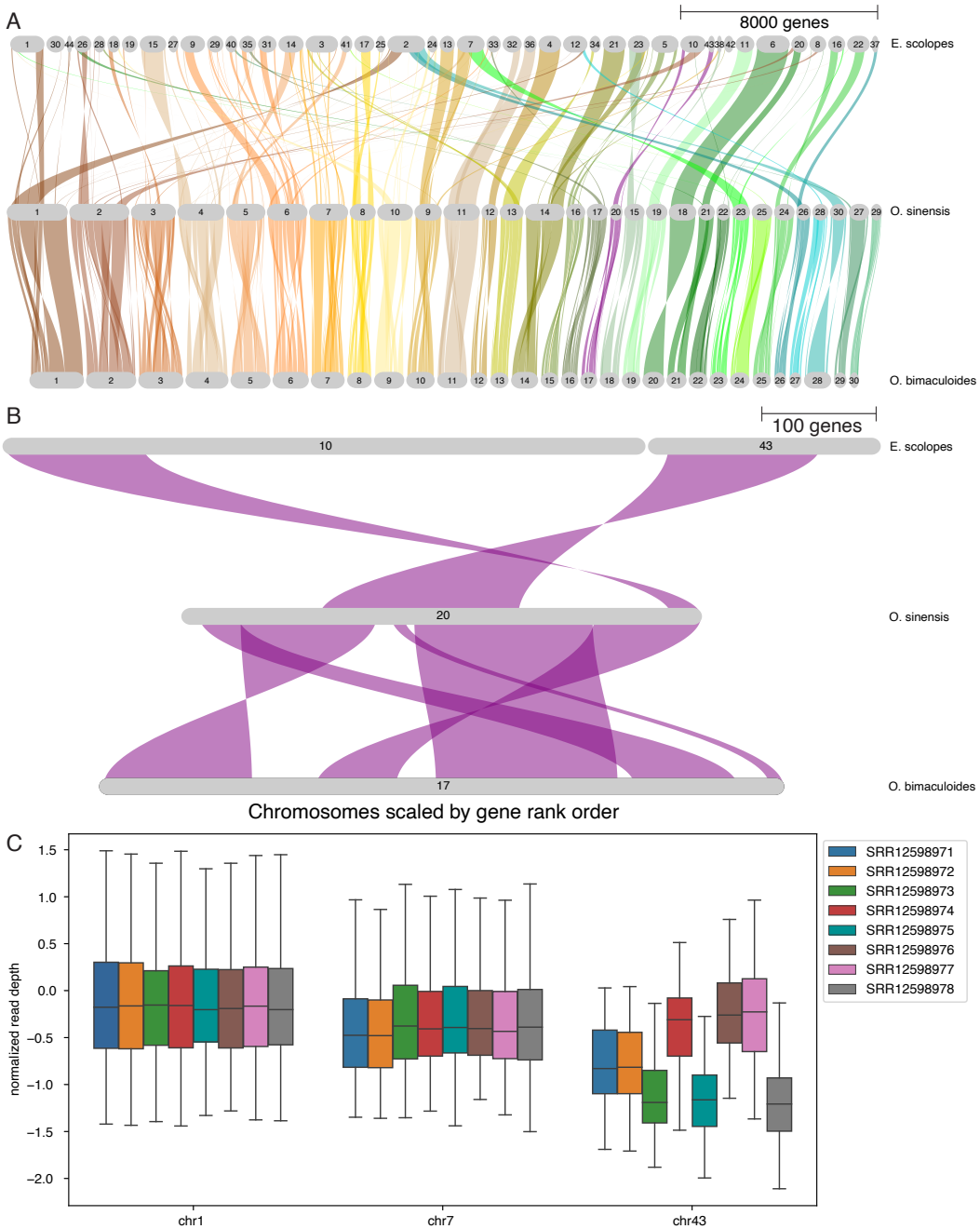

Figure 3: **A.** Riparian synteny plot generated using orthogroups between the chromosomes of squid and octopus. Ribbons connected to the *O.bimaculoides* Z chromosome are highlighted in purple. **B.** Riparian plot demonstrating gene-based synteny between octopus and squid sex chromosomes reveals conservation of the Z chromosome among *O. bimaculoides*, *O. sinesis*, and *E. scolopes*. **C.** Normalized coverage of short read libraries mapped to representative chromosomes in *E. scolopes*. Chromosome 43 is the only chromosome with such variation in coverage. SRA identifiers for each library are shown in the legend. Data was generated by (*16*)

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) (*21*). 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 S2).

