

A chromosome-level genome assembly of the disco clam, *Ctenoides ales*

Kyle E. McElroy ¹, Rick Masonbrink ², Sivanandan Chudalayandi,² Andrew J. Severin ², Jeanne M. Serb ¹

¹Department of Ecology, Evolutionary, and Organismal Biology, Iowa State University, Ames, IA 50011, USA

²Genome Informatics Facility, Iowa State University, Ames, IA 50011, USA

Corresponding author: Kyle E. McElroy, Department of Ecology, Evolutionary, and Organismal Biology, Iowa State University, Ames, IA 50011, USA.

Email: kmcelroy@iastate.edu

The bivalve subclass Pteriomorpha, which includes the economically important scallops, oysters, mussels, and ark clams, exhibits extreme ecological, morphological, and behavioral diversity. Among this diversity are five morphologically distinct eye types, making Pteriomorpha an excellent setting to explore the molecular basis for the evolution of novel traits. Of pteriomorphian bivalves, Limida is the only order lacking genomic resources, greatly limiting the potential phylogenomic analyses related to eyes and phototransduction. Here, we present a limid genome assembly, the disco clam, *Ctenoides ales* (*C. ales*), which is characterized by invaginated eyes, exceptionally long tentacles, and a flashing light display. This genome assembly was constructed with PacBio long reads and Dovetail Omni-C™ proximity-ligation sequencing. The final assembly is ~2.3Gb and over 99% of the total length is contained in 18 pseudomolecule scaffolds. We annotated 41,064 protein coding genes and reported a BUSCO completeness of 91.9% for metazoa_obd10. Additionally, we report a complete and annotated mitochondrial genome, which also had been lacking from Limida. The ~20Kb mitogenome has 12 protein coding genes, 22 tRNAs, 2 rRNA genes, and a 1,589 bp duplicated sequence containing the origin of replication. The *C. ales* nuclear genome size is substantially larger than other pteriomorphian genomes, mainly accounted for by transposable element sequences. We inventoried the genome for opsins, the signaling proteins that initiate phototransduction, and found that, unlike its closest eyed-relatives, the scallops, *C. ales* lacks duplication of the rhabdomeric G_q-protein-coupled opsin that is typically used for invertebrate vision. In fact, *C. ales* has uncharacteristically few opsins relative to the other pteriomorphian families, all of which have unique expansions of xenopsins, a recently discovered opsin subfamily. This chromosome-level assembly, along with the mitogenome, is a valuable resource for comparative genomics and phylogenetics in bivalves and particularly for the understudied but charismatic limids.

Keywords: mollusca; bivalve; opsin; transposable element; mitochondria

Introduction

The bivalve subclass Pteriomorpha includes many of the most economically important bivalves such as mussels, oysters, and scallops. Among the numerous morphological innovations in this clade are nonhomologous pallial eye types, including pigmented cups and compound eyes in Arcidae, cap eyespots in Ostreidae, mirror eyes in Pectinidae, and invaginated eyes in Limidae (Audino et al. 2020). The many origins of eyes in Pteriomorpha make this clade a compelling setting to study how novel traits arise and whether divergence in genetic architecture underlies this evolution.

To date, over 30 pteriomorphian genomes have been sequenced including at least one chromosome-level assembly from Arcida (Bai et al. 2019), Mytilida (Yang et al. 2021), Ostreida (Peng et al. 2020; Takeuchi et al. 2022), and Pectinida (Kenny et al. 2020). Outside of scallops (Pectinidae), no other eyed species have been sequenced, therefore limiting the possibility of comprehensive phylogenomic comparative studies on the evolution of eyes in this taxonomic lineage. Here, we present an annotated chromosome-level genome assembly of *Ctenoides ales* (*C. ales*) (Finlay, 1927), an eyed limid species. This genomic data will be

valuable for exploring the repeated evolution of eyes in Pteriomorpha and represent the first genomic data from Limida, a diverse order of about 200 extant species (MolluscBase) commonly referred to as flame scallops, file clams, or file shells.

The charismatic *C. ales* (Fig. 1a) is known as the “disco clam” or “electric flame scallop” for its flashing mantle display. Despite its bioluminescent-like appearance, this presumed antipredator display is actually the result of light reflecting from silica nanospheres incorporated into the mantle tissue (Dougherty et al. 2014). Limids are also known for their brightly colored mantle tissue lining the two valves, which is a source of chemical deterrent from predators (Dougherty et al. 2019). Long tentacles are used not only for sensory perception, but also for swimming and chemical defense (Mikkelsen and Bieler 2003; Donovan et al. 2004; Dougherty et al. 2019). At the base of these tentacles in many limid species are multiple “invaginated” eyes embedded in the mantle tissue (Bell and Mpitso 1968; Mpitso 1973; Morton 2000). Similar to scallops (Gorman and McReynolds 1969), limid eyes have two distinct retinas, a “proximal” and a “distal” retina that are made up of rhabdomeric and ciliary photoreceptor cells, respectively (Speiser et al. 2023), with opposing responses to light

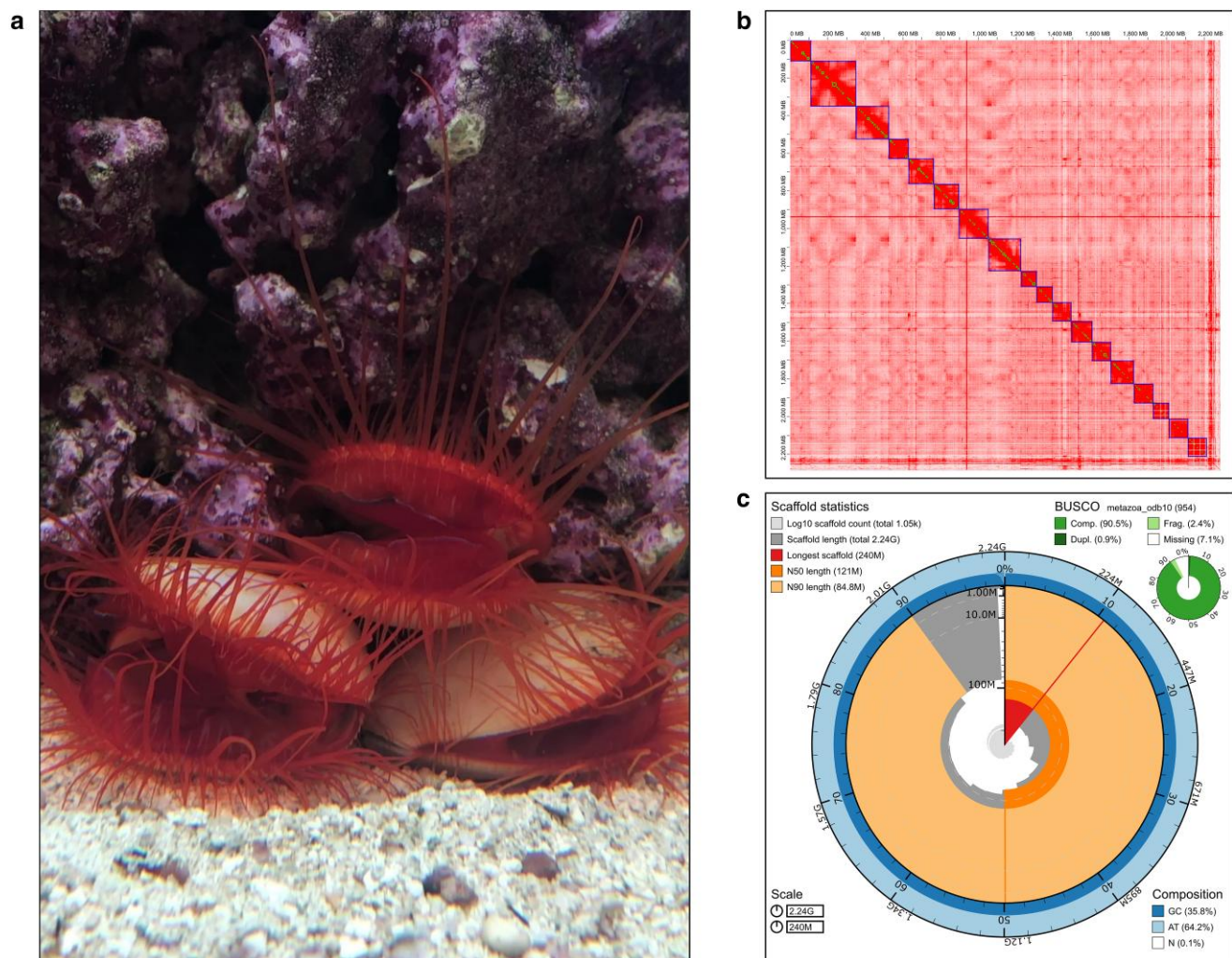


Fig. 1. *Ctenoides ales* and summary of its genome assembly. a) Adult *Ctenoides ales* in aquarium setting, flashing light display visible in middle individual. Image credit: Jeanne M Serb. b) Hi-C contact map for *C. ales*, highlighting the 18 chromosomes recovered from the genome assembly. Darker shades (red) indicates higher density of contact, large (blue) and small (green) boxes denote chromosome and contigs, respectively. c) Snail plot summarizing key assembly statistics for final *C. ales* assembly with BUSCO results (–long –augustus parameters enabled).

(Mpitosos 1973). However, the morphological characterization of limid eyes is complicated by differing interpretations and taxonomically narrow studies. Whether limids have spatial vision is still unknown (Speiser et al. 2023). Morphological and behavioral analyses suggest that *C. ales* has poor visual resolution and is unable to distinguish light directionality (Dougherty et al. 2017), in contrast to its closest eyed relatives, the scallops, equipped with image-forming eyes (Land 1965). Genomic comparisons across a variety of eyed and eyeless pteriomorphians should yield great insights into the molecular evolutionary process underlying the emergence of novel photoreceptive organs.

Gene duplication is an important source of new genetic information that can be used in the evolution of novel traits (Ohno 1970; Lynch and Conery 2000; Zhang 2003; Birchler and Yang 2022). Opsin duplication is important for the path to wider spectral sensitivity, particularly in vertebrates (e.g. Escobar-Camacho et al. 2020; reviewed in Hagen et al. 2023) and arthropods (Briscoe 2001; Briscoe et al. 2010; Bentley et al. 2016). Opsins are transmembrane G-protein-coupled receptor (GPCR) proteins that form photopigments by binding to a chromophore, typically a vitamin A-derived retinol, which upon light absorption starts a

phototransduction cascade. Opsins are a phylogenetically diverse and widespread protein family with a variety of both ocular and extraocular light-dependent functions and light-independent functions (Terakita 2005; Shichida and Matsuyama 2009; Terakita and Nagata 2014; Moraes et al. 2021). Opsin classification is based on the specific G-protein an opsin activates (e.g. G_q , G_t , G_i , G_o , G_s), the type of photoreceptor cell where it is expressed (e.g. ciliary vs rhabdomic), and its phylogenetic placement (e.g. tetraopsins) (reviewed in Shichida and Matsuyama 2009; Porter et al. 2011). Although the phrase “visual opsin” has recently come into question (Feuda et al. 2022), typically vertebrates rely on ciliary (c-type) opsins that couple with G_t proteins for vision, bilaterian invertebrates use rhabdomic (r-type) G_q -opsins, and cnidarians use a distinct group of G_s -opsins. Transcriptome sequencing has revealed multiple duplications of the rhabdomic opsins (r-opsins) in scallops (Porath-Krause et al. 2016), which have a complex, image-forming mirror eye, raising the possibility that opsin duplication is characteristic of eyed pteriomorphian lineages. We explore this hypothesis by scanning the *C. ales* genome assembly for opsins and phylogenetically analyzing them in the broader context of Pteriomorphian opsin evolution.

Materials and methods

Sample collection and DNA extraction

Live *C. ales* individuals, acquired through the aquaria pet trade, were dissected from their shells, then whole animals were immediately flash frozen in liquid nitrogen and stored at -80°C . Frozen adductor muscle tissue (400 mg) was homogenized into a fine power in liquid nitrogen and then DNA was extracted using a QIAGEN Genomic-tip protocol for tissue. Briefly, powdered tissue was lysed in QIAGEN G2 buffer at 50°C with RNase A and Proteinase K for 2 hours, then the lysate was passed through a QIAGEN Genomic-tip by gravity. DNA was precipitated with 0.7 volumes of isopropanol, pelleted, and washed with 70% ethanol before resuspension.

