Z-chromosome dosage compensation and a male-specific lncRNA in octopus

Natali Papanicolaou[1,#], Sebastian Wettersten[1], Alexander Kloosterman[1], Björn Reinius[1, #]

[1]Department of Medical Biochemistry and Biophysics, Karolinska Institutet, Sweden
 # Correspondence: bjorn.reinius@ki.se & natali.papanicolaou@ki.se

Abstract

Sex chromosomes typically present as two copies of a large, gene-rich differentiated chromosome in the homogametic sex (e.g. XX in female mammals, ZZ in male birds) but as a single copy in the heterogametic sex (XY in male mammals, ZW in female birds). This creates a significant gene-copy imbalance of the X and Z chromosomes between the sexes as well as between the single sex chromosome and the autosomes in the heterogametic sex, making its resolution essential. Dosage compensation and its molecular mechanisms have primarily been studied in select organisms, such as primates, rodents, birds, fruit flies, and *C. elegans*, while the extent and mechanisms of dosage compensation remain largely unexplored in the vast majority of animal taxa.

Recently, coleoid cephalopods, including octopuses, squids, and cuttlefish, were reported to possess a ZZ/Z0 sex chromosome system[1], where males carry two copies of the Z chromosome while females carry only one. However, whether dosage compensation occurs in these species remains unexplored. Here, we present evidence that Z-chromosome dosage compensation is achieved in octopus. In two species (*Octopus vulgaris* and *Octopus sinensis*), we observed male-to-female ratios of Z-chromosome expression exceeding 1 but significantly below 2, indicating partial dosage compensation, comparable to that achieved for the Z chromosome in birds. Additionally, Z-to-autosomal expression ratios were close to 1. Furthermore, we identified a Z-linked, male-specific long non-coding RNA (lncRNA) in *O. vulgaris* ("*Zmast*"), with conserved sequence in annotated Z-linked loci of *O. sinensis*—notably expressed in a male-specific manner also in this species.

Our results provide the first evidence of sex chromosome dosage compensation in octopus and introduce the intriguing possibility of a long non-coding RNA playing a sex-specific role in gene regulating.

Introduction

Genetic sex determination in animals typically depends on specialized sex chromosomes that differ in copy number between male and female cells. This is exemplified by well-characterized systems such as the XX/XY system in mammals and *Drosophila melanogaster*, where female cells carry two large X chromosomes, while male cells have a single X and a degenerated Y. Conversely, the ZW/ZZ system found in birds, some reptiles, and fish mirrors the XX/XY system, with males being the homogametic sex,

carrying two large Z chromosomes and females carrying a single Z chromosome and a degenerated W[2]. Each of these systems present the hemizygous sex (XY males in mammals, ZW females in birds and Z0 females in cephalopods) with a considerable gene-dosage imbalance arising from the presence of only a single copy of the non-degenerated chromosome and the autosomes. The significance of this imbalance has driven the evolution of various dosage-compensation mechanisms, including X-chromosome upregulation in *Drosophila*[3], X-chromosome inactivation together with X-upregulation in mammals[4], and Z-chromosome upregulation in birds[5], aligning with the theoretical propositions originally put forth by Susumu Ohno[3]. Collectively, these examples underscore the importance of resolving gene dosage imbalances across animal species. X-chromosome dosage compensation in eutherians (e.g. mouse and human) remains the most deeply studied dosage-compensation system, with the female-specific long non-coding RNA *Xist* identified as the key regulator, functioning as a X-chromosome silencer in cis.

Recently, a study on coleoid cephalopod genomes and their evolution, uncovered that various species of celoid cephalopods including octopuses, squids and cuttlefish possess a ZZ/Z0 sex chromosome system with males carrying two copies of the Z chromosome and females carrying only a single copy, with no evidence for the presence of a degenerated W chromosome[1]. Whether dosage compensation takes place remains unknown.

