Supplemental Methods

Collection and Preparation

After extraction, the DNA samples were sheared with a 1.5 blunt end needles (Jensen Global, Santa Barbara, CA, USA) and run through pulse field gel electrophoresis, in order to separate large DNA molecules, for 16 hours. Samples were checked with a spectrophotometer (Synergy H1 Hybrid Reader, BioTek Instruments Inc., Winooski, VT) and Qubit® fluorometer for quantification of our genomic DNA.

Illumina and Pacbio Hybrid Assembly, Genome Size Estimation, and Quality Assessment

All computational and bioinformatics analyses were conducted on the High Performance Computing (HPC) Cluster located that University of California, Irvine. Sequence data generated from two lanes of Illumina HiSeq 2500 were concatenated and raw sequence reads were assembled through PLATANUS v1.2.1 (*1*), which accounts for heterozygous diploid sequence data. Parameters used for PLATANUS was -m 256 (memory) and -t 48 (threads) for this initial assembly. Afterwards, contigs assembled from PLATANUS and reads from 40 SMRT cells of PacBio sequencing were assembled with a hybrid assembler DBG2OLC v1.0 (*2*). We used the following parameters in DBG2OLC: k 17 KmerCovTh 2 MinOverlap 20 AdaptiveTh 0.01 LD1 0 and RemoveChimera 1 and ran pbdagcon with default parameters. Without Illumina sequence reads, we also conducted a PacBio reads only assembly with FALCON v0.3.0 (https://github.com/PacificBiosciences/FALCON) with default parameters in order to assemble PacBio reads into contiguous sequences. The parameters we used were input type = raw, length_cutoff = 4000, length_cutoff_pr = 8000, with different cluster settings 32, 16, 32, 8, 64, and 32 cores, concurrency setting jobs were 32, and the remaining were default parameters. After our FALCON assembly, we used the outputs from FALCON and DBG2OLC as input for QUICKMERGE v1.0 (*3*), a metassembler and assembly gap filler developed for long molecule-based assemblies. Several different parameters in QUICKMERGE v1.0 were conducted as suggested by the authors until an optimal assembly was obtained with HCO 10 C 3.0 -l 2400000 -ml 5000. With the final genome assembly, we processed our genome through REPEATMASKER v.4.0.6 (*4*) to mask repetitive elements with the parameter -species teleostei. We also estimated transposable elements (TE) content by using the fugu repeat database in REPEATMASKER. Genome size was estimated with only Illumina sequences and tandem repeats were detected with bioinformatic tools (electronic supplemental material). We used BUSCO v3 (*5*) to estimate the completeness of our genome assembly with the vertebrata and Actinopterygii gene set [consists of 2,586 Benchmarking Universal Single-Copy Orthologs (BUSCOs)] to estimate completeness of our *C. violaceus* genome.

RNA-Seq Tissue Extraction and Sequencing

Five individual *C. violaceus* were collected during the fall of 2015 for our transcriptomic analyses and annotation of the *C. violaceus* genome. We extracted brain, gill, gonads (testes), heart, liver, mid intestine, proximal intestine, pyloric caeca, and spleen from each individual and preserved the tissues in RNAlater® (Ambion, Austin, TX, USA). All individuals had digesta in their guts during dissection (i.e., they had all eaten), and digesta was removed prior to tissue fixation in RNAlater®. Total RNA was extracted using a Trizol protocol, and sample quality (RNA Integrity Number ≥8.2) confirmed using an Agilent bioanalyzer 2100 (RNA nano chip; Agilent Technologies). We used an Illumina TruSeq Sample Preparation v2 (Illumina) kit with

AMPure XP beads (Beckman Coulter Inc.) and SuperScriptTM III Reverse Transcriptase (Invitrogen) to prepare our tissue samples for Illumina sequencing. See Supplemental Table S8 for adaptor indexes used for the transcriptomic sequence data, which was performed using multiplexed samples at 10 nM in 10 µl, and sequenced on two lanes on an Illumina HiSeq 2500 (100 bp Paired Ends) at UC Irvine's GHTF.

Transcript Assembly, Annotation, and Heatmap Generation of All Tissues and Genes Associated with Diet

Reads were normalized, assembled, annotated and gene expression of all transcripts were measured from all nine tissue types and TRINITY V2.3.2 (*6*) was used to identify how many reads mapped back to the *C. violaceus* genome (Supplemental Figure S12 and Table S8). Differentially expressed genes for all tissue types were viewed with a heatmap that was generated with the CUMMERBUND R package (http://compbio.mit.edu/cummeRbund/; Supplemental Figure S13). Candidate genes which pertained to glycolytic, lipid metabolism/gluconeogenesis, ketone degradation, glucosidases (both α and β), proteases, and lipases were identified in the *C. violaceus* transcriptome by scanning the annotation of CUFFLINKS assembled transcripts and used to generate our heatmap.

Genome Size Estimation and Identification of Tandem Repeats

The c-value has been estimated for *Cebidichthys violaceus* (*7*), which is 0.81. Based on this c-value, the estimate of the genome size is ~792 Mb. In addition, we estimated the genome size using only Illumina sequences by using JELLYFISH v2.2.0 (*8*). We selected multiple k-mers (25, 27, 29, 31) for counting and generating a histogram of the k-mer frequencies. We used a perl script (written by Joseph Ryan) to estimate genome size based on k-mer sizes and peak values determined from histograms generated in JELLYFISH.

We used tandem repeats finder (trf v4.07b; *9*) to identify tandem repeats throughout the unmasked genome. We used the following parameters in trf 1 1 2 80 5 200 2000 -d -h to identify repeats. Once the largest repeats were identified, we used period size of the repeats multiplied by the number of copies of the repeat to generate the largest fragments. This method was used to identify repetitive regions which can possibly represent centromere or telomere regions of the *C. violaceus* genome.

Transcript Assembly for all Tissues and Annotation

The following pipeline was used to assemble and measure expression of all transcripts from all nine tissue types (Supplemental Figure S12). Prior to assembly, all raw reads were trimmed with TRIMMOMATIC v0.35 (*10*). Afterwards, trimmed reads were normalized using a perl script provided by TRINITY v r2013-02-16 (*11*). Prior to aligning transcriptomic reads to the genome, the final masked assembled genome was prepared with BOWTIE2-BUILD v2.2.7 (*12*) for a BOWTIE index and then all (normalized) reads from each tissue type were mapped using TOPHAT v2.1.0 (*13*) to our assembled masked genome using the following parameters -I 1000 -i 20 -p 4. Afterwards, aligned reads from each tissue were indexed with SAMTOOLS v1.3 (*14*) as a BAM file. Once indexed through SAMTOOLS, transcripts were assembled by using CUFFLINKS v2.2.1 (*15*) with an overlap-radius 1. All assemblies were merged using CUFFMERGE and then differential expression was estimated with CUFFDIFF, both programs are part of the CUFFLINKS package. All differential expression analyses and plots were produced in R (https://www.r-project.org/) using CUMMERBUND tool located on the bioconductor website (https://www.bioconductor.org/). Once all transcripts were assembled, we ran REPEATMASKER v.4.0.6 with the parameter -species teleostei to mask repetitive elements within our transcriptomes.

All masked transcripts were annotated with the trinotate annotation pipeline (https://trinotate.github.io/), which uses Swiss-Prot (*16*), Pfam (*17*), eggNOG (*18*), Gene Ontology (*19*), SignalP (*20*), and Rnammer (*21*). We also processed our transcripts through BLASTX against the UniProt database (downloaded on September 26th, 2017) with the following parameters: num_threads 8, evalue 1e-20, and max_target_seqs 1. The BLASTX output was processed through trinity analyze_blastPlus_topHit_coverage.pl script to count the amount of transcripts of full length or near full length. To provide a robust number of full-length transcripts, the assembled genome was processed through AUGUSTUS v3.2.1 (*22*) without hints using default parameters for gene predictions using a generalized hidden Markov model in order to identify genes throughout the genome, and predicted transcripts were also masked for repetitive elements through REPEATMASKER.

Comparative Analysis for Syntenic Regions across Teleosts fishes

We selected the following fish genomes: zebrafish (*Danio rerio*), stickleback (*Gasterosteus aculeatus*), spotted gar (*Lepisosteus oculatus*), and the Japanese medaka (*Oryzias latipes*) to identify syntenic regions with our *C. violaceus* genome assembly. All genomes were masked using REPEATMASKER v.4.0.6 with the parameter -species teleostei. To identify syntenic regions we used contigs from our *C. violaceus* which represent 1MB or larger and then concatenated the remaining contigs. We used SATSUMA v3.1.0 (*23*) with the following parameters -n 4 -m 8 for identifying syntenic regions between zebrafish, stickleback, spotted gar, Japanese medaka and the *C. violaceus* genome. Afterwards, we developed circos plots to view syntenic regions shared between species using CIRCOS v0.63-4 (*24*).

Identification of Orthologs Across Teleost Fishes and Identification of Syntenic Regions

The following teleost and non-teleost genomes were taken from ENSEMBL release 89 for our comparative analysis of orthologs and phylogeny of fishes; *Poecilia formosa* (Amazon molly), *Astyanax mexicanus* (blind cave fish), *Gadus morhua* (Atlantic Cod), *Takifugu rubripes* (Fugu), *Oryzias latipes* (Japanese medaka), *Xiphophorus maculatus* (platyfish), *Lepisosteus oculatus* (spotted gar), *Gasterosteus aculeatus* (stickleback), *Tetraodon nigroviridis* (green spotted puffer), *Oreochromis niloticus* (Nile tilapia), *Danio rerio* (zebrafish), *Latimeria chalumnae* (coelacanth).

We used INPARANOID v4.0 (*25*) to conduct 78 possible pairwise comparisons, where N is number of taxa [(N(N-2))/2= possible pairwise comparisons]. From the outputs of INPARANOID, we used QUICKPARANOID (http://pl.postech.ac.kr/QuickParanoid/) to identify orthologous clusters from all 13 species.

Supplemental Results and Discussion

I - Genome and Transcriptome Assembly

From PacBio sequencing, we were able to generate ~29,700 Mb of sequence data from 40 SMRT cells, this represents ~37X coverage based on the c-value estimated for *C. violaceus*

(792 Mb; *7*). From two lanes of Illumina sequencing we were able to generate 84,539 Mb which represents ~107X coverage of the genome. From the PLATANUS assembly with Illumina only sequence data, our N50 was 2,760 bp. When combining PLATANUS assembled contigs and 40 SMRT cells of PacBio for a hybrid assembly in DBG2OLC, we managed to obtain an N50 of 2.21Mb. Afterwards, when using FALCON (PacBio only reads) we managed to obtain an N50 of 2.45Mb. We used the FALCON assembly and the DBG2OLC assembly and through QUICKMERGE, we obtained an N50 of 6.69 Mb. Following two rounds of QUIVER, and PILON, we assembled our final draft genome which composed of 467 contigs, and obtained N25, N50, and N75 values of 15.17, 6.71, and 1.85 (Mb) respectively which were composed of 593,001,491 base pairs (Supplemental Figure S4). Genomic regions were masked and a summary of the identities of repetitive elements are present in Supplemental Table S4.

II - Estimation of Genome Size, Completeness Assessment, and Tandem Repeats Throughout the Genome

By using JELLYFISH we estimated the genome size based on an average of four k-mer size counts (25, 27, 29, 31) 656,598,967 base pairs based with a standard deviation of 4,138,853 base pairs (Supplemental Figure S5, Table S7). Through BUSCO v3 there were 97% (2,508 genes out of 2,586) complete orthologs detected which included 1.3% duplicated orthologs. In addition, there were 1.1% (28 BUSCOs) partial orthologs present and 1.9% (50 BUSCOs) of orthologs were not detected in the *C. violaceus* genome (Supplemental Table S3).

In identifying large genomic regions of tandem repeats, we were able to identify about 38,448 repeated loci where the repetitive sequence (period) range from 1 to 1,983 and number of repeats identified were 1.8 to 14,140.8bps. The largest repeat locus (period size multiplied by the repeat amount) was 109,161bps with a period of 90bp and repeat amount of 1,212.9bps. We saw an increase in size for 35 loci which had repeat locus the size of 32,594.1bps and greater (Supplemental Figure S6-S7) as compared to any other repetitive locus.

III – Assembly and Annotation of Nine Tissue Transcriptomes

From five individuals that we selected for our transcriptomic analyses, the total reads mapped back to the genome ranged from 67.37% (liver) to 84.8% (heart; Supplemental Table S8). The range of transcripts present in each tissue type ranged from 20,008 (liver) to 78,629 (gill) transcripts (Supplemental Table S8).

From the TUXEDO package, there were 101,922 transcripts estimated from the nine tissues. When evaluating Fragments Per Kilobase per Million mapped Reads (FPKM) for each of the nine transcriptomes, we see the lowest median with the liver and the highest median with the gill tissue (Supplemental Figure S14). All other tissues appeared to have a similar profile. In addition, we see that the gill tissue had a short Quartile group 2 as compared to the other tissue types (except liver tissue; Supplemental Figure S14). When we look at the differentially expressed genes across all tissue types, we see that there are a cluster of genes highly expressed in the liver as compared to some of the tissues (Supplemental Figure S13). By using getSig in the CUMMERBUND package we evaluated which genes are significantly regulated, with the highest number (1,383) between liver and brain, as these tissues are highly specialized. When we look at Jensen-Shannon Distances (Supplemental Figures S16-S17), we see that pyloric caeca and middle intestine have similar expression profiles, brain and gonad have similar profiles, and the proximal intestine has a very different expression profile.

From the assembled transcripts in TRINITY and predicted transcripts from AUGUSTUS, there were a total of 105,167,222 and 44,120,550 bases with 0.08% and 2.15% of bases masked respectively (Supplemental Tables S9-10). When annotated, there were 65,535 transcripts in TRINOTATE and there were only 26,356 transcripts with a BLASTX hit identified. When conducting a BLASTX on our transcripts to identify full length transcripts in our dataset, we were only able to obtain 5,199 transcripts which had an 80% hit coverage (Supplemental Tables S12). When using only using AUGUSTUS (without hints) and identified 29,525 genes. There were 29,485 genes which had 60 amino acids or greater in our transcriptomic dataset.

IV - Expression Profiles of Candidate Genes for Digestion and Metabolism

We were able to identify candidate genes associated with digestion, fermentation, ketone degradation in our transcriptomic assembly (Fig. 2a & b in the main manuscript) and view differential gene expression patterns across the nine tissues where we have transcriptomic dataset (Supplemental Figure S13). We were not able to distinguish between *amy2a* and *amy2b* in our transcriptomic assembly. Therefore, we used the AUGUSTUS gene prediction from the genome as a transcriptome reference to detect the *amy2a* and *amy2b* genes and mapped our transcriptome reads back to this reference transcriptome dataset. From this dataset, we were able to detect *amy2a* and *amy2b* gene expression profiles. As expected, we see high expression profiles for genes associated with digestion and metabolism in the pyloric caeca, proximal intestine, mid intestine, and liver (Fig. 2a & b).

V - Evaluation of Candidate Genes Associated with Digestion

From our MUMMER and BLAST search for pancreatic amylase, we have identified three tandem copies of amylase (*amy2a*) and (*amy2b*), as opposed to the six haploid copies detected in the German et al. (*26*). We identified amylase on contig 440 and we see two hypothetical proteins between the three amylase genes and a transposase near *amy2b* (Fig. 3b; Supplemental Figure S26). In addition, each amylase gene is preceded by a 4.3K20bp DNA element encoding a transposase (Fig. 3b, Supplemental Figure S26). The three tandem amylase loci differ from the estimated six haploid copies (based on gene dosage curves using RT-qPCR) proposed to be present in the *C. violaceus* genome by German et al. (*26*). Upon further inspection, we have reached the conclusion that the per cell gene count of the German et al. (*26*) study is the diploid copy number, not the haploid copy number (*C. violaceus* is a diploid, vertebrate). Hence, the copy number based on gene dosage curves is three for amylase in general, with roughly two copies for *amy2a* and one copy for *amy2b* (*26*) which agrees completely with what is observed in our genomic assembly. Although there is the possibility for copy number variation amongst individuals within a population, as there is for human salivary amylase (*27*) and dog pancreatic amylase (*28*), that is not what was observed by German et al. (*26*), as that would entail different methodology and more robust sampling of *C. violaceus* individuals.

We only selected one gene copy of *amy2a* because they are identical, and *amy2b* when estimating selection in DATAMONKEY. When testing all 11 branches for seven taxa in aBSREL, we see only one branch under episodic diversifying selection (*C. violaceus*, *amy2b*; Fig. 3c). We do not see this pattern of positive selection in any of the other branch with a significant p-value.

In MEME, we identified three sites with episodic positive selection with a p-value threshold of 0.05 (sites: 41, 256, and 279; Fig. 3d) and in GARD there was no evidence of recombination.

Our analyses of aminopeptidase (also known as alanyl aminopeptidase, E.C. 3.4.11.2) genes have revealed some interesting results. Fishes appear to have five aminopeptidase genes, which varies from the one (aminopeptidase N) seen in mammals. As we probed genomes of sufficient quality (e.g., using http://ensembl.org), it became clear that fish aminopeptidase loci show signatures of retention following whole genome duplication (WGD) events (*29*-*32*). The website http://ohnologs.curie.fr lists aminopeptidase a (Supplemental Figure S21) and aminopeptidase b (Supplemental Figure S22b & c) in *Danio rerio* and *Oryzias latipes* as being ohnologs from the vertebrate WGD (*31*-*32*). This same website then lists aminopeptidase b and aminopeptidase N (Supplemental Figure S22c) as ohnologs from the Teleost-specific WGD event (*29*-*30*). Our synteny maps for other fishes, including *C. violaceus*, support this contention, especially among aminopeptidase b and aminopeptidase N, as there are other shared genes (e.g., *svp2b*) among the separate loci for aminopeptidase b and aminopeptidase N (Supplemental Figure S22). Indeed, our limited phylogenetic analysis suggests that aminopeptidase a is sister to all other aminopeptidase genes (Supplemental Figure S24). Aminopeptidase b and Ey are more related (Supplemental Figure S22b), and they are sister to a clade that includes aminopeptidase N and Ey-like (Supplemental Figures S22 and S24). The evolutionary history of aminopeptidases clearly requires more work, but our preliminary analysis suggests that the history of aminopeptidases in fishes may involve WGD events. What this means for digestion in fishes with different diets should be explored, as all of the aminopeptidase genes show elevated gut expression, except aminopeptidase N (Fig. 2), which is the name of the alanyl aminopeptidase in mammals. Aminopeptidase activity can be plastic in fishes fed different diets (*33*-*35*), and it doesn’t always appear to only be elevated in fishes consuming more protein. Thus, each aminopeptidase protein necessitates investigation into how their functions may vary in the gut and how this may matter for fishes with different diets.

When examining aminopeptidase a (*anpepa*) codons for positive selection, we did not identify any branches under episodic diversifying selection and identified four sites under episodic positive selection (sites: 194, 412, 445, and 593; Supplemental Figure S21).

We found no branches under episodic diversifying selection in aminopeptidase b, N, and Ey (Supplemental Figure S23). For aminopeptidase Ey-like (*anpep-Ey-like*) we found one branch which leads to *C. violaceus* and *A. purpurescens* under episodic diversifying selection and one site with episodic positive selection with a p-value threshold of 0.05 (site: 38; Supplemental Figure S23, Table S13). We also found one site under episodic positive selection in *anpepb* with a threshold of 0.05 (site: 156) and one site in *anpep N* (site: 351).

For Phospholipase B1, plb1-1, we found no branches under selection and six sites under selection and a p-value threshold of 0.05 (sites: 66, 97, 289, 438, 800, and 821; Supplemental Figure S27, Table S13). As for *plb1-2*, we found one branch which leads to *C. violaceus* and *A. purpurescens* under episodic diversifying selection with three sites episodic positive selection with a p-value threshold of 0.05 (sites: 183, 230 and 476). Lastly, we did not see any branches or sites under selection for plb1-3. When evaluating Phospholipase Group 12 B (*pg12b1* & 2; Supplemental Figure S28) we found no branches under positive diversifying selection. Only for pg12b-2, we found nine sites under selection with a p-value threshold of 0.05 (sites: 25, 31, 32, 33, 37, 83, 142, 146, and 185). When evaluating *cel* and *cel*-like and we did not detect any branches under episodic diversifying selection. We only found three sites under episodic positive selection for cel-1 with a p-value threshold of 0.05 (sites: 64, 258, and 355). For chymotrypsin A (*chymo A*), we did not detect any branches under episodic diversifying selection or sites under episodic positive selection. For chymotrypsin B (*ctrb*), we see two branches with episodic diversifying selection (*C. violaceus* and *A. purpurescens*). There is one site with episodic positive selection (site: 112; Supplemental Figure S18, Table S13). For chymotrypsin-like, we did not detect any branches under episodic diversifying selection or sites under episodic positive selection (Supplemental Figure S19). As for trypsin (*tryp3-1 & -2*), we did not detect any branches under episodic diversifying selection or sites under episodic positive selection except for *tryp3-2* which has one site under episodic positive selection with a threshold of 0.05 (site: 91; Supplemental Figure S20). All candidate genes were evaluated for recombination with gard, which some loci had putative breakpoints, but with the KH (Kishino–Hasegawa test, at P = 0.1) test, there were 0 breakpoints with significant topological incongruence (Supplemental Table S13).

We focused on the loci that encode for *cel* and were able to identify the four tandem copies of *cel* on contig 445 (Supplemental Figure S29). We estimated selection and only see evidence episodic selection of sites on *cel*-1 genes (Fig. 4e in the main manuscript).

With regards to elevated lipase activity in the taxa consuming more fiber in their diets, it is known that in industry settings, adding microcrystalline fiber to lipolytic reactions acts to stabilize the lipase proteins and can increase lipase activities (*36-37*). However, these fibrous compounds added directly to the reactive environment. In our case, when measuring lipase activities, we homogenize the tissues separate from gut contents, and then centrifuge the homogenates to get the supernatant, which would contain only those enzymes that are soluble and not bound to larger molecules, like fiber. Hence, the elevated lipase activities measured *in* *vitro* in our investigations cannot be coming from the potential effects of fiber on the lipase proteins themselves, even if these interactions might act to aid lipolytic action *in vivo* within the gut environment. We are, therefore, confident that the increased lipase activities in the algae-eating fishes are due to the molecular differences in CEL proteins (Fig. 4; Supplemental Figure S29-S31), and any differences in gene expression.

VIII - Orthologs and Phylogenetic Analyses

We constructed a phylogenetic tree using maximum likelihood using thirteen fish taxa including the *Cebidichthys violaceus* which included 30 loci and 33,508 bases with 1,000 bootstrap replicates (Fig. 1; Supplemental Table S5). We used JMODELTEST v2.1.0 and with AICc we detected that GTR+I+G was the best model selected for our phylogenetic analyses. All 30 loci used for our phylogenetic analyses were extracted from orthologs detected in INPARANOID (Supplemental Table S6).

VI - Syntenic regions across multiple fish species

We have 114 contigs which have a 1MB or greater, and we pooled all contigs which had less than 1MB (353 contigs) were merged together in our synteny analyses. When we compare our assembled genome to the *G. aculeatus* genome, we see multiple homologous regions between the two species, and multiple loci from each linkage group of the *G. aculeatus* genome represented in the *C. violaceus* genome; Supplemental Figure S8). When comparing the *O. latipes, D. rerio, L. oculatus*, and genomes to our *C. violaceus* genome (Supplemental Figures

S9-S11), we also see each chromosome/linkage group represented in the *C. violaceus* genome and strong synteny between *O. latipes* and *C. violaceus* whereas we less homologous strands between the *C. violaceus* genome and the *D. rerio* or *L. oculatus* genomes.

VIII - Opsin Gene Copies and Selection

After reviewing our BUSCO analyses for gene duplicates present in our *C. violaceus* genome, we identified three Opsin Short Wave Sensitive (*opn1sw*) genes in tandem on contig 443 (Supplemental Figure S32). We find this interesting because *C. violaceus* endures a period of time out of water during low tide, in which Horn and Riegle (*38*) showed that a large *C. violaceus* (~24 cm SL; 92 g) can survive out of water for 37 hours. The ability to survive out of water may require adaptations of vision when exposed to air during low tides, in which we further evaluated the *opn1sw* gene copies for signatures of positive selection. In addition, we identified and compared gene copy numbers of short-wave opsin genes from *Danio rerio*, *Oreochromis niloticus*, and *Gasterosteus aculeatus* genomes, which have one or two gene copies present (Supplemental Figure S32). In addition, we estimated selection by using the DATAMONKEY server v2.0 by using GARD, aBSREL, and MEME. With GARD, we observed evidence of recombination breakpoints, (locations: 196 and 319) but there are 0 breakpoints with significant topological incongruence (p=0.01). From our aBSREL analysis, we observed one branch of episodic diversifying selection out of 11 leading to *opn1sw2a* along with an *opn1sw2* identified in *G. aculeatus* with a corrected p-value of 0.0172. In addition, we detected episodic positive/diversifying selection at 7 sites. The sites under selection with a p-value less than 0.05 were the following: sites 4, 95, 165, 202, 224, and 326, and there was one site under selection with a p-value less than 0.01 (site 337; Supplemental Figure S32). From this analysis, evaluation of *opn1sw* gene sequences from subtidal and intertidal stichaeids can elucidate how vision may play an important role for intertidal prickleback species.