PacBio library and sequencing

DNA samples were quantified using Qubit 2.0 Fluorometer (Life Technologies, Carlsbad, CA, USA). The PacBio SMRTbell library (~20 kb) for PacBio Sequel was constructed using SMRTbell Express Template Prep Kit 2.0 (PacBio, Menlo Park, CA, USA) using the manufacturer's recommended protocol. The library was bound to polymerase using the Sequel II Binding Kit 2.0 (PacBio). Sequencing was performed on PacBio Sequel II 8 M SMRT cells.

Dovetail Omni-C library preparation and sequencing

Frozen adductor muscle tissue (50 mg) from the same individual used for PacBio sequencing was pulverized to a powder in liquid nitrogen then used as input to prepare the Dovetail Omni-C library. Chromatin was fixed in place with formaldehyde in the nucleus and then extracted. Fixed chromatin was digested with DNase I, and chromatin ends were repaired and ligated to a biotinylated bridge adapter followed by proximity ligation of adapter containing ends. After proximity ligation, crosslinks were reversed, and the DNA purified. Purified DNA was treated to remove biotin that was not internal to ligated fragments. Sequencing libraries were generated using NEBNext Ultra enzymes and Illumina-compatible adapters. Biotin-containing fragments were isolated using streptavidin beads before PCR enrichment of each library. The library was sequenced on an Illumina HiSeqX platform to produce approximately 30x sequence coverage. Then HiRise used MQ > 50 reads for scaffolding.

RNA isolation and sequencing

Flash frozen whole-animal tissue was stored at -80°C until preparation for RNA isolation. For every animal used, a sample of adductor, mantle, and eyes was prepared. Isolated tissue was first ground into a homogenous powder using an OPS Diagnostics LLC CryoGrinder™ System and then stored at -80°C until RNA isolation. Total RNA was isolated from ground tissue powder using an E.Z.N.A Total RNA Kit II (Omega BIO-TEK). RNA purity was assessed with a Nanodrop ND-1000 Spectrophotometer and ND-1000 3.2.1 software. RNA integrity and quantity were determined with an Agilent 2100 Bioanalyzer (Agilent Technologies). Libraries for sequencing were prepared with the NEBNext Ultra Directional RNA Library Prep Kit (Illumina), following poly(A) mRNA enrichment with Oligo dT Beads, and then sequenced (2 × 150 bp reads) on an Illumina NovaSeq 6000 at Iowa State University's DNA Facility.

Draft assembly

Approximately 290.8 Gb of PacBio Continuous Long Reads (CLR) reads were used as an input for assembly by WTDBG2 v2.5

(Ruan and Li 2020) with genome size set to 1.7 Gb, minimum read length 20,000, and minimum alignment length 8,192. Additionally, realignment was enabled with the -R option and read type was set with the option -x sq. Blast results of the WTDBG2 output assembly against the NCBI NT database were used as input for Blobtools v1.1.1 (Laetsch and Blaxter 2017) and scaffolds identified as possible contamination were removed from the assembly. Finally, purge_dups v1.2.3 (Guan et al. 2020) was used to remove haplotigs and contig overlaps.

HiRise scaffolding

The de novo WTDBG2 assembly and Dovetail Omni-C library reads were used as input data for the proximity ligation-based genome-scaffolding pipeline, HiRise (Putnam et al. 2016). Dovetail Omni-C library sequences were aligned to the draft input assembly using (<https://github.com/lh3/bwa>) (Li 2013). The separations of Dovetail Omni-C read pairs mapped within draft scaffolds were analyzed by HiRise to produce a likelihood model for genomic distance between read pairs, and the model was used to identify and break putative misjoins, to score prospective joins, and make joins above a threshold.

Final assembly

Dovetail assembly scaffolds were scaffolded using Juicer v1.5.7 (Durand et al. 2016), 3D-DNA v180114 (Dudchenko et al. 2017), Juicebox v1.11.08 (Robinson et al. 2018), based on BWA 0.7.17 alignments of Omni-C reads. Contigs that could not be placed in pseudomolecules were subjected to redundancy filtering using coordinates from two criteria: mapping contigs to the pseudomolecules with Minimap v2.2 and repeats identified as described in the following. These coordinates were merged using Bedtools (Quinlan and Hall 2010) merge, and when a contig could achieve 90% identity across 90% of the contig length to a pseudomolecule, it was removed. Additionally, unplaced contigs were again assessed as potential contamination with Blobtools v2.2.0, based on blastn v2.11.0 (Camacho et al. 2009) hits against the NCBI nt database (downloaded May 30, 2020) and mapping coverage from Minimap v2.2 (Li 2018) alignments of PacBio subreads. A final assessment of genome completeness following contig elimination was measured with BUSCO v5.1.2 (Waterhouse et al. 2018) using metazoa_odb10, mollusca_odb10, and eukaryota_odb10.

Transposable element characterization

De novo repeat identification was conducted with RepeatModeler v2.0.2 (Flynn et al. 2020), utilizing RepeatScout v1.0.6 (Price et al. 2005), RECON v1.08 (Bao and Eddy 2002), LTR_Retrieve v2.9.0 (Ou and Jiang 2018), and the Extensive de novo TE Annotator (EDTA) v2.0.1 (Ou et al. 2019), which uses three pipelines for TE discovery based on structural characteristics of terminal inverted repeated (TIR) DNA transposons, long terminal repeat (LTR) retrotransposon, and helitrons. Next, TEsorter v1.4.6 (Zhang et al. 2022) was used to classify sequences from the RepeatModeler and EDTA output based on REXdb v. 3 HMM profiles and CD-HIT-EST v4.8.1 was used to combine classified sequences and reduce redundancy between the two sets of repeats (Li and Godzik 2006; Fu et al. 2012). Because SINEs lack any coding sequences and no structural detection pipeline is included in EDTA, the sequences annotated as SINEs by RepeatModeler were queried against Repbase online (Bao et al. 2015) with CENSOR and kept sequences matching SINEs with scores >250. We combined the TEsorter classified sequences with SINEs and the remaining EDTA sequences identified via structural features (e.g. LTRs) into a single TE library for each

species, and then masked each genome with its specific TE library using RepeatMasker v4.1.2-P (Smit et al. 2015). We used the RepeatMasker script buildSummary.pl to summarize the TE content of each genome and calcDivergenceFromAlign.pl to collect the Kimura substitution levels of TE copies to generate a repeat landscape. For comparison, we performed these analyses on the *C. ales* genome and other available pteriomorphian genome assemblies (see what follows).

Gene prediction and functional annotation

Gene annotation was performed with BRAKER v2.1.2 (Brůna et al. 2021) using AUGUSTUS v3.3.2 (Stanke and Waack 2003; Stanke et al. 2006) and GeneMark v4.38 (Brůna et al. 2020). First, AUGUSTUS was trained via BUSCO v5.1.2 (–long –augustus –auto-lineage-euk) run on the *C. ales* genome assembly. Then, RNA-seq reads were mapped to a softmasked *C. ales* genome assembly with STAR v2.5.3a (Dobin et al. 2013) to generate splice-aware alignments for GeneMark-ET training in the BRAKER2 pipeline. The resulting gene annotations were further refined using Mikado v2.3.2 (Venturini et al. 2018) with Transdecoder (Haas et al. 2013) and BLAST + v2.11.0. Briefly, high-quality splice-junctions from the RNA-seq mapping results identified by Porticullis v1.2.2 (Mapleson et al. 2018) and protein alignments based on 283,363 Bivalvia proteins downloaded from Uniprot and mapped to the *C. ales* genome with Genomethreader v1.7.3 (Gremme et al. 2005) were used to identify the best transcript models with Mikado. Functional gene annotations were created via Diamond v2.0.4 (Buchfink et al. 2015) searches to NCBI NR (downloaded May 3, 2021), Uniprot/Swissprot (downloaded May 28, 2022), and Interproscan v5.38 (Jones et al. 2014).

Phylogenetic analysis

Protein sequences were collected from annotated bivalve genomes, including 11 other pteriomorphian species and 2 outgroups (*Mercenaria mercenaria* and *Sinonovacula constricta*) to phylogenetically compare the protein coding content of the *Ctenoides ales* genome (Supplementary Table 1). A total of 508,333 proteins were analyzed with OrthoFinder v2.5.4 (Emms and Kelly 2019). From these results, 1,156 proteins were identified as single-copy and present in all 14 species analyzed. We used these protein sequences to produce a species tree. The amino acid sequences identified as single-copy and complete for all 14 species were aligned using mafft v7.481 (–auto) (Katoh et al. 2002; Katoh and Standley 2013). Next, we trimmed the alignments with trimal v1.4.rev15 (–automated1) (Capella-Gutiérrez et al. 2009). Then we used IQtree2 v2.1.3 (Nguyen et al. 2015; Minh et al. 2020) with modelfinder (–m MFP) (Kalyaanamoorthy et al. 2017) to generate maximum-likelihood trees for each trimmed protein alignment and a summary file of protein substitution models used. The substitution model results were combined with a gene partition file generated with catsequences (V1.3; 10.5281/zenodo.4409153) and used as input for a partitioned IQtree2 ML analysis (Chernomor et al. 2016). Branch support was evaluated by ultrafast bootstrap (Hoang et al. 2018), SH approximate likelihood ratio test, and approximate Bayes test features in IQtree2 (–B 1000 –alrt 1000 –abayes) (Anisimova et al. 2011).

A second species tree was generated to include multiple taxa per pteriomorphian family where high-quality genomes were available but not genome annotations (i.e. additional Arcidae). This was done to account for variation within taxonomic families in downstream characterization of transposable element and opsin content. As in McElroy et al. (2023), BUSCO was used to predict

conserved protein sequences to construct a species tree independent of genome annotations. We downloaded genome assemblies from 12 pteriomorphian species, three representatives from the families Arcidae, Pectinidae, Ostreidae, and Mytilidae, along with four other bivalve species as outgroups (accessions and assembly statistics listed in Supplementary Table 2). We ran BUSCO v5.2.2 on each of these 16 genome assemblies using the metazoa_odb10 database (BUSCO scores listed in Supplementary Fig. 1, Supplementary Table 2). We then used the 177 BUSCO amino acid sequences identified as single-copy and complete for all 16 species as input for a maximum-likelihood analysis following the same methods described for the first species tree.

Synteny between *C. ales* and *P. maximus*

Synteny was determined using i-ADHoRe v3.0.01 (Proost et al. 2012) using the OrthoFinder2 results and their respective assemblies and annotations. Synteny was only performed for scaffolds larger than 1 Mb. The following parameters were included in the i-ADHoRe config: blast_table = black.blastTable, prob_cutoff = 0.001, anchor_points = 3, number_of_threads = 36, visualizeAlignment = false, output_path = out_5, alignment_method = nw, gap_size = 25, cluster_gap = 50, level_2_only = true, and q_value = 0.9. The function dashbio.Circos in the dashbio library was used to visualize the synteny in a circos plot.

Opsin identification and analysis

We identified opsins via the Phylogenetically Informed Annotation tool (PIA; Speiser et al. 2014) from the *C. ales* gene models and the homology-based de novo gene prediction with the BITACORA pipeline v1.3 (Vizueta et al. 2020) using GeMoMa (Keilwagen et al. 2016, 2018) and a database of molluscan opsin sequences described in McElroy et al. (2023). For PIA, we used the modified version from <https://github.com/MartinGuehmann/PIA2> and the Light Interacting Toolkit (LIT_1.1; opsins using r_opsin_20_rtrans.fas for opsin classification). For further opsins classification and phylogenetic comparison, we included McElroy et al. (2023) opsins for the following species: *Argopecten irradians*, *Pecten maximus*, *Mizuhopecten (Patinopecten) yessoensis*, *Scapharca broughtonii*, *Scapharca kagoshimensis*, *Tegillarca granosa*, *Perna viridis*, *Mytilus coruscus*, *Mytilus galloprovincialis*, *Crassostrea gigas*, and *Crassostrea virginica*. We additionally generated opsin models for *Ostrea edulis*. For all opsins, we manually inspected the sequences to ensure high quality (i.e. intact GPCR Class A 7tm_1 domain and containing the K296 position). For outgroup sequences, we used melatonin receptors and the opsin-like sequences from Placozoa, “placopsins” (Feuda et al. 2012), along with the more recently described GPCR relative of opsins found in lophotrochozoans, the “pseudopsins” (De Vivo et al. 2023). We aligned the opsin and outgroup amino acid sequence for the 13 pteriomorphian species using mafft [–maxiterate 1000 –genafpair]. We then generated a maximum-likelihood tree in IQtree2 with modelfinder (best-fit model Q.yeast + F + R8 according to Bayesian Information Criterion) that was evaluated with ultrafast bootstrap, SH approximate likelihood ratio test, and approximate Bayes test features in IQtree2 (–B 1000 –alrt 1000 –abayes).