Here, we investigated, for the first time, whether dosage compensation occurs in coleoid cephalopods by leveraging publicly available RNA-sequencing datasets. We provide evidence of partial Z-chromosome dosage compensation in octopus and pinpoint a Z-encoded lncRNA expressed in a male-specific manner, suggestive of a sex-specific role in the regulation of chromosome Z.

Results

To investigate sex chromosome dosage compensation in octopus, we re-analyzed publicly available RNA-sequencing data from sex-matched tissues (Sub-oesophageal mass - SEB, Supra-oesophageal mass - SEM, optic lobe – OL and Anterior Arm – ARM) of *Octopus vulgaris*[4] (**Methods**). Given the ZZ/Z0 sex chromosome system, we began by calculating gene-wise Male-to-Female (M:F) expression ratios for each chromosome and tissue to identify potential outlier chromosomes. Indeed, in three out of four tissues examined, *O.vulgaris* chr16 displayed markedly increased M:F expression ratios compared to other chromosomes (p-value < 0.05, MWU for SEB, SEM and ARM, p-value > 0.05, MWU, for OL) (**Figure 1A-B**). In summary, shifted M:F ratios observed across tissues imply a sexually dimorphic dosage specifically of chromosome 16, suggesting that this chromosome constitutes Z in *Octopus vulgaris*.

The observed M:F ratios significantly deviated from the expected value of 2 that would follow in the absence of any dosage compensation in a Z/Z0 system. Deviation from this expected M:F ratio can result from various gene-regulatory modes, including increased transcription from the single female Z chromosome[5–7], inactivation of one Z in males, reduced transcription from the two male Z chromosomes[8], reduced[9] or

increased[10] RNA degradation in females and males respectively, or a combination of these modes, as previously observed in other species.

Based on our findings, we proceeded to compare chromosome expression ratios between the putative Z and autosomes in *O.vulgaris*. As female octopuses are expected to be hemizygous for chromosome Z, in absence of dosage compensation, its expression is expected to be considerably lower than those of autosomal pairs of similar gene content. Interestingly, we observed no significant difference between Z-chromosome-expression and autosomal expression in any female tissue examined **(Figure 1C**), providing evidence towards increased transcriptional output from the single female Z chromosome to achieve expression levels matching those of autosomes, similar to X-upregulation in mammals[6,7,9,11,12] and Z-upregulation in birds[5].

To determine whether *O.vulgaris* chr16 has sequence similarity to known Z chromosomes in other octopus species[1], we ran GENESPACE[13] using publicly available assemblies of *Octopus bimaculoides*[14] (GCA_001194135.2) and *Octopus sinensis*[15] (GCA_006345805.1). We observed a near-perfect conservation of syntenic blocks between *O.vulgaris* chr16 and *O.sinensis* chr LG20 (**Figure 2A**), the latter of which was recently identified as a Z chromosome[1], thereby validating *O. vulgaris* chr16 as the Z chromosome. Interestingly, we observed a similar degree of synteny between *O.vulgaris* chr16 and *O.bimaculoides* chr16, but not chr17 as recently suggested in Coffing et al[1]. We believe that this is a naming discrepancy, possibly due to a large number of unplaced scaffolds in the publicly accessible assembly version of the *O.bimaculoides* genome (ASM119413v2) and not due to disagreement between the syntenic alignments.

We proceeded to uncover whether any genes on *O.vulgaris'* putative Z displayed distinct expression signatures between male and female samples. Strikingly, we observed that a lncRNA, OCTVUL_1NCB338440 encoded by the putative Z chromosome, showed high, male-specific expression in all tissues examined (**Figure 2B**)—thus we called it **Z-ma**le-**s**pecific **t**ranscript (*Zmast*). Coding potential calculation[16] indeed classified this transcript as "*noncoding (weak)*" (alike *Xist*). *Zmast* is encoded in an 18kb-long locus, located between 19.5589–19.5769 Mb, annotated to be transcribed into two isoforms, T1 (18007 nt-long, 19.5590-19.5770, containing one exon (7098 nt-long; chrZ: 19.559–19.566Mb) and T2 (1969 nt-long; chrZ: 19.575–19.577 Mb), containing two overlapping exons (exon 1: 19.575004–19.576972 Mb; exon 2: 19.576719–19.576972 Mb) based on current assembly. Using BLAST, we found that a portion of the *Zmast* lncRNA is conserved in three segments, annotated as LOC115222586, LOC118767237 and LOC118767236, in *O.sinensis*, an octopus species that is believed to have diverged from *O.vulgaris* approximately 2.5MYA[17] (**Figure 2C-D**). Analysis of RNA-sequencing reads and junctions suggested that these segments belong to one and the same transcript. To explore the potential male-specific expression of *Zmast* transcripts in *O.sinensis*, we obtained publicly available datasets of RNA-sequencing of gonadal tissue (testis and ovary tissues) from *O.sinensis*[18].