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 S2). 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's 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 (*22*), which was input into GenomeScope v.2.0 (*23*) to generate genome size estimation plots. Estimation with 21-mers estimated the genome to be 2.00Gb with 0.838% heterozygosity. Estimation with 31-mers estimated the genome to be 2.03Gb with 0.653% heterozygosity (Fig. S5).

Initial genome assembly was conducted using HiFiasm v.0.15.5-r352 (*24*) with HiFi reads and standard parameters. Hifiasm did not sufficiently remove duplicates, so purge_dups v.1.2.5 (*25*) 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 (*26*) were used to identify Hi-C contact points on the genome. 3D de novo assembly (3D-DNA) pipeline v.190716 (*27*) was used to correct genome mis-assemblies, anchor

order of chromosomes, and orient genomic fragments.

Genome annotation

Transposable elements were identified using Repeat Modeler (*28*) and were annotated with Repeat Masker v.4.1.5 (*29*). Protein-coding genes were annotated using both existing RNA-seq data and our newly generated Iso-seq data. Using existing RNA-seq database, we ran Braker v.2.1.6 (*30*) to annotate. Separately, we generated annotations with the Iso-seq reads using SQANTI3 v.5.1 (*31*). 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.

Multiple species alignment

We constructed whole genome alignments of eight cephalapod species with available reference genomes (Table S4) using progressive cactus (v2.3.0 via Docker container) (*10*). We used the whole genome alignments and phastCons v.1.5 (*32*) 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 compare coverage of male and female sequencing data, we aligned short reads with bwa-mem2 (*33*) and long reads with minimap2 (*34*). Resulting sam files were sorted with samtools sort and depth was calculated with samtools depth. Finally, the mean depth was calculated in 1Mb windows using mapBed.

Assessing sequence divergence among scaffolds

We converted the hierarchical alignment (hal) generated by progressive cactus to maf with hal2maf. We filtered the full alignment into 1Mb windows along the genome using the phast utility msa_view, where windows where made with bedtools makewindows v2.31.0 (*35*). Filtered alignments with multiple sequences per species using maftools mafDuplicateFilter v.0.1 (*36*). We used phyloFit to construct phylogenies on each resulting 1Mb alignment region, only using alignments that had less than 40% of positions with gaps. 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 (*37*).

Synteny analysis

Synteny analyses between all chromosomes of *O. bimaculoides*, *O. sinensis*, and *E. scolopes* were conducted using the R package GENESPACE v.1.2.3 (*13*). *O. bimaculoides* was used as the reference species for all synteny maps.

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 S4.

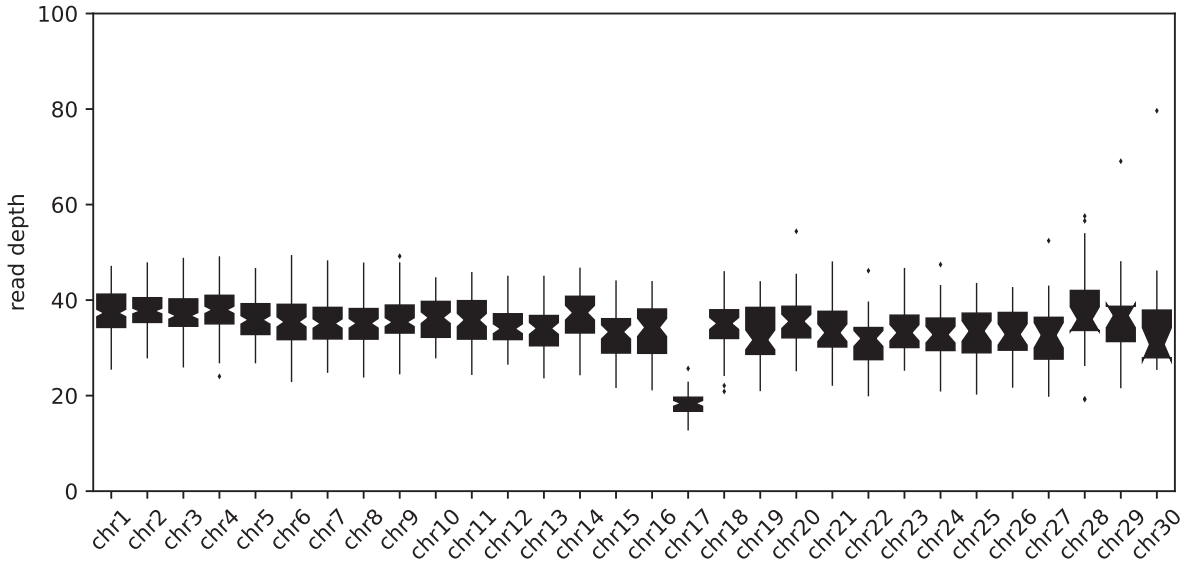

Figure S1. Read depth of HiFi sequencing data across chromosomes from *O. bimaculoides.*