References

1. Kajitani R et al. 2014 Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun short reads. Gen. Res. 24, 1384–1395. (doi:10.1101/gr.170720.113)

2. Ye C, Hill CM, Wu S, Ruan J, Ma Z. 2016 DBG2OLC: Efficient Assembly of Large Genomes Using Long Erroneous Reads of the Third Generation Sequencing Technologies. Sci. Rep. 6. (doi:10.1038/srep31900)

3. Chakraborty M, Baldwin-Brown JG, Long AD, Emerson JJ. 2016 Contiguous and accuratede novoassembly of metazoan genomes with modest long read coverage. Nucl. Acid. Res., gkw654. (doi:10.1093/nar/gkw654)

4. Smit AF. 2004 Repeat-Masker Open-3.0. http://www. repeatmasker. org.

5. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015 BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformat. 31, 3210–3212. (doi:10.1093/bioinformatics/btv351)

6. Haas BJ et al. 2013 De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nature Prot. 8, 1494–1512. (doi:10.1038/nprot.2013.084)

7. Hinegardner R, Rosen DE. 1972 Cellular DNA content and the evolution of teleostean fishes. American Nat.106, 621-644

8. Marçais G, Kingsford C. 2011 A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics 27, 764–770. (doi:10.1093/bioinformatics/btr011)

9. Benson G. 1999 Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Research 27, 573–580. (doi:10.1093/nar/27.2.573)

10. Bolger AM, Lohse M, Usadel B. 2014 Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120. (doi:10.1093/bioinformatics/btu170)

11. Grabherr MG et al. 2011 Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature Biotechnology 29, 644–652. (doi:10.1038/nbt.1883)

12. Langmead B, Salzberg SL. 2012 Fast gapped-read alignment with Bowtie 2. Nature Methods 9, 357–359. (doi:10.1038/nmeth.1923)

13. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. 2013 TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biology 14, R36. (doi:10.1186/gb-2013-14-4-r36)

14. Li H et al. 2009 The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079. (doi:10.1093/bioinformatics/btp352)

15. Trapnell C et al. 2012 Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nature Protocols 7, 562–578. (doi:10.1038/nprot.2012.016)

16. Boeckmann B. 2003 The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. Nucleic Acids Research 31, 365–370. (doi:10.1093/nar/gkg095)

17. Bateman A. 2004 The Pfam protein families database. Nucleic Acids Research 32, 138D–141. (doi:10.1093/nar/gkh121)

18. Powell S et al. 2013 eggNOG v4.0: nested orthology inference across 3686 organisms. Nucleic Acids Research 42, D231–D239. (doi:10.1093/nar/gkt1253)

19. Ashburner M et al. 2000 Gene Ontology: tool for the unification of biology. Nature Genetics 25, 25–29. (doi:10.1038/75556)

20. Petersen TN, Brunak S, von Heijne G, Nielsen H. 2011 SignalP 4.0: discriminating signal peptides from transmembrane regions. Nature Methods 8, 785–786. (doi:10.1038/nmeth.1701)

21. Lagesen K, Hallin P, Rødland EA, Stærfeldt H-H, Rognes T, Ussery DW. 2007 RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Research 35, 3100–3108. (doi:10.1093/nar/gkm160)

22. Stanke M, Keller O, Gunduz I, Hayes A, Waack S, Morgenstern B. 2006 AUGUSTUS: ab initio prediction of alternative transcripts. Nucleic Acids Research 34, W435–W439. (doi:10.1093/nar/gkl200)

23. Grabherr MG, Russell P, Meyer M, Mauceli E, Alfoldi J, Di Palma F, Lindblad-Toh K. 2010 Genome-wide synteny through highly sensitive sequence alignment: Satsuma. Bioinformatics 26, 1145–1151. (doi:10.1093/bioinformatics/btq102)

24. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. 2009 Circos: An information aesthetic for comparative genomics. Genome Research 19, 1639–1645. (doi:10.1101/gr.092759.109)

25. O'Brien KP. 2004 Inparanoid: a comprehensive database of eukaryotic orthologs. Nucl. Acid. Res. 33, D476–D480. (doi:10.1093/nar/gki107)

26. German DP, Foti DM, Heras J, Amerkhanian H, Lockwood BL. 2016 Elevated Gene Copy Number Does Not Always Explain Elevated Amylase Activities in Fishes. Physiological and Biochemical Zoology 89, 277–293. (doi:10.1086/687288)

27. Perry GH et al. 2007 Diet and the evolution of human amylase gene copy number variation. Nature Genetics 39, 1256–1260. (doi:10.1038/ng2123)

28. Axelsson E et al. 2013 The genomic signature of dog domestication reveals adaptation to a starch-rich diet. Nature 495, 360–364. (doi:10.1038/nature11837)

29. Christoffels A, Koh EGL, Chia J, Brenner S, Aparicio S, Venkatesh B. 2004 Fugu Genome Analysis Provides Evidence for a Whole-Genome Duplication Early During the Evolution of Ray-Finned Fishes. Molecular Biology and Evolution 21, 1146–1151. (doi:10.1093/molbev/msh114)

30. Glasauer SMK, Neuhauss SCF. 2014 Whole-genome duplication in teleost fishes and its evolutionary consequences. Molecular Genetics and Genomics 289, 1045–1060. (doi:10.1007/s00438-014-0889-2)

31. Kasahara M. 2007 The 2R hypothesis: an update. Current Opinion in Immunology 19, 547–552. (doi:10.1016/j.coi.2007.07.009)

32 Ohno S. 1970 Evolution by gene duplication. New York, Springer-Verlag.

33. German DP, Horn MH, Gawlicka A. 2004 Digestive Enzyme Activities in Herbivorous and Carnivorous Prickleback Fishes (Teleostei: Stichaeidae): Ontogenetic, Dietary, and Phylogenetic Effects. Physiological and Biochemical Zoology 77, 789–804. (doi:10.1086/422228)

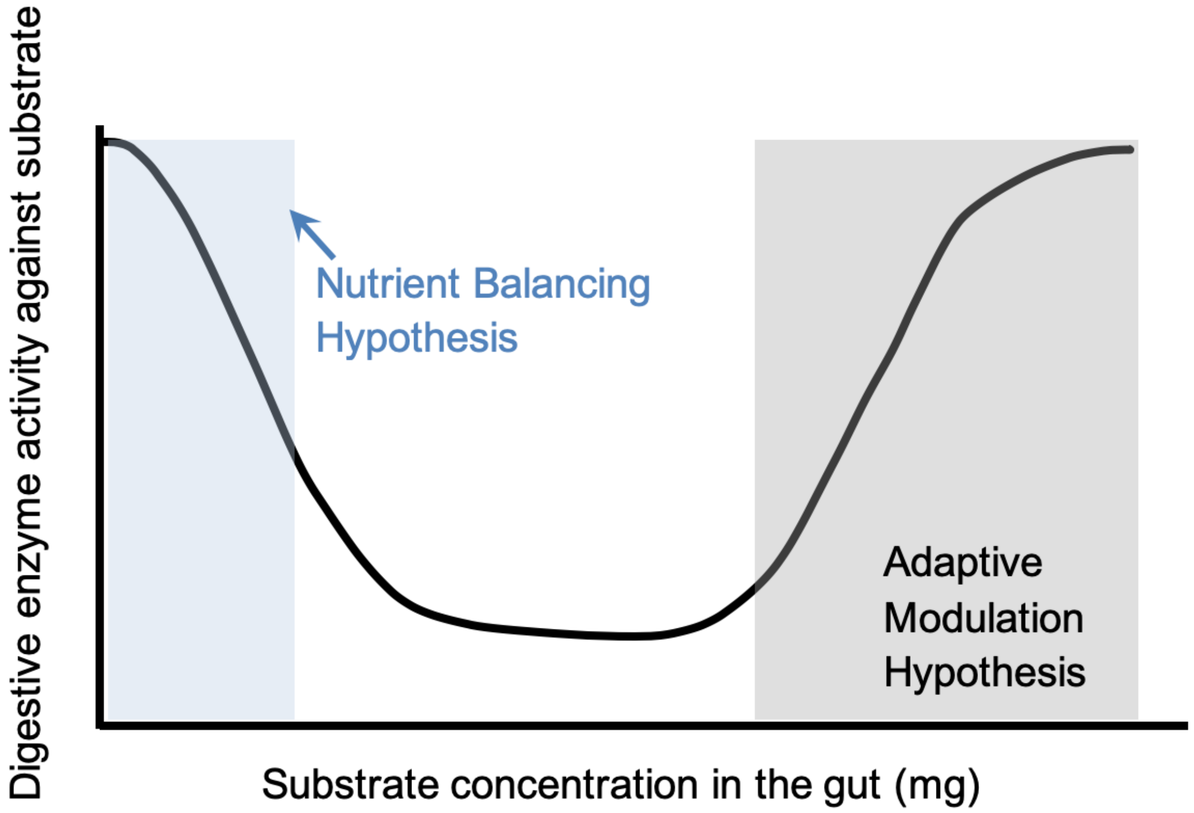
34 Harpaz S, Uni Z. 1999 Activity of intestinal mucosal brush border membrane enzymes in relation to the feeding habits of three aquaculture fish species. Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology 124, 155–160. (doi:10.1016/s1095-6433(99)00106-3)

35. Leigh SC, Nguyen-Phuc B-Q, German DP. 2017 The effects of protein and fiber content on gut structure and function in zebrafish (*Danio rerio*). Journal of Comparative Physiology B 188, 237–253. (doi:10.1007/s00360-017-1122-5)

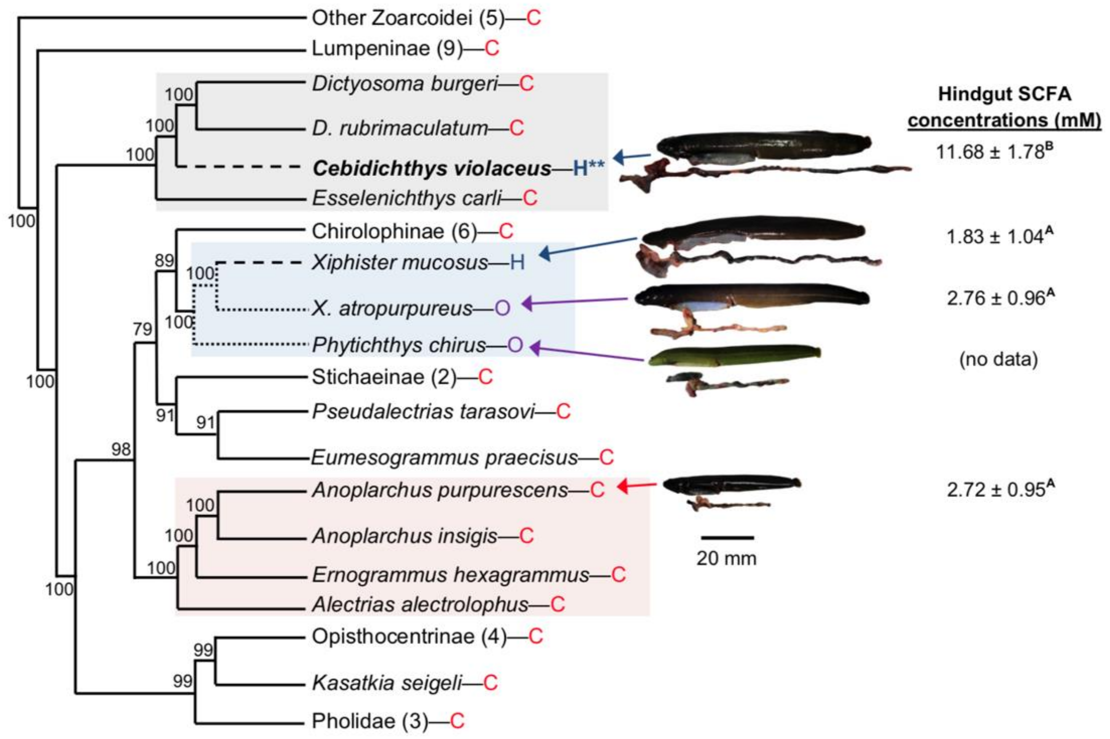
36. Cai W, Liang X, Yuan X, Liu L, He S, Li J, Li B, Xue M. 2018 Different strategies of grass carp (*Ctenopharyngodon idella*) responding to insufficient or excessive dietary carbohydrate. Aquaculture 497, 292–298. (doi:10.1016/j.aquaculture.2018.07.042)

37. Kim JH, Park S, Kim H, Kim HJ, Yang Y-H, Kim YH, Jung S-K, Kan E, Lee SH. 2017 Alginate/bacterial cellulose nanocomposite beads prepared using *Gluconacetobacter xylinus* and their application in lipase immobilization. Carbohydrate Polymers 157, 137–145. (doi:10.1016/j.carbpol.2016.09.074)

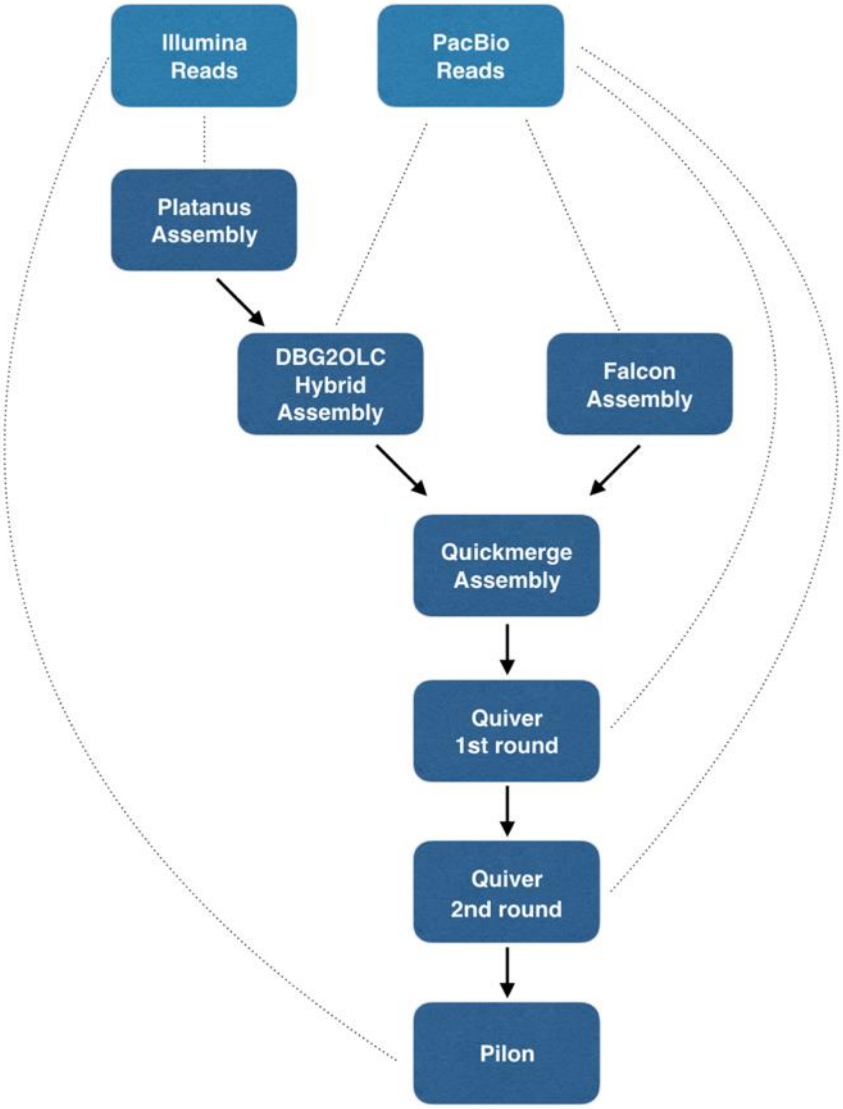
38. Horn MH, Riegle KC. 1981 Evaporative water loss and intertidal vertical distribution in relation to body size and morphology of stichaeoid fishes from California. Journal of Experimental Marine Biology and Ecology 50, 273–288. (doi:10.1016/0022-0981(81)90054-x)



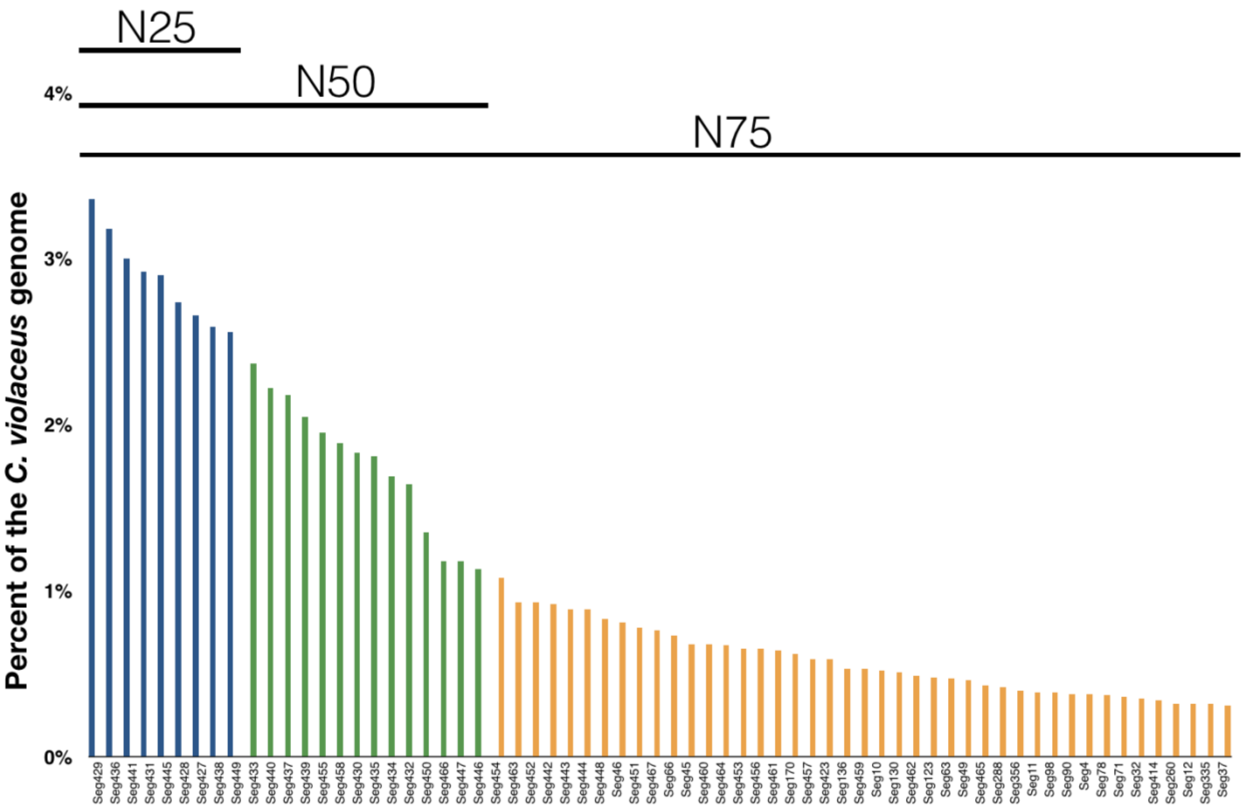
Supplemental Figure S1: Theoretical activity level of a digestive enzyme in an animal's gut as a function of ingested concentrations of substrate for that enzyme. Two hypotheses are commonly invoked in the literature to explain patterns of digestive enzyme activities observed frequently in nature. The Adaptive Modulation Hypothesis (Karasov, 1992) suggests a positive correlation between substrate concentration and enzyme activities; an abundant substrate should invoke more enzyme activity ensure adequate digestion of that substrate. This is commonly seen for carbohydrases, like amylase (German et al. 2016). The Nutrient Balancing Hypothesis (Clissold et al. 2010) suggests that an animal should invest in elevated enzyme activities against limiting nutrients that are low in concentration to ensure acquisition of these important nutrients. This is seen in transporter density for some minerals (e.g., iron), and for lipases to digest lipids in herbivorous fishes (German et al. 2004, 2015).



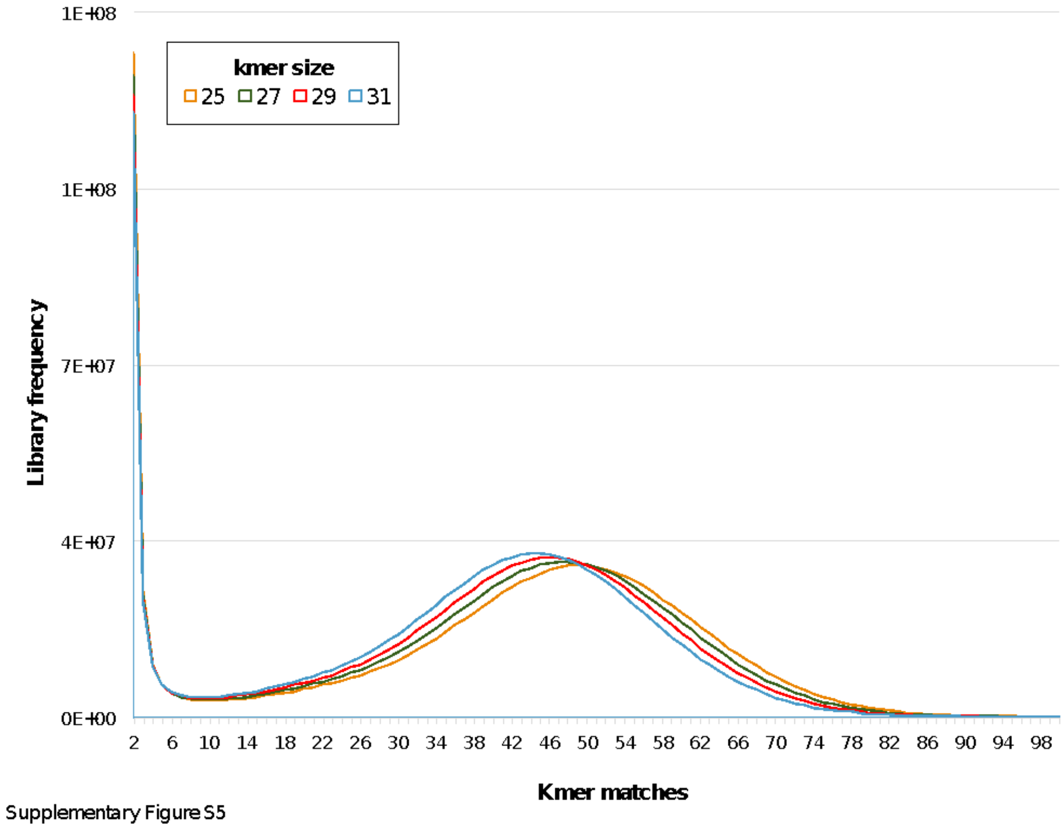
Supplemental Figure S2: Phylogenetic relationships of the polyphyletic family Stichaeidae based on 2,100 bp of *cytb*, *16s*, and *tomo4c4* genes (Kim et al. 2014). Bayesian posterior probabilities are indicated on nodes. *Cebidichthys violaceus* is bolded, and photos of *C. violaceus* and other studied taxa are shown with their digestive systems beneath their bodies. Note the differences in gut size. H=herbivory, O=omnivory, C=carnivory. Evolution of herbivory (— — — —) and omnivory (............) are shown. Numbers in parentheses show number of taxa evaluated at that branch. Boxes highlight alleged families or subfamilies within the polyphyletic family Stichaeidae, with Cebidichthyidae (top), Xiphisterinae (middle), and Alectriinae (bottom) all highlighted. Hindgut short chain fatty acid (SCFA) concentrations are mean ± standard deviation, and were compared with ANOVA ($F_{3,33} = 127.92$; $P < 0.001$). SCFA data sharing a superscript letter are not significantly different (from German et al. 2015).