Indeed, we observed high expression in all three orthologous segments in testis and the lack thereof in ovaries (**Figure 2C**).

Although the transcriptomic profiles of female and male gonads are expected to be dissimilar and therefore imperfect for calculating M:F ratios, we nonetheless calculated such rations for the gonads a did observe a tendency for increased M:F ratios for the Z chromosome (LG20) (**Figure 3A**). Using Z-to-autosome (Z:A) ratios calculated within tissues, we observed that in both male and female samples, the transcriptional output of Z was comparable to autosomes (p=0.41, Mann-Whitney-U test), pointing to dosage compensation between the single female Z and the pairs of autosomes (**Figure 3B**) similar to our previous observations in *O.vulgaris* tissues.

In summary, we found evidence for Z-chromosome dosage compensation in octopus and identified a Z-encoded male-specific transcript (*Zmast*), expressed specifically in males, pointing to a potential sex-specific role of this lncRNA in octopus.

Discussion

The recent revelation of a ZZ/Z0 sex chromosome system in coleoid cephalopods including octopus, cuttlefish, squid, and the chambered nautilus by Coffing and colleagues marks a landmark discovery that pushed the evolutionary age of sex chromosomes to ~480MYA[1]. The revealed hemizygosity begs the question of whether the Z-dosage discrepancy in female cephalopod cells is resolved at the transcriptional level similarly to other sex chromosome systems investigated across the animal kingdom. Using publicly available RNA-sequencing datasets from matched *O.vulgaris* somatic tissues[4] and *O.sinensis*[18] adult gonadal tissues, we found that Male-to-Female ratios of Z-chromosome gene expression are significantly lower than 2, which is the expected ration in the absence of any dosage-compensating mechanism, while being above 1, thereby indicating partial dosage compensation. Interestingly, this observation bears similarity with the partial Z-chromosome dosage compensation known in birds[5]. Furthermore, we observed near-balanced Z:autosome ratios within samples, suggesting that the transcriptional output of the single female Z chromosome is upregulated to match the two autosomes, something that was previously observed in mammals[6,7,12], birds[5], *Drosophila melanogaster*[19,20] and *Caenorabditis elegans*[11]—thus we conclude that sex-chromosome upregulation also occurs in octopus.

Curiously, we identified an lncRNA, "*Zmast*" (OCTVUL_1NCB338440), located on *O.vulgaris* chromosome Z (annotated as chr16 in GCA_951406725.2 genome assembly), displaying high male-specific expression, which is conserved in *O.sinensis* in three shorter annotated non-coding transcripts (LOC115222586, LOC118767237 and LOC118767236) and exclusively expressed in males also in this species. Our current analysis suggests that these annotated segments represent same transcribed unit. LncRNAs have long been implicated as master regulators of sex-chromosome dosage compensation in various species, such as *rhoX1* and *rhoX2* responsible for X-chromosome upregulation in

Drosophila[20], *Xist*[21,22] and *Rsx*[23], responsible for X-chromosome inactivation in eutherians and marsupials respectively. More recently, the lncRNA *MAYEX* was shown to be responsible for X-chromosome upregulation of the single male X in the green anole lizard[24]. The possibility that lncRNAs may be involved in coeloid cephalopod sex chromosome dosage compensation, therefore, does not stretch credulity. If this lncRNA is involved in dosage compensation, it could potentially act by either downregulating the transcriptional output of both Z chromosomes, or by preferentially inactivating one of the two Z chromosomes, mirroring mechanisms seen in mammals[21]. The function and cellular localization of *Zmast* should be investigated experimentally in following studies. The observation of Z-chromosome dosage compensation in octopus extends the general concept of the need for sex-chromosome dosage compensation across taxa.