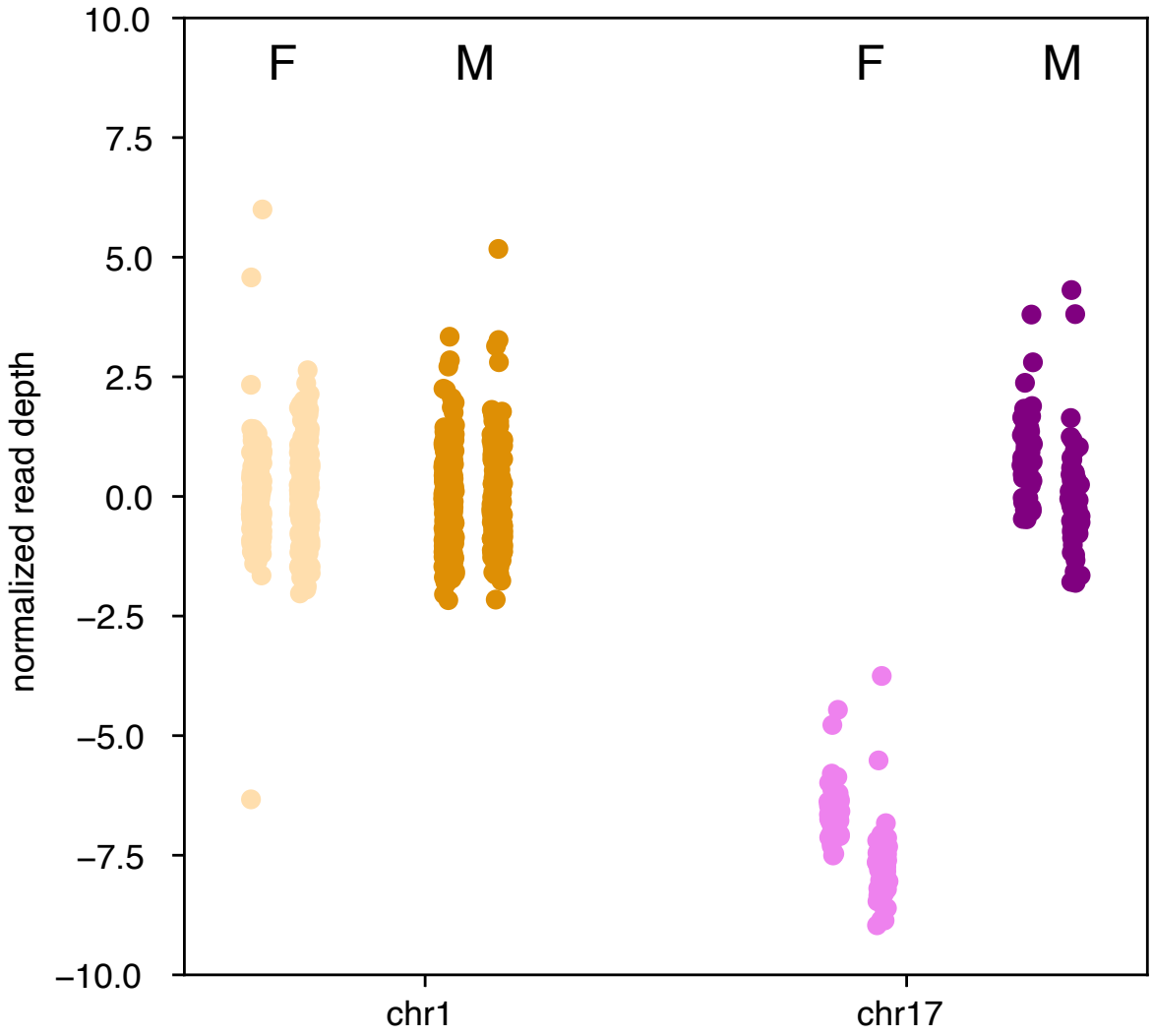

Figure S2. Normalized depth of known sex individuals of *O. bimaculoides*. Includes two male (darker shades; labelled 'M') and two female (lighter shades; labelled 'F') with short-read sequencing. *O. bimaculoides* individuals at chromosomes 1 and 17 are colored to reflect designation in main text. Both females show hemizygous coverage at chromosome 17, whereas both males show the same dosage at chromosome 17 as at chromosome 1. See Table S2 for library information.