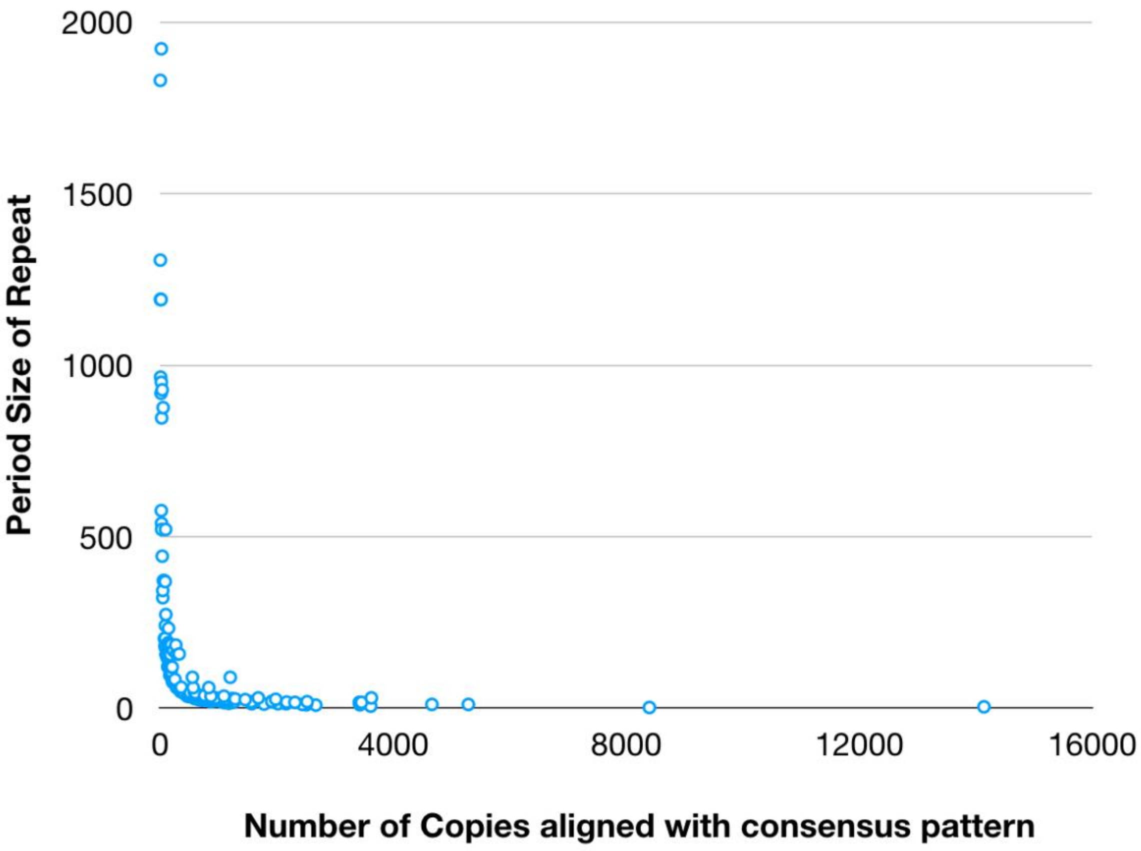
Supplemental Figure S3: Flowchart of our final genome assembly from Illumina (two lanes of PE 100bp) and Pacific Biosciences (40 single molecule real-time [SMRT] cells) sequence reads. Light blue boxes indicate raw sequence reads and dark blue boxes indicate bioinformatic programs used for the assembly. (…………) indicates the type of sequence information that was used for the assembly method. Arrows indicate the next step taken to proceed in the genome assembly of *Cebidichthys violaceus*.



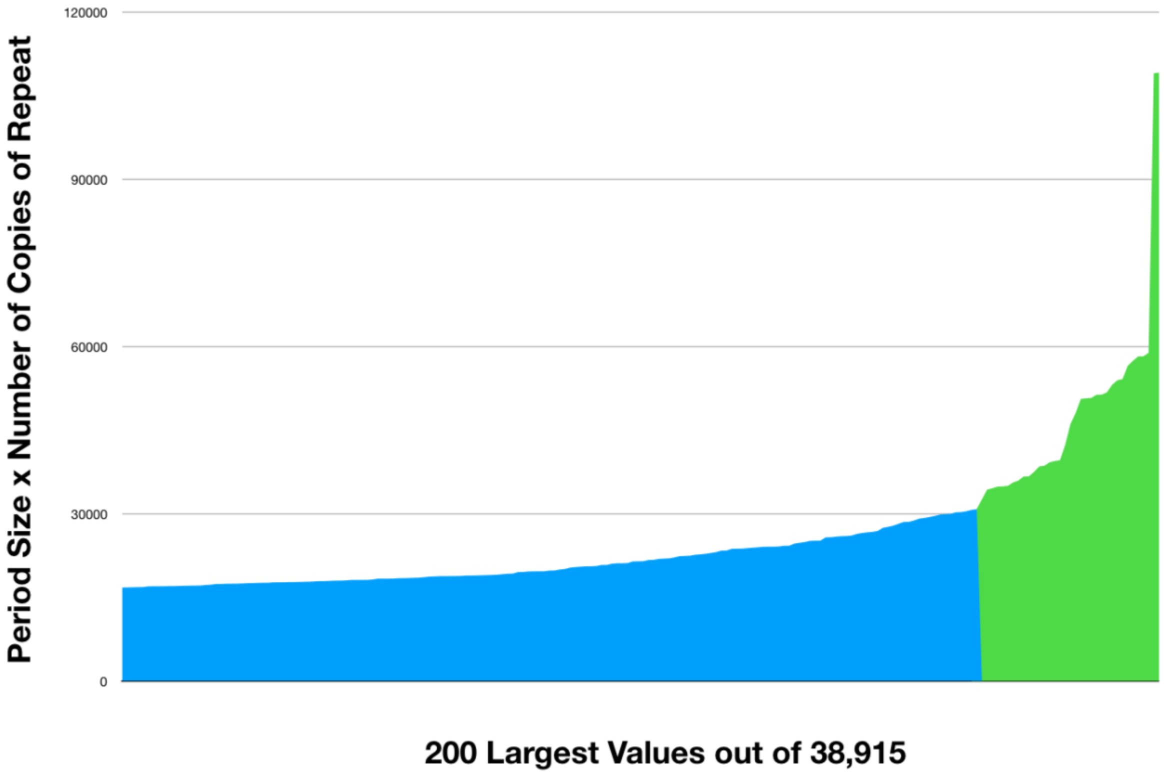
Supplemental Figure S4: Summarization of contig lengths of the *Cebidichthys violaceus* genome. N25 (blue bars), N50 (blue + green bars), and N75 (blue + green + orange bars) values was estimated for the *C. violaceus* genome. There are 66 contigs which represent the N75 of the *C. violaceus* genome. Contig ID is labeled along the x-axis.



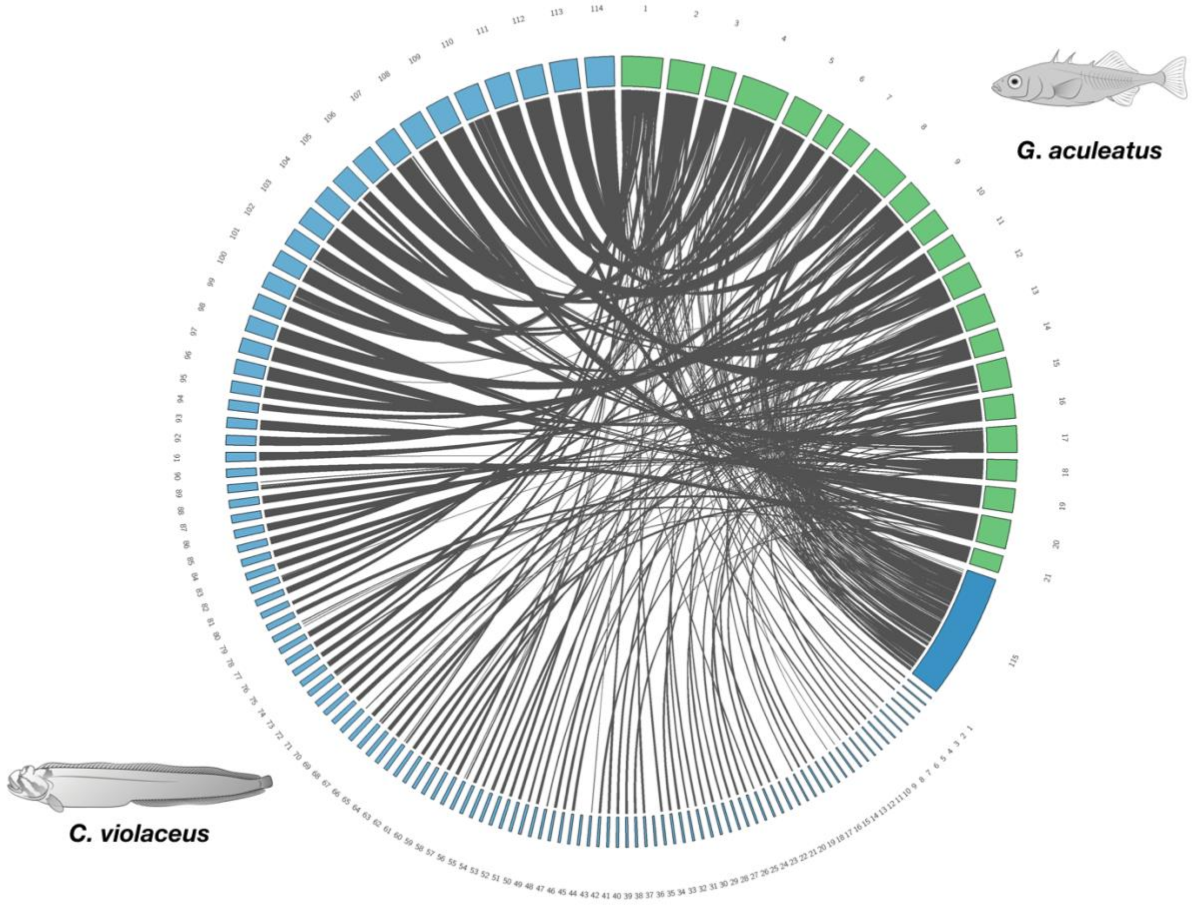
Supplemental Figure S5: K-mer frequency for *Cebidichthys violaceus*. Use of raw Illumina (only) reads from *C. violaceus* gDNA to estimate the *C. violaceus* genome size. K-mer sizes of 25, 27, 29, and 31 were selected to generate histograms.



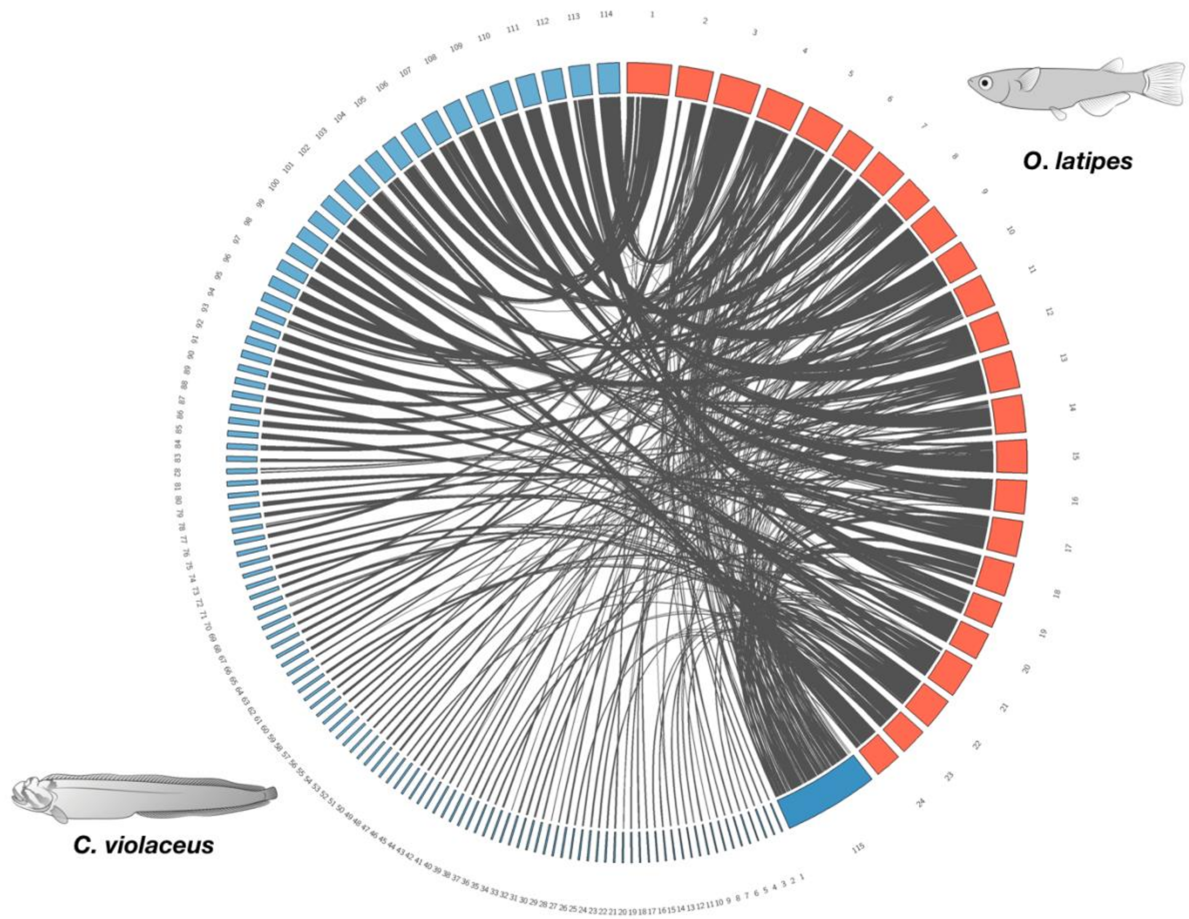
Supplemental Figure S6: Period size and copies of pattern of tandem repeats identified in the *Cebidichthys violaceus* genome. The length of the tandem repeat sequence and the size of the repeat found.



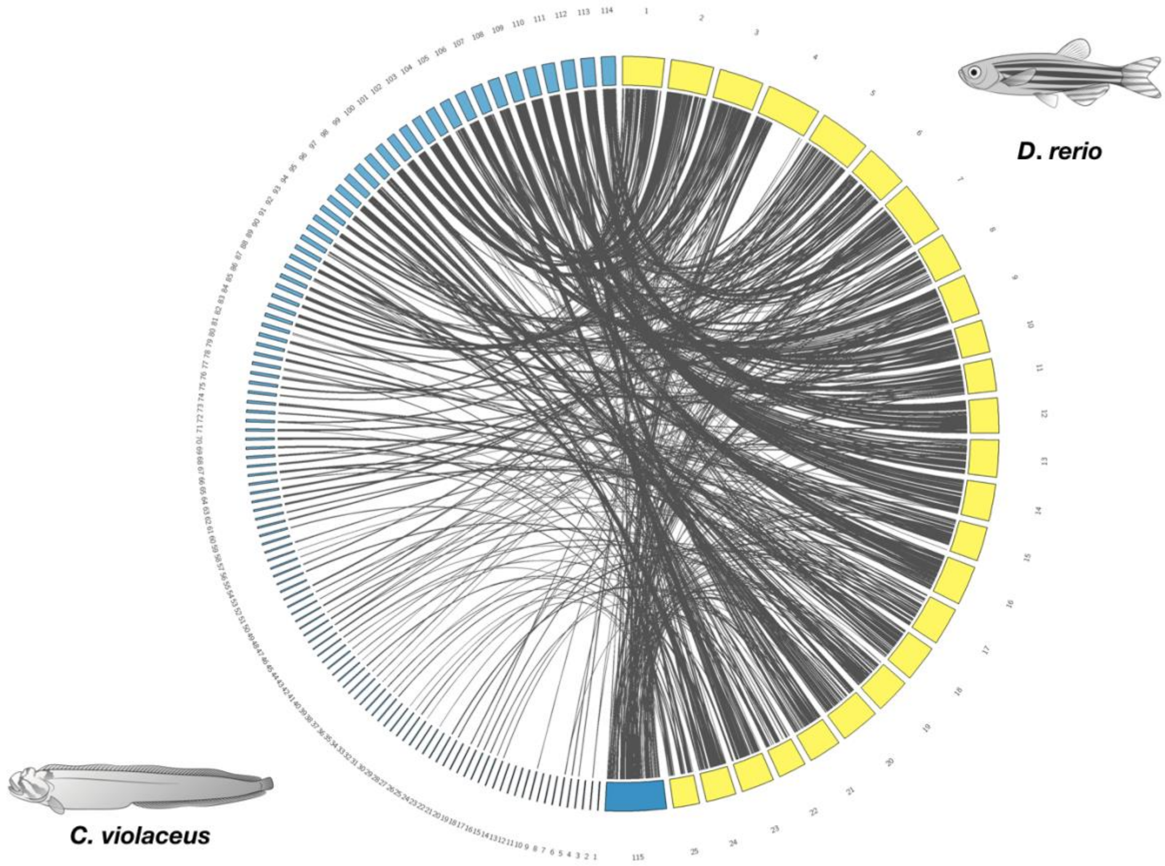
Supplemental Figure S7: Histogram of period size multiplied by number of copies of repeat. Total of the 200 longest repeats identified, where blue indicates values of 30,877 base pairs (bps) or less. Green indicates values greater than 30,877 bps.



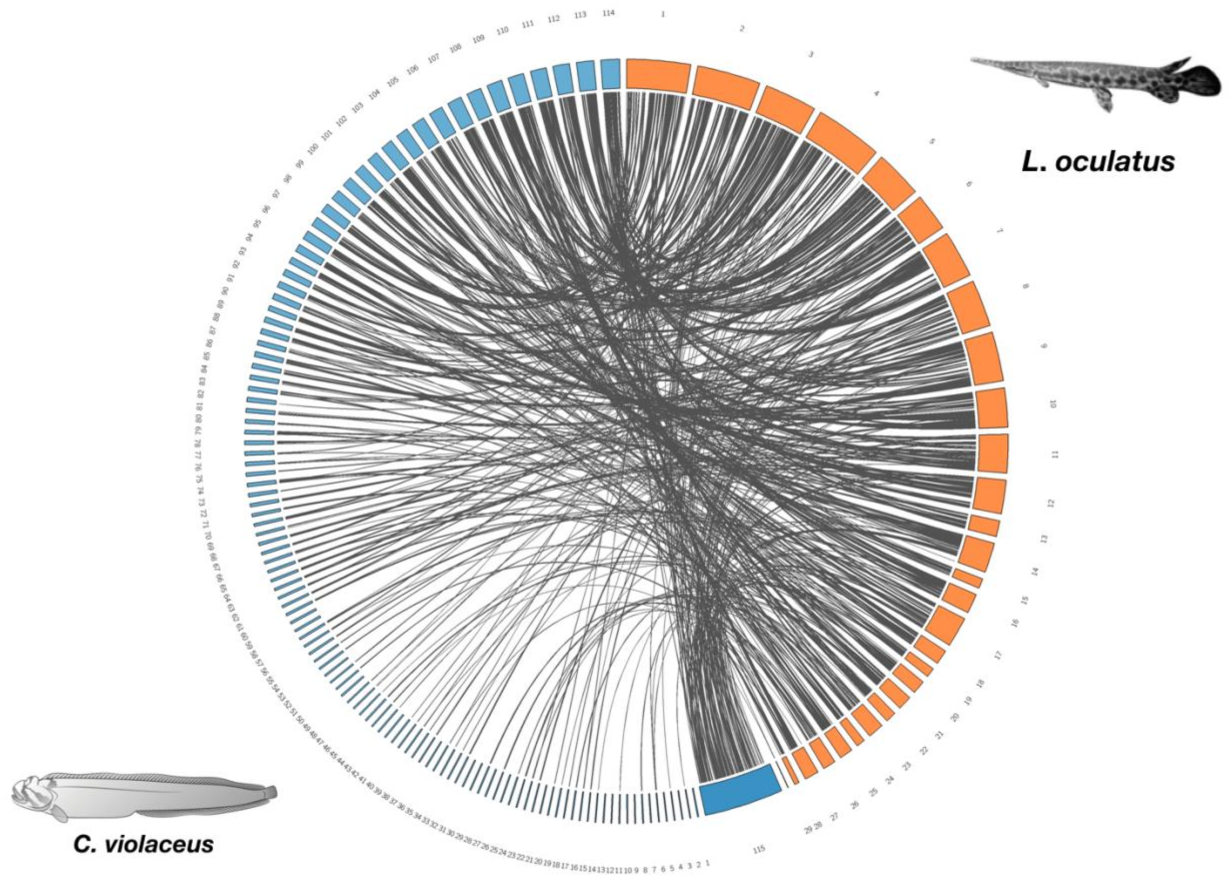
Supplemental Figure S8: Circos plot showing synteny between the assembled genome *Cebidichthys violaceu*s and *Gasterosteus aculeatus* (three-spined stickleback). There are 21 chromosomes (green boxes) which represent the three-spined stickleback genome. There are 114 blue boxes which are 1 MB or greater that represent the *C. violaceus* genome. There are 353 contigs that are less than 1 MB which were concatenated into box labeled as 115. Gray strands indicate syntenic regions between the two genomes. Both *G. aculeatus* and *C. violaceus* illustrations were drawn by Andrea Dingeldein.



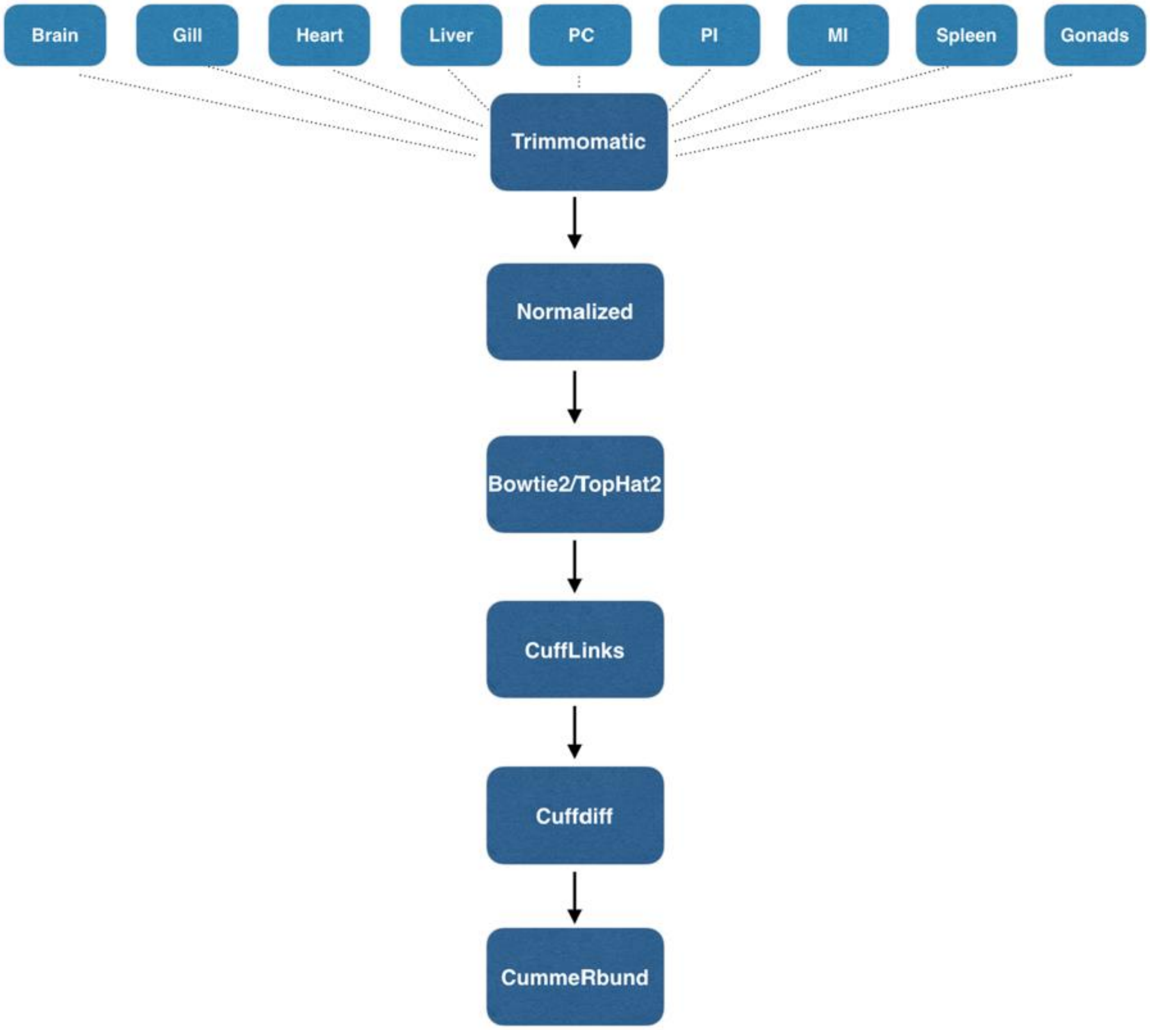
Supplemental Figure S9: Circos plot showing synteny between the assembled genome *Cebidichthys violaceus* and *Oryzias latipes* (Japanese rice fish). There are 24 chromosomes (red boxes) which represent the Japanese rice fish genome. There are 114 blue boxes which are 1 MB or greater that represent the *C. violaceus* genome. There are 353 contigs that are less than 1 MB which were concatenated into box labeled as 115. Gray strands indicate syntenic regions between the two genomes. Both *O. latipes* and *C. violaceus* illustrations were drawn by Andrea Dingeldein.