Methods

Processing of publicly available RNA-seq datasets. *Octopus vulgaris* and *Octopus sinensis* bulk tissue RNA-seq fastq files[4,18] were obtained from ENA under accession numbers: E-MTAB-3957 and SRP499429. The latest available assemblies of the *O.vulgaris* and *O.sinensis*[15] genome was obtained from NCBI (assemblies: GCA_951406725.2, GCA_006345805.1).

To improve mapping accuracy, STAR (v.2.7.2a) indexes for the respective genomes were created prior to mapping using both genome fasta and GTF files (STAR –runMode genomeGenerate, --sjdbGTFfile *.gtf). Raw fastq files were aligned to the respective genomes and transcriptomes and reads were quantified using STAR (v.2.7.2a) with the following options:

--runMode alignReads,
--clip3pAdapterSeq AGATCGGAAGAGCACACGTCTGAACTCCAGTCA AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
--quantMode TranscriptomeSAM GeneCounts,
--outSAMtype BAM Unsorted SortedByCoordinate

Gene expression read counts were TPM normalized by dividing gene counts by the effective gene length in kb (RPK), dividing the total gene-length-normalized counts by 10^6 to calculate the scaling factor and dividing the RPK by the scaling factor.

Expression ratio calculations

Gene-wise Male:Female expression ratios were calculated for expressed genes (TPM >1) per tissue for *O.vulgaris* and *O.sinensis* data. Chromosome:Autosome expression ratios were calculated per sample, for expressed genes (TPM >1) and after exclusion of the top 5% expressed genes per chromosome and tissue to avoid bias from highly expressed tissue-specific genes.

Nucleotide BLAST (BLASTN)

Discontiguous megablast was used to compare *O.vulgaris Zmast* lncRNA exonic and intronic sequences to genomic sequences of *O.sinensis* with default settings. Sequences producing significant alignments with percentage identity more than 90%

and query cover more than 80% were considered for further analysis.

Synteny calculation

The latest available assemblies of the *Octopus bimaculoides*, *Octopus sinensis* and *Octopus vulgaris* genome was obtained from NCBI (assemblies: GCF_001194135.2, GCF_006345805.1 and GCA_951406725.2). From these assemblies the genomic gff files and the translated cds were used for analysis. Analysis was performed using R Statistical Software (v4.1.2) and the R package GENESPACE (v1.3.1)[13]. Default GENESPACE workflow and output was used to plot the syntenic block analysis for all chromosomes. Region of interest syntenic block plot was made using *O. vulgaris* as reference with prior GENESPACE output, without rendering background and using chromosome 16 as the region of interest.