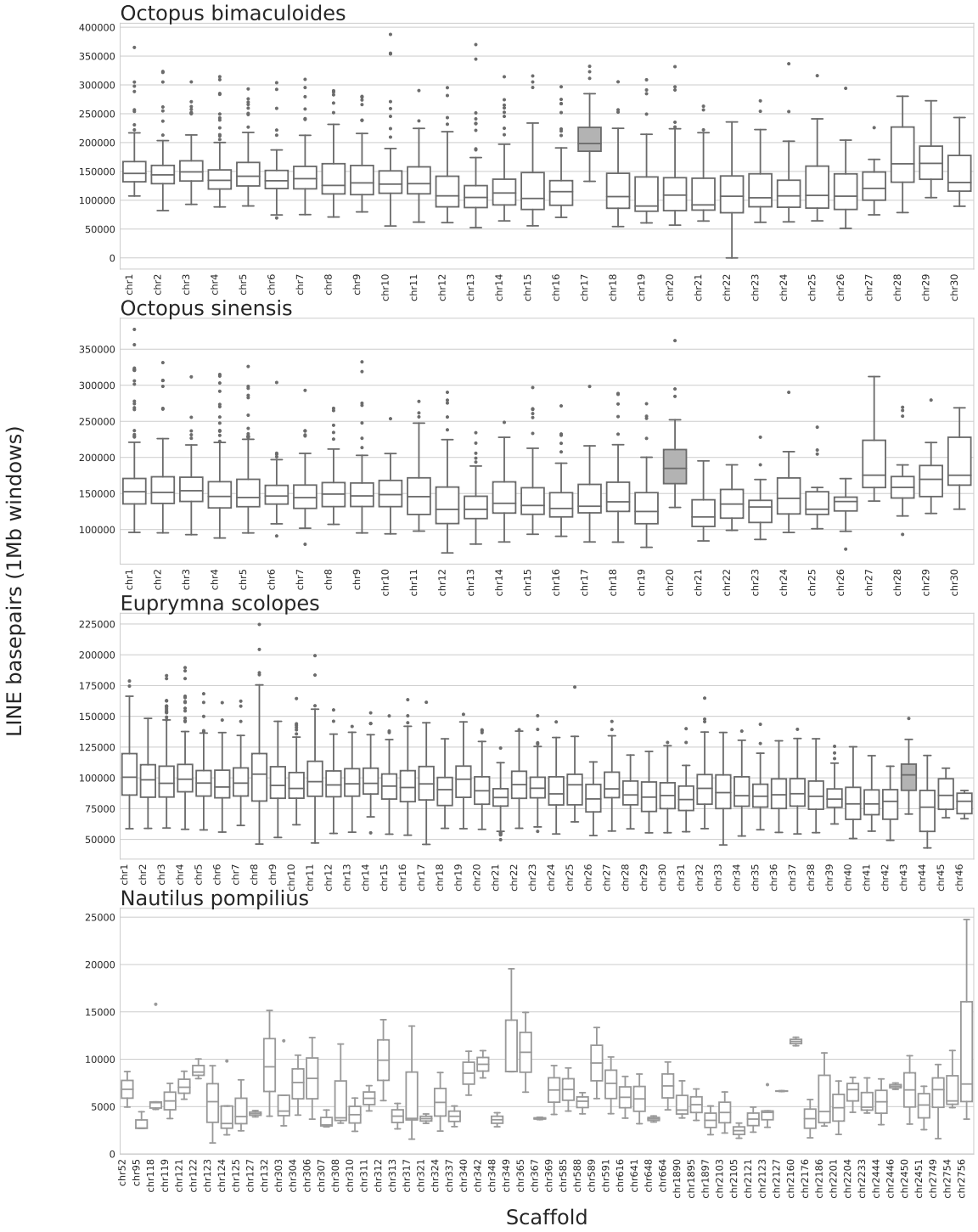

Figure S3. LINE content on chromosomes arms for four cephalod sequences. Data as in Figure 4 but with inclusion of *Nautilus* (bottom panel)