Mitogenome assembly and assembly

To characterize the mitochondrial genome, a partial *C. ales* cytochrome c oxidase subunit I (COI (MF540379.1) was queried against the final genome assembly using BLASTn. This search yielded no returns. Next, the draft, unfiltered WTDBG2 assembly was searched, returning a single contig approximately 18 kb long

Table 1. Summary statistics of *Ctenoides ales* genome and annotation.

Assembly stats		Annotation stats	
Number of scaffolds	1,049	Number of genes	41,064
Length of assembly	2,237,305,496 bp	Number of mRNA	44,726
Longest scaffold	240,195,547 bp	Mean gene length	23,950 bp
Scaffold L50/N50	7/121,434,358 bp	Mean CDS length	1,346 bp
Scaffold L90/N90	16/84,817,633 bp	Mean exon per gene	7.2
Scaffold CG Content	35.77%	Mean introns per gene	6.2
Scaffold N content	0.09%	Mean exon length	188 bp
Number of contigs	8,191	Mean intron length	4,375 bp
Longest contig	7,630,288 bp	Total CDS length	50,641,044 bp
Contig L50/N50	508/1,207,478 bp		
Contig L90/N90	2476/133,914 bp		

containing the partial COI sequence. The PacBio reads were mapped to the mitochondrial contig with minimap2 v2.14-r883 (Li 2018). The aligned reads were then extracted and assembled with Flye v2.9 (Kolmogorov et al. 2019), which generated a single circular contig 20,859 bp long. This contig was then annotated on the MITOS [Genetic Code 5: Invertebrate Mitochondrial (Bernt et al. 2013)] and MITOS2 (Donath et al. 2019) web servers (RefSeq 89 Metazoa; Genetic Code 5: Invertebrate Mitochondrial). Additionally, the mitochondria genome was scanned for open reading frames (ORFs) with ORFinder on NCBI [https://www.ncbi.nlm.nih.gov/orffinder/ (Genetic Code 5: Invertebrate Mitochondrial; “ATG” and alternative start codons)]. The ORFinder results were used to characterize the complete coding sequences for each of the 12 proteins identified. ARWEN v1.2 (Laslett and Canbäck 2008) was additionally used to evaluate tRNAs predicted by MITOS and potentially identify the 2 tRNAs classified as “missing” by MITOS annotation. The PacBio and forward reads from the Omni-C Illumina sequences were mapped [minimap2 and bwa-mem2 (Vasimuddin et al. 2019), respectively] to evaluate the apparent duplication of the OH sequence. To visualize the duplication by coverage of the short-read mapping, a version of the mitogenome was generated with 1 copy of the duplicated region hard masked to prevent reads from aligning to both duplication regions.

Results and discussion

Genome assembly and completeness analysis

Dovetail performed a de novo genome assembly of *Ctenoides ales*. Approximately 290 billion bases of PacBio CLR were used in the assembly (14,188,342 reads, average length = 20.5Kb). The WTDBG2 assembler generated a draft assembly, which was subsequently scaffolded with 602 million Omni-C reads (Supplementary Table 3 for intermediate stats). These reads were then used to further scaffold the Dovetail genome using Juicer, 3D-DNA, and manually corrected using Juicebox. After eliminating haplotigs and contaminants, the final assembly contained 1,049 scaffolds and 2.2 billion bases. Among these scaffolds, the 18 largest correspond to chromosomes, accounting for 99.16% of the total nucleotide content (Fig. 1b). The assembly’s scaffold N50 value is 121 Mb, with the longest scaffold measuring 240 Mb (Fig. 1c). The completeness of this assembly is exemplified by the 98.56% mapping rate of the PacBio long reads against it, which amounts to approximately 100x coverage. Furthermore, the assembly exhibits a BUSCO completeness score of 91.9% using the metazoa_odb10 gene set (Fig. 1c) and an 85.4% score based on the mollusca_odb10 gene set. This BUSCO-based assembly quality is in line with other pteriomorphian assemblies (Supplementary Fig. 1), including scallops; however, the genome is considerably larger. For

example, the *Pecten maximus* genome released in 2019 by the Wellcome Sanger Institute is chromosomal with 19 chromosomes and 3,983 scaffolds containing 918 million bases (Kenny et al. 2020), making our *C. ales* genome assembly over two times larger (see Supplementary Table 2 for assembly size comparisons). We further examined the genome size disparity between *C. ales* and other pteriomorphians in the context of gene content, genome duplication, and transposable elements.

Gene content and phylogenetic analysis

RNA-seq data were generated at the Iowa State University DNA Facility using eye, mantle, and adductor muscle tissues from 10 *C. ales* specimens and used to annotate genes with BRAKER2, which resulted in 43,799 genes. Further refinement of these gene models with Mikado resulted in 41,064 gene models with an average gene length of 23,950 bp (see Table 1 for further details). We ran BUSCO in protein mode on our annotation results, yielding high completeness scores for metazoa_odb10 (C:89.2%, F:6.0%, M:4.8%, n:954) and mollusca_odb10 (C:81.5%, F:4.6%, M:13.9%, n:5295) that were comparable to the scores for the genome assembly. We aligned amino acid sequences from predicted genes against the NR and Swissprot databases and assigned functional information to 77.7% of the genes. The metazoa BUSCO scores of the gene annotation indicate a high degree of completeness at 89.2%, which is qualitatively similar to the results from the genome assembly. These gene model predictions suggest that *C. ales* may have nearly 50% more genes than is typical for some pteriomorphians. The first scallop genomes, *Mizuhopecten (Patinopecten) yessoensis* and *Chlamys farreri*, were described as having 26,415 (Wang et al. 2017) and 28,602 (Li et al. 2017), respectively. While the scallop *Pecten maximus* was initially reported to have 67,741 protein coding genes (Kenny et al. 2020), Zeng et al. (2021) identified 26,995 genes. Arcidae gene reports tend to have similar numbers of genes as scallops, e.g. 24,045 in *Scapharca broughtonii* (Bai et al. 2019) and 24,398 in *Tegillarca granosa* (Bao et al. 2021). The number of protein coding genes predicted in *C. ales* is more comparable with genome annotations from Mytilidae and Ostreidae genomes, which tend to contain between 30,000 and 40,000 genes (e.g. Peñaloza et al. 2021; Yang et al. 2021). Notably, transposable elements (TEs) do not appear to make up a substantial portion of the predicted genes as only 1,825 genes (~4% of total) have annotation terms from likely TEs (“transposase,” “reverse transcriptase,” “helicase,” “LINE,” “integrase,” and “RNase H”). Therefore, the relatively high number of genes predicted in this genome are not likely inflated by TEs annotated as genes. Variation in gene model prediction strategies may account for some of the differences in gene totals, as fragmentation of lowly expressed genes may have occurred. Additionally, factors such as genome size

Table 2. Transposable element content in *Ctenoides ales* genome.

Class	Family	Count	bp	Genome %
TIR	Total	4,918,972	920,785,650	41.2
	CACTA	1,271,100	182,680,333	8.2
	Merlin	601	277,138	0.0
	MuDR_Mutator	2,187,690	440,892,520	19.7
	P	367	56,747	0.0
	PIF_Harbinger	261,453	34,285,779	1.5
	PiggyBac	638	284,541	0.0
	Sola2	223	93,940	0.0
	Tc1_Mariner	464,116	121,158,874	5.4
	hAT	732,784	141,055,778	6.3
MITE	Total	289,357	35,456,827	1.6
	CACTA	2,4778	3,052,430	0.1
	MuDR_Mutator	21,2367	23,590,962	1.1
	PIF_Harbinger	8,601	713,658	0.0
	Tc1_Mariner	2,775	218,873	0.0
	hAT	4,0836	7,880,904	0.4
DNA (TIR & MITE)	Total	5,208,329	956,242,477	42.7
Helitron		79,069	10,440,877	0.5
Maverick		36,882	18,385,871	0.8
DIRS		17,555	10,406,517	0.5
LTR	Total	1,174,081	309,813,718	13.8
	Bel-Pao	65,336	19,418,066	0.9
	Copia	119,994	17,947,770	0.8
	Gypsy	297,782	132,963,481	5.9
	unknown	690,969	139,484,401	6.2
LINE		129,954	59,550,410	2.7
Penelope		247,032	90,026,897	4.0
SINE	5S	622	87,775	0.0
unknown		20,928	4,675,207	0.2
Total TEs		6,780,946	1,420,396,484	65.2

TIR, terminal inverted repeat; MITE, miniature inverted-repeat transposable element; DIRS, Dictyostelium Intermediate Repeat Sequence; LINE, long interspersed nuclear element; SINE, short interspersed nuclear element.

and evolutionary processes may also contribute to these differences.

To phylogenetically compare the protein coding content of the *Ctenoides ales* genome, we collected protein sequences from annotated bivalve genomes, including 11 pteriomorphians and 2 out-group species (*Mercenaria mercenaria* and *Sinonovacula constricta*). We analyzed a total of 508,333 proteins with OrthoFinder2 (Emms and Kelly 2015; Emms and Kelly 2019), which resulted in clustering 91.5% of the proteins into 465,163 orthogroups (summarized in [Supplementary Table 4](#)). Only 7.4% of the proteins belong to species-specific orthogroups. For *C. ales*, 33,033 of the 39,079 protein sequences analyzed were placed into orthogroups, 5,230 of which were specific orthogroups to *C. ales*, meaning that the majority (71.1%) of protein-coding genes from our annotation are shared with other species (e.g. [Supplementary Fig. 2](#)). A lack of close relatives may have influenced these results, as species with the highest proportion of genes in species-specific orthogroups are those that are the sole representatives of a family (*C. ales*: 13.4%, *M. mercenaria*: 19.5%, and *P. fuctata*: 16.3%). We used the 1,156 single-copy orthologs found in all 14 species to reconstruct the species phylogeny using maximum-likelihood analysis with IQTREE2. This phylogeny had high support values at all nodes and met expectations of species relationships, including *C. ales* representing Limidae as a sister lineage to Pectinidae (e.g. [Audino et al. 2020](#)) ([Fig. 2b](#)). Together, these analyses of the *C. ales* gene content are indicative of a high-quality annotation for this first limid genome. Furthermore, although the gene count for *C. ales* is relatively high within Pteriomorphia, protein coding sequence accounts for 50Mb, or 2.3%, of the total genome assembly length, therefore contributing very little to its large genome size (e.g. the scallops *P. maximus* and *Mi. yessoensis* have about 41Mb of coding sequence, each). More genome assemblies from limid species

will be needed to clarify if the *Ctenoides* genome is broadly representative of species in Limida for gene content and determine whether gene expansions in this lineage predated, accompanied, or followed its genome size increase.

Synteny shows no whole-genome duplication events

To investigate the potential involvement of whole-genome duplication (WGD) in the larger genome size of *Ctenoides ales* than scallops and other pteriomorphian species, we conducted a synteny comparison with the king scallop, *Pecten maximus*. The expectation was to observe a majority of chromosomes exhibiting synteny from one chromosome in *P. maximus* to two chromosomes in *C. ales*, indicating a WGD event. We found no evidence of WGD. Instead, the results revealed a one-to-one correspondence for 12 chromosome pairs (*Ctenoides* (C):*Pecten* (P)): C2:P24, C3:P30, C4:P22, C5:P21, C6:P18, C7:P17, C8:P15, C9:P32, C12:P26, C14:P19, C15:P25, and C18:P29 ([Fig. 2a](#)).