References

1. Coffing, G. C. *et al.* Cephalopod Sex Determination and its Ancient Evolutionary Origin. 2024.02.21.581452 Preprint at https://doi.org/10.1101/2024.02.21.581452 (2024).
2. Graves, J. A. M. Evolution of vertebrate sex chromosomes and dosage compensation. *Nat. Rev. Genet.* **17**, 33–46 (2016).
3. Ohno, S. Various Consequences of the Dosage Compensation by X-inactivation. in *Sex Chromosomes and Sex-Linked Genes* (ed. Ohno, S.) 101–112 (Springer, Berlin, Heidelberg, 1966). doi:10.1007/978-3-642-88178-7_9.
4. Petrosino, G. *et al.* Identification of LINE retrotransposons and long non-coding RNAs expressed in the octopus brain. *BMC Biol.* **20**, 116 (2022).
5. Papanicolaou, N. *et al.* Multi-layered dosage compensation of the avian Z chromosome. 2024.08.20.608780 Preprint at https://doi.org/10.1101/2024.08.20.608780 (2024).
6. Larsson, A. J. M., Coucoravas, C., Sandberg, R. & Reinius, B. X-chromosome upregulation is driven by increased burst frequency. *Nat. Struct. Mol. Biol.* **26**, 963–969 (2019).
7. Lentini, A. *et al.* Elastic dosage compensation by X-chromosome upregulation. *Nat. Commun.* **13**, 1854 (2022).
8. Ercan, S. & Lieb, J. D. C. elegans dosage compensation: A window into mechanisms of domain-scale gene regulation. *Chromosome Res.* **17**, 215–227 (2009).
9. Deng, X. *et al.* Mammalian X Upregulation Is Associated with Enhanced Transcription Initiation, RNA Half-Life, and MOF-Mediated H4K16 Acetylation. *Dev. Cell* **25**, 55–68 (2013).
10. Fallahshahroudi, A. *et al.* A male-essential microRNA is key for avian sex chromosome dosage compensation. 2024.03.06.581755 Preprint at https://doi.org/10.1101/2024.03.06.581755 (2024).
11. Deng, X. *et al.* Evidence for compensatory upregulation of expressed X-linked genes in mammals, Caenorhabditis elegans and Drosophila melanogaster. *Nat. Genet.* **43**, 1179–1185 (2011).
12. Kharchenko, P. V., Xi, R. & Park, P. J. Evidence for dosage compensation between the X chromosome and autosomes in mammals. *Nat. Genet.* **43**, 1167–1169 (2011).
13. Lovell, J. T. *et al.* GENESPACE tracks regions of interest and gene copy number variation across multiple genomes. *eLife* **11**, e78526 (2022).
14. Albertin, C. B. *et al.* Genome and transcriptome mechanisms driving cephalopod evolution. *Nat. Commun.* **13**, 2427 (2022).
15. Li, F. *et al.* Chromosome-level genome assembly of the East Asian common octopus (Octopus sinensis) using PacBio sequencing and Hi-C technology. *Mol. Ecol. Resour.* **20**, 1572–1582 (2020).
16. Kong, L. *et al.* CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Res.* **35**, W345–W349 (2007).
17. Destanović, D. *et al.* A chromosome-level reference genome for the common octopus, Octopus vulgaris (Cuvier, 1797). *G3 GenesGenomesGenetics* **13**, jkad220 (2023).
18. Li, F. *et al.* Gonadal Transcriptome Sequencing Analysis Reveals the Candidate Sex-Related Genes and Signaling Pathways in the East Asian Common Octopus, Octopus sinensis. *Genes* **15**, 682 (2024).
19. Deng, X. *et al.* Evidence for compensatory upregulation of expressed X-linked genes in mammals, Caenorhabditis elegans and Drosophila melanogaster. *Nat. Genet.* **43**, 1179–1185 (2011).
20. Lucchesi, J. C. & Kuroda, M. I. Dosage Compensation in Drosophila. *Cold Spring Harb. Perspect. Biol.* **7**, a019398 (2015).
21. Brown, C. J. *et al.* A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. *Nature* **349**, 38–44 (1991).
22. Kay, G. F. *et al.* Expression of Xist during mouse development suggests a role in the initiation of X chromosome inactivation. *Cell* **72**, 171–182 (1993).
23. Grant, J. *et al.* Rsx is a metatherian RNA with Xist-like properties in X-chromosome inactivation. *Nature* **487**, 254–258 (2012).
24. Tenorio, M. *et al.* MAYEX is an old long noncoding RNA recruited for X chromosome dosage compensation in a reptile. *Science* **385**, 1347–1354 (2024).