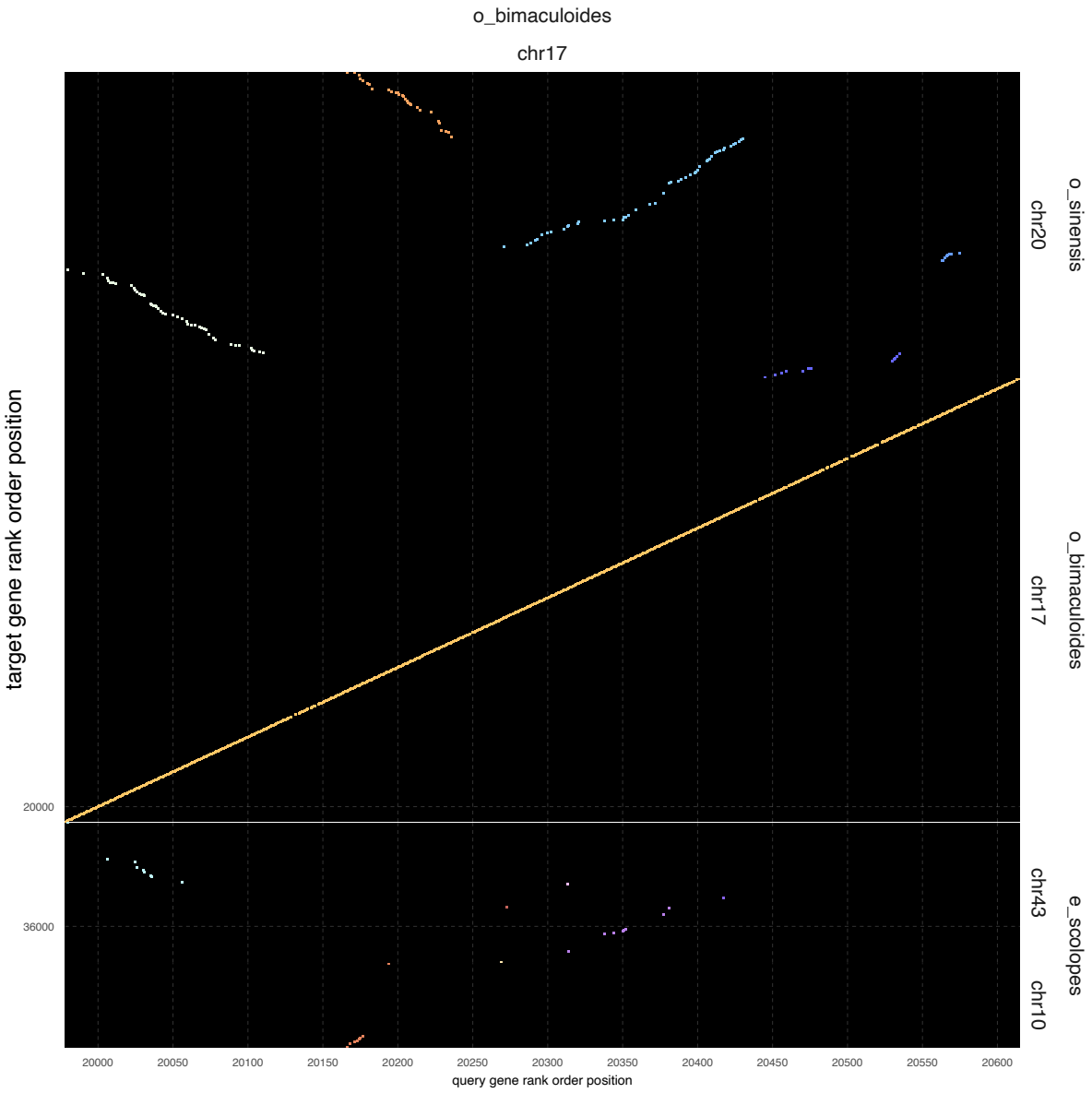

Figure S4. Dotplot of syntenic hits between the sex chromosomes of *O. bimaculoides*, *O. sinensis*, and *E. scolopes*.