Supplemental Figure S10: Circos plot showing synteny between the assembled genome *Cebidichthys violaceus* and *Danio rerio* (Zebrafish). There are 25 chromosomes (yellow boxes) which represent the zebrafish genome. There are 114 blue boxes which are 1 MB or greater that represent the *C. violaceus* genome. There are 353 contigs that are less than 1 MB which were concatenated into box labeled as 115. Gray strands indicate syntenic regions between the two genomes. Both *D. rerio* and *C. violaceus* illustrations were drawn by Andrea Dingeldein.



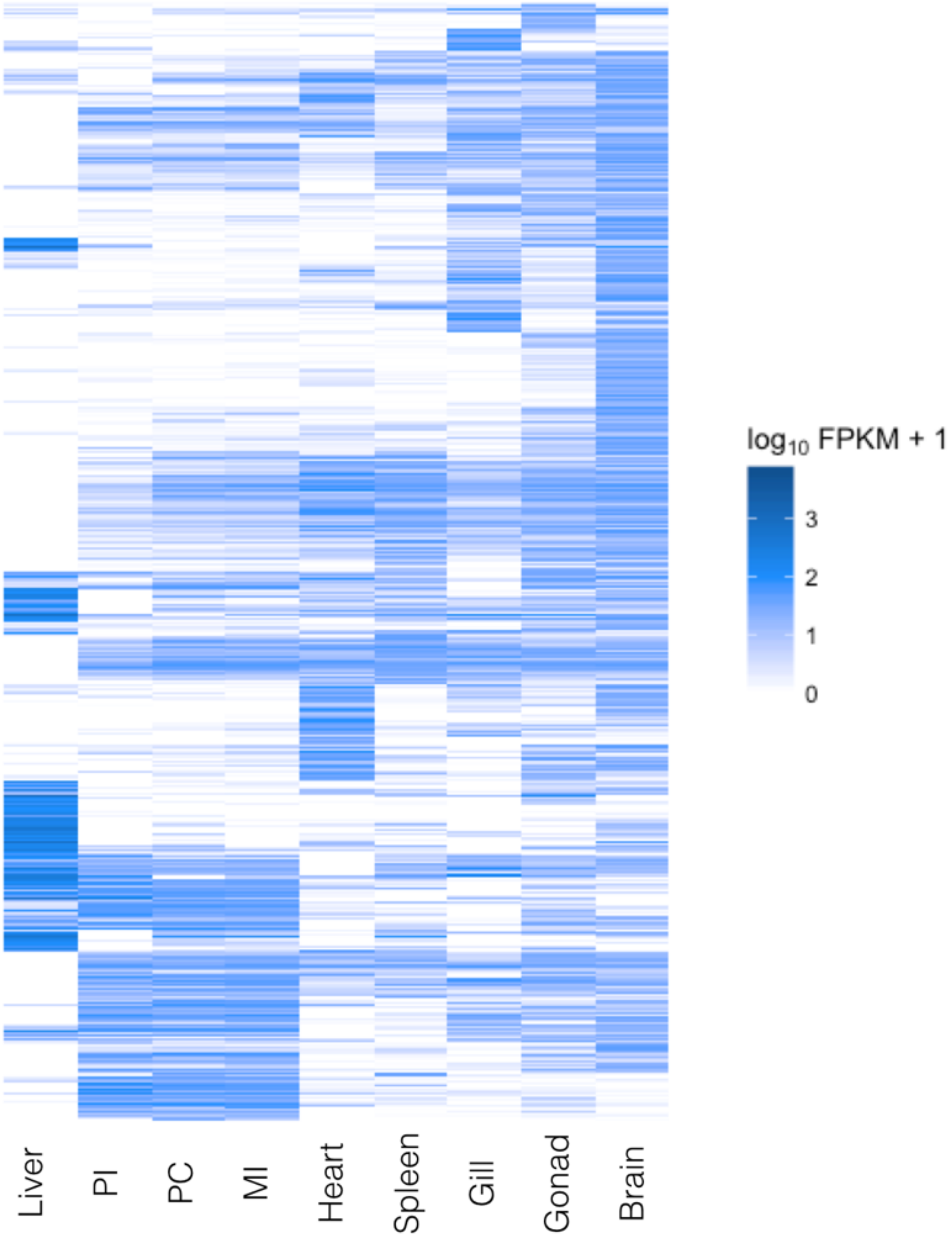
Supplemental Figure S11: Circos plot showing synteny between the assembled genome *Cebidichthys violaceus* and *Lepisosteus oculatus* (Spotted gar). There are 29 linkage groups (orange boxes) which represent the Spotted gar genome. There are 114 blue boxes which are 1 MB or greater that represent the *C. violaceus* genome. There are 353 contigs that are less than 1 MB which were concatenated into box labeled as 115. Gray strands indicate syntenic regions between the two genomes. *Lepisosteus oculatus* photo was taken by David Solomon and the *C. violaceus* illustration was drawn by Andrea Dingeldein.



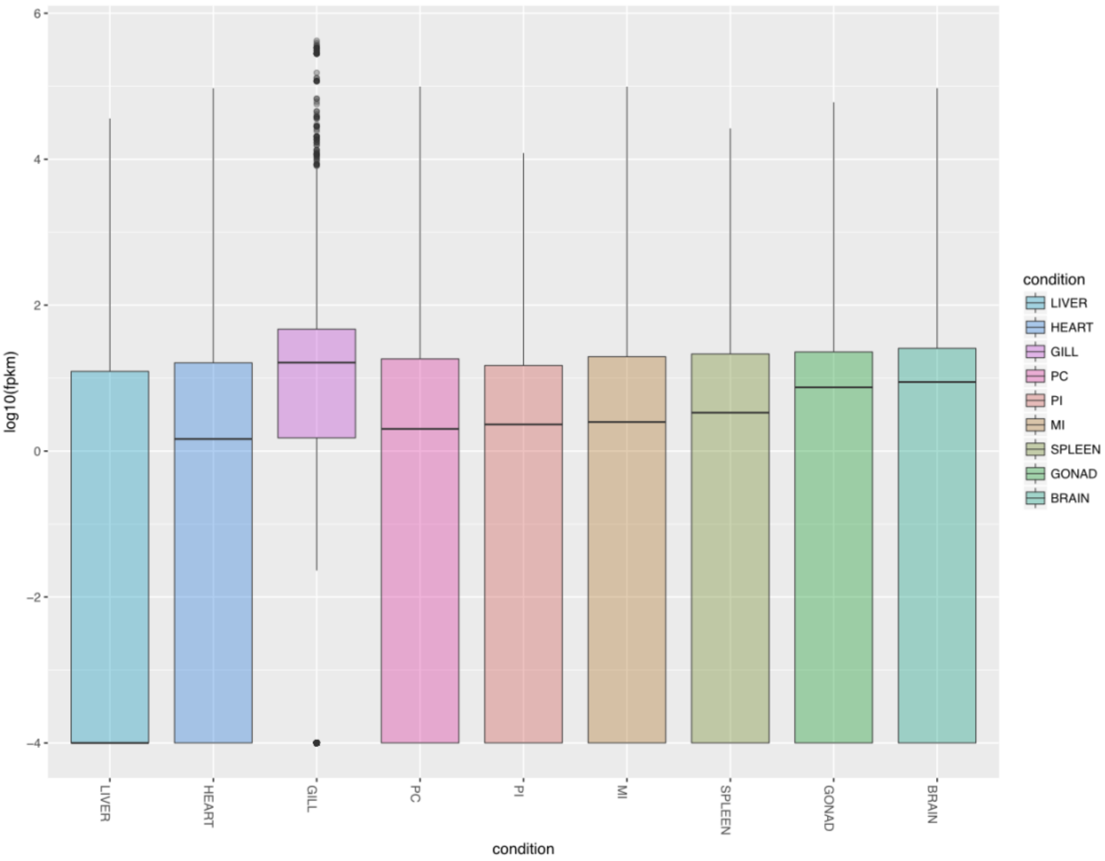
Supplemental Figure S12: Flowchart of our genome guided transcriptome assembly from nine tissues using Illumina (two lanes of PE 100bp).

Light blue boxes indicate raw sequence reads for nine tissues: liver, heart, gill, pyloric caeca (PC), proximal intestine (PI), middle intestine (MI), spleen, gonad (testes), and brain.

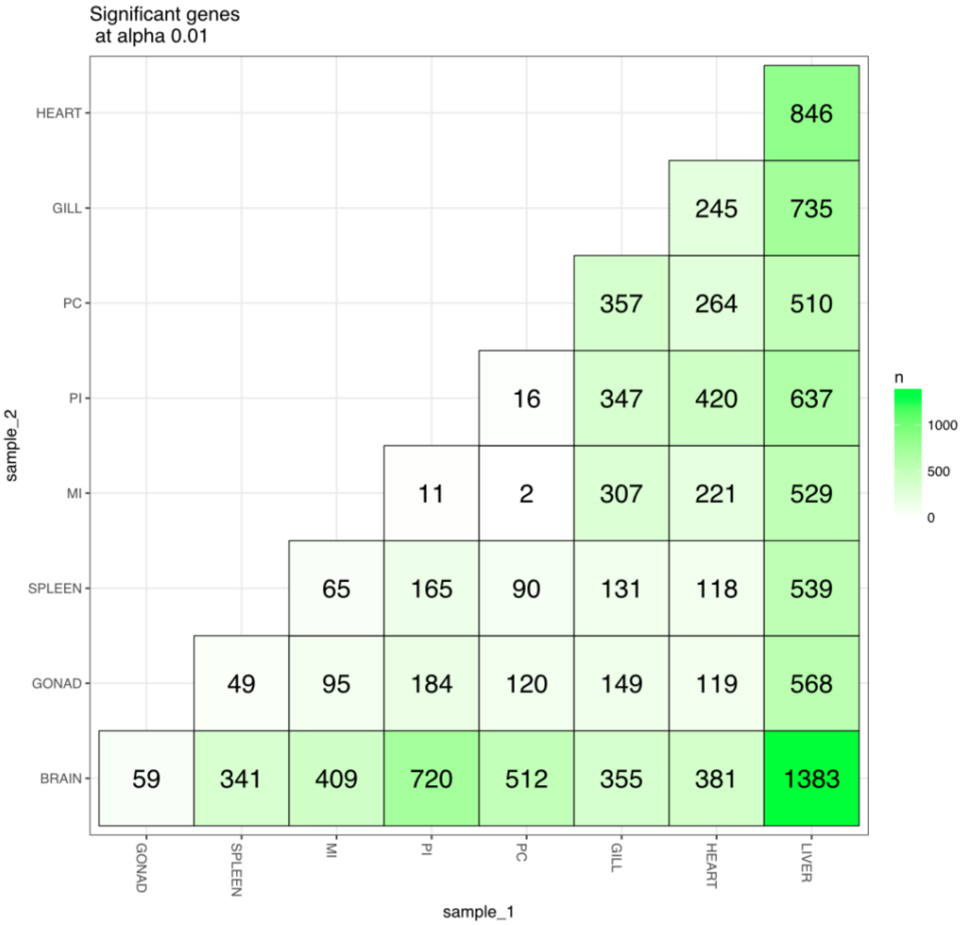
Dark blue boxes indicate the bioinformatic program used in the pipeline for trimming/cleaning reads, normalizing reads, assembling transcripts with our assemble genome as a reference, and estimating differential gene expression (DEGs) and analysis of DEGs. (…………) indicates the type of sequence information that was used for the start of the assembly method. Arrows indicate the next step taken to proceed in the transcriptome assembly and analysis.



Supplemental Figure S13: Heatmap showing Differentially Expressed Genes for nine tissues of *Cebidichthys violaceus*. Heatmap was generated with the csHeatmap feature in cummerbund, where dark blue represents a high FPKM value and white indicates a low FPKM value. There were 15,490 differentially expressed genes across all nine tissue types.

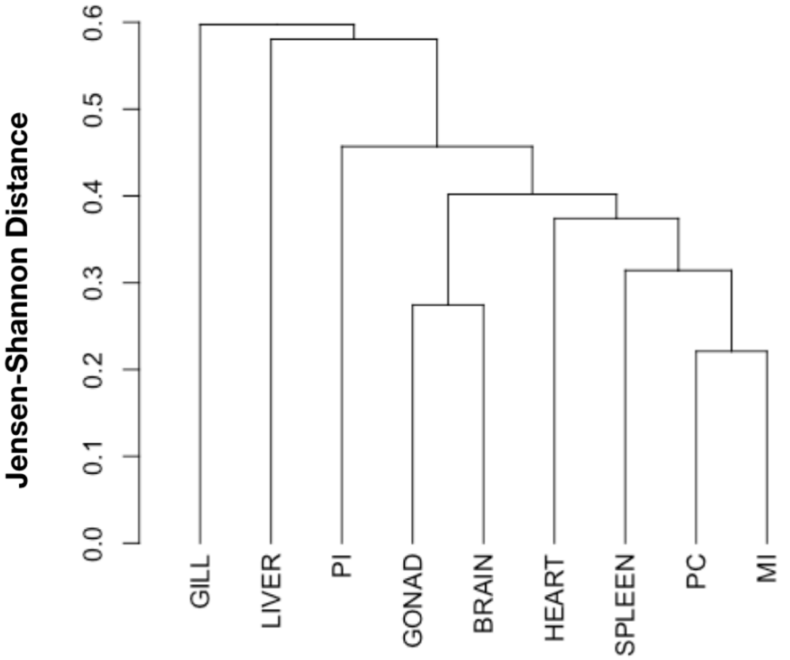


Supplemental Figure S14: Boxplots visualized for all nine tissue types displaying summary statistics of Fragments Per Kilobase of transcript per Million mapped reads (FPKM). Boxplots were generated with the csBoxplot function in cummerbund for nine tissues.

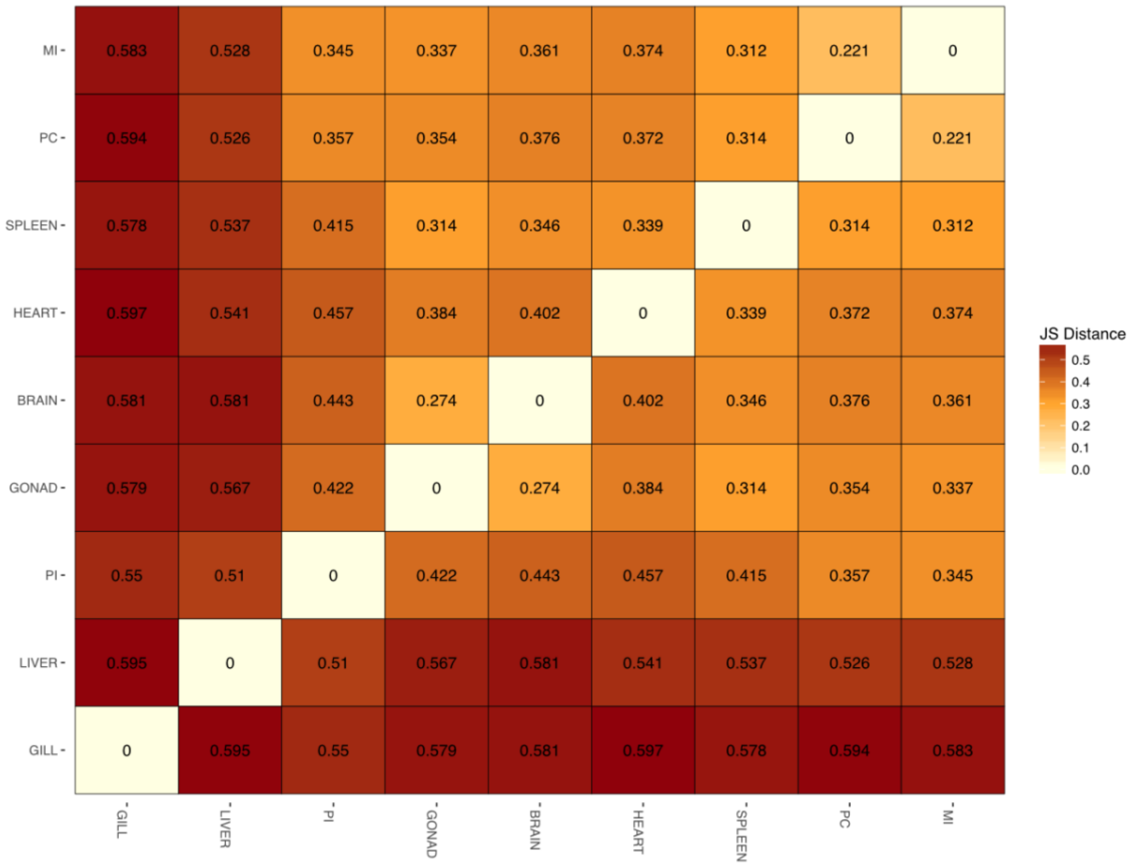


sample_2	GONAD	SPLEEN	MI	PI	PC	GILL	HEART	LIVER
HEART								846
GILL							245	735
PC						357	264	510
PI					16	347	420	637
MI				11	2	307	221	529
SPLEEN			65	165	90	131	118	539
GONAD		49	95	184	120	149	119	568
BRAIN	59	341	409	720	512	355	381	1383

Supplemental Figure S15: Significant features of genes for all nine tissue types. Significant features (alpha value 0.01) of genes between tissue types were estimated by using the sigmatrix function in CUMMERBUND. The darker green shades indicate a higher significant features of genes identified, whereas a lighter green/white shade indicates a less significant features of genes identified.

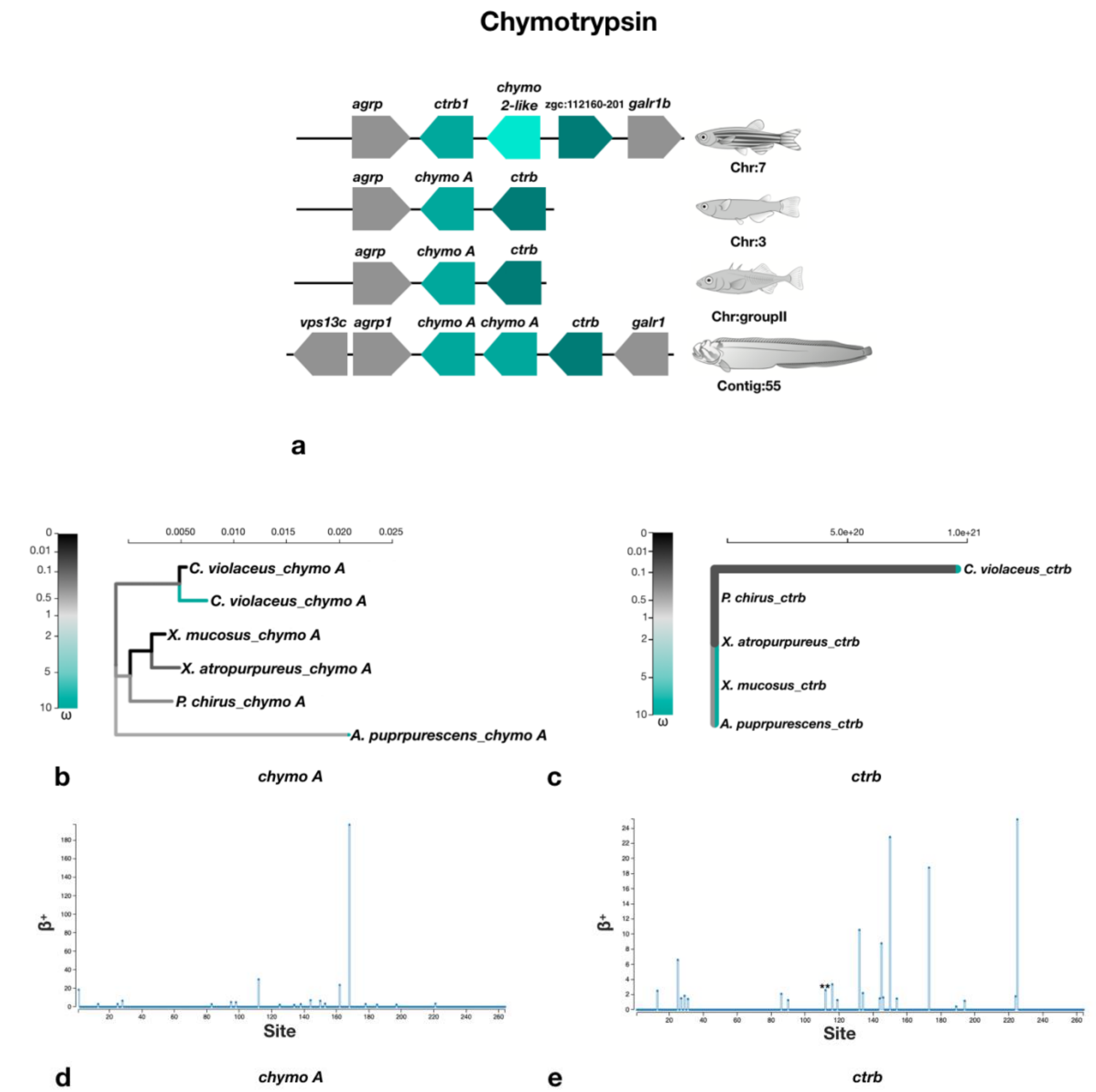


Supplemental Figure S16: Dendrogram of all nine tissue transcriptomes to determine relationships of each tissue type. A dendrogram was constructed of all nine tissue types by using Jensen-Shannon (JS) distances as shown.



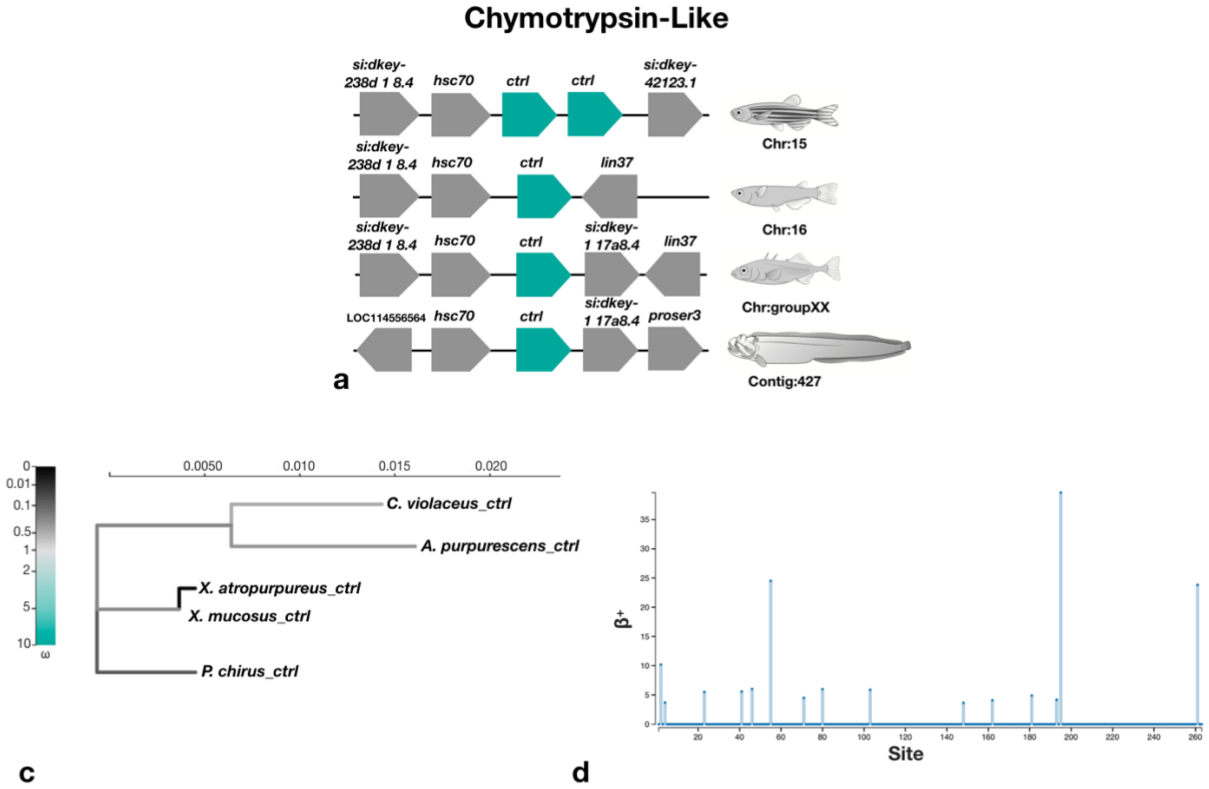
	GILL	LIVER	PI	GONAD	BRAIN	HEART	SPLEEN	PC	MI
MI	0.583	0.528	0.345	0.337	0.361	0.374	0.312	0.221	0
PC	0.594	0.526	0.357	0.354	0.376	0.372	0.314	0	0.221
SPLEEN	0.578	0.537	0.415	0.314	0.346	0.339	0	0.314	0.312
HEART	0.597	0.541	0.457	0.384	0.402	0	0.339	0.372	0.374
BRAIN	0.581	0.581	0.443	0.274	0	0.402	0.346	0.376	0.361
GONAD	0.579	0.567	0.422	0	0.274	0.384	0.314	0.354	0.337
PI	0.55	0.51	0	0.422	0.443	0.457	0.415	0.357	0.345
LIVER	0.595	0	0.51	0.567	0.581	0.541	0.537	0.526	0.528
GILL	0	0.595	0.55	0.579	0.581	0.597	0.578	0.594	0.583

Supplemental Figure S17: Distance matrix of all nine tissue transcriptomes based on Jensen-Shannon (JS). Plot was constructed with csDistHeat function in CUMMERBUND. Dark red indicates an increased JS distance between the pairwise comparison. Lighter red/white indicates less distance between the pairwise comparison.