Figure 1

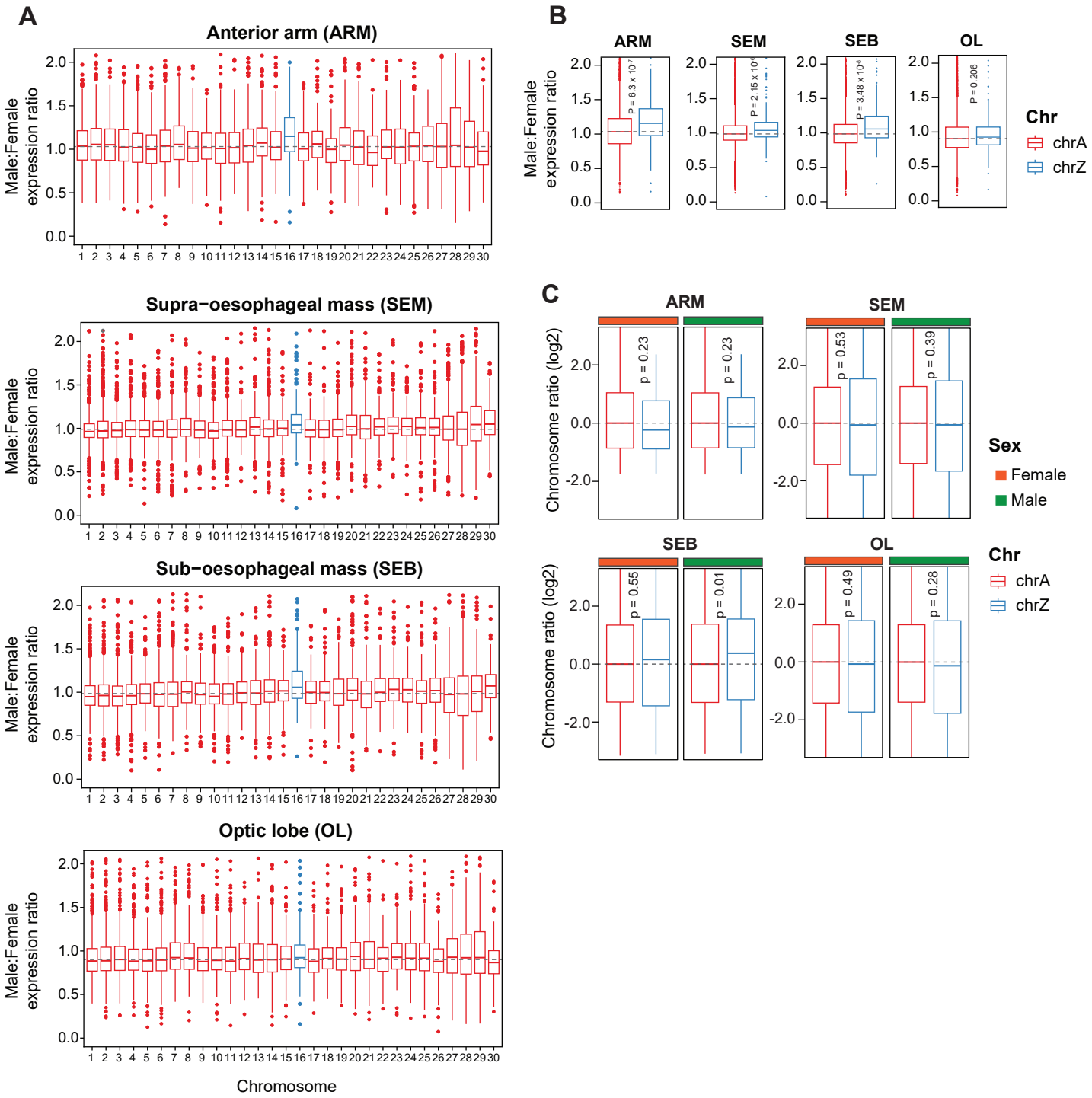

Figure 1. Z-chromosome dosage compensation in *O.vulgaris*.
(A) Gene-wise Male:Female (M:F) expression ratios per chromosome and tissue in *O.vulgaris*, calculated for expressed genes (TPM >1). Dashed line in gray represents median autosomal M:F ratio. Autosomal chromosomes shown in red and the putative Z in blue. **(B)** Male:Female ratios for autosomes and Z chromosomes per tissue in *O.vulgaris* for expressed genes (TPM >1). Mann-Whitney-U test used for significance testing. Dashed line in gray represents median autosomal M:F ratio. Autosomal chromosomes shown in red and putative Z in blue. **(C)** Chromosome:autosome ratios (log2 scale) calculated for expressed genes (TPM >1) for *O.vulgaris* tissues. Dashed line in gray denotes median autosomal ratio. Sex is denoted by the colored bars over the individual panels, with orange denoting female and green denoting male samples. Autosomal chromosomes shown in red and putative Z in blue. Mann-Whitney-U test used for significance testing.