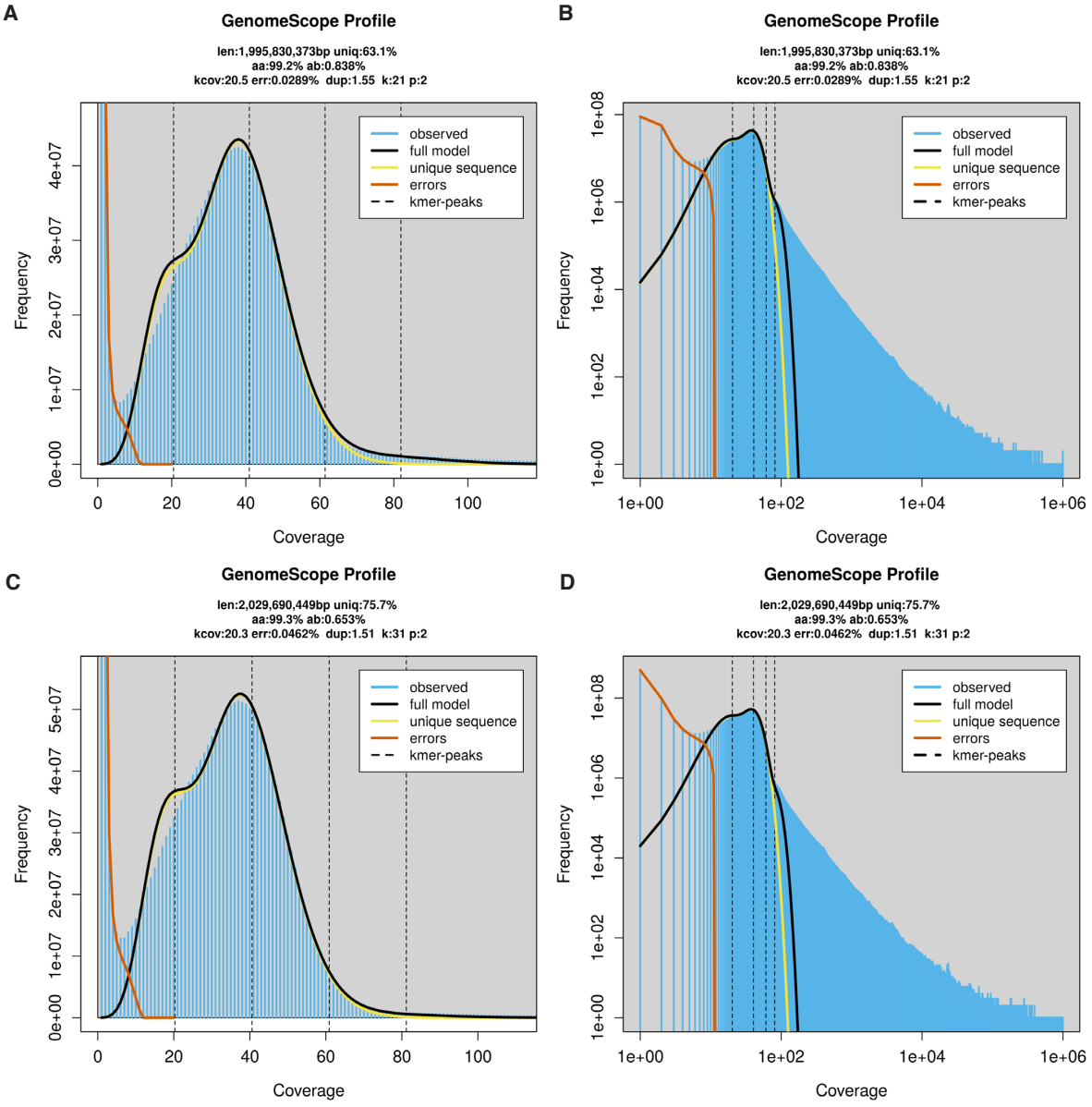

Figure S5.K-mer based genome size prediction of O. bimaculoides. **A** and **B** are 21-mer. **C** and **D** are 31-mer.