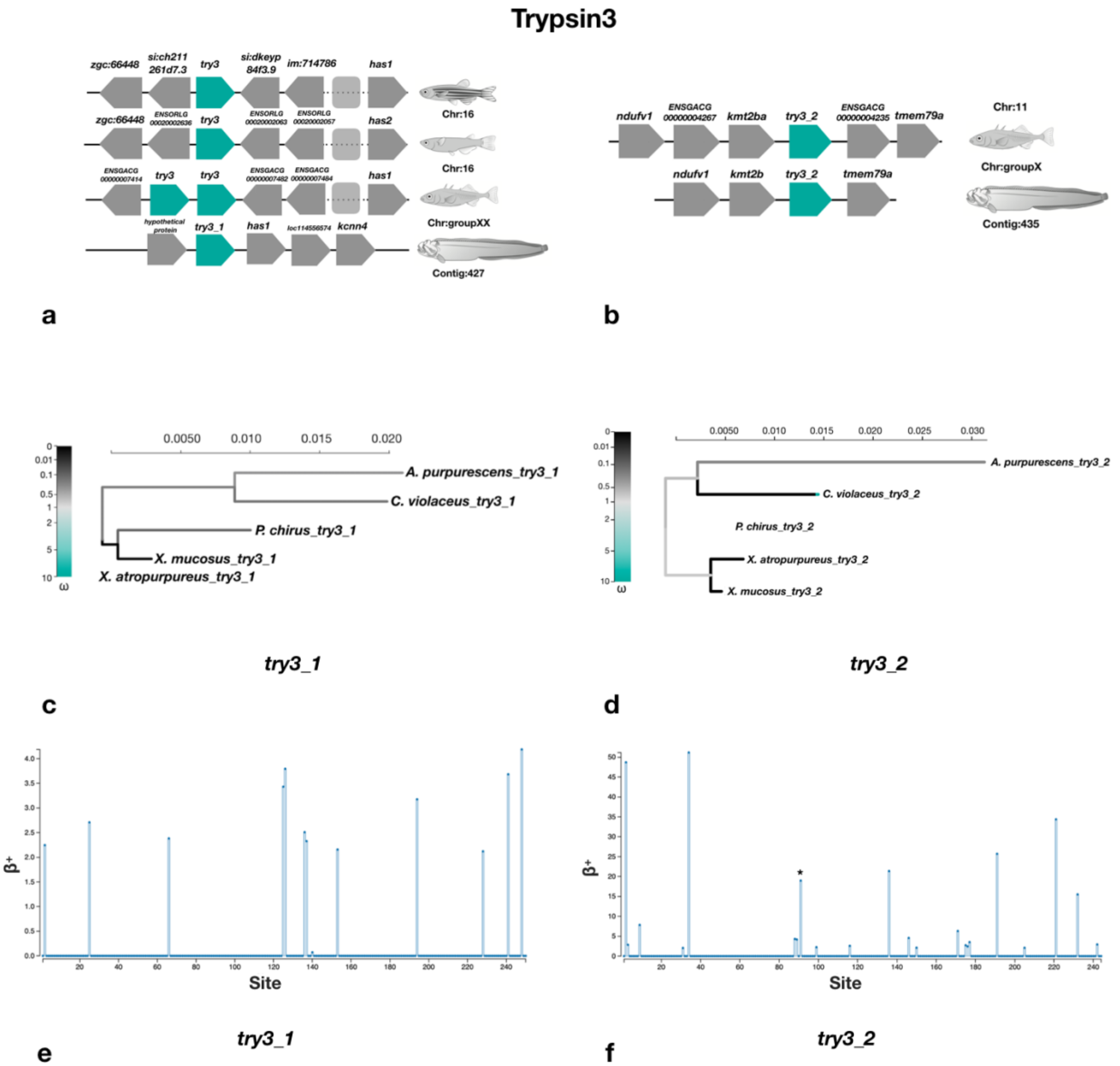
Supplemental Figure 18: Phylogenetic relationship of chymotrypsin in stichaeids, gene copy number, and molecular evolution of chymotrypsin.

a, Synteny map for chymotrypsin genes from *Danio rerio*, *Oryzias latipes*, *Gasterosteus aculeatus,* and *C. violaceus*. **c-d**, An adaptive branch-site Random Effects Likelihood (aBSREL) test for episodic diversification phylogenetic tree constructed for chymotrypsin genes from *C. violaceus* (*chymo A* and *ctrb*) and other intertidal stichaeid species. Branches thicker than the other branches have a P<0.05 (corrected for multiple comparisons) to reject the null hypothesis of all ω on that branch (neutral or negative selection only). A thick branch is considered to have experienced diversifying positive selection. **e-f**, The output of Mixed Effects Model of Evolution (MEME) to detect episodic positive/diversifying selection at sites. β+ is the non-synonymous substitution rate at a site for the positive/neutral evolution throughout the sequence of the chymotrypsin A gene. **e**, MEME output for chymotrypsin B gene. ** is an indication that the positive/diversifying site is statistically significant with a p-value < 0.01.



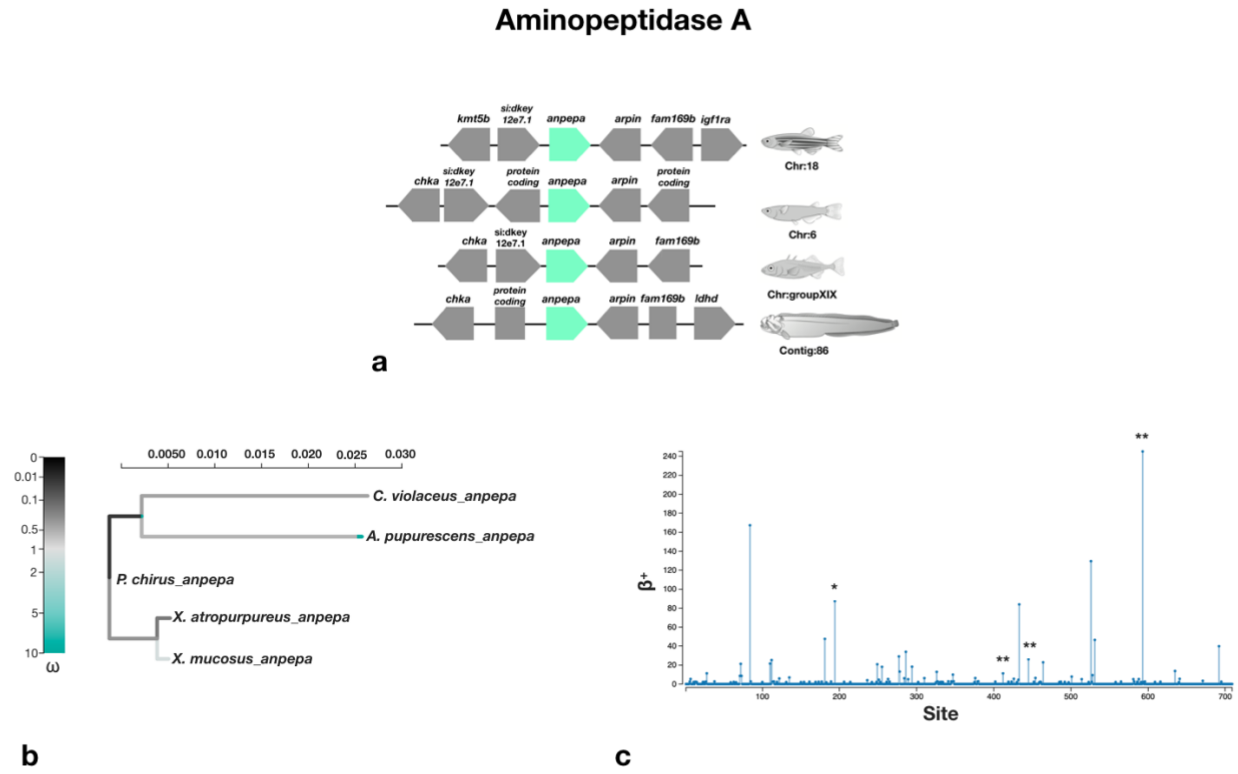
Supplemental Figure 19: Phylogenetic relationship of chymotrypsin-like gene in stichaeids, gene copy number, and molecular evolution of chymotrypsin-like (*ctrl*).

a, A maximum likelihood tree was generated for chymotrypsin-like genes from *Cebidichthys violaceus* and other intertidal stichaeid species: *Anoplarchus purpurescens*, *Phytichthys chirus, Xiphister mucosus*, and *Xiphister atropurureus*. *Gasterosteus aculeatus* chymotrypsin gene was used as an outgroup for our phylogenetic analysis. **b**, Synteny map for amylase genes from *Danio rerio*, *Oryzias latipes*, *Gasterosteus aculeatus,* and *C. violaceus*. **c**, An adaptive branch-site Random Effects Likelihood (aBSREL) test for episodic diversification phylogenetic tree constructed for the chymotrypsin gene from *C. violaceus* (*ctrl*) and other intertidal stichaeid species. Branches thicker than the other branches have a $P<0.05$ (corrected for multiple comparisons) to reject the null hypothesis of all ω on that branch (neutral or negative selection only). A thick branch is considered to have experienced diversifying positive selection. **d**, The output of Mixed Effects Model of Evolution (MEME) to detect episodic positive/diversifying selection at sites. β_+ is the non-synonymous substitution rate at a site for the positive/neutral evolution throughout the sequence of the gene. There were zero sites under positive/diversifying selection.

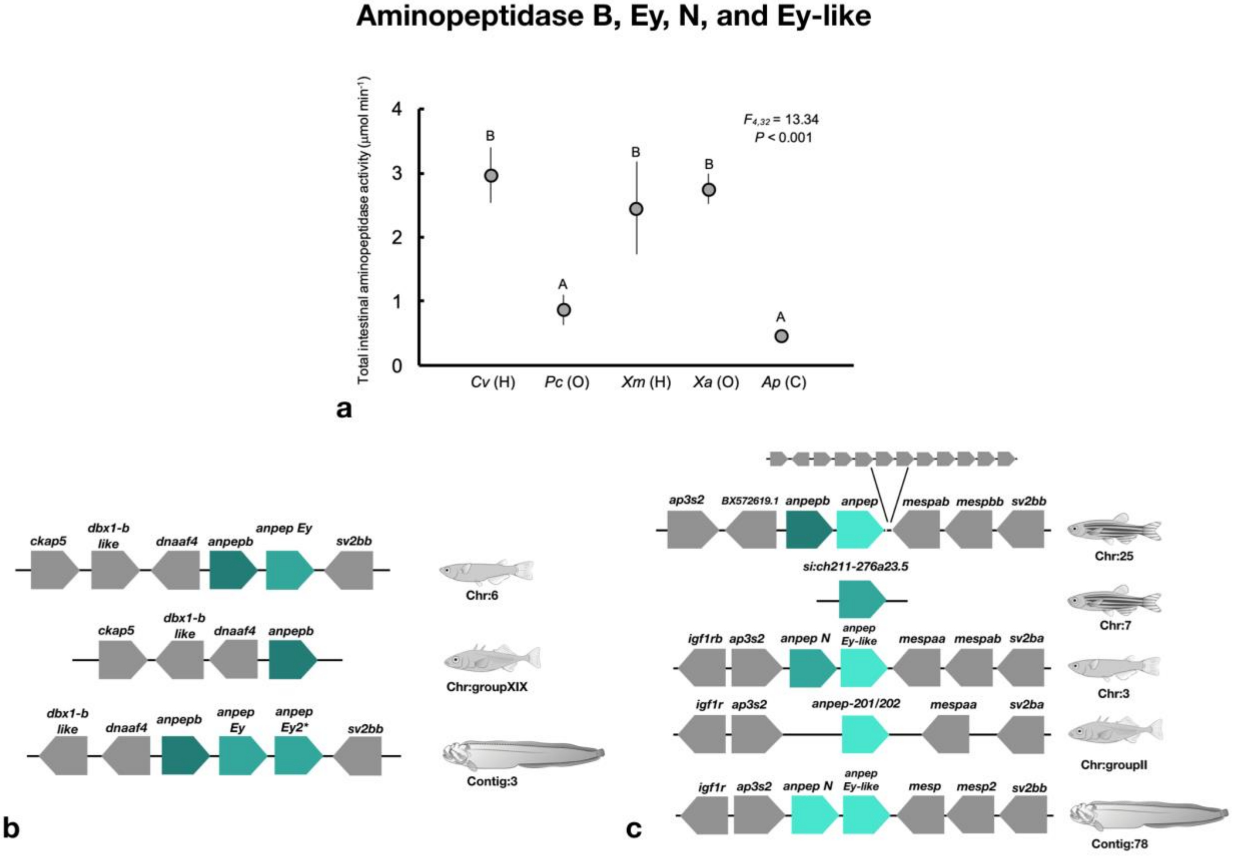


Supplemental Figure 20: Phylogenetic relationship of trypsin in stichaeids, gene copy number, and molecular evolution of trypsin (*try3*).

a-b, Synteny map for trypsin genes from *Danio rerio*, *Oryzias latipes*, *Gasterosteus aculeatus,* and *C. violaceus*. Clear boxes indicate multiple loci present in this region. **c-d**, An adaptive branch-site Random Effects Likelihood (aBSREL) test for episodic diversification phylogenetic tree constructed for trypsin genes from *C. violaceus* (*try3*) and other intertidal stichaeid species. Branches thicker than the other branches have a P<0.05 (corrected for multiple comparisons) to reject the null hypothesis of all ω on that branch (neutral or negative selection only). A thick branch is considered to have experienced diversifying positive selection. **e-f**, The output of Mixed Effects Model of Evolution (MEME) to detect episodic positive/diversifying selection at sites. β_+ is the non-synonymous substitution rate at a site for the positive/neutral evolution throughout the sequence of the gene. * is an indication that the positive/diversifying site is statistically significant with a p-value < 0.05.

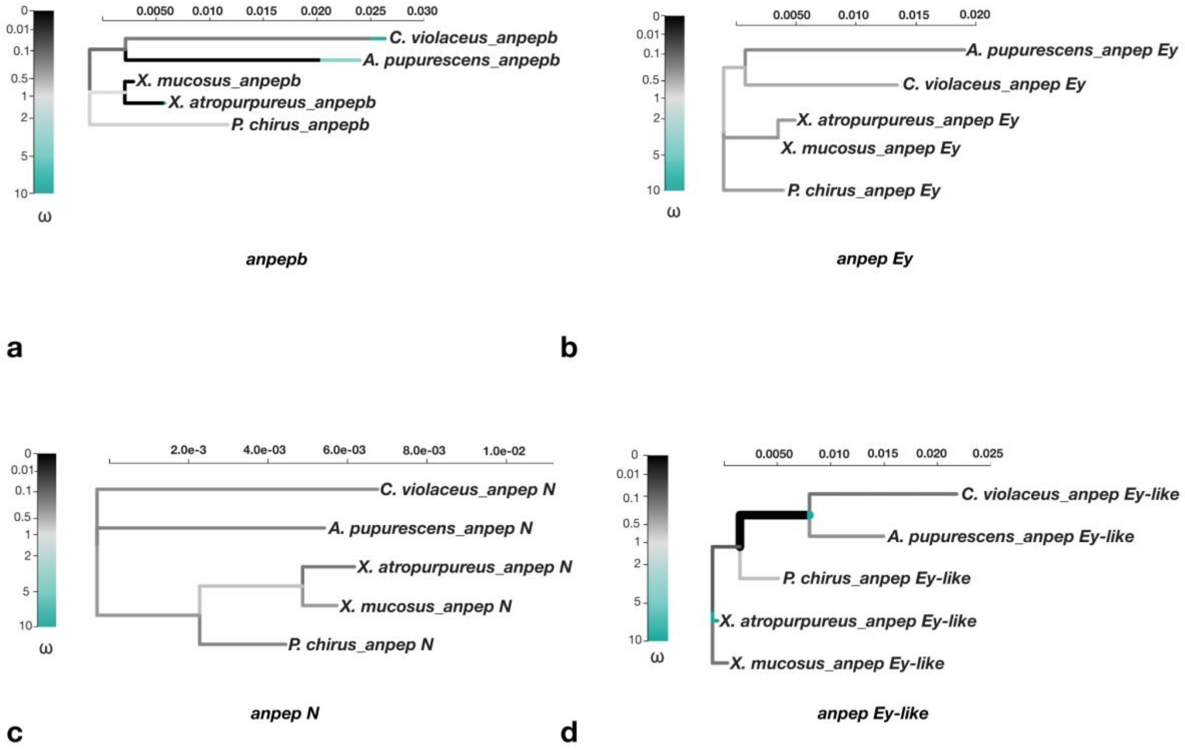


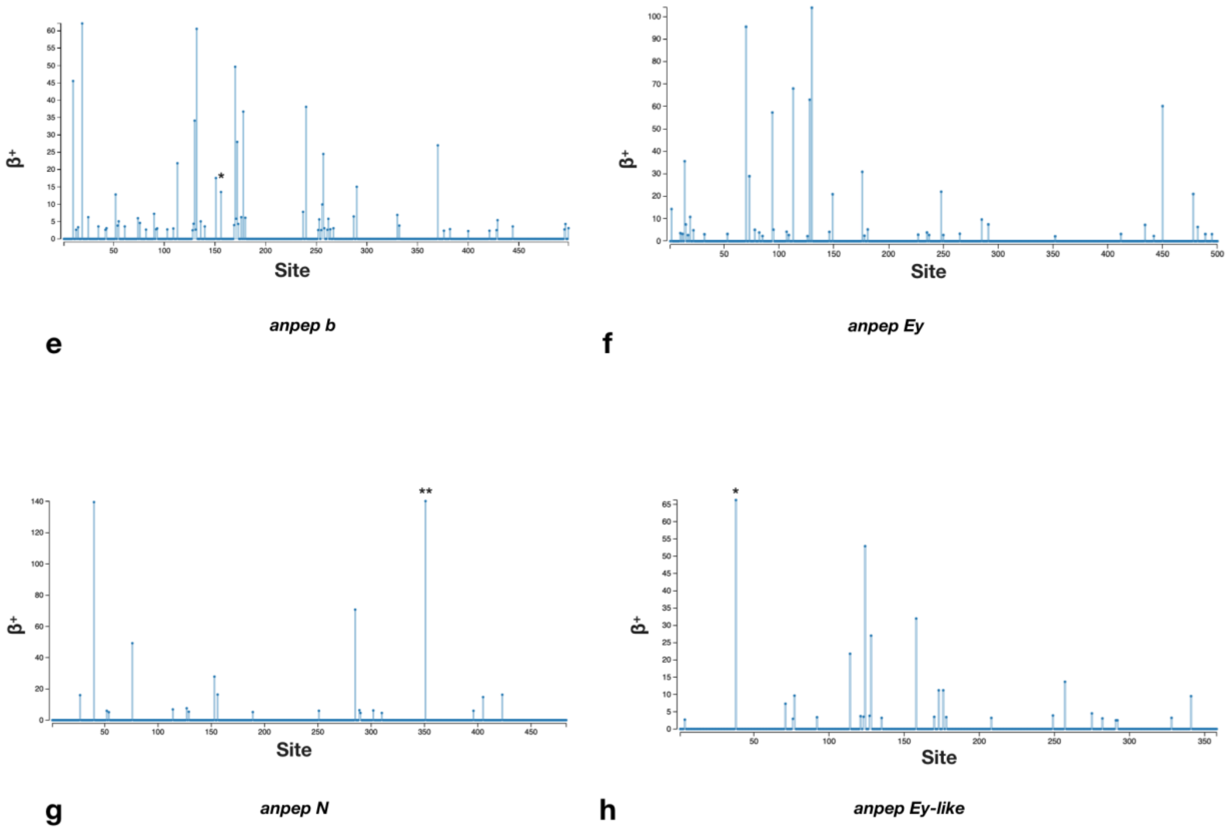
Supplemental Figure 21: Gene copy number and molecular evolution of alanyl aminopeptidase a (*anpepa*). a, Synteny map for *anpepa* genes from *Danio rerio*, *Oryzias latipes*, *Gasterosteus aculeatus,* and *C. violaceus*. **b**, An adaptive branch-site Random Effects Likelihood (aBSREL) test for episodic diversification phylogenetic tree constructed for the *anpepa* gene from *C. violaceus* and other intertidal stichaeid species. ω is the ratio of nonsynonymous to synonymous substitutions. The color gradient represents the magnitude of the corresponding ω. Branches thicker than the other branches have a P<0.05 (corrected for multiple comparisons) to reject the null hypothesis of all ω on that branch (neutral or negative selection only). A thick branch is considered to have experienced diversifying positive selection. **c**, The output of Mixed Effects Model of Evolution (MEME) to detect episodic positive/diversifying selection at sites. β_+ is the non-synonymous substitution rate at a site for the positive/neutral evolution throughout the sequence of the gene. ** is an indication that the positive/diversifying site is statistically significant with a p-value < 0.01 and * is for p-value < 0.05.



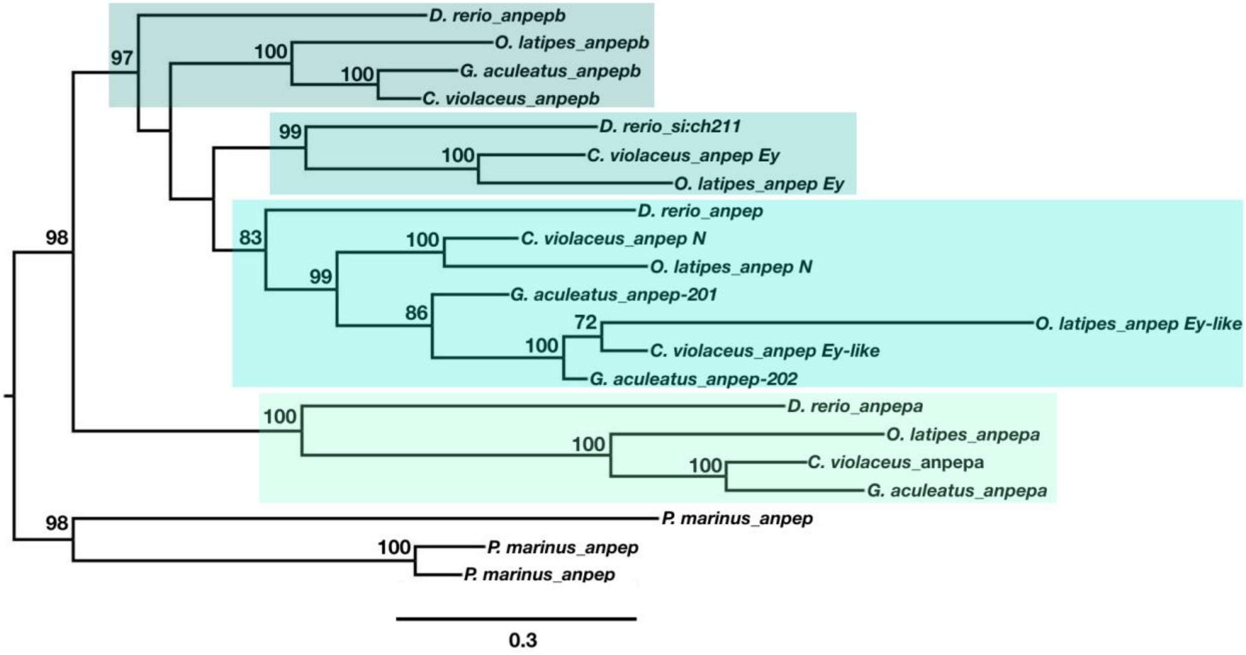
Supplemental Figure 22: Enzyme activity and gene copy number of alanyl aminopeptidase (*anpepb*, *anpep Ey*, *anpep N*, and *anpep Ey-like*). a, Total gut standardized aminopeptidase activity for *Cebidichthys violaceus* (Cv) and other intertidal stichaeid species: *Phytichthys chirus* (Pc), *Xiphister mucosus* (Xm), *Xiphister atropurpureus* (Xa), and *Anoplarchus purpurescens* (Ap). H = herbivory, O = Omnivory, and C = Carnivory. Values are mean ± standard deviation with n = 6 for Cv, Xm, Xa, and Ap; and n = 9 for Pc (German et al. 2015). Interspecific comparisons were made for aminopeptidase with ANOVA, where circles that share a letter are not significantly different. **b-c**, Synteny maps for aminopeptidase genes from *Danio rerio*, *Oryzias latipes*, *Gasterosteus aculeatus,* and *C. violaceus*.