Considering *C. ales* possessing one less chromosome, a fusion event was anticipated. The circos plot demonstrated C1 mapping to P20 and P23, but an inverse pattern was also identified where C15 and C17 both mapped to P16 ([Fig. 2a](#)). This observation highlights the occurrence of genomic rearrangement since the divergence of these species from their common ancestor. Genomic rearrangement is further supported by the diminished or fragmented synteny observed between chromosomes C10, C11, and C13 in *C. ales* and P27, P28, P31, and P33 in *P. maximus* ([Fig. 2a](#)). The circos plot's syntenic relationships reveal a high degree of synteny between the two species, accompanied by some genomic rearrangement. However, there was no evidence of a WGD event to account for the substantially larger genomic content in *C. ales*.

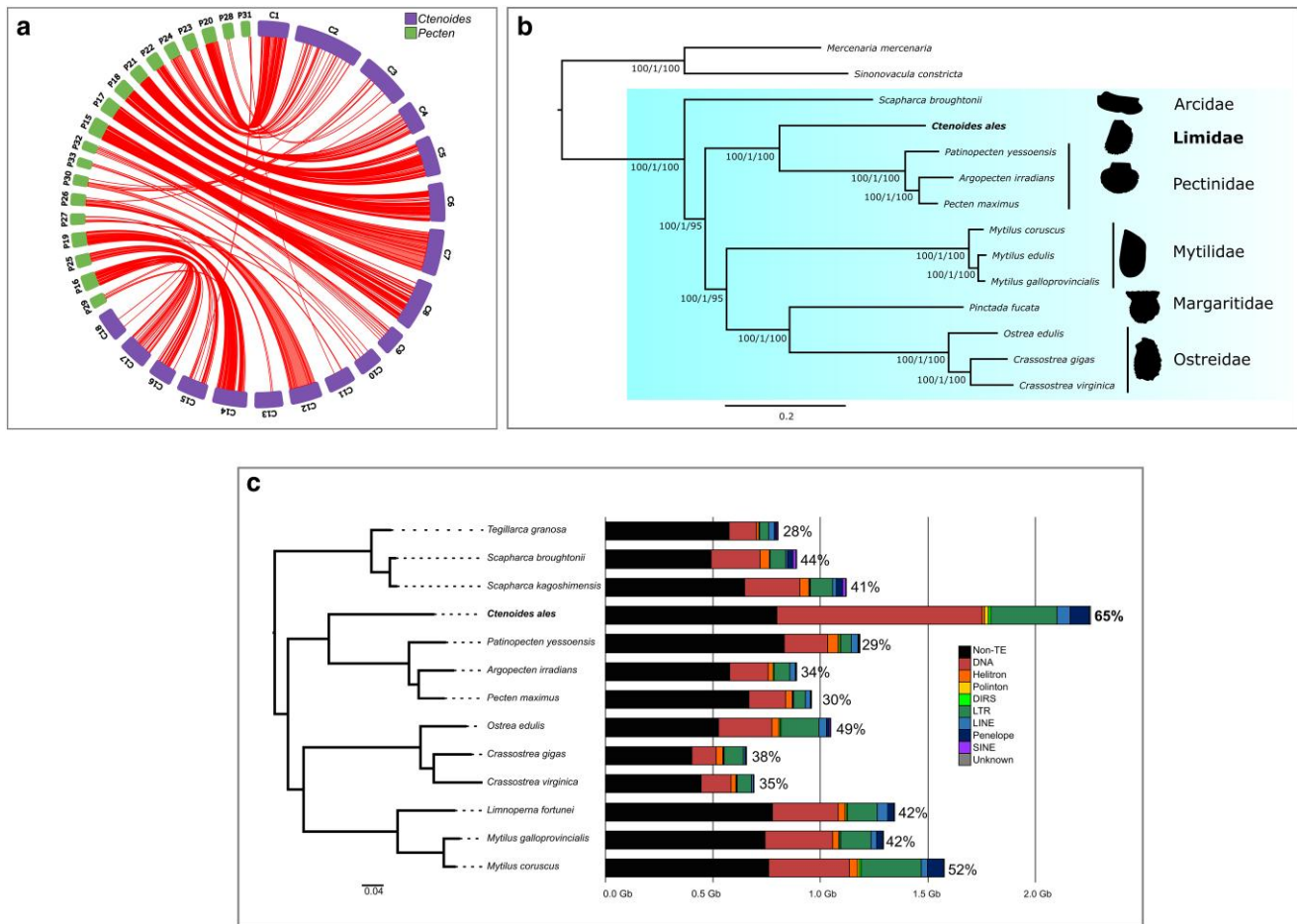


Fig. 2. Comparative genomics with other bivalves. a) Circos plot of syntenic regions (red arcs) between *C. ales* (in purple) and *P. maximus* (in green). Scaffold names were shortened to their unique numbering and the name portion replaced with a C for *Ctenoides* or a P for *Pecten*. b) Phylogenetic placement of *Ctenoides ales* among bivalves. Maximum-likelihood species tree generated with IQtree2 based on a partitioned amino acid supermatrix from 1,156 single-copy orthologs identified from OrthoFinder2 recovered from all 14 species displayed. Branch values are SH-aLRT % support (with 1,000 replicates)/aBayes probability/UFBoot support % (with 1,000 replicates). c) Genome assemblies for 13 pteriomorphian species characterized by TE amount by distinct types of TEs and “non-TE” portion of the genome in black; % given per species is the genomic proportion of TEs.

In general, high degrees of synteny have been observed across much of Pteriomorpha, including between scallop species (Han *et al.* 2022), between scallop and ark clam genomes (Bao *et al.* 2021), and scallop vs Mytilidae comparisons (Yang *et al.* 2021). The most extensive genome rearrangements in Pteriomorpha are in the oysters, which have notably fewer chromosomes (e.g. Gundappa *et al.* 2022) than groups such as scallops. Genomic evidence of past WGD has been reported from pteriomorphian genomes (Corrochano-Fraile *et al.* 2022), but these duplication events have not been placed phylogenetically, making it unclear whether these apparent WGD are shared across Pteriomorpha or occurred in parallel among separate lineages. Our analysis of the first limid genome indicates that no genome duplication has occurred in either Pectinidae or Limidae since their split. Furthermore, while the chromosomes of *C. ales* are much larger than *P. maximus*, gene order has largely been preserved between these lineages.

Transposable elements contribute to large genome size in *Ctenoides*

The *C. ales* genome is twice as large as some genomes from Pectinidae (Fig. 2a), despite no evidence of WGD. To determine if the size difference was due to repetitive elements, we characterized repeat content from *C. ales* and 12 other pteriomorphians

(Supplementary Table 5) with RepeatModeler, EDTA, and RepeatMasker. We found that much of the difference in genome size across Pteriomorpha is attributable to transposable elements. Based on our estimation, about 65% of the *C. ales* genome is made up of TEs, with DNA transposons making up 40% of the genome (summarized in Table 2). The TE content of *C. ales* is about twice that of the Pectinidae, and greater than any of the 12 other pteriomorphian genomes analyzed here (Fig. 2a). The repeat landscape of the *C. ales* genome reveals a large, ancient burst of TE activity, mainly in DNA transposons and a more recent burst in LTR Gypsy elements (Fig. 2b). We also observed far more “intact” (i.e. full structural and coding components) long terminal repeat (LTR) elements from the EDTA analysis than the other species analyzed, which also reflects recent activity of these TEs, such that they have not degraded or been excised from the genome (Supplementary Table 6).

Several classes of TEs are far more abundant in *C. ales* than other pteriomorphians, particularly DNA transposons and LTR retrotransposons, and, to a lesser degree, LINES and Penelope retrotransposons. Helitrons are less abundant in *C. ales* than other species analyzed here (Supplementary Table 5). SINEs appear to have had very little activity outside of Arcida (Fig. 2d, Supplementary Fig. 3), including *C. ales*. Based on the repeat landscapes, there has likely been greater TE activity more recently in

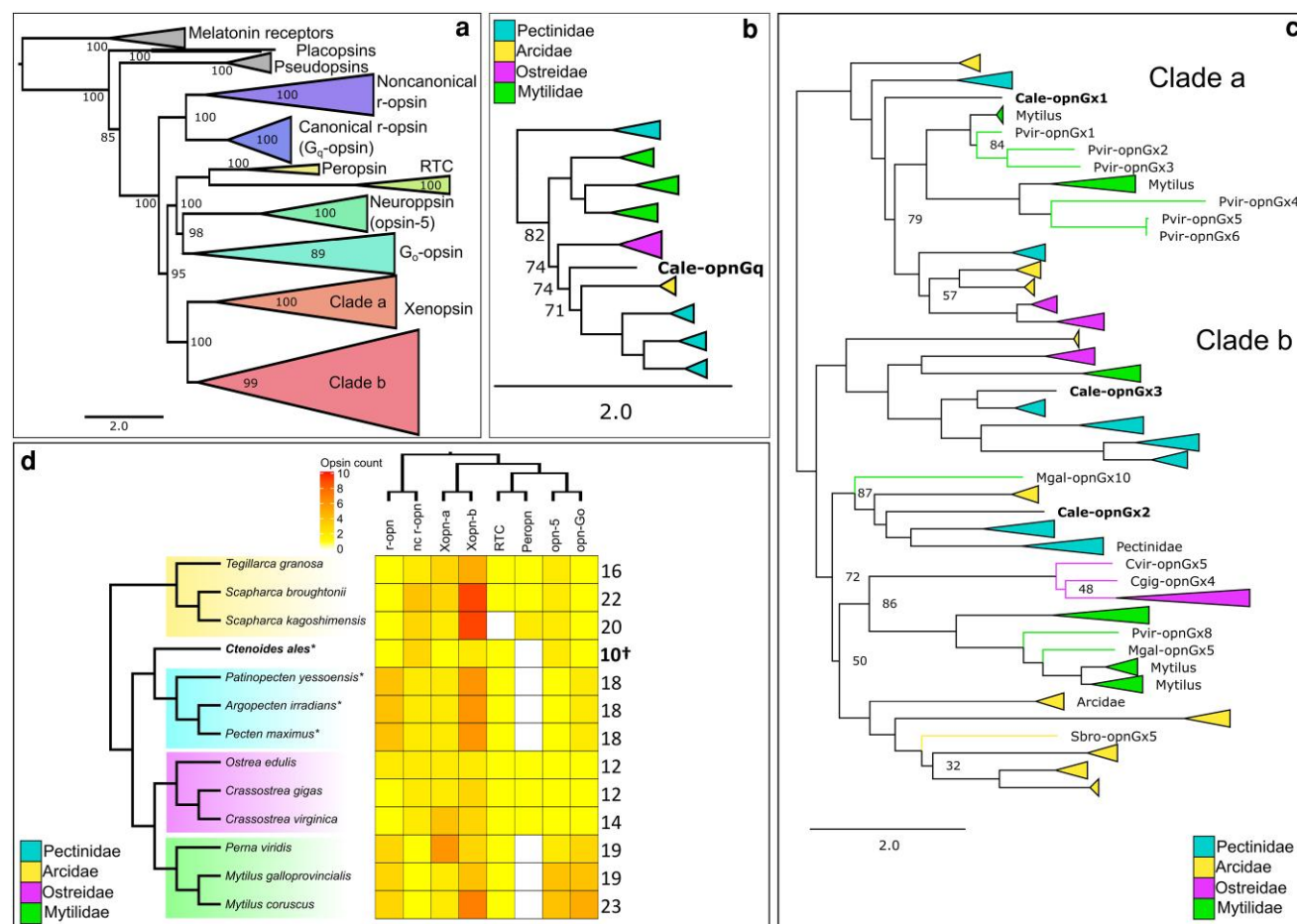


Fig. 3. Summary of opsin content in *Ctenoides ales* and other pteriomorphian bivalves. a) ML phylogenetic tree of opsins from 13 pteriomorphian genomes, including *C. ales*. Ultrafast bootstrap (UFboot) support is shown (Supplementary Fig. 4). Subtrees of b) "canonical" G_q-coupled r-opsins ("opnGq"), and c) xenopsins ("opnGx"). In b) and c), UFboot support < 90 displayed; branches collapsed and color coded according to shared opsin duplicates within families; *C. ales* opsins noted in bold as "Cale-opn". d) Heatmap reflects number of genes from each of 8 opsin groups (r-opn: rhabdomeric G_q-coupled opsin; nc r-opn: noncanonical r-opsin; Xopn-a/b: xenopsin clades a and b; RTC: retinochrome; Peropn: peropsin; opn-5: neuropsin; opn-G_o: G_o-coupled opsin). Phylogenetic relationship among opsin groups based on Fig. 3a. Total numbers of opsins per species reported right of heatmap. Shading around taxa according to family color coding in previous panels. *eyed species. †*C. ales* count includes the partial noncanonical r-opsin and G_o-opsin.

the *C. ales* genome than *P. maximus* (Supplementary Fig. 3). We generally observed similar TE content within families, e.g. in Pectinidae *A. irradians*, *P. maximus*, and *M. yessoensis* differed by 5% across the three species. Within-family differences in TE content still typically varied less between species in the same genus, e.g. *Crassostrea*, but *M. galloprovincialis* appears to have more TEs than *M. coruscus* (Fig. 2a). With *C. ales* being the first limid genome sequenced, it will be important to sequence more broadly from Limida to have a thorough account of how and when this genomic expansion occurred.