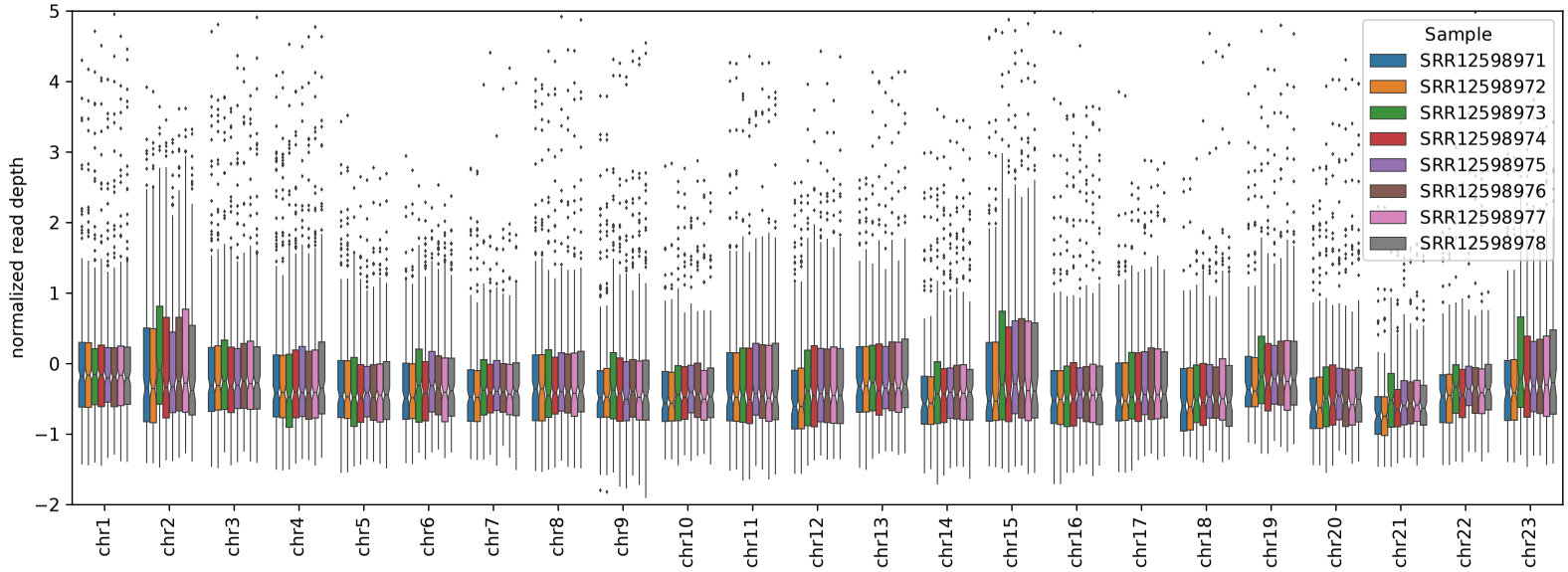

Figure S6. Normalized coverage of short read libraries mapped to chromosomes 1 through 23 in *E. scolopes*. Chromosome 43 is the only chromosome with distinct variation in coverage. SRA identifiers for each library are shown in the legend. Data was generated by (*16*).