Aminopeptidase B, Ey, N, and Ey-like

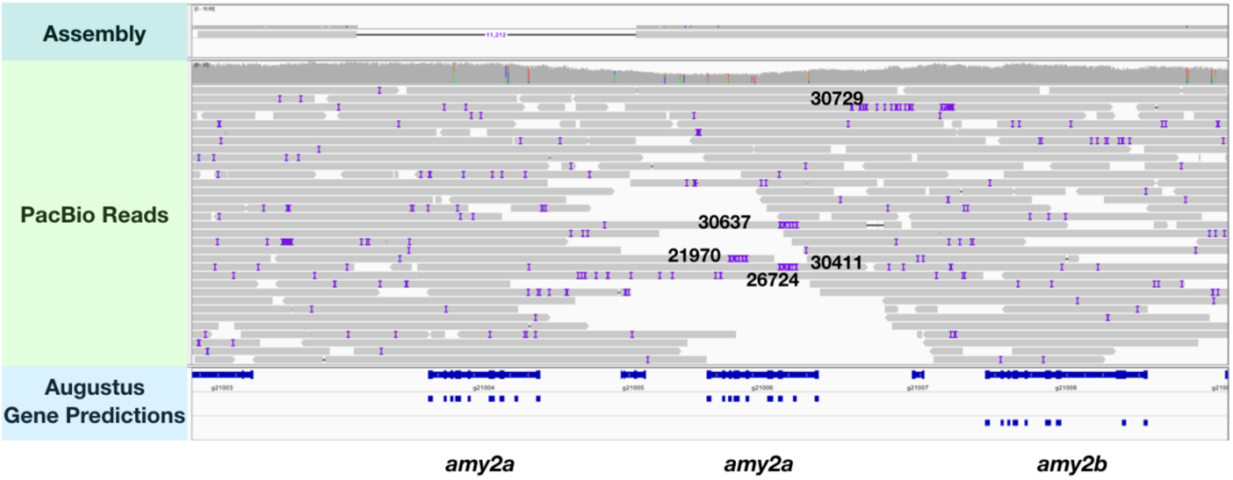




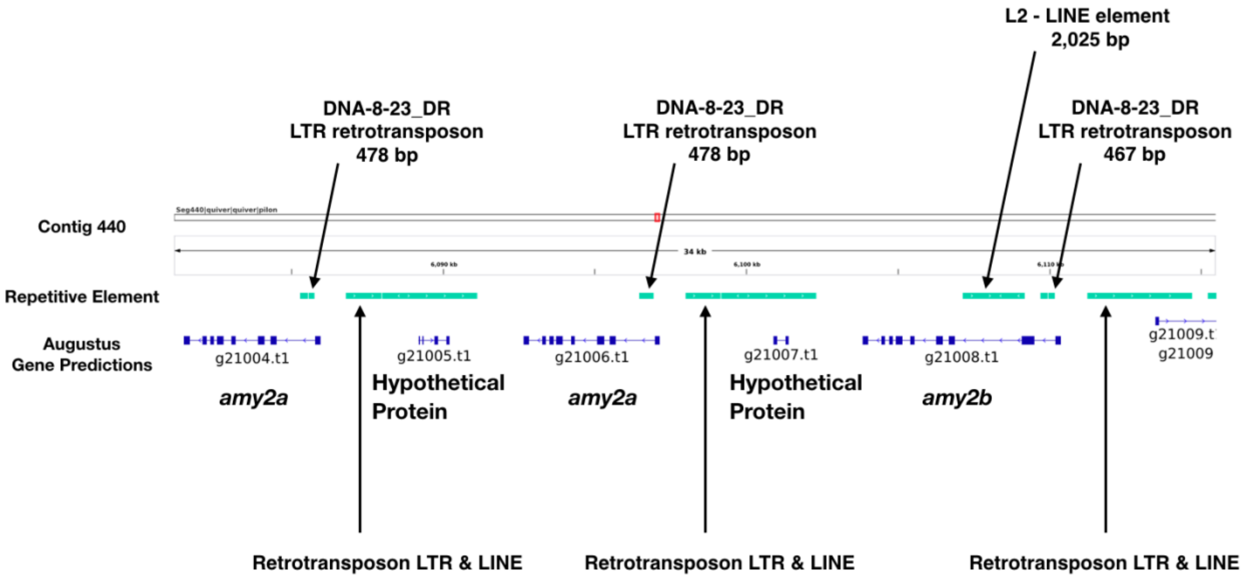
Supplemental Figure 23: Estimation of selection analyses of alanyl aminopeptidase (*anpepb*, *anpep Ey*, *anpep N*, and *anpep Ey-like*) genes in stichaeids. a-d, An adaptive branch-site Random Effects Likelihood (aBSREL) test for episodic diversification phylogenetic tree constructed for *anpep* genes from *C. violaceus* and other intertidal stichaeid species. Branches thicker than the other branches have a P<0.05 (corrected for multiple comparisons) to reject the null hypothesis of all ω on that branch (neutral or negative selection only). A thick branch is considered to have experienced diversifying positive selection. **e-h**, The output of Mixed Effects Model of Evolution (MEME) to detect episodic positive/diversifying selection at sites. ** is an indication that the positive/diversifying site is statistically significant with a p-value < 0.01 and * is for p-value < 0.05.



Supplemental Figure 24: Phylogenetic relationship of alanyl aminopeptidase genes in fishes (including *Cebidichthys violaceus*). A maximum likelihood (ML) tree was constructed with 1,000 bootstrap replicates in PhyML v3.1 based alanyl aminopeptidase sequences from *C. violaceus* and other fish species (e.g. *Danio rerio*, *Gasterosteus aculeatus*, *Oryzias latipes*, and *Petromyzon marinus*). Alanyl aminopeptidase sequences from *P. marinus* were used as an outgroup.

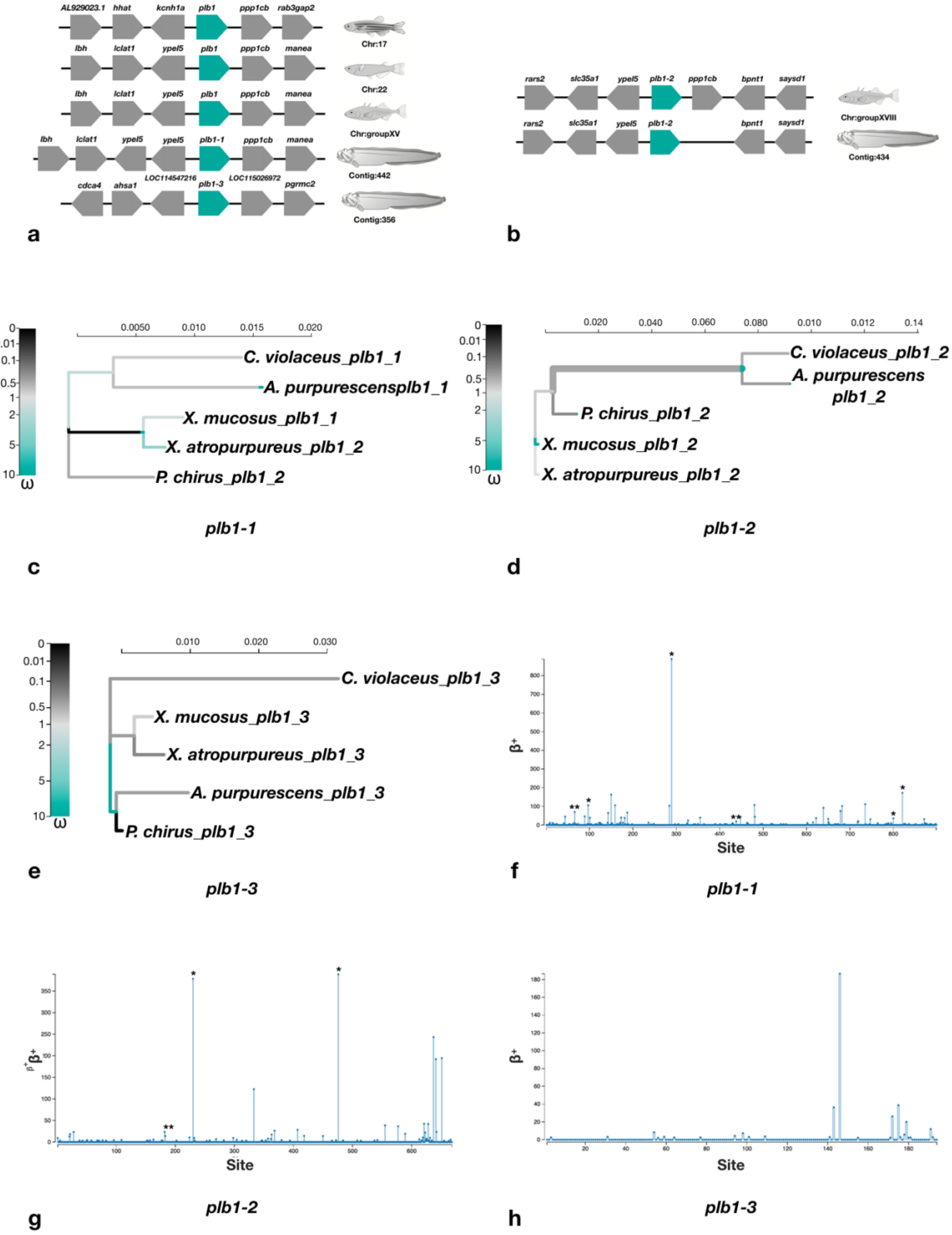


Supplemental Figure S25: Pacific Biosciences (PacBio) reads mapped to *Cebidichthys violaceus* genome assembly on contig 440. PacBio reads with read ID numbers labeled which span regions of the three amylase loci (two *amy2a* loci and the *amy2b*) on contig 440. AUGUSTUS gene predictions are used to reference where *amy2* loci are located on the *C. violaceus* genome.

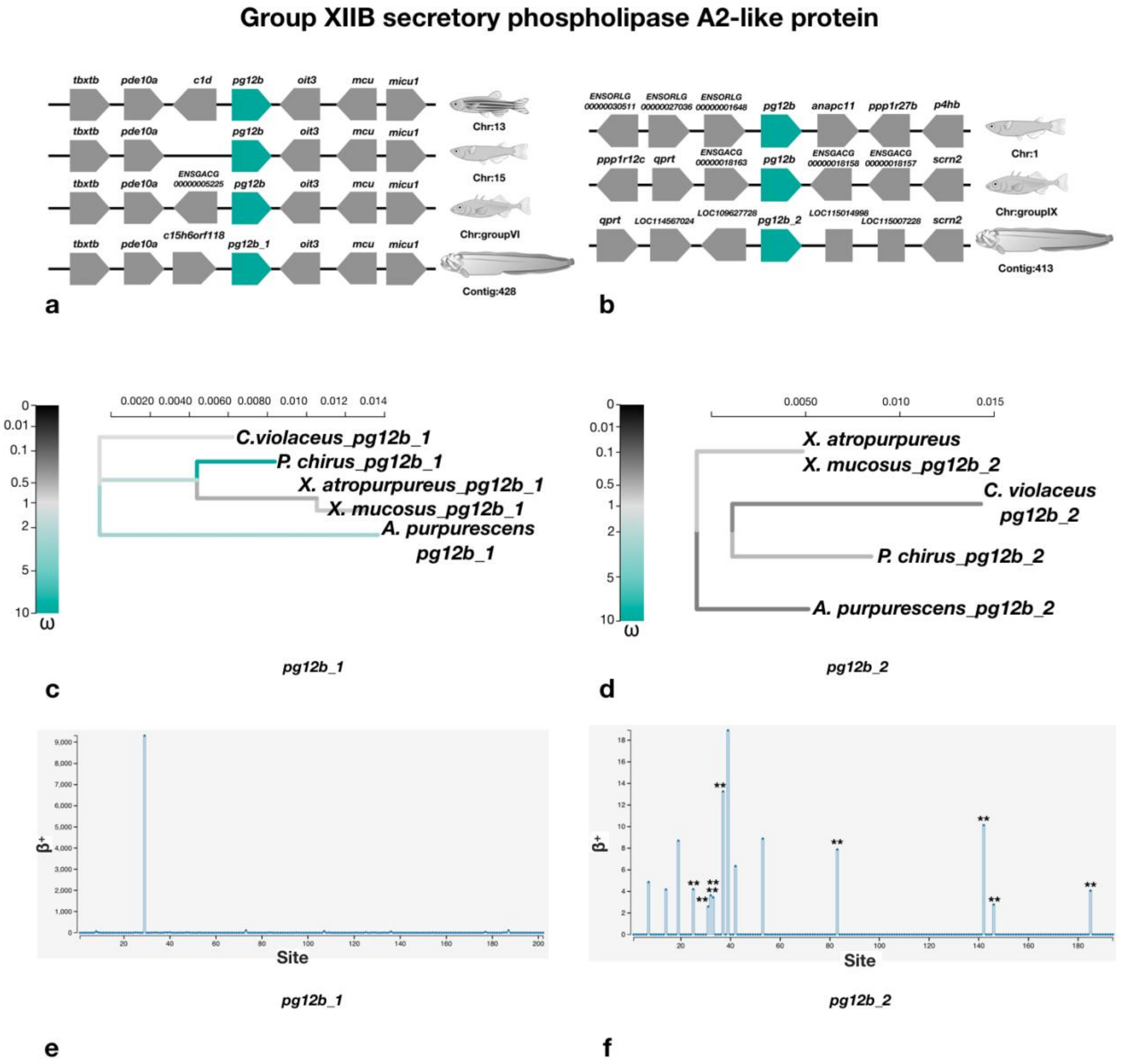


Supplemental Figure S26: Repetitive elements identified adjacent to amylase loci. Light green bars are repetitive elements identified on contig 440 and blue bars represent exons of amylase loci (AUGUSTUS gene prediction).

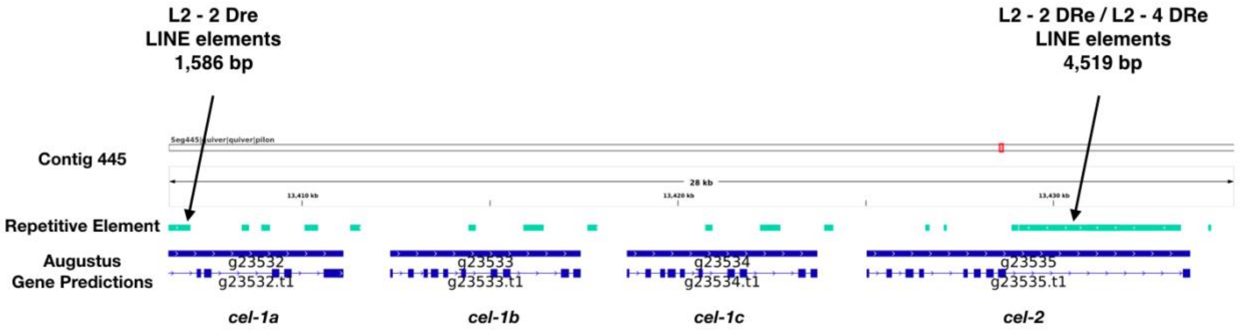
Phospholipase B1



Supplemental Figure 27: Gene copy number, and molecular evolution of Phospholipase B1 (*plb*). **a-b**, Synteny map for phospholipase B1 (*plb1*) genes from *Danio rerio*, *Oryzias latipes*, *Gasterosteus aculeatus,* and *C. violaceus*. **c-e**, An adaptive branch-site Random Effects Likelihood (aBSREL) test for episodic diversification phylogenetic tree constructed for *plb1* genes from *C. violaceus* and other intertidal stichaeid species. Branches thicker than the other branches have a $P<0.05$ (corrected for multiple comparisons) to reject the null hypothesis of all ω on that branch (neutral or negative selection only). A thick branch is considered to have experienced diversifying positive selection. **f-h,** The output of Mixed Effects Model of Evolution (MEME) to detect episodic positive/diversifying selection at sites. β_{+} is the non-synonymous substitution rate at a site for the positive/neutral evolution throughout the sequence of the gene. ** is an indication that the positive/diversifying site is statistically significant with a p-value < 0.01 and * is for p-value < 0.05.

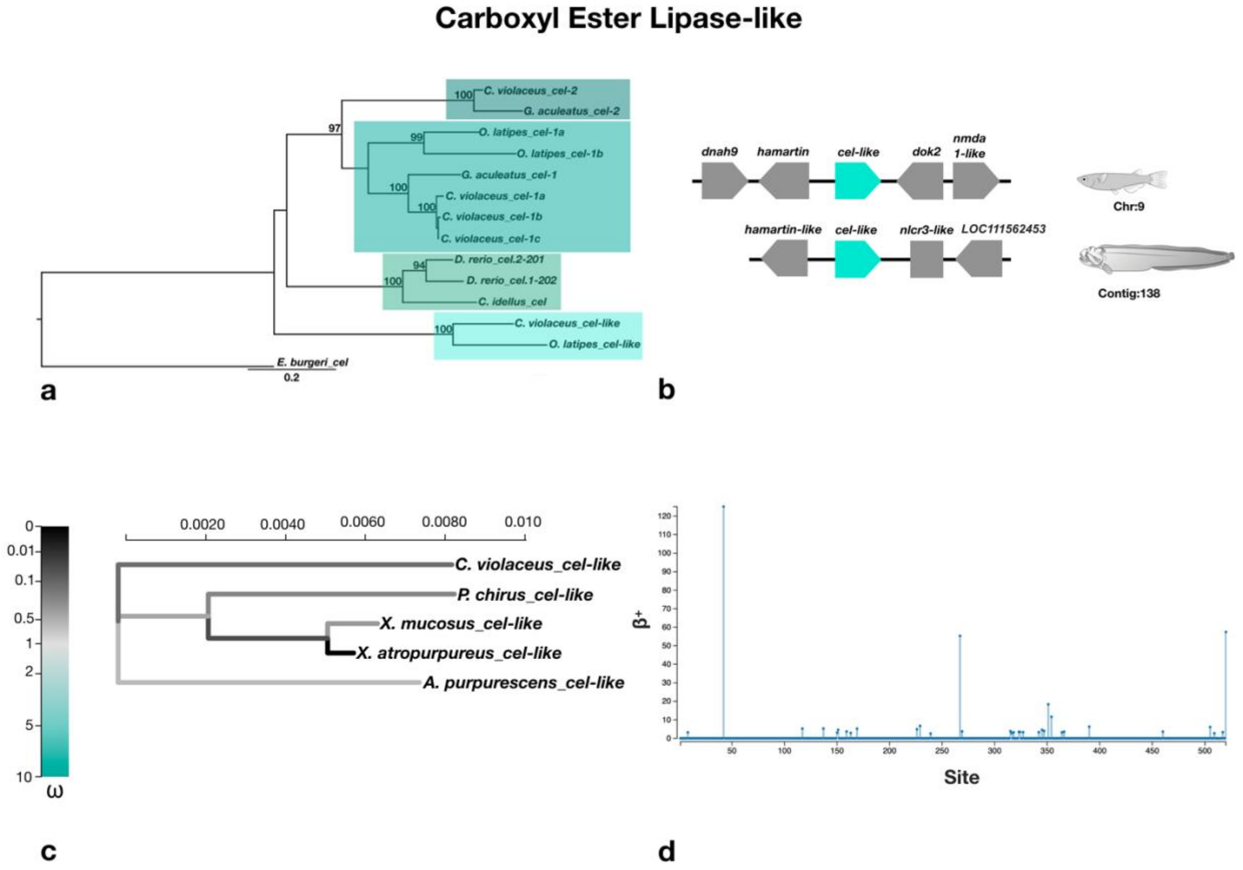


Supplemental Figure 28: Phylogenetic relationships, gene copy number, and molecular evolution of secretory phospholipase Group 12B (*pg12b*) a-b, Synteny map for genes from *Danio rerio*, *Oryzias latipes*, *Gasterosteus aculeatus,* and *C. violaceus*. **c-d**, An adaptive branch-site Random Effects Likelihood (aBSREL) test for episodic diversification phylogenetic tree constructed for *pg12b* genes from *C. violaceus* and other intertidal stichaeid species. Branches thicker than the other branches have a P<0.05 (corrected for multiple comparisons) to reject the null hypothesis of all ω on that branch (neutral or negative selection only). A thick branch is considered to have experienced diversifying positive selection. **e-f**, The output of Mixed Effects Model of Evolution (MEME) to detect episodic positive/diversifying selection at sites. β+ is the non-synonymous substitution rate at a site for the positive/neutral evolution throughout the sequence of the gene. ** is an indication that the positive/diversifying site is statistically significant with a p-value < 0.01.

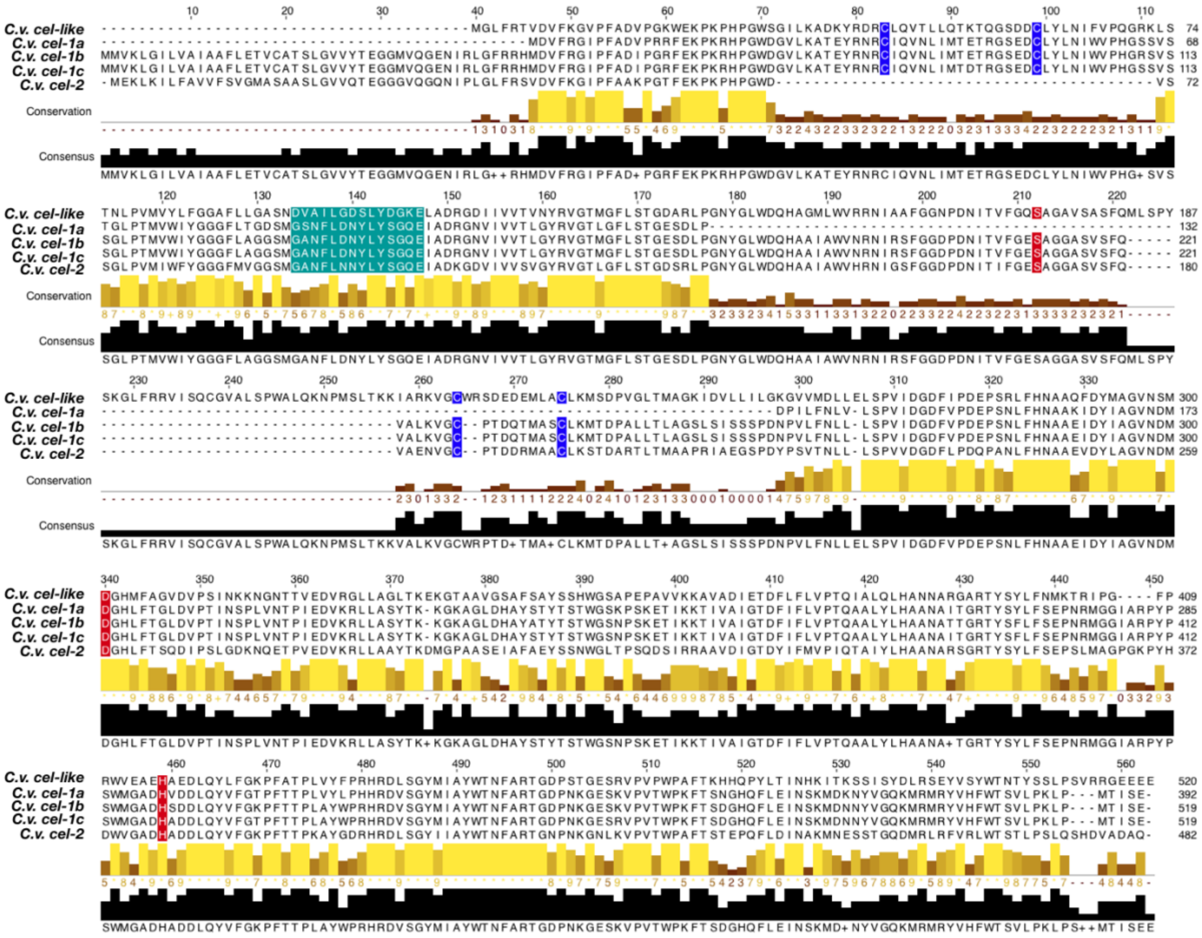


Supplemental Figure S29: Repetitive elements identified adjacent to carboxyl ester lipase (*cel*) loci.