TEs are important drivers of genome size evolution (Kidwell 2002) and can account for drastic variation in genome size across closely related species (e.g. Naville et al. 2019; Wong et al. 2019), but the factors influencing differential accumulation and loss of TEs across taxa are still an ongoing area of research. Population genetic theory (Lynch 2007) and empirical data (e.g. Szitenberg et al. 2016) point toward genetic drift as a powerful force influencing TE accumulation. Therefore, changes in effective population size could explain differences in TE content across taxa. Other factors, such as decreased TE silencing (Liu et al. 2022), may also contribute to TE expansion. The genome expansion we found in *C. ales* along with recent evidence for highly variable TE content across bivalves (Martellosi et al. 2023) highlight the importance of this

genomically understudied group of animals for exploring the genomic and biological influences on TE evolution.

Relatively few opsins in the *C. ales* genome

Using the Phylogenetically Informed Analysis tool with the light-detection toolkit (Speiser et al. 2014), we identified 6 opsins from our genome-wide annotation (1 of each: canonical r-opsin, noncanonical r-opsin, xenopsin, G_o-opsin, retinochrome, and neuropsin) and an additional 2 complete xenopsins from the BITACORA pipeline output. We also found 2 partial opsin sequences (a noncanonical r-opsin and a G_o-coupled opsin) from the BITACORA output containing the K296 residue but a truncated seven-transmembrane protein domain. To add evolutionary context for these *C. ales* opsins, we generated an ML phylogenetic tree that included opsins from 12 other pteriomorphian species (Fig. 3a, Supplementary Fig. 4). As outgroup proteins in this analysis, we included melatonin receptors and the placozoan opsin-like sequences "placopsins" (Feuda et al. 2012), along with the recently described group of closely related GPCRs in lophotrochozoans, "pseudopsins" (De Vivo et al. 2023).

The opsins in the *C. ales* genome belong to the major opsin groups present in mollusks: r-opsin (canonical and noncanonical), xenopsin, neuropsin, G_o-opsin, and retinochrome (Ramirez et al.

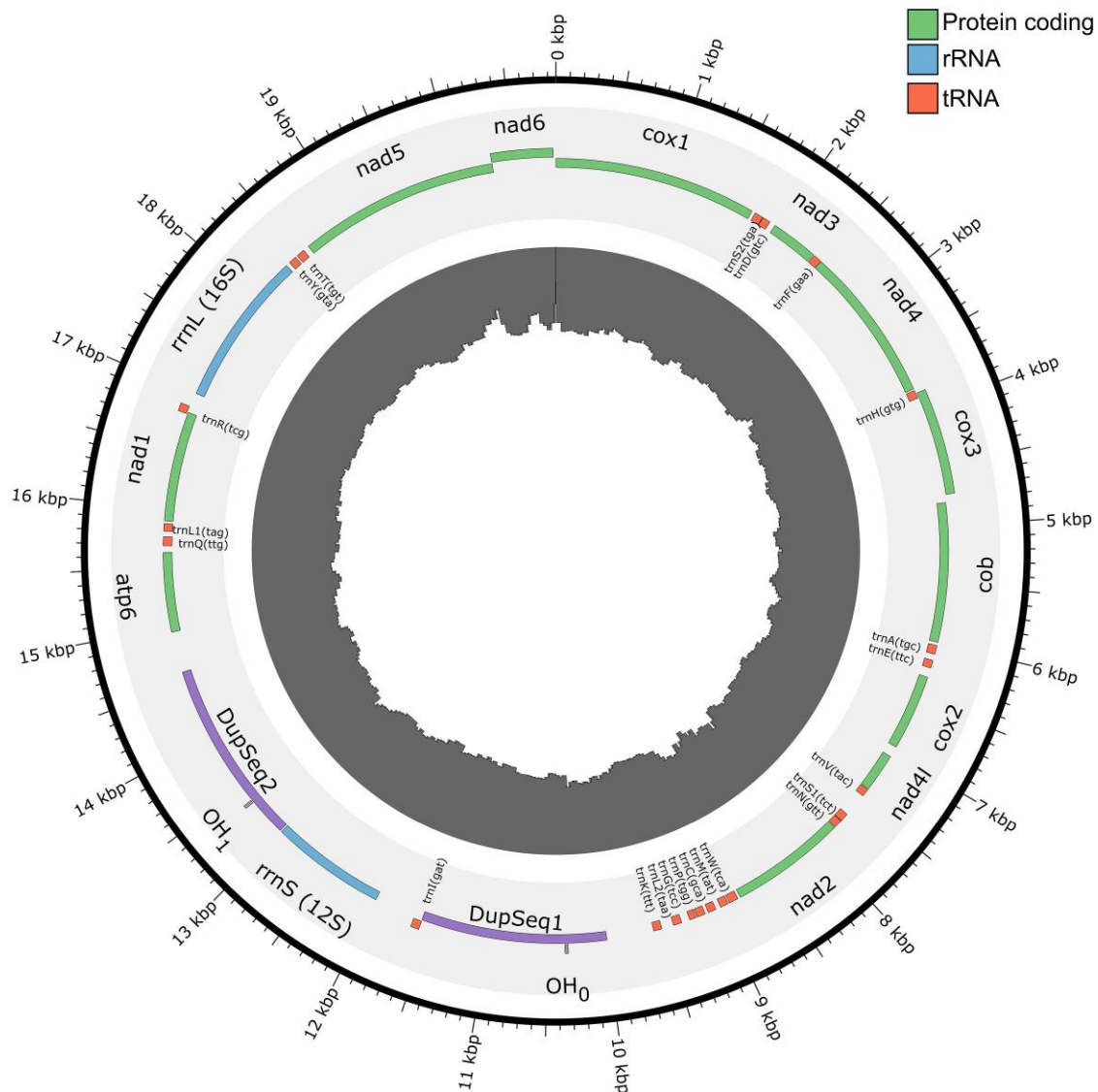


Fig. 4. Annotation of *C. ales* mitochondrial genome. Protein-coding, rRNA, and tRNA sequences color coding in legend, duplicated sequence with heavy-strand origin of replication (OH) in purple. Inner histogram of Illumina coverage (2052–5282X), measured in 100 bp windows step size of 50 bp.

2016) (Fig. 3d). There does not appear to be a peropsin (Fig. 3d) in the *C. ales* genome, which is also absent in Pectinidae and Mytilidae, but present in Ostreidae and Arcidae. The absence of perops in *C. ales* is likely a shared loss with Pectinidae. Compared with other species in Pteriomorpha, the limid, *C. ales* has a small opsin repertoire (Fig. 3d). Unlike other families in Pteriomorpha, this representative limid lacks any lineage-specific opsin duplications (Fig. 3d), even considering the partial noncanonical r-opsin and G_o-opsin (Supplementary Fig. 4). The xenopsin group is particularly expansion prone in Pteriomorpha, as an independent series of lineage-specific duplications are present in genomes from Pectinidae, Arcidae, Mytilidae, and Ostreidae (McElroy et al. 2023). However, the three *C. ales* xenopsins are each phylogenetically located in separate clades of this opsin subfamily (Fig. 3c), making lineage-specific paralogous duplication highly unlikely. Very little is known about xenopsins, as they were only recently recognized as a distinct group of opsins sister to the cnidarian “cnidopsins” (Ramirez et al. 2016). Xenopsins are only found in mollusks, other lophotrochozoans, and rotifers

(Ramirez et al. 2016; Döring et al. 2020) and may be expressed in eyes along with r-opsins and c-opsins—depending on the species (Matsuo et al. 2019; Döring et al. 2020).

The “canonical” G_q-coupled r-opsins are characteristically used in vision across the invertebrate clades of bilaterians. Previously, duplications of this opsin were identified in the bay scallop, *Argopecten irradians* (Porath-Krause et al. 2016), leading to the hypothesis that r-opsin expansion may be a common feature of eye evolution in bivalves. Here, we found only a single r-opsin in the *C. ales* genome vs the four in each of the scallop genomes (Fig. 3b). This result indicates that scallop r-opsin duplication all occurred after the split from Limida. It also demonstrates that recruitment of additional r-opsins is not necessary for eye evolution in Pteriomorpha. In fact, with eyeless mytilid species also having multiple r-opsin duplications (Fig. 3b), the role of opsins in ocular vs nonocular processes requires particular attention in Pteriomorpha. The lack of opsin duplication in the *C. ales* genome contributes to the growing evidence for opsin evolution being unrelated to visual complexity in mollusks (De Vivo et al. 2023; McElroy et al. 2023).

Functional assays of *C. ales* opsins, including in vitro protein expression, as has been performed with scallop opsins (Smedley et al. 2022), as well as tissue-specific RNA-seq combined with in situ hybridization and/or immunohistochemistry, will be valuable next steps in determining whether opsins from *C. ales* form photopigments and are expressed in eyes and other light-sensitive tissues.

First mitogenome from Limida

Mollusks exhibit some of the most variable genomic architecture, molecular functions, and patterns of inheritance for mitochondria in metazoans (reviewed in Ghiselli et al. 2021a). Pteriomorphian bivalves are known for dynamic mitochondrial (mt) genome evolution, with Arcidae having repeatedly evolved some of the largest bilaterian mt genomes (Kong et al. 2020), and Pectinidae with frequent gene order rearrangements (Malkócs et al. 2022) and species with exceptionally large mt genomes (e.g. La Roche et al. 1990). Currently, around 300 complete or nearly complete mitochondrial genomes are publicly available on NCBI GenBank, none of which are from Limida. We present the first mt genome assembly for this order of bivalves, which should be a valuable resource for future phylogenetic analyses.

We assembled a 20,859 bp circular contig from the PacBio reads, representing a complete mitochondrial genome sequence. Using a combination of MITOS2, ARWEN, and ORF identification, we annotated 12 complete protein coding genes, the 12S and 16S rRNA genes, and 22 tRNA genes in the *C. ales* mitogenome (Fig. 4). The only typical metazoan protein coding gene not annotated was *atp8*, which is commonly absent in bivalve mitogenomes (Serb and Lydeard 2003). However, recent analyses support the presence of *atp8* sequences in Pectinidae (Malkócs et al. 2022) and Mytilidae (Zhao et al. 2022).

An interesting feature of this mitogenome is the apparent duplication of the heavy-strand origin of replication (“OH” annotation in MITOS, vs “OL” for light-strand). We found a 1,589 bp sequence duplicated on either side of the 12S rRNA gene that includes an OH annotation from MITOS. The two sequences are 99.8% identical, only differing at three positions. Many (907/4,230) PacBio reads span these repeated sequences (Supplementary Fig. 5). We also observed a 2X relative coverage of one repeat (here, “DupSeq1”) vs the rest of the mt genome when we mapped Illumina reads to an assembly with the second repeat (“DupSeq2”) hard masked (i.e. all nucleotides reported as Ns), which prevented reads from alignment in that region (Supplementary Fig. 6). These results support the accuracy of the mitogenome assembly in having a large duplicated sequence. This duplicated sequence containing OH likely represents the mitogenome “control region,” which regulates replication and transcription (Boore 1999). Duplication of control regions has occurred in a variety of taxa, including birds (Schirtzinger et al. 2012), snakes (Jiang et al. 2007), and velvet worms (Braband et al. 2010), but little is known about the genetic architecture and/or evolutionary pressures underlying and maintaining duplicated control regions. The mitogenome of *C. ales* and limids more broadly may offer insights into the evolution of control regions and further support bivalves as an emerging system for studying mitochondria (Ghiselli et al. 2021b).