Figure 2

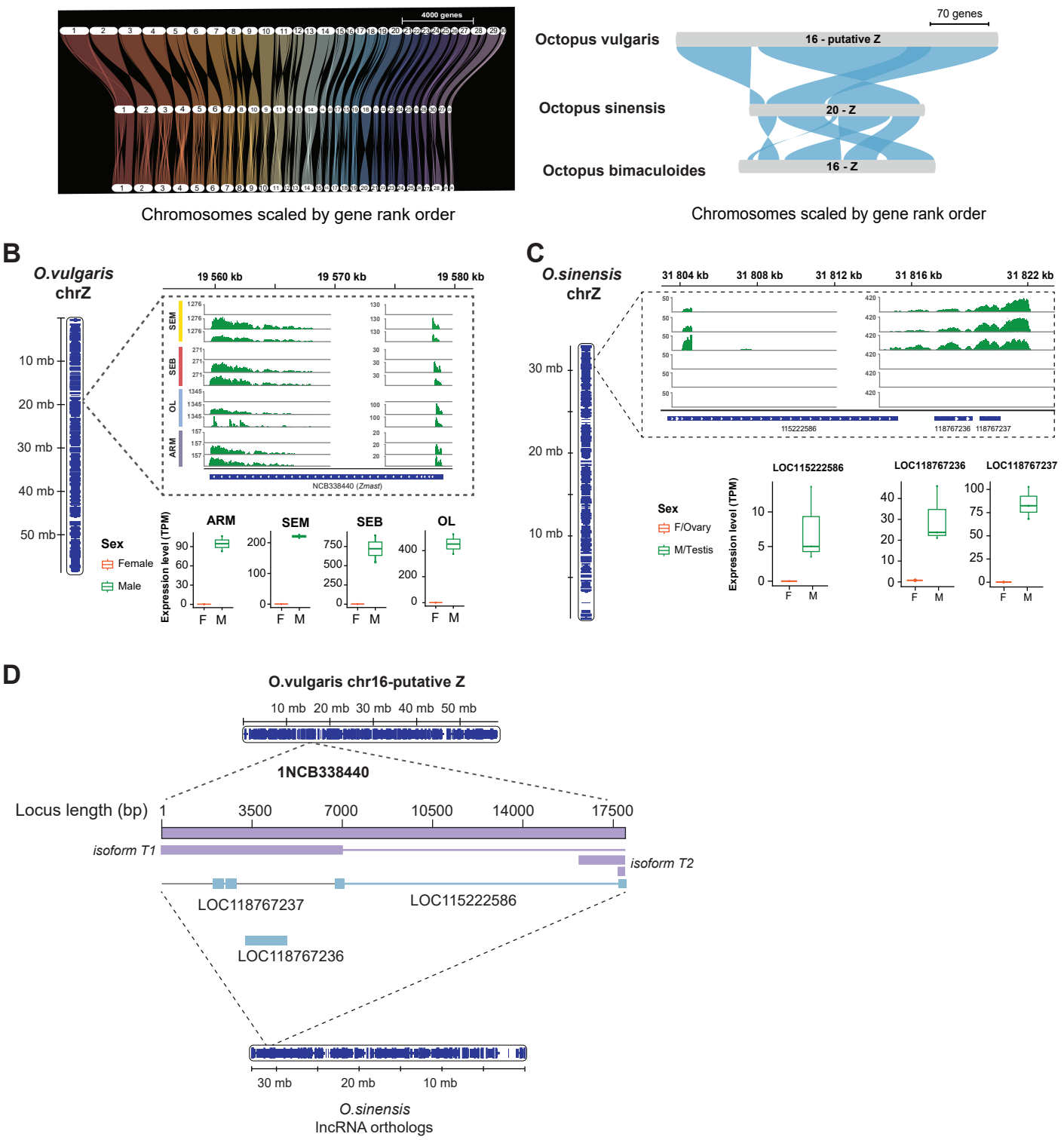

Figure 2. Identification of a male-specific Z-linked lncRNA.
(A) GENESPACE output if chromosomal syntenic alignments between O.vulgaris, O.sinensis and O.bimaculoides. **(B)** Expression of Z-chromosome male-specific transcript (Zmast) in O.vulgaris tissues. Left: O.vulgaris chrZ with length shown in Mb. Right top: Genomic tracks of RNA-seq reads across Zmast locus with male samples shown in green and female samples in orange by tissue sample (shown on the left). Right bottom: Expression of Zmast lncRNA in TPM per tissue and sex. **(C)** Expression of Zmast orthologous segments in O.sinensis tissues. Left: O.sinensis chrZ with length shown in Mb. Right top: Genomic tracks of RNA-seq reads across Zmast ortholog lncRNA loci with male samples shown in green and female samples in orange. Right bottom: Expression of Zmast ortholog lncRNAs in TPM per tissue. **(D)** Sequence alignments between O.vulgaris Zmast lncRNA and O.sinensis Zmast ortholog.