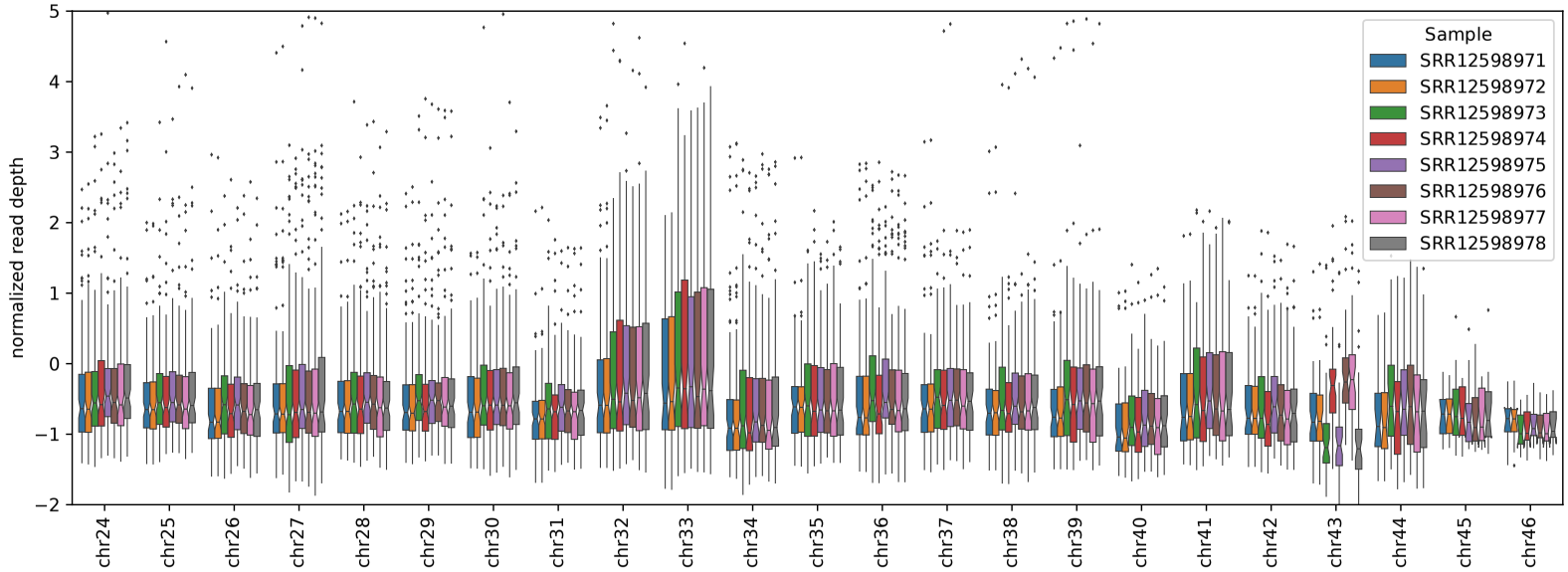

Figure S7. Normalized coverage of short read libraries mapped to chromosomes 24 through 46 in *E. scolopes*. Chromosome 43 is the only chromosome with distinct variation in coverage. SRA identifiers for each library are shown in the legend. Data was generated by (*16*).

	Albertin et al. 2015	**Albertin et al. 2022**	**Our Assembly**
Size	2.34 Gb	2.37 Gb	2.34 Gb
Number of scaffolds	151,674	370,699	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 (*5*), a subsequent reassembly of those same sequencing data (*6*) and the present assembly we introduce here.

Individual	Tissue	Data type	Total bases	Number of reads or read pairs	Depth	Notes
Ob1 (female adult)	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 (female adult)	optic lobe	Illumina NovaSeq 6000	90.3Gb	567M	71x	–
Obmac_M_3 (male adult)	optic lobe	Illumina NovaSeq 6000	76.2Gb	479M	60x	–
Obmac_F_10 (female adult)	gill	Illumina NovaSeq 6000	80.7Gb	507M	64x	–
Obmac_M_14 (male adult)	testes	Illumina NovaSeq 6000	66.0Gb	408M	52x	–

Table S2. Raw sequencing data collected

octo_gene_id	human_gene_id	human_gene_symbol	e-val	function	human reproductive mRNA expression?	human reproductive protein expression?
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)
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)
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)
obimac_0008638.1	no hit	-	-	-	-	-
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)
obimac_0008642.1	no hit	-	-	-	-	-
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
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
obimac_0008648.1	NP_001011547.2	SLC5A9	0.00E+00	Electrogenic Na(+)-coupled sugar symporter	ovary, uterus, prostate, testis	none
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
obimac_0008668.1	NP_001287785.1	FOSL1	2.00E-13	FOS Like 1, AP-1 Transcription Factor Subunit	ovary, uterus, prostate, testis	ovary
obimac_0008926.1	AAH14937.1	FHIP1A	3.00E-129	protein localization to perinuclear region of cytoplasm	ovary, uterus, prostate, testis	testis (fetal)
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)
obimac_0008956.1	no hit	-	-	-	-	-
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)
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
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)
obimac_0008989.8	NP_001381782.1	KLC1	0.00E+00	Kinesin Light Chain 1	ovary, uterus, placenta, prostate, testis	
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)

Table S3. Table of blast results from orthologous region genes of chromosome Z to humans

Species	**Source**	**Accession ID**	**Citation**
Nautilus pompilius	CNCB	GWHBECW00000000	(*38*)
Euprymna scolopes	NCBI	GCA_024364805.1	(*39*)
Sepia pharaonis	NCBI	GCA_903632075.3	(*40*)
Architeuthis dux	NCBI	GCA_006491835.1	(*41*)
Octopus minor	gigadb	N/A	(*42*)
Hapalochlaena maculosa	NCBI	GCA_015501135.1	(*43*)
Octopus bimaculoides	NCBI	PRJNA1076263	This Article
Octopus sinensis	NCBI	GCF_006345805.1	(*44*)

Table S4. Source of whole genome assemblies by species and the associated publications

.