Conclusion

In this study, we report a high-quality, chromosome-level assembly for *Ctenoides ales*, the first genome sequenced from the bivalve order Limida. The genome of *C. ales* is noticeably larger than other

pteriomorphian bivalves, largely due to a substantial number of transposable elements. We also find that this species has relatively few opsins, compared with other pteriomorphians, indicating that opsin diversification is not guaranteed to accompany the evolution of specialized adult eyes in bivalves. Additionally, we present the complete mitochondrial genome, another first for Limida.

Data availability

All raw read data have been uploaded to the NCBI SRA database, and the genome assembly has been submitted to NCBI GenBank. Raw reads and assembly are associated with the NCBI BioProject PRJNA1078364. The nuclear and mitochondrial genome assemblies with their annotations, the data underlying species and opsin phylogenies, and scripts used in this study are available on Figshare <https://doi.org/10.6084/m9.figshare.25290208>.

Supplemental material available at G3 online.

Acknowledgments

We are grateful to Justin Polson at Crowned Aquatics Pet Supply for providing care for the *Ctenoides* specimens. We thank Dr. Joel Sharbrough for helpful suggestions on analyzing the mitogenome.

Funding

This project was supported by the National Science Foundation (DEB 1754331) to JMS and AJS and a Dovetail Tree of Life Grant Award to JMS. The research reported in this paper is partially supported by the HPC@ISU equipment at Iowa State University, some of which has been purchased through funding provided by National Science Foundation under MRI grant numbers 1726447 and 2018594.

Author contribution

KEM and AJS wrote the paper. JMS and AJS designed the research study. AJS, KEM, RM, SC, JMS, and KEM performed the experiments and analyzed the data. JMS, RM, and SC provided critical review of the paper. All authors approved the final version of the paper for submission.

Conflicts of interest

The authors declare no conflict of interest.

Literature cited

- Anisimova M, Gil M, Dufayard J-F, Dessimoz C, Gascuel O. 2011. Survey of branch support methods demonstrates accuracy, power, and robustness of fast likelihood-based approximation schemes. *Syst Biol.* 60(5):685–699. doi:10.1093/sysbio/syr041.
- Audino JA, Serb JM, Marian JEAR. 2020. Hard to get, easy to lose: evolution of mantle photoreceptor organs in bivalves (Bivalvia, Pteriomorphia). *Evolution.* 71(9):2105–2120. doi:10.1111/evo.14050.
- Bai CM, Xin LS, Rosani U, Wu B, Wang QC, Duan XK, Liu ZH, Wang CM. 2019. Chromosomal-level assembly of the blood clam, *Scapharca (Anadara) broughtonii*, using long sequence reads and Hi-C. *GigaScience.* 8(7):1–8. doi:10.1093/gigascience/giz067.

- Bao Z, Eddy SR. 2002. Automated de novo identification of repeat sequence families in sequenced genomes. *Genome Res.* 12(8): 1269–1276. doi:[10.1101/gr.88502](https://doi.org/10.1101/gr.88502).
- Bao W, Kojima KK, Kohany O. 2015. Repbase update, a database of repetitive elements in eukaryotic genomes. *Mob DNA.* 6(1):11. doi:[10.1186/s13100-015-0041-9](https://doi.org/10.1186/s13100-015-0041-9).
- Bao Y, Zeng Q, Wang J, Zhang Z, Zhang Y, Sufang W, Wong N-K, Yuan W, Huang Y, Zhang W, et al. 2021. Genomic insights into the origin and evolution of molluscan red-bloodedness in the blood clam *tegillarca granosa*. *Mol Biol Evol.* 38(6):2351–2365. doi:[10.1093/molbev/msab030](https://doi.org/10.1093/molbev/msab030).
- Bell AL, Mpitsos GJ. 1968. Morphology of the eye of the flame fringe clam, *Lima scabra*. *Biol Bull.* 135:414–415.
- Bentley MA, Holland PWH, Feuda R, Marle F. 2016. Conservation, duplication, and divergence of five opsin genes in insect evolution. *Genome Biol Evol.* 8(3):579–587. doi:[10.1093/gbe/evw015](https://doi.org/10.1093/gbe/evw015).
- Bernt M, Donath A, Jühling F, Externbrink F, Florentz C, Fritzsche G, Pütz J, Middendorf M, Stadler PF. 2013. MITOS: improved de novo metazoan mitochondrial genome annotation. *Mol Phylogenet Evol.* 69(2):313–319. doi:[10.1016/j.ympev.2012.08.023](https://doi.org/10.1016/j.ympev.2012.08.023).
- Birchler JA, Yang H. 2022. The multiple fates of gene duplications: deletion, hypofunctionalization, subfunctionalization, neofunctionalization, dosage balance constraints, and neutral variation. *Plant Cell.* 34(7):2466–2474. doi:[10.1093/plcell/koac076](https://doi.org/10.1093/plcell/koac076).
- Boore JL. 1999. Animal mitochondrial genomes. *Nucleic Acids Res.* 27(8):1767–1780. doi:[10.1093/nar/27.8.1767](https://doi.org/10.1093/nar/27.8.1767).
- Braband A, Podsiadlowski L, Cameron SL, Daniels S, Mayer G. 2010. Extensive duplication events account for multiple control regions and pseudo-genes in the mitochondrial genome of the velvet worm *Metaperipatus inae* (Onychophora, Peripatopsidae). *Mol Phylogenet Evol.* 57(1):293–300. doi:[10.1016/j.ympev.2010.05.012](https://doi.org/10.1016/j.ympev.2010.05.012).
- Briscoe AD. 2001. Functional diversification of lepidopteran opsins following gene duplication. *Mol Biol Evol.* 18(12):2270–2279. doi:[10.1093/oxfordjournals.molbev.a003773](https://doi.org/10.1093/oxfordjournals.molbev.a003773).
- Briscoe AD, Bybee SM, Bernard GD, Yuan F, Sison-Mangus MP, Reed RD, Warren AD, Llorente-Bousquets J, Chiao C-C. 2010. Positive selection of a duplicated UV-sensitive visual pigment coincides with wing pigment evolution in *Heliconius* butterflies. *Proc Nat Acad Sci U S A.* 107(8):3628–3633. doi:[10.1073/pnas.0910085107](https://doi.org/10.1073/pnas.0910085107).
- Brůna T, Hoff KJ, Lomsadze A, Stanke M, Borodovsky M. 2021. BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genomics Bioinform.* 3(1):lqaa108. doi:[10.1093/nargab/lqaa108](https://doi.org/10.1093/nargab/lqaa108).
- Brůna T, Lomsadze A, Borodovsky M. 2020. GeneMark-EP+: eukaryotic gene prediction with self-training in the space of genes and proteins. *NAR Genomics Bioinform.* 2(2):lqaa026. doi:[10.1093/nargab/lqaa026](https://doi.org/10.1093/nargab/lqaa026).
- Buchfink B, Xie C, Huson DH. 2015. Fast and sensitive protein alignment using DIAMOND. *Nat Methods.* 12(1):59–60. doi:[10.1038/nmeth.3176](https://doi.org/10.1038/nmeth.3176).
- Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. *BMC Bioinform.* 10(1):421. doi:[10.1186/1471-2105-10-421](https://doi.org/10.1186/1471-2105-10-421).
- Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. Trimal: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinform.* 25(15):1972–1973. doi:[10.1093/bioinformatics/btp348](https://doi.org/10.1093/bioinformatics/btp348).
- Chernomor O, von Haeseler A, Minh BQ. 2016. Terrace aware data structure for phylogenomic inference from supermatrices. *Syst Biol.* 65(6):997–1008. doi:[10.1093/sysbio/syw037](https://doi.org/10.1093/sysbio/syw037).
- Corrochano-Fraile A, Davie A, Carboni S, Bekaert M. 2022. Evidence of multiple genome duplication events in *Mytilus* evolution. *BMC Genomics.* 23(1):340. doi:[10.1186/s12864-022-08575-9](https://doi.org/10.1186/s12864-022-08575-9).
- De Vivo G, Crocetta F, Ferretti M, Feuda R, D'Aniello S. 2023. Duplication and losses of opsin genes in lophotrochozoan evolution. *Mol Biol Evol.* 40(4):msad066. doi:[10.1093/molbev/msad066](https://doi.org/10.1093/molbev/msad066).
- Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-Seq aligner. *Bioinformatics.* 29(1):15–21. doi:[10.1093/bioinformatics/bts635](https://doi.org/10.1093/bioinformatics/bts635).
- Donath A, Jühling F, Al-Arab M, Bernhart SH, Reinhardt F, Stadler PF, Middendorf M, Bernt M. 2019. Improved annotation of protein-coding genes boundaries in metazoan mitochondrial genomes. *Nucleic Acids Res.* 47(20):10543–10552. doi:[10.1093/nar/gkz833](https://doi.org/10.1093/nar/gkz833).
- Donovan DA, Elias JP, Baldwin J. 2004. Swimming behavior and morphometry of the file shell *Limaria fragilis*. *Mar Fresh Behav Physiol.* 37(1):7–16. doi:[10.1080/10236240310001603774](https://doi.org/10.1080/10236240310001603774).
- Döring CC, Kumar S, Tumu SC, Kourtesis I, Hausen H. 2020. The visual pigment xenopsin is widespread in protostome eyes and impacts the view on eye evolution. In: Larhammar D, Tautz D, editors. *eLife.* 9:e55193. doi:[10.7554/eLife.55193](https://doi.org/10.7554/eLife.55193).
- Dougherty LF, Dubielzig RR, Schobert CS, Teixeira LB, Li J. 2017. Do you see what I see? Optical morphology and visual capability of 'disco' clams (*Ctenoides ales*). *Biol Open.* 6(5):648–653. doi:[10.1242/bio.024570](https://doi.org/10.1242/bio.024570).
- Dougherty LF, Johnsen S, Caldwell RL, Marshall NJ. 2014. A dynamic broadband reflector built from microscopic silica spheres in the 'disco' clam *Ctenoides ales*. *J Royal Soc Int.* 11(98):20140407. doi:[10.1098/rsif.2014.0407](https://doi.org/10.1098/rsif.2014.0407).
- Dougherty LF, Niebergall AK, Broeckling CD, Schauer KL, Li J. 2019. Brightly coloured tissues in limid bivalves chemically deter predators. *Royal Soc Open Sci.* 6(10):191298. doi:[10.1098/rsos.191298](https://doi.org/10.1098/rsos.191298).
- Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand NC, Shamim MS, Machol I, Lander ES, Aiden AP, et al. 2017. De novo assembly of the *aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science.* 356(6333):92–95. doi:[10.1126/science.aal3327](https://doi.org/10.1126/science.aal3327).
- Durand NC, Shamim MS, Machol I, Rao SSP, Huntley MH, Lander ES, Aiden EL. 2016. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst.* 3:95–98. doi:[10.1016/j.cels.2016.07.002](https://doi.org/10.1016/j.cels.2016.07.002).
- Emms DM, Kelly S. 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* 16(1):157. doi:[10.1186/s13059-015-0721-2](https://doi.org/10.1186/s13059-015-0721-2).
- Emms DM, Kelly S. 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* 20(1):238. doi:[10.1186/s13059-019-1832-y](https://doi.org/10.1186/s13059-019-1832-y).
- Escobar-Camacho D, Carleton KL, Narain DW, Pierotti MER. 2020. Visual pigment evolution in Characiformes: the dynamic interplay of teleost whole-genome duplication, surviving opsins and spectral tuning. *Mol Ecol.* 29(12):2234–2253. doi:[10.1111/mec.15474](https://doi.org/10.1111/mec.15474).
- Feuda R, Hamilton SC, McInerney JO, Pisani D. 2012. Metazoan opsin evolution reveals a simple route to animal vision. *Proc Nat Acad Sci U S A.* 109(46):18868–18872. doi:[10.1073/pnas.1204609109](https://doi.org/10.1073/pnas.1204609109).
- Feuda R, Menon AK, Göpfert MC. 2022. Rethinking opsins. *Mol Biol Evol.* 39(3):msac033. doi:[10.1093/molbev/msac033](https://doi.org/10.1093/molbev/msac033).
- Finlay HJ. 1927. New specific names for Austral Mollusca. *Trans New Zealand Inst.* 57:488–533.
- Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, Smit AF. 2020. RepeatModeler2 for automated genomic discovery of transposable element families. *Proc Nat Acad Sci U S A.* 117(17):9451–9457. doi:[10.1073/pnas.1921046117](https://doi.org/10.1073/pnas.1921046117).
- Fu L, Niu B, Zhu Z, Wu S, Li W. 2012. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics.* 28(23):3150–3152. doi:[10.1093/bioinformatics/bts565](https://doi.org/10.1093/bioinformatics/bts565).