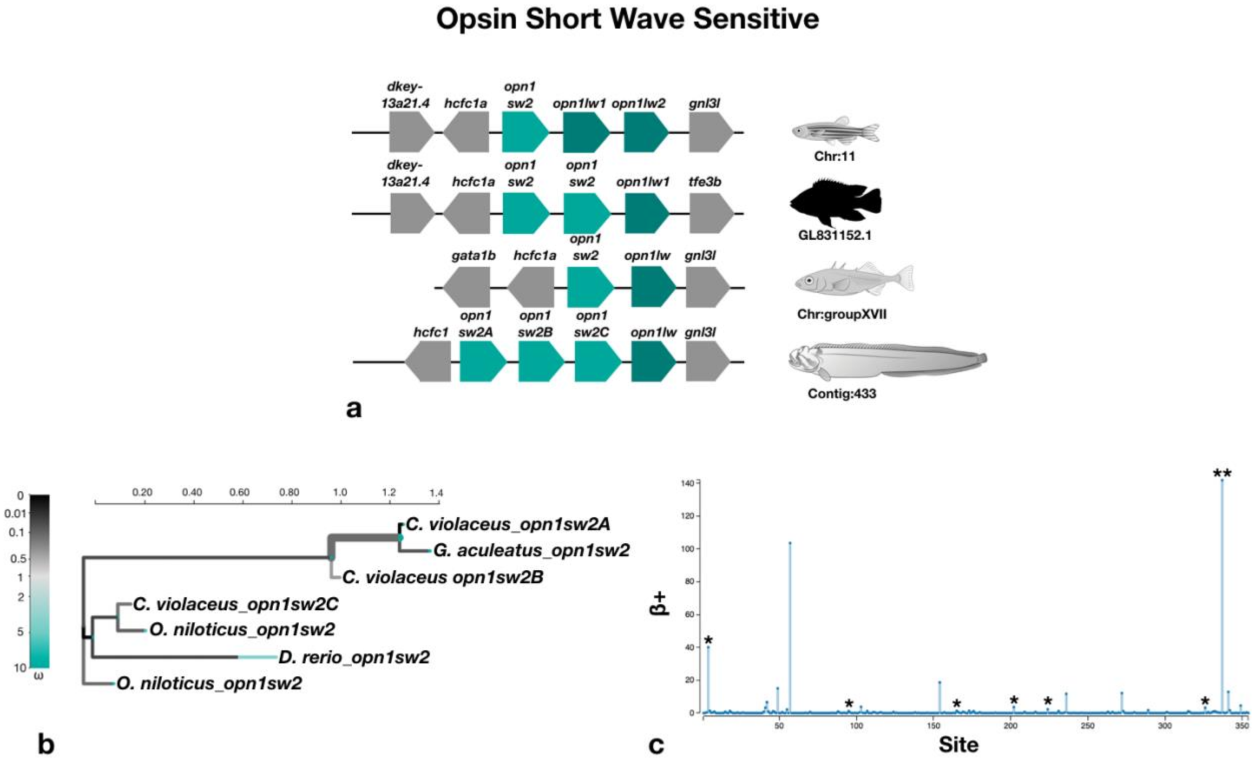
Light green bars are repetitive elements identified on contig 445 and blue bars represent exons of *cel* loci (AUGUSTUS gene prediction).



Supplemental Figure 30: Molecular analyses of carboxyl ester lipase-like genes in stichaeids a, A maximum likelihood (ML) phylogenetic tree of carboxyl ester lipase genes using *Danio rerio*, *Oryzias latipes*, *Gasterosteus aculeatus, Ctenopharyngodon idella, Eptatretus burgeri*, and *C. violaceus.* This tree was constructed with 1,000 bootstrap replicates and a TIM2+I+G model. **b,** gene copy number of *cel-like* loci. **c,** estimation of lineage-specific selection of stichaeids *cel-like* loci using aBSREL. Branches thicker than the other branches have a P<0.05 (corrected for multiple comparisons) to reject the null hypothesis of all ω on that branch (neutral or negative selection only). **d,** estimation of site-level episodic selection in stichaeids *cel-like* loci using MEME**.** There were zero sites under positive/diversifying selection.



Supplemental Figure S31: Alignment of carboxyl ester lipase (*cel*) amino acid sequences. *Cebidichthys violaceus cel* sequences viewed in JALVIEW v2.10.5 (http://www.jalview.org/). Amino acids highlighted in blue indicates disulfide bonds, red indicates active sites present, and teal indicates the bile salt-binding site.



Supplemental Figure S32: Gene copy number and molecular evolution of Opsin Short Wave Sensitive (*opn1sw*) genes. a, Synteny map for *opn1sw* genes from *Danio rerio*, *Oreochromis niloticus*, *Gasterosteus aculeatus*, and *Cebidichthys violaceus*. *D. rerio*, *G. aculeatus*, and *C. violaceus* were drawn by Andrea Dingeldein. *O. niloticus* illustration was drawn by Milton Tan. **b**, An adaptive Branch-Site Random Effects Likelihood (aBSREL) test for episodic diversification was estimated and represented as a phylogenetic tree for *opn1sw* genes from *C. violaceus* and three other fishes. Branches thicker than the other branches have a P<0.05 (corrected for multiple testing) to reject the null hypothesis of all ω on that branch (neutral or negative selection only). A thick branch is considered to have experienced diversifying positive selection. c, The output of Mixed Efffects Model of Evolution (MEME) to detect episodic positive/diversifying selection at sites. β+ is the non-synonymous substitution rate at a site for the positive/neutral evolution throughout the sequence of the gene. ** is an indication that the positive/diversifying site is statistically significant with a p-value < 0.01 and * is for p-value < 0.05.

Supplemental Table S1. Fish genomes available online, including all relevant data about each submission. Natural diets indiciated by color with red being carnivore, purple omnivore, and green herbivore.