Figure 3

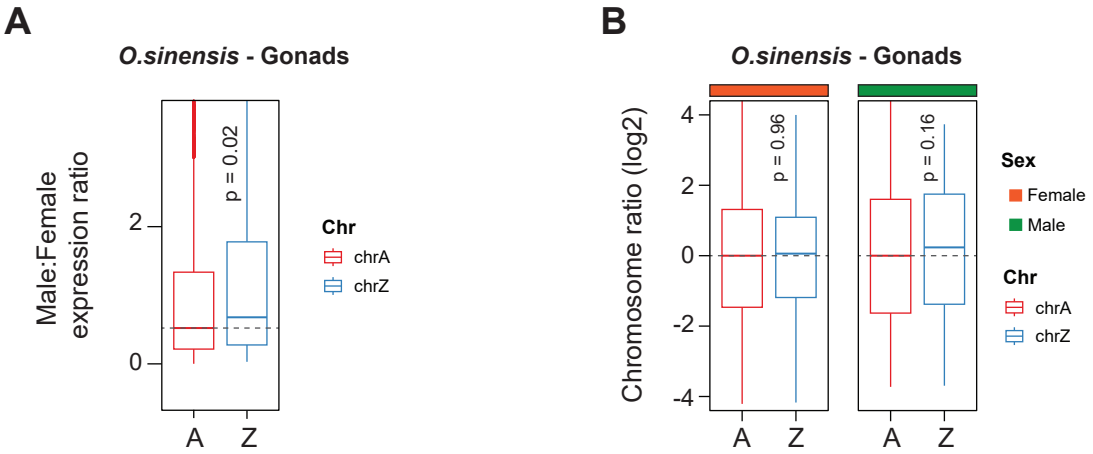

Figure 3. Z-chromosome expression in *O.sinensis* gonadal tissue.
(A) Chromosome:autosome ratios (log2 scale) calculated for expressed genes (TPM >1) for *O.sinensis* tissues. Dashed line in gray denotes median autosomal ratio. Sex is denoted by the colored bars over the individual panels, with orange denoting female and green denoting male samples. Autosomal chromosomes shown in red and putative Z in blue. Mann-Whitney-U test used for significance testing. **(B).** Male:Female expression ratios for autosomes and Z chromosomes per tissue in *O.sinensis* for expressed genes (TPM >1). Mann-Whitney-U test used for significance testing. Dashed line in gray represents median autosomal M:F ratio. Autosomal chromosomes shown in red and putative Z in blue.