- Ghiselli F, Gomes-dos-Santos A, Adema CM, Lopes-Lima M, Sharbrough J, Boore JL. 2021a. Molluscan mitochondrial genomes break the rules. *Philosoph Trans Royal Soc B: Biol Sci.* 376(1825): 20200159. doi:[10.1098/rstb.2020.0159](https://doi.org/10.1098/rstb.2020.0159).
- Ghiselli F, Iannello M, Piccinini G, Milani L. 2021b. Bivalve molluscs as model systems for studying mitochondrial biology. *Int Comparat Biol.* 61(5):1699–1714. doi:[10.1093/icb/icab057](https://doi.org/10.1093/icb/icab057).
- Gorman ALF, McReynolds JS. 1969. Hyperpolarizing and depolarizing receptor potentials in the scallop eye. *Science.* 165(3890):309–310. doi:[10.1126/science.165.3890.309](https://doi.org/10.1126/science.165.3890.309).
- Gremme G, Brendel V, Sparks ME, Kurtz S. 2005. Engineering a software tool for gene structure prediction in higher organisms. *Inf Soft Tech.* 47(15):965–978. doi:[10.1016/j.infsof.2005.09.005](https://doi.org/10.1016/j.infsof.2005.09.005).
- Guan D, McCarthy SA, Wood J, Howe K, Wang Y, Durbin R. 2020. Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics.* 36(9):2896–2898. doi:[10.1093/bioinformatics/btaa025](https://doi.org/10.1093/bioinformatics/btaa025).
- Gundappa MK, Peñaloza C, Regan T, Boutet I, Tanguy A, Houston RD, Bean TP, Macqueen DJ. 2022. Chromosome-level reference genome for European flat oyster (*ostrea edulis* L.). *Evol App.* 15(11): 1713–1729. doi:[10.1111/eva.13460](https://doi.org/10.1111/eva.13460).
- Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, et al. 2013. De novo transcript sequence reconstruction from RNA-Seq using the Trinity platform for reference generation and analysis. *Nat Proto.* 8(8): 1494–1512. doi:[10.1038/nprot.2013.084](https://doi.org/10.1038/nprot.2013.084).
- Hagen JFD, Roberts NS, Johnstone RJ. 2023. The evolutionary history and spectral tuning of vertebrate visual opsins. *Dev Biol.* 493: 40–66. doi:[10.1016/j.ydbio.2022.10.014](https://doi.org/10.1016/j.ydbio.2022.10.014).
- Han W, Liu L, Wang J, Wei H, Yuli L, Lijing Z, Guo Z, Yajuan L, Liu T, Zeng Q, et al. 2022. Ancient homomorphy of molluscan sex chromosomes sustained by reversible sex-biased genes and sex determiner translocation. *Nat Ecol Evol.* 6(12):1891–1906. doi:[10.1038/s41559-022-01898-6](https://doi.org/10.1038/s41559-022-01898-6).
- Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. 2018. UFBoot2: improving the ultrafast bootstrap approximation. *Mol Biol Evol.* 35(2):518–522. doi:[10.1093/molbev/msx281](https://doi.org/10.1093/molbev/msx281).
- Jiang ZJ, Castoe TA, Austin CC, Burbrink FT, Herron MD, McGuire JA, Parkinson CL, Pollock DD. 2007. Comparative mitochondrial genomics of snakes: extraordinary substitution rate dynamics and functionality of the duplicate control region. *BMC Evol Biol.* 7(1): 123. doi:[10.1186/1471-2148-7-123](https://doi.org/10.1186/1471-2148-7-123).
- Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, et al. 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics.* 30(9):1236–1240. doi:[10.1093/bioinformatics/btu031](https://doi.org/10.1093/bioinformatics/btu031).
- Kalyanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermini LS. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods.* 14(6):587–589. doi:[10.1038/nmeth.4285](https://doi.org/10.1038/nmeth.4285).
- Katoh K, Misawa K, Kuma K, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucl Acids Res.* 30(14):3059–3066. doi:[10.1093/nar/gkf436](https://doi.org/10.1093/nar/gkf436).
- Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* 30(4):772–780. doi:[10.1093/molbev/mst010](https://doi.org/10.1093/molbev/mst010).
- Keilwagen J, Hartung F, Paulini M, Twardziok SO, Grau J. 2018. Combining RNA-Seq data and homology-based gene prediction for plants, animals and fungi. *BMC Bioinformat.* 19(1):189. doi:[10.1186/s12859-018-2203-5](https://doi.org/10.1186/s12859-018-2203-5).
- Keilwagen J, Wenk M, Erickson JL, Schattat MH, Grau J, Hartung F. 2016. Using intron position conservation for homology-based gene prediction. *Nucl Acids Res.* 44(9):e89–e89. doi:[10.1093/nar/gkw092](https://doi.org/10.1093/nar/gkw092).
- Kenny NJ, McCarthy SA, Dudchenko O, James K, Betteridge E, Corton C, Dolucan J, Mead D, Oliver K, Omer AD, et al. 2020. The gene-rich genome of the scallop *Pecten maximus*. *GigaScience.* 9(5): g1aa037. doi:[10.1093/gigascience/g1aa037](https://doi.org/10.1093/gigascience/g1aa037).
- Kidwell MG. 2002. Transposable elements and the evolution of genome size in eukaryotes. *Genetica.* 115(1):49–63. doi:[10.1023/A:1016072014259](https://doi.org/10.1023/A:1016072014259).
- Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol.* 37(5): 540–546. doi:[10.1038/s41587-019-0072-8](https://doi.org/10.1038/s41587-019-0072-8).
- Kong L, Li Y, Kocot KM, Yang Y, Qi L, Li Q, Halanynch KM. 2020. Mitogenomics reveals phylogenetic relationships of Arcoida (Mollusca, Bivalvia) and multiple independent expansions and contractions in mitochondrial genome size. *Mol Phylogenet Evol.* 150:106857. doi:[10.1016/j.ympev.2020.106857](https://doi.org/10.1016/j.ympev.2020.106857).
- Laetsch DR, Blaxter ML. 2017. BlobTools: interrogation of genome assemblies. *F1000Res.* 6:1287. doi:[10.12688/f1000research.12232.1](https://doi.org/10.12688/f1000research.12232.1).
- Land M. 1965. Image formation by a concave reflector in the eye of the scallop, *Pecten maximus*. *J Physiol.* 179(1):138–153. doi:[10.1113/jphysiol.1965.sp007653](https://doi.org/10.1113/jphysiol.1965.sp007653).
- La Roche J, Snyder M, Cook DI, Fuller K, Zouros E. 1990. Molecular characterization of a repeat element causing large-scale size variation in the mitochondrial DNA of the sea scallop *placopecten magellanicus*. *Mol Biol Evol.* 7(1):45–64. doi:[10.1093/oxfordjournals.molbev.a040586](https://doi.org/10.1093/oxfordjournals.molbev.a040586).
- Laslett D, Canbäck B. 2008. ARWEN: a program to detect tRNA genes in metazoan mitochondrial nucleotide sequences. *Bioinformatics.* 24(2):172–175. doi:[10.1093/bioinformatics/btm573](https://doi.org/10.1093/bioinformatics/btm573).
- Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv:1303.3997v2*.
- Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics.* 34(18):3094–3100. doi:[10.1093/bioinformatics/bty191](https://doi.org/10.1093/bioinformatics/bty191).
- Li W, Godzik A. 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics.* 22(13):1658–1659. doi:[10.1093/bioinformatics/btl158](https://doi.org/10.1093/bioinformatics/btl158).
- Li Y, Sun X, Hu X, Xun X, Zhang J, Guo X, Jiao W, Zhang L, Liu W, Wang J, et al. 2017. Scallop genome reveals molecular adaptations to semi-sessile life and neurotoxins. *Nat Commun.* 8(1):1721. doi:[10.1038/s41467-017-01927-0](https://doi.org/10.1038/s41467-017-01927-0).
- Liu X, Majid M, Yuan H, Chang H, Zhao L, Nie Y, He L, Liu X, He X, Huang Y. 2022. Transposable element expansion and low-level piRNA silencing in grasshoppers may cause genome gigantism. *BMC Biol.* 20(1):243. doi:[10.1186/s12915-022-01441-w](https://doi.org/10.1186/s12915-022-01441-w).
- Lynch M. 2007. *The Origins of Genome Architecture*. Sunderland, MA: Sinauer Associates.
- Lynch M, Conery JS. 2000. The evolutionary fate and consequences of duplicate genes. *Science.* 290(5494):1151–1155. doi:[10.1126/science.290.5494.1151](https://doi.org/10.1126/science.290.5494.1151).
- Malkócs T, Viricel A, Becquet V, Evin L, Dubillot E, Pante E. 2022. Complex mitogenomic rearrangements within the pectinidae (Mollusca: Bivalvia). *BMC Ecol Evol.* 22(1):29. doi:[10.1186/s12862-022-01976-0](https://doi.org/10.1186/s12862-022-01976-0).
- Mapleson D, Venturini L, Kaithakottil G, Swarbreck D. 2018. Efficient and accurate detection of splice junctions from RNA-Seq with portcullis. *GigaScience.* 7(12):giy131. doi:[10.1093/gigascience/giy131](https://doi.org/10.1093/gigascience/giy131).

- Martelossi J, Nicolini F, Subacchi S, Pasquale D, Ghiselli F, Luchetti A. 2023. Multiple and diversified transposon lineages contribute to early and recent bivalve genome evolution. *BMC Biol.* 21(1):145. doi:[10.1186/s12915-023-01632-z](https://doi.org/10.1186/s12915-023-01632-z).
- Matsuo R, Koyanagi M, Nagata A, Matsuo Y. 2019. Co-expression of opsins in the eye photoreceptor cells of the terrestrial slug *limax valentianus*. *J Comparat Neurol.* 527(18):3073–3086. doi:[10.1002/cne.24732](https://doi.org/10.1002/cne.24732).
- McElroy KE, Audino JA, Serb JM. 2023. Molluscan genomes reveal extensive differences in photopigment evolution across the phylum. *Mol Biol Evol.* 40(12):msad263. doi:[10.1093/molbev/msad263](https://doi.org/10.1093/molbev/msad263).
- Mikkelsen PM, Bieler R. 2003. Systematic revision of the western atlantic file clams, *Lima* and *Ctenoides* (bivalvia : limoidea : limidae). *Invert Systemat.* 17(5):667–710. doi:[10.1071/IS03007](https://doi.org/10.1071/IS03007).
- Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol.* 37(5):1530–1534. doi:[10.1093/molbev/msaa015](https://doi.org/10.1093/molbev/msaa015).
- Moraes MN, de Assis LVM, Provencio I, de Castrucci AML. 2021. Opsins outside the eye and the skin: a more complex scenario than originally thought for a classical light sensor. *Cell Tissue Res.* 385(3):519–538. doi:[10.1007/s00441-021-03500-0](https://doi.org/10.1007/s00441-021-03500-0).
- Morton B. 2000. The pallial eyes of *ctenoides floridanus* (bivalvia: limoidea). *J Moll Stud.* 66(4):449–455. doi:[10.1093/mollus/66.4.449](https://doi.org/10.1093/mollus/66.4.449).
- Mpitosos GJ. 1973. Physiology of vision in the mollusk *Lima scabra*. *J Neurophysiol.* 36(2):371–383. doi:[10.1152/jn.1973.36.2.371](https://doi.org/10.1152/jn.1973.36.2.371).
- Naville M, Henriot S, Warren I, Sumic S, Reeve M, Volff JN, Chourrout D. 2019. Massive changes of genome size driven by expansions of non-autonomous transposable elements. *Curr Biol.* 29(7):1161–1168. doi:[10.1016/j.cub.2019.01.080](https://doi.org/10.1016/j.cub.2019.01.080).
- Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating Maximum-likelihood phylogenies. *Mol Biol Evol.* 32(1):268–274. doi:[10.1093/molbev/msu300](https://doi.org/10.1093/molbev/msu300).
- Ohno S. 1970. *Evolution by Gene Duplication*. New York, NY: Springer.
- Ou S, Jiang N. 2018. LTR_retriever: a highly accurate and sensitive program for identification of long terminal repeat retrotransposons. *Plant Physiol.* 176(2):14101422. doi:[10.1104/pp.17.01310](https://doi.org/10.1104/pp.17.01310).
- Ou S, Su W, Liao Y, Chougule K, Agda JRA, Hellinga AJ, Lugo CSB, Elliott TA, Ware D, Peterson T, et al. 2019. Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. *Genome Biol.* 20(1):275. doi:[10.1186/s13059-019-1905-y](https://doi.org/10.1186/s13059-019-1905-y).
- Peñaloza C, Gutierrez AP, Eöry L, Wang S, Guo X, Archibald AL, Bean TP, Houston RD. 2021. A chromosome-level genome assembly for the pacific oyster *crassostrea gigas*. *GigaScience.* 10(3):giab020. doi:[10.1093/gigascience/giab020](https://doi.org/10.1093/gigascience/giab020).
- Peng J, Li Q, Xu L, Wei P, He P, Zhang X, Zhang L, Guan J, Zhang X, Lin Y, et al. 2020. Chromosome-level analysis of the *Crassostrea hongkongensis* genome reveals extensive duplication of immune-related genes in bivalves. *Mol Ecol Res.* 20(4):980–994. doi:[10.1111/1755-0998.13157](https://doi.org/10.1111/1755-0998.13157).
- Porath-Krause AJ, Pairett AN, Faggionato D, Birla BS, Sankar K, Serb JM. 2016. Structural differences and differential expression among rhabdomeric opsins reveal functional change after gene duplication in the bay scallop, *Argopecten irradians* (Pectinidae). *BMC Evol Biol.* 16(1):250. doi:[10.1186/s12862-016-0823-9](https://doi.org/10.1186/s12862-016-0823-9).
- Porter ML, Blasic JR, Bok MJ, Cameron EG, Pringle T, Cronin TW, Robinson PR. 2011. Shedding new light on opsin evolution. *Proc Royal Soc B: Biol Sci.* 279(1726):3–14. doi:[10.1098/rspb.2011.1819](https://doi.org/10.1098/rspb.2011.1819).
- Price AL, Jones NC, Pevzner PA. 2005. De novo identification of repeat families in large genomes. *Bioinformatics.* 21(Suppl 1):i351–i358. doi:[10.1093/bioinformatics/bti1018](https://doi.org/10.1093/bioinformatics/bti1018).
- Proost S, Fostier J, De Witte D, Dhoedt B, Demeester P, Van de Peer Y, Vandepoele K. 2012. i-ADHoRe 3.0—fast and sensitive detection of genomic homology in extremely large data sets. *Nucl Acids Res.* 40(2):e11. doi:[10.1093/nar/gkr955](https://doi.org/10.1093/nar/gkr955).
- Putnam NH, O'Connell BL, Stites JC, Rice BJ, Blanchette M, Calef R, Troll CJ, Fields A, Hartley PD, Sugnet CW, et al. 2016. Chromosome-scale shotgun assembly using an in vitro method for long-range linkage. *Genome Res.* 26(3):342–350. doi:[10.1101/gr.193474.115](https://doi.org/10.1101/gr.193474.115).
- Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics.* 26(6):841–842. doi:[10.1093/bioinformatics/btq033](https://doi.org/10.1093/bioinformatics/btq033).
- Ramirez MD, Pairett AN, Pankey MS, Serb JM, Speiser DI, Swafford AJ, Oakley TH. 2016. The last common ancestor of most bilaterian animals possessed at least nine opsins. *Genome Biol Evol.* 8(12):3640–3652. doi:[10.1093/gbe/evw248](https://doi.org/10.1093/gbe/evw248).
- Robinson JT, Turner D, Durand NC, Thorvaldsdóttir H, Mesirov JP, Aiden EL. 2018. Juicebox.js provides a cloud-based visualization system for Hi-C data. *Cell Syst.* 6(2):256–258.e1. doi:[10.1016/j.cels.2018.01.001](https://doi.org/10.1016/j.cels.2018.01.001).
- Ruan J, Li H. 2020. Fast and accurate long-read assembly with wtdbg2. *Nat Methods.* 17(2):155–158. doi:[10.1038/s41592-019-0669-3](https://doi.org/10.1038/s41592-019-0669-3).
- Schirtzinger EE, Tavares ES, Gonzales LA, Eberhard JR, Miyaki CY, Sanchez JJ, Hernandez A, Mueller H, Graves GR, Fleischer RC, et al. 2012. Multiple independent origins of mitochondrial control region duplications in the order psittaciformes. *Mol Phylogenet Evol.* 64(2):342–356. doi:[10.1016/j.ympev.2012.04.009](https://doi.org/10.1016/j.ympev.2012.04.009).
- Serb JM, Lydeard C. 2003. Complete mtDNA sequence of the north American freshwater mussel, *lampsilis ornata* (unionidae): an examination of the evolution and phylogenetic utility of mitochondrial genome organization in bivalvia (Mollusca). *Mol Biol Evol.* 20(11):1854–1866. doi:[10.1093/molbev/msg218](https://doi.org/10.1093/molbev/msg218).
- Shichida Y, Matsuyama T. 2009. Evolution of opsins and phototransduction. *Philosoph Trans Royal Soc B: Biol Sci.* 364(1531):2881–2895. doi:[10.1098/rstb.2009.0051](https://doi.org/10.1098/rstb.2009.0051).
- Smedley GD, McElroy KE, Feller KD, Serb JM. 2022. Additive and epistatic effects influence spectral tuning in molluscan retinochrome opsin. *J Exp Biol.* 225(10):jeb242929. doi:[10.1242/jeb.242929](https://doi.org/10.1242/jeb.242929).
- Smit AF, Hubley R, Green P. 2015. RepeatMasker Open-4.0. <http://www.repeatmasker.org>.
- Speiser DI, Chappell DR, Audino JA, Kingston AC, Serb JM. 2023. *Distributed Visual Systems in Pteriomorphian Bivalves*. New York, NY: Springer. p. 117–145. doi:[10.1007/978-3-031-23216-9_5](https://doi.org/10.1007/978-3-031-23216-9_5).
- Speiser DI, Pankey MS, Zaharoff AK, Battelle BA, Bracken-Grissom HD, Breinholt JW, Bybee SM, Cronin TW, Garm A, Lindgren AR, et al. 2014. Using phylogenetically-informed annotation (PIA) to search for light-interacting genes in transcriptomes from non-model organisms. *BMC Bioinform.* 15(1):350. doi:[10.1186/s12859-014-0350-x](https://doi.org/10.1186/s12859-014-0350-x).
- Stanke M, Schöffmann O, Morgenstern B, Waack S. 2006. Gene prediction in eukaryotes with a generalized hidden markov model that uses hints from external sources. *BMC Bioinform.* 7(1):62. doi:[10.1186/1471-2105-7-62](https://doi.org/10.1186/1471-2105-7-62).
- Stanke M, Waack S. 2003. Gene prediction with a hidden markov model and a new intron submodel. *Bioinform.* 19(suppl_2):ii215–ii225. doi:[10.1093/bioinformatics/btg1080](https://doi.org/10.1093/bioinformatics/btg1080).
- Szitenberg A, Cha S, Opperman CH, Bird DM, Blaxter ML, Lunt DH. 2016. Genetic drift, not life history or RNAi, determine long-term

- evolution of transposable elements. *Genome Biol Evol.* 8(9): 2964–2978. doi:[10.1093/gbe/evw208](https://doi.org/10.1093/gbe/evw208).
- Takeuchi T, Suzuki Y, Watabe S, Nagai K, Masaoka T, Fujie M, Kawamitsu M, Satoh N, Myers EW. 2022. A high-quality, haplotype-phased genome reconstruction reveals unexpected haplotype diversity in a pearl oyster. *DNA Res.* 29(6):dsac035. doi:[10.1093/dnares/dsac035](https://doi.org/10.1093/dnares/dsac035).
- Terakita A. 2005. The opsins. *Genome Biol.* 6(3):1–9. doi:[10.1186/gb-2005-6-3-213](https://doi.org/10.1186/gb-2005-6-3-213).
- Terakita A, Nagata T. 2014. Functional properties of opsins and their contribution to light-sensing physiology. *Zoolog Sci.* 31(10): 653–659. doi:[10.2108/zs140094](https://doi.org/10.2108/zs140094).
- Vasimuddin Md, Sanchit M, Heng L, Srinivas A. 2019. Efficient architecture-aware acceleration of BWA-MEM for multicore systems. In: IEEE Parallel and Distributed Processing Symposium (IPDPS). doi:[10.1109/IPDPS.2019.00041](https://doi.org/10.1109/IPDPS.2019.00041).
- Venturini L, Caim S, Kaithakottil GG, Mapleson DL, Swarbreck D. 2018. Leveraging multiple transcriptome assembly methods for improved gene structure annotation. *GigaScience.* 7(8):giy093. doi:[10.1093/gigascience/giy093](https://doi.org/10.1093/gigascience/giy093).
- Vizueta J, Sánchez-Gracia A, Rozas J. 2020. Bitacora: a comprehensive tool for the identification and annotation of gene families in genome assemblies. *Mol Ecol Res.* 20(5):1445–1452. doi:[10.1111/1755-0998.13202](https://doi.org/10.1111/1755-0998.13202).
- Wang S, Zhang J, Jiao W, Li J, Xun X, Sun Y, Guo X, Huan P, Dong B, Zhang L, et al. 2017. Scallop genome provides insights into evolution of bilaterian karyotype and development. *Nat Ecol Evol.* 1(5): 1–12. doi:[10.1038/s41559-017-0120](https://doi.org/10.1038/s41559-017-0120).
- Waterhouse RM, Seppey M, Simao FA, Manni M, Ioannidis P, Klioutchnikov G, Kriventseva EV, Zdobnov EM. 2018. BUSCO applications from quality assessments to gene prediction and phylogenomics. *Mol Biol Evol.* 35(3):543–548. doi:[10.1093/molbev/msx319](https://doi.org/10.1093/molbev/msx319).
- Wong WY, Simakov O, Bridge DM, Cartwright P, Bellantuono AJ, Kuhn A, Holstein TW, David CN, Steele RE, Martínez DE. 2019. Expansion of a single transposable element family is associated with genome-size increase and radiation in the genus *Hydra*. *Proc Nat Acad Sci U S A.* 116(46):22915–22917. doi:[10.1073/pnas.1910106116](https://doi.org/10.1073/pnas.1910106116).
- Yang J-L, Feng D-D, Liu J, Xu J-K, Chen K, Li Y-F, Zhu Y-T, Liang X, Lu Y. 2021. Chromosome-level genome assembly of the hard-shelled mussel *Mytilus coruscus*, a widely distributed species from the temperate areas of east Asia. *GigaScience.* 10(4):giab024. doi:[10.1093/gigascience/giab024](https://doi.org/10.1093/gigascience/giab024).
- Zeng Q, Liu J, Wang C, Wang H, Zhang L, Hu J, Bao L, Wang S. 2021. High-quality reannotation of the king scallop genome reveals no ‘gene-rich’ feature and evolution of toxin resistance. *Computat Struct Biotech J.* 19:4954–4960. doi:[10.1016/j.csbj.2021.08.038](https://doi.org/10.1016/j.csbj.2021.08.038).
- Zhang J. 2003. Evolution by gene duplication: an update. *Trends Ecol Evol.* 18(6):292–298. doi:[10.1016/S0169-5347\(03\)00033-8](https://doi.org/10.1016/S0169-5347(03)00033-8).
- Zhang R-G, Li G-Y, Wang X-L, Dainat J, Wang Z-X, Ou S, Ma Y. 2022. TESorter: an accurate and fast method to classify LTR-retrotransposons in plant genomes. *Horticul Res.* 9: uhac017. doi:[10.1093/hr/uhac017](https://doi.org/10.1093/hr/uhac017).
- Zhao B, Gao S, Zhao M, Lv H, Song J, Wang H, Zeng Q, Liu J. 2022. Mitochondrial genomic analyses provide new insights into the “missing” atp8 and adaptive evolution of mytilidae. *BMC Genomics.* 23(1):738. doi:[10.1186/s12864-022-08940-8](https://doi.org/10.1186/s12864-022-08940-8).

Editor: A. Rokas