Organism	Name	Link	Submitter	Date	Genome representation	Assembly level	Version status	RefSeq category	Total sequence length	Total assembly gap length	Gaps between scaffolds	Number of scaffolds	Scaffold N50	Scaffold L50	Number of contigs	Contig N50	Contig L50	Total number of chromosomes and plasmids	Number of component sequences (WGS or clone)
Xiphophorus maculatus (southern platyfish)	X_maculatus-5.0-male	https://www.ncbi.nlm.nih.gov/assembly/?term=X_maculatus-5.0-male	The Genome Institute, Washington University at St. Louis	12/07/2017	full	Chromosome	latest	representative genome	704,321,165	3,327,905	0	102	31,535,491	11	258	9,181,372	25	25	102
Astatotilapia calliptera (eastern happy)	fAstCal1.1	https://www.ncbi.nlm.nih.gov/assembly/?term=fAstCal1.1	SC	01/17/2018	full	Scaffold	latest	representative genome	883,159,802	1,319,763	0	338	12,523,454	19	733	4,533,513	56	0	338
Oryzias latipes (Japanese medaka)	ASM223469v1	https://www.ncbi.nlm.nih.gov/assembly/?term=ASM223469v1	The University of Tokyo	07/27/2017	full	Chromosome	latest	na	744,414,398	318,000	0	24	32,853,055	11	342	3,516,609	67	24	24
Oreochromis niloticus (Nile tilapia)	ASM185804v2	https://www.ncbi.nlm.nih.gov/assembly/?term=ASM185804v2	University of Maryland	11/21/16	full	Chromosome	latest	representative genome	1,009,856,516	4,230,000	0	2,567	37,007,722	12	2,990	3,090,215	93	24	2,567
Oryzias latipes (Japanese medaka)	ASM223467v1	https://www.ncbi.nlm.nih.gov/assembly/?term=ASM223467v1	The University of Tokyo	07/27/2017	full	Chromosome	latest	representative genome	734,057,086	491,000	0	25	31,218,526	11	516	2,530,934	88	25	25
Lates calcarifer (barramundi perch)	ASB_FALCONassembly_v1	https://www.ncbi.nlm.nih.gov/assembly/?term=ASB_FALCONassembly_v1	TEMASEK LIFE SCIENCES LABORATORY	03/02/2017	full	Contig	latest	na	605,008,470	0					1,422	1,921,842	78	0	1,422
Astyanax mexicanus (Mexican tetra)	Astyanax_mexicanus-2.0	https://www.ncbi.nlm.nih.gov/assembly/?term=Astyanax_mexicanus-2.0	Washington University School of Medince	09/25/2017	full	Chromosome	latest	representative genome	1,335,239,194	43,642,763	0	2,415	35,377,769	16	3,030	1,767,240	198	25	2,415
Oryzias latipes (Japanese medaka)	ASM223471v1	https://www.ncbi.nlm.nih.gov/assembly/?term=ASM223471v1	The University of Tokyo	07/27/2017	full	Chromosome	latest	na	677,633,405	717,000	0	24	28,873,095	11	741	1,450,098	134	24	24
Maylandia zebra (zebra mbuna)	M_zebra_UMD2	https://www.ncbi.nlm.nih.gov/assembly/?term=M_zebra_UMD2	University of Maryland	01/10/2018	full	Chromosome	latest	representative genome	957,200,631	64,100	0	1,690	32,647,892	13	2,331	1,407,087	189	23	1,690
Lates calcarifer (barramundi perch)	ASB_HGAPassembly_v1	https://www.ncbi.nlm.nih.gov/assembly/?term=ASB_HGAPassembly_v1	TEMASEK LIFE SCIENCES LABORATORY	03/02/2017	full	Scaffold	latest	na	668,464,831	11,462	0	3,807	1,191,366	119	3,917	1,066,117	139	0	3,807
Lates calcarifer (barramundi perch)	ASM164080v1	https://www.ncbi.nlm.nih.gov/assembly/?term=ASM164080v1	Temasek Life Sciences Laboratory	05/09/2016	full	Scaffold	latest	representative genome	668,481,366	11,462	0	3,808	1,191,366	119	3,918	1,066,117	139	1	3,808
Seriola quinqueradiata (Japanese amberjack)	Squ_2.0	https://www.ncbi.nlm.nih.gov/assembly/?term=Squ_2.0	Natinal Research Insititute of Aquaculture, FRA	07/05/2017	full	Scaffold	latest	representative genome	639,269,536	92,800	0	384	5,610,255	28	1,312	872,227	239	0	384
Danio rerio (zebrafish)	GRCz11	https://www.ncbi.nlm.nih.gov/assembly/?term=GRCz11	Genome Reference Consortium	05/09/2017	full	Chromosome	latest	reference genome	1,679,203,469	4,691,662	925	2,848	4,737,936	71	20,680	854,399	358	26	
Seriola rivoliana (Almaco jack)	ASM299450v1	https://www.ncbi.nlm.nih.gov/assembly/?term=ASM299450v1	Iowa State University	03/12/2018	full	Scaffold	latest	representative genome	666,141,578	4,357,646	0	1,343	9,509,606	23	3,939	740,108	263	1	1,343
Labrus bergylta (ballan wrasse)	BallGen_V1	https://www.ncbi.nlm.nih.gov/assembly/?term=BallGen_V1	National Institute of nutrition and seafood research	05/03/2016	full	Scaffold	latest	representative genome	805,480,521	137,791	0	13,466	794,648	252	13,723	703,847	282	0	13,466
Symphodus melops (corkwing wrasse)	ASM281910v1	https://www.ncbi.nlm.nih.gov/assembly/?term=ASM281910v1	University of Agder	12/05/2017	full	Contig	latest	representative genome	614,569,099	0					5,060	461,652	343	0	
Amphiprion ocellaris (clown anemonefish)	AmpOce1.0	https://www.ncbi.nlm.nih.gov/assembly/?term=AmpOce1.0	Deakin University	11/16/17	full	Scaffold	latest	representative genome	880,720,895	545,170	0	6,405	401,715	631	7,803	324,210	756	1	6,405
Seriola dumerili (greater amberjack)	Sdu_1.0	https://www.ncbi.nlm.nih.gov/assembly/?term=Sdu_1.0	Genomic breeding group, Research Center for Aquatic Breeding, National Research Institute of aquaculture, FRA	08/03/2017	full	Scaffold	latest	representative genome	677,686,174	5,558,255	0	34,656	5,812,906	29	41,188	249,509	771	1	34,656
Oncorhynchus tshawytscha (Chinook salmon)	Otsh_v1.0	https://www.ncbi.nlm.nih.gov/assembly/?term=Otsh_v1.0	Fisheries and Oceans Canada	01/16/2018	full	Chromosome	latest	representative genome	2,425,713,975	66,667,345	1,916	15,946	1,728,323	336	69,485	133,169	4,051	35	15,946
Larimichthys crocea (large yellow croaker)	Larimichthys_crocea_chromosome_1.0	https://www.ncbi.nlm.nih.gov/assembly/?term=Larimichthys_crocea_chromosome_1.0	JIMEI UNIVERSITY	01/18/2018	full	Chromosome	latest	na	689,173,177	2,809,384	0	599	20,042,208	14	9,930	130,633	1,616	24	599
Esox lucius (northern pike)	Eluc_V3	https://www.ncbi.nlm.nih.gov/assembly/?term=Eluc_V3	Ben F Koop, Jong S Leong	01/18/2017	full	Chromosome	latest	representative genome	904,497,253	12,103,098	192	1,211	7,945,253	29	18,750	125,635	1,991	26	1,211
Gasterosteus aculeatus (three-spined stickleback)	ASM18067v1	https://www.ncbi.nlm.nih.gov/assembly/?term=ASM18067v1		02/01/2006	full	Contig	latest	representative genome	446,610,788	0					16,957	83,204	1,459	0	16,957
Miichthys miiuy (Mi-iuy croaker)	ASM159371v1	https://www.ncbi.nlm.nih.gov/assembly/?term=ASM159371v1	Zhejiang Ocean University	03/17/2016	full	Scaffold	latest	representative genome	619,300,777	25,190,473	0	6,294	1,145,539	95	20,386	81,271	1,741	0	6,294
Melanogrammus aeglefinus (haddock)	MAEA_1	https://www.ncbi.nlm.nih.gov/assembly/?term=MAEA_1	CEES	03/19/2018	full	Scaffold	latest	representative genome	652,790,733	17,210,136	0	8,420	209,126	880	15,188	77,605	2,164	0	
Ictalurus punctatus (channel catfish)	IpCoco_1.2	https://www.ncbi.nlm.nih.gov/assembly/?term=IpCoco_1.2	Auburn University	06/21/2016	full	Chromosome	latest	representative genome	783,274,721	11,381,915	634	9,975	7,726,806	31	34,544	77,201	2,838	30	
Maccullochella peelii (Murray cod)	mcod_v1	https://www.ncbi.nlm.nih.gov/assembly/?term=mcod_v1	Monash University Malaysia	05/09/2017	full	Scaffold	latest	representative genome	633,241,041	5,672,993	0	18,198	109,974	1,605	31,008	70,439	2,510	0	18,198
Larimichthys crocea (large yellow croaker)	L_crocea_1.0	https://www.ncbi.nlm.nih.gov/assembly/?term=L_crocea_1.0	BGI	04/21/2015	full	Scaffold	latest	representative genome	678,938,134	17,630,028	0	6,014	1,034,540	200	25,556	68,287	2,735	1	25,556
Scleropages formosus (Asian bonytongue)	ASM162424v1	https://www.ncbi.nlm.nih.gov/assembly/?term=ASM162424v1	BGI-SZ	04/20/2016	full	Scaffold	latest	na	746,544,453	9,581,315	0	14,183	1,870,178	115	36,404	63,526	3,378	0	36,404
Scleropages formosus (Asian bonytongue)	ASM162425v1	https://www.ncbi.nlm.nih.gov/assembly/?term=ASM162425v1	BGI-SZ	04/20/2016	full	Scaffold	latest	na	738,407,480	10,148,826	0	10,955	1,625,668	129	33,737	60,472	3,551	0	33,737
Oncorhynchus kisutch (coho salmon)	Okis_V1	https://www.ncbi.nlm.nih.gov/assembly/?term=Okis_V1	University of Victoria	03/06/2017	full	Chromosome	latest	representative genome	2,369,932,239	109,511,248	1,566	22,813	1,266,128	433	97,074	58,118	8,653	31	22,813
Kryptolebias marmoratus (mangrove rivulus)	ASM164957v1	https://www.ncbi.nlm.nih.gov/assembly/?term=ASM164957v1	Sungkyunkwan University	05/22/2016	full	Scaffold	latest	representative genome	680,366,784	38,525,168	0	3,073	2,229,659	80	26,350	57,884	2,931	1	3,073
Pygocentrus nattereri (red-bellied piranha)	Pygocentrus_nattereri-1.0.2	https://www.ncbi.nlm.nih.gov/assembly/?term=Pygocentrus_nattereri-1.0.2	McDonnell Genome Institute	07/12/2016	full	Scaffold	latest	representative genome	1,285,352,492	33,265,717	0	283,518	1,440,044	199	325,620	57,732	5,642	1	
Salmo salar (Atlantic salmon)	ICSASG_v2	https://www.ncbi.nlm.nih.gov/assembly/?term=ICSASG_v2	International Cooperation to Sequence the Atlantic Salmon Genome	06/10/2015	full	Chromosome	latest	representative genome	2,966,890,203	347,999,900	9,418	241,573	1,366,254	350	368,060	57,618	9,869	30	241,573
Poecilia formosa (Amazon molly)	Poecilia_formosa-5.1.2	https://www.ncbi.nlm.nih.gov/assembly/?term=Poecilia_formosa-5.1.2	Aquatic Genome Models	10/28/13	full	Scaffold	latest	representative genome	748,923,461	34,726,196	0	3,985	1,574,226	130	31,058	57,472	3,547	0	31,058
Salvelinus alpinus (Arctic char)	ASM291031v2	https://www.ncbi.nlm.nih.gov/assembly/?term=ASM291031v2	University of Victoria	02/12/2018	full	Chromosome	latest	representative genome	2,169,553,147	146,619,173	1,447	16,702	1,018,695	493	97,014	55,619	8,430	40	16,702
Dicentrarchus labrax (European seabass)	seabass_V1.0	https://www.ncbi.nlm.nih.gov/assembly/?term=seabass_V1.0	MPI-PZ	04/09/2014	full	Scaffold	latest	representative genome	675,917,103	7,655,826	0	25	26,439,989	10	37,781	54,134	3,423	0	37,781
Seriola lalandi dorsalis (yellowtail amberjack)	Sedor1	https://www.ncbi.nlm.nih.gov/assembly/?term=Sedor1	Iowa State University	12/08/2017	full	Scaffold	latest	representative genome	732,509,836	16,134,094	0	99,598	1,269,737	153	133,262	51,346	3,701	1	99,598
Stegastes partitus (bicolor damselfish)	Stegastes_partitus-1.0.2	https://www.ncbi.nlm.nih.gov/assembly/?term=Stegastes_partitus-1.0.2	Aquatic Genome Models	05/13/2014	full	Scaffold	latest	representative genome	800,491,834	50,754,421	0	5,818	411,659	565	42,060	43,010	4,606	0	42,060
Poecilia mexicana (shortfin molly)	P_mexicana-1.0	https://www.ncbi.nlm.nih.gov/assembly/?term=P_mexicana-1.0	The Genome Institute at Washington University School of Medicine (WUGSC)	11/13/15	full	Scaffold	latest	representative genome	801,711,499	121,879,375	0	18,105	275,316	877	50,601	39,840	4,911	0	
Hippocampus comes (tiger tail seahorse)	H_comes_QL1_v1	https://www.ncbi.nlm.nih.gov/assembly/?term=H_comes_QL1_v1	South China Sea Institute of Oceanology, Chinese Academy of Sciences	12/08/2016	full	Scaffold	latest	representative genome	493,775,940	34,408,932	0	37,377	2,034,572	63	60,478	39,546	3,292	1	
Leuciscus waleckii (Amur ide)	Amur ide genome	https://www.ncbi.nlm.nih.gov/assembly/?term=Amur ide genome	CHINESE ACADEMY OF FISHERY SCIENCE	06/28/2016	full	Scaffold	latest	representative genome	752,538,629	14,275,045	0	4,888	21,959,719	14	38,277	38,877	5,488	0	4,888
Paramormyrops kingsleyae (bony fishes)	PKINGS_0.1	https://www.ncbi.nlm.nih.gov/assembly/?term=PKINGS_0.1	Michigan State University	01/10/2018	full	Scaffold	latest	representative genome	799,421,083	63,873,835	0	4,667	1,731,158	134	47,999	37,656	5,450	0	4,667
Paralichthys olivaceus (Japanese flounder)	Flounder_ref_guided_V1.0	https://www.ncbi.nlm.nih.gov/assembly/?term=Flounder_ref_guided_V1.0	Tsinghua University	01/24/2017	full	Scaffold	latest	representative genome	643,911,827	122,144,871	0	9,525	10,546,925	21	32,249	36,690	4,071	1	9,525
Poecilia latipinna (sailfin molly)	P_latipinna-1.0	https://www.ncbi.nlm.nih.gov/assembly/?term=P_latipinna-1.0	The Genome Institute at Washington University School of Medicine (WUGSC)	11/13/15	full	Scaffold	latest	representative genome	815,144,743	135,364,427	0	17,988	279,200	882	54,625	33,278	5,928	0	
Scleropages formosus (Asian bonytongue)	ASM162426v1	https://www.ncbi.nlm.nih.gov/assembly/?term=ASM162426v1	BGI-SZ	04/20/2016	full	Scaffold	latest	representative genome	777,359,276	37,826,889	0	4,819	5,965,175	40	55,899	30,793	7,063	1	55,899
Paralichthys olivaceus (Japanese flounder)	ParOli_1.1	https://www.ncbi.nlm.nih.gov/assembly/?term=ParOli_1.1	Yellow Sea Fisheries Research Institute, CAFS	03/15/2017	full	Chromosome	latest	na	545,775,252	11,528,182	281	7,202	3,817,360	40	38,614	30,544	5,224	24	38,614
Oryzias melastigma (Indian medaka)	Om_v0.7.RACA	https://www.ncbi.nlm.nih.gov/assembly/?term=Om_v0.7.RACA	Sungkyunkwan University	02/07/2018	full	Scaffold	latest	representative genome	779,469,774	41,226,186	0	8,603	23,737,187	14	56,275	30,057	6,936	1	8,603
Oreochromis niloticus (Nile tilapia)	Orenil1.1	https://www.ncbi.nlm.nih.gov/assembly/?term=Orenil1.1	Broad Institute	02/08/2012	full	Chromosome	latest	na	927,679,487	111,611,440	232	5,909	2,766,223	96	77,754	29,493	6,912	22	77,754
Sinocyclocheilus grahami (bony fishes)	SAMN03320097.WGS_v1.1	https://www.ncbi.nlm.nih.gov/assembly/?term=SAMN03320097.WGS_v1.1	BGI, Shenzhen	12/16/15	full	Scaffold	latest	representative genome	1,750,287,761	182,848,512	0	31,277	1,156,368	416	168,074	29,353	15,557	1	168,074
Tetraodon nigroviridis (spotted green pufferfish)	ASM18073v1	https://www.ncbi.nlm.nih.gov/assembly/?term=ASM18073v1	Genoscope	05/14/2004	full	Scaffold	latest	representative genome	342,403,326	30,004,609	0	25,773	734,039	102	41,566	29,054	2,685	0	25,773
Periophthalmus magnuspinnatus (bony fishes)	PM.fa	https://www.ncbi.nlm.nih.gov/assembly/?term=PM.fa	BGI-shenzhen	12/02/2014	full	Scaffold	latest	representative genome	701,696,780	13,607,820	0	26,060	296,161	577	76,770	28,254	6,580	0	76,770
Cynoglossus semilaevis (tongue sole)	Cse_v1.0	https://www.ncbi.nlm.nih.gov/assembly/?term=Cse_v1.0	Beijing Genomics Institute	01/28/2014	full	Chromosome	latest	representative genome	470,199,494	24,157,720	1,538	31,181	509,861	289	62,912	27,008	4,682	23	62,912
Larimichthys crocea (large yellow croaker)	ASM74293v1	https://www.ncbi.nlm.nih.gov/assembly/?term=ASM74293v1	Zhejiang Ocean University	08/22/2014	full	Scaffold	latest	na	648,390,617	29,477,058	0	10,265	495,736	355	51,577	25,717	6,933	0	51,577
Danio rerio (zebrafish)	WGS31	https://www.ncbi.nlm.nih.gov/assembly/?term=WGS31	Wellcome Trust Sanger Institute	05/24/2010	full	Scaffold	latest	na	1,411,763,065	10,851,421	0	32,031	613,723	633	119,179	24,925	16,539	0	119,119
Danio rerio (zebrafish)	Zebrafish Genome Assembly WGS32	https://www.ncbi.nlm.nih.gov/assembly/?term=Zebrafish Genome Assembly WGS32	SC	12/23/15	full	Scaffold	latest	na	1,442,815,812	45,002,846	0	64,482	3,091,434	118	249,050	24,356	14,173	0	249,040
Amphilophus citrinellus (Midas cichlid)	Midas_v5	https://www.ncbi.nlm.nih.gov/assembly/?term=Midas_v5	UNIVERSITY OF KONSTANZ	09/07/2014	full	Scaffold	latest	representative genome	844,902,565	59,780,352	0	6,637	1,216,136	188	67,543	23,456	9,697	0	6,637
Mola mola (ocean sunfish)	ASM169857v1	https://www.ncbi.nlm.nih.gov/assembly/?term=ASM169857v1	BGI-shenzhen	08/08/2016	full	Scaffold	latest	representative genome	639,451,992	13,278,132	0	5,552	8,766,736	19	51,826	23,239	8,277	0	51,826
Pundamilia nyererei (bony fishes)	PunNye1.0	https://www.ncbi.nlm.nih.gov/assembly/?term=PunNye1.0	Broad Institute	12/22/11	full	Scaffold	latest	representative genome	830,133,247	131,372,167	0	7,236	2,525,540	93	68,053	22,622	8,701	0	68,053
Clupea harengus (Atlantic herring)	ASM96633v1	https://www.ncbi.nlm.nih.gov/assembly/?term=ASM96633v1	BGI	03/30/2015	full	Scaffold	latest	representative genome	807,711,962	82,755,438	0	6,915	1,860,920	113	73,682	22,275	8,566	1	73,682
Xiphophorus maculatus (southern platyfish)	Xiphophorus_maculatus-4.4.2	https://www.ncbi.nlm.nih.gov/assembly/?term=Xiphophorus_maculatus-4.4.2	The Genome Institute, Washington University at St. Louis	01/06/2012	full	Scaffold	latest	na	729,662,853	76,849,145	0	20,640	1,303,070	150	67,077	22,273	8,468	1	67,070
Monopterus albus (swamp eel)	M_albus_1.0	https://www.ncbi.nlm.nih.gov/assembly/?term=M_albus_1.0	Wuhan University	01/17/2017	full	Scaffold	latest	representative genome	684,144,148	54,828,134	0	20,622	2,106,322	87	71,879	22,239	8,438	1	
Haplochromis burtoni (Burton's mouthbrooder)	AstBur1.0	https://www.ncbi.nlm.nih.gov/assembly/?term=AstBur1.0	Broad Institute	12/22/11	full	Scaffold	latest	representative genome	831,411,547	132,461,904	0	8,001	1,194,190	181	69,074	21,886	8,930	0	69,074
Coryphaenoides rupestris (roundnose grenadier)	ASM289596v1	https://www.ncbi.nlm.nih.gov/assembly/?term=ASM289596v1	Molecular Ecology Group, Department of Biosciences, Durham University	01/23/2018	full	Scaffold	latest	representative genome	829,208,733	39,093,420	0	47,680	159,738	1,171	82,633	20,848	10,225	0	47,680
Cyprinodon variegatus (sheepshead minnow)	C_variegatus-1.0	https://www.ncbi.nlm.nih.gov/assembly/?term=C_variegatus-1.0	Aquatic Genome Models	10/01/2015	full	Scaffold	latest	representative genome	1,035,184,475	135,643,709	0	9,259	835,301	365	110,959	20,803	11,158	1	
Boleophthalmus pectinirostris (great blue-spotted mudskippe	BP.fa	https://www.ncbi.nlm.nih.gov/assembly/?term=BP.fa	BGI-shenzhen	12/02/2014	full	Scaffold	latest	representative genome	955,752,150	65,847,082	0	16,620	2,375,582	89	108,947	20,437	12,346	1	108,947
Nothobranchius furzeri (turquoise killifish)	Nfu_20140520	https://www.ncbi.nlm.nih.gov/assembly/?term=Nfu_20140520	LEIBNIZ INSTITUTE FOR AGE RESEARCH - FRITZ LIPMANN	12/11/15	full	Chromosome	latest	representative genome	1,242,518,059	385,687,973	116	6,013	15,858,201	25	74,941	19,950	12,601	20	68,911
Seriola quinqueradiata (Japanese amberjack)	Squ_1.0	https://www.ncbi.nlm.nih.gov/assembly/?term=Squ_1.0	Faculty of Bioresources, Mie university	06/12/2017	full	Scaffold	latest	na	861,629,195	22,986,614	0	74,780	791,853	327	191,594	19,672	7,882	0	74,780
Oncorhynchus tshawytscha (Chinook salmon)	CHI06	https://www.ncbi.nlm.nih.gov/assembly/?term=CHI06	Mathomics and Columbia River Inter-Tribal Fish Commission (CRITFC)	12/11/17	full	Chromosome	latest	na	2,362,280,314	181,934,675	12,690	115,115	153,278	3,679	234,121	19,113	28,254	34	234,121
Sinocyclocheilus rhinocerous (bony fishes)	SAMN03320098_v1.1	https://www.ncbi.nlm.nih.gov/assembly/?term=SAMN03320098_v1.1	BGI, Shenzhen	12/14/15	full	Scaffold	latest	representative genome	1,655,786,410	134,307,852	0	164,173	945,738	486	314,963	18,758	23,365	1	
Morone saxatilis (striped sea-bass)	SBDraft1	https://www.ncbi.nlm.nih.gov/assembly/?term=SBDraft1	Striped Bass Genomics	06/16/2016	full	Scaffold	latest	representative genome	585,166,552	2,051,914	0	35,012	29,984	5,488	69,860	17,868	9,291	0	35,012
Notothenia coriiceps (black rockcod)	NC01	https://www.ncbi.nlm.nih.gov/assembly/?term=NC01	Antarctic Fish Genome Project	07/29/2014	full	Scaffold	latest	representative genome	636,613,682	13,263,011	0	38,657	217,655	458	72,571	17,492	9,130	1	72,571
Sinocyclocheilus anshuiensis (bony fishes)	SAMN03320099.WGS_v1.1	https://www.ncbi.nlm.nih.gov/assembly/?term=SAMN03320099.WGS_v1.1	BGI, Shenzhen	12/14/15	full	Scaffold	latest	representative genome	1,632,718,266	119,821,887	0	85,682	1,284,143	364	254,423	17,271	25,856	1	254,423
Periophthalmodon schlosseri (giant mudskipper)	PS.fa	https://www.ncbi.nlm.nih.gov/assembly/?term=PS.fa	BGI-Shenzhen	12/02/2014	full	Scaffold	latest	representative genome	679,761,122	4,785,098	0	46,662	39,308	4,620	85,749	16,946	11,471	0	85,749
Fundulus heteroclitus (mummichog)	Fundulus_heteroclitus-3.0.2	https://www.ncbi.nlm.nih.gov/assembly/?term=Fundulus_heteroclitus-3.0.2	The Genome Institute at Washington University School of Medicine	01/21/2015	full	Scaffold	latest	representative genome	1,021,898,560	89,411,150	0	10,180	1,252,252	221	120,723	16,688	14,866	1	120,723
Acanthochromis polyacanthus (spiny chromis)	ASM210954v1	https://www.ncbi.nlm.nih.gov/assembly/?term=ASM210954v1	King Abdullah University of Science and technology	04/25/2017	full	Scaffold	latest	representative genome	991,584,656	148,733,072	0	30,414	334,400	789	159,493	16,099	13,227	0	30,414
Kryptolebias marmoratus (mangrove rivulus)	ASM166395v1	https://www.ncbi.nlm.nih.gov/assembly/?term=ASM166395v1	Washington State University	06/17/2016	full	Scaffold	latest	na	653,959,781	24,677,176	0	40,757	111,539	1,319	88,360	16,024	10,303	0	40,757
Xiphophorus couchianus (Monterrey platyfish)	Xiphophorus_couchianus-4.0.1	https://www.ncbi.nlm.nih.gov/assembly/?term=Xiphophorus_couchianus-4.0.1	McDonnell Genome Institute - Washington University School of Medicine	11/17/15	full	Scaffold	latest	representative genome	708,396,389	151,639,013	0	25	29,398,440	11	74,316	14,608	11,049	0	74,316
Lates calcarifer (barramundi perch)	ASM101014v1	https://www.ncbi.nlm.nih.gov/assembly/?term=ASM101014v1	James Cook University	05/18/2015	full	Scaffold	latest	na	589,095,062	4,238,362	0	22,775	310,513	528	105,873	14,534	11,729	0	22,775
Oncorhynchus mykiss (rainbow trout)	Omyk_1.0	https://www.ncbi.nlm.nih.gov/assembly/?term=Omyk_1.0	USDA/ARS	06/02/2017	full	Chromosome	latest	representative genome	2,178,999,613	251,515,894	7,839	139,800	1,670,138	259	559,855	13,827	32,576	30	139,800
Sebastes rubrivinctus (flag rockfish)	SRub1.0	https://www.ncbi.nlm.nih.gov/assembly/?term=SRub1.0	University of Southern California	10/22/13	full	Scaffold	latest	representative genome	756,296,653	6,781,638	0	68,206	30,046	7,100	136,109	13,541	15,750	0	136,109
Sebastes nigrocinctus (tiger rockfish)	ASM47523v3	https://www.ncbi.nlm.nih.gov/assembly/?term=ASM47523v3	USC	12/20/16	full	Scaffold	latest	representative genome	746,044,620	87,060,894	0	15,872	116,274	1,779	89,356	13,471	13,686	0	15,872
Neolamprologus brichardi (lyretail cichlid)	NeoBri1.0	https://www.ncbi.nlm.nih.gov/assembly/?term=NeoBri1.0	Broad Institute	12/22/11	full	Scaffold	latest	representative genome	847,910,432	161,986,530	0	9,099	4,430,025	50	118,197	13,047	13,507	1	118,197
Xiphophorus hellerii (green swordtail)	Xiphophorus_hellerii-3.0.1	https://www.ncbi.nlm.nih.gov/assembly/?term=Xiphophorus_hellerii-3.0.1	McDonnell Genome Institute - Washington University School of Medicine	11/13/15	full	Scaffold	latest	representative genome	733,802,474	173,853,923	0	25	29,445,952	11	107,971	11,299	14,325	0	107,971
Micropterus floridanus (Florida bass)	ASM259238v1	https://www.ncbi.nlm.nih.gov/assembly/?term=ASM259238v1	Auburn University	10/19/17	full	Contig	latest	representative genome	1,001,521,525	0					249,768	10,978	23,349	0	
Sebastes aleutianus (rougheye rockfish)	ASM191080v2	https://www.ncbi.nlm.nih.gov/assembly/?term=ASM191080v2	USC	12/20/16	full	Scaffold	latest	representative genome	899,650,391	285,260,829	0	10,489	340,062	703	110,635	10,838	14,984	0	10,489
Oryzias latipes (Japanese medaka)	ASM31367v1	https://www.ncbi.nlm.nih.gov/assembly/?term=ASM31367v1	Medaka genome sequencing project	05/09/2007	full	Chromosome	latest	na	869,801,494	169,431,611	118	7,306	6,078,048	42	134,429	9,628	21,697	24	134,429
Oncorhynchus mykiss (rainbow trout)	AUL_PRJEB4421_v1	https://www.ncbi.nlm.nih.gov/assembly/?term=AUL_PRJEB4421_v1	Genoscope CEA	04/29/2014	full	Chromosome	latest	na	1,877,559,617	419,186,132	120	79,942	383,627	1,014	221,128	9,390	43,966	28	192,415
Scartelaos histophorus (walking goby)	SH.fa	https://www.ncbi.nlm.nih.gov/assembly/?term=SH.fa	BGI-shenzhen	12/02/2014	full	Scaffold	latest	representative genome	695,008,792	6,428,304	0	156,044	15,105	11,573	209,353	8,806	21,136	0	209,353
Nothobranchius furzeri (turquoise killifish)	NotFur1	https://www.ncbi.nlm.nih.gov/assembly/?term=NotFur1	Stanford University	02/20/2015	full	Scaffold	latest	na	1,022,966,143	78,432,621	0	46,717	117,689	2,489	241,887	8,598	32,672	0	
Thunnus orientalis (Pacific bluefin tuna)	Thunnus_orientalis_ver_Ba_1.0	https://www.ncbi.nlm.nih.gov/assembly/?term=Thunnus_orientalis_ver_Ba_1.0	National Research Institute of Fisheries Science	06/27/2013	full	Contig	latest	representative genome	684,497,465	0					133,062	8,235	23,876	0	133,062
Takifugu flavidus (sansaifugu)	version 1 of Takifugu flavidus genome	https://www.ncbi.nlm.nih.gov/assembly/?term=version 1 of Takifugu flavidus genome	IOCAS	05/30/2013	full	Scaffold	latest	representative genome	378,032,400	63,143,527	0	34,332	315,240	297	133,909	8,233	8,829	0	133,909
Austrofundulus limnaeus (bony fishes)	Austrofundulus_limnaeus-1.0	https://www.ncbi.nlm.nih.gov/assembly/?term=Austrofundulus_limnaeus-1.0	Center for Life in Extreme Environments at Portland State University, Portland, OR	07/28/2015	full	Scaffold	latest	representative genome	866,963,281	171,917,903	0	29,785	1,098,383	184	168,369	8,097	24,012	0	168,369
Cyprinus carpio (common carp)	ASM127010v1	https://www.ncbi.nlm.nih.gov/assembly/?term=ASM127010v1	ZF-screens B.V.	08/18/2015	full	Scaffold	latest	na	1,380,095,472	18,433,641	0	80,028	66,838	5,709	427,338	7,872	51,239	0	80,028
Anguilla japonica (Japanese eel)	Ajaponica_ver_D2	https://www.ncbi.nlm.nih.gov/assembly/?term=Ajaponica_ver_D2	Tohoku National Fisheries Research Institute	10/25/17	full	Scaffold	latest	representative genome	1,151,067,796	117,435,651	0	195,366	472,078	534	412,396	7,661	30,923	0	195,366
Pimephales promelas (fathead minnow)	FHM_SOAPdenovo	https://www.ncbi.nlm.nih.gov/assembly/?term=FHM_SOAPdenovo	DuPont	06/10/2014	full	Scaffold	latest	representative genome	1,219,326,373	408,157,010	0	73,057	60,380	5,505	215,176	7,468	28,848	0	73,057
Anguilla rostrata (American eel)	ASM160608v1	https://www.ncbi.nlm.nih.gov/assembly/?term=ASM160608v1	Universite Laval	04/04/2016	full	Scaffold	latest	representative genome	1,413,032,609	222,780,344	0	79,209	86,641	4,171	307,316	7,355	47,986	1	79,209
Anoplopoma fimbria (sablefish)	AnoFim1.0	https://www.ncbi.nlm.nih.gov/assembly/?term=AnoFim1.0	University of Victoria	11/25/13	full	Contig	latest	representative genome	699,326,415	0					208,506	5,156	39,748	0	208,506
Scleropages formosus (Asian bonytongue)	aro_v2	https://www.ncbi.nlm.nih.gov/assembly/?term=aro_v2	Monash University Malaysia	10/13/15	full	Scaffold	latest	na	708,403,365	20,052,501	0	42,110	58,849	3,326	253,924	5,013	40,264	0	42,110
Dicentrarchus labrax (European seabass)	ASM18081v1	https://www.ncbi.nlm.nih.gov/assembly/?term=ASM18081v1	European seabass sequencing consortium	02/12/2010	partial	Contig	latest	na	98,228,070	0					36,166	3,461	9,751	0	36,166
Mora moro (bony fishes)	Mora mora assembly	https://www.ncbi.nlm.nih.gov/assembly/?term=Mora mora assembly	CEES	03/18/2018	full	Scaffold	latest	representative genome	344,961,111	828,204	0	100,621	4,433	21,716	125,652	3,267	30,846	0	100,621
Anguilla japonica (Japanese eel)	japanese_eel_genome_v1_25_oct_2011_japonica_c401b400k25m200_sspacepremiumk3a02n24_extra.final.scaffolds	https://www.ncbi.nlm.nih.gov/assembly/?term=japanese_eel_genome_v1_25_oct_2011_japonica_c401b400k25m200_sspacepremiumk3a02n24_extra.final.scaffolds	ZF-screens B.V.	03/18/2014	full	Scaffold	latest	na	1,151,137,423	126,169,453	0	323,740	52,849	4,463	822,110	3,215	77,948	1	822,110
Pagrus major (red seabream)	Pmaj_1.0	https://www.ncbi.nlm.nih.gov/assembly/?term=Pmaj_1.0	Center for Marine Environmental Studies, Ehime University	12/28/17	full	Scaffold	latest	representative genome	875,465,402	58,815,581	0	886,260	4,644	35,607	1,164,424	2,822	60,234	0	886,260
Danio rerio (zebrafish)	CG2v1.0	https://www.ncbi.nlm.nih.gov/assembly/?term=CG2v1.0	University of Chicago	12/29/15	full	Scaffold	latest	na	1,228,672,559	70,675,980	0	73,493	34,289	9,949	1,136,583	2,585	125,485	0	73,493
Anguilla anguilla (European eel)	Anguilla_anguilla_v1_09_nov_10	https://www.ncbi.nlm.nih.gov/assembly/?term=Anguilla_anguilla_v1_09_nov_10	ZF-screens B.V.	05/20/2014	full	Scaffold	latest	representative genome	1,018,701,900	133,573,415	0	501,148	59,657	2,979	865,467	2,544	84,717	0	501,148
Gadus morhua (Atlantic cod)	GadMor_May2010	https://www.ncbi.nlm.nih.gov/assembly/?term=GadMor_May2010	Genofisk	08/26/2011	full	Scaffold	latest	representative genome	824,311,139	216,281,269	0	427,427	393,166	428	554,869	2,311	64,437	0	554,869
Oryzias latipes (Japanese medaka)	ASM15182v1	https://www.ncbi.nlm.nih.gov/assembly/?term=ASM15182v1	Medaka genome sequencing project	04/26/2007	full	Scaffold	latest	na	662,701,370	77,546,629	0	82,496	25,745	7,701	346,141	2,223	76,136	0	346,141
Cottus rhenanus (bony fishes)	ASM145555v1	https://www.ncbi.nlm.nih.gov/assembly/?term=ASM145555v1	Cold Spring Harbor Laboratory	12/02/2015	full	Scaffold	latest	representative genome	563,609,416	75,741,692	0	164,693	7,249	21,086	490,620	2,129	61,259	0	164,693
Pseudopleuronectes yokohamae (marbled flounder)	Pyoko_1.0	https://www.ncbi.nlm.nih.gov/assembly/?term=Pyoko_1.0	Tohoku university	11/24/14	full	Contig	latest	representative genome	547,831,023	0					525,502	1,994	69,628	0	525,502
Sebastes minor (bony fishes)	ASM191076v2	https://www.ncbi.nlm.nih.gov/assembly/?term=ASM191076v2	USC	12/20/16	full	Scaffold	latest	representative genome	681,652,711	21,325,619	0	166,448	7,676	25,679	812,852	1,901	101,393	0	166,448
Squalius pyrenaicus (bony fishes)	pp_cds_nonRedudant	https://www.ncbi.nlm.nih.gov/assembly/?term=pp_cds_nonRedudant	Centre for Ecology, Evolution and Environmental Changes (cE3c)	07/08/2015	partial	Contig	latest	na	48,139,320	0					40,926	1,710	8,213	0	40,926
Pimephales promelas (fathead minnow)	FHM_SGA	https://www.ncbi.nlm.nih.gov/assembly/?term=FHM_SGA	DuPont	06/11/2014	full	Scaffold	latest	na	957,809,772	144,304,306	0	810,921	15,414	11,492	1,007,221	1,668	106,890	0	810,921
Cyprinodon nevadensis pectoralis (Amargosa pupfish)	ASM77601v1	https://www.ncbi.nlm.nih.gov/assembly/?term=ASM77601v1	University of Colorado, Boulder	11/13/14	full	Scaffold	latest	representative genome	1,011,849,000	178,825,625	0	96,516	83,166	3,383	1,303,755	1,416	135,030	0	96,516
Sebastes steindachneri (bony fishes)	ASM191078v2	https://www.ncbi.nlm.nih.gov/assembly/?term=ASM191078v2	USC	12/20/16	full	Scaffold	latest	representative genome	648,011,071	26,095,549	0	279,232	4,288	43,142	1,089,366	1,311	135,731	0	279,232
Maylandia zebra (zebra mbuna)	ASM15091v1	https://www.ncbi.nlm.nih.gov/assembly/?term=ASM15091v1	Cichlid Genome Consortium	07/09/2008	partial	Scaffold	latest	na	79,168,277	2,206,835	0	65,094	1,359	21,794	88,991	1,211	24,283	0	87,341
Nothobranchius furzeri (turquoise killifish)	ASM18203v2	https://www.ncbi.nlm.nih.gov/assembly/?term=ASM18203v2	Dept. of Genome Analysis	12/16/15	partial	Scaffold	latest	na	113,234,354	1,508	0	119,834	1,191	39,237	119,922	1,191	39,215	0	119,834
Rhamphochromis esox (bony fishes)	ASM15093v1	https://www.ncbi.nlm.nih.gov/assembly/?term=ASM15093v1	Cichlid Genome Consortium	07/09/2008	full	Scaffold	latest	representative genome	71,295,074	1,497,497	0	55,751	1,324	19,496	78,130	1,126	22,210	0	75,444
Labeotropheus fuelleborni (blue mbuna)	ASM15087v1	https://www.ncbi.nlm.nih.gov/assembly/?term=ASM15087v1	Cichlid Genome Consortium	07/09/2008	full	Scaffold	latest	representative genome	70,858,381	1,584,106	0	58,245	1,204	19,680	81,167	1,070	23,041	0	78,388
Mchenga conophoros (bony fishes)	ASM15085v1	https://www.ncbi.nlm.nih.gov/assembly/?term=ASM15085v1	Cichlid Genome Consortium	07/09/2008	full	Scaffold	latest	representative genome	73,425,564	2,057,889	0	61,923	1,329	18,520	85,821	1,066	22,269	0	84,024
Pampus argenteus (silver pomfret)	PamArg1.0	https://www.ncbi.nlm.nih.gov/assembly/?term=PamArg1.0	Kuwait Institute for Scientific Research (KISR)	06/03/2014	full	Scaffold	latest	representative genome	350,448,509	4,502,242	0	298,139	1,586	62,360	532,813	1,001	95,981	0	298,139
Nothobranchius kuhntae (Beira killifish)	ASM17385v1	https://www.ncbi.nlm.nih.gov/assembly/?term=ASM17385v1	Leibniz Institute for Age Research - Fritz Lipmann Institute, Jena, Germany	03/17/2009	partial	Scaffold	latest	na	5,234,607	84	0	5,934	990	1,929	5,941	990	1,928	0	5,934
Nothobranchius furzeri (turquoise killifish)	ASM18221v1	https://www.ncbi.nlm.nih.gov/assembly/?term=ASM18221v1	Leibniz Institute for Age Research - Fritz Lipmann Institute, Jena, Germany	03/17/2009	partial	Scaffold	latest	na	5,252,863	178	0	5,617	985	1,948	5,633	985	1,948	0	5,617
Melanochromis auratus (golden mbuna)	ASM15089v1	https://www.ncbi.nlm.nih.gov/assembly/?term=ASM15089v1	Cichlid Genome Consortium	07/09/2008	full	Scaffold	latest	representative genome	68,238,634	1,761,364	0	63,297	1,063	21,993	86,145	977	24,884	0	84,132
Anguilla japonica (Japanese eel)	Anguilla_japonica_LG1_1.0	https://www.ncbi.nlm.nih.gov/assembly/?term=Anguilla_japonica_LG1_1.0	National Research Institute of Fisheries Science	02/03/2018	full	Contig	latest	na	35,636,560	0					82,081	524	22,596	0	82,081
Cyprinus carpio (common carp)	common carp genome	https://www.ncbi.nlm.nih.gov/assembly/?term=common carp genome	CHINESE ACADEMY OF FISHERY SCIENCE	11/03/2014	full	Chromosome	latest	representative genome	SEE PAPER										
Poecilia reticulata (guppy)	Guppy_female_1.0+MT	https://www.ncbi.nlm.nih.gov/assembly/?term=Guppy_female_1.0+MT	Max Planck Institute for Developmental Biology	04/28/2014	full	Chromosome	latest	representative genome	SEE PAPER										
Takifugu rubripes (torafugu)	FUGU5	https://www.ncbi.nlm.nih.gov/assembly/?term=FUGU5 Synonyms fr3	The Fugu Genome Sequencing Consortium	10/13/11	full	Chromosome	latest	representative genome	SEE PAPER										
Carnivore																			
Omnivore																			
Herbivore																			

Supplemental Table S2: Genome Sequencing Information

Pacific Biosciences	
Number of SMRT Cells	40
Amount Polymerase in picomolar (pM)*	150 (17); 200 (1); 300 (13); 400 (9)
Total Number of Reads	2,421,941
Average N50 (bp)	17,102.78
Total data (Mb)	29,700
Illumina (100 Paired End Sequencing)	
Number of Lanes	2
Number of Reads from the 1st Lane (Both Reads 1 and 2)	422,313,916
Number of Reads from the 2nd Lane (Both Reads 1 and 2)	423,075,242
Total data (Mb)	84,539

* values within parentheses indicate the amount of SMRT cells used

Supplemental Table S3: BUSCO v3 Estimation on the *Cebidichthys violaceus* genome

	BUSCO V3
Complete BUSCOs	2508
Complete BUSCOs and single-copy BUSCOs	2474
Complete BUSCOs and duplicated BUSCOs	34
Fragmented BUSCOs	28
Missing BUSCOs	50
Total BUSCO groups searched	2586

Supplemental Table S4: RepeatMasker for the *C. violaceus* assembled genome

	Number of Elements	Length Occupied (bp)	Percentage of Sequence (%)		
Retroelements	11715	5340689	0.9		
SINEs:	1985	223655	0.04		
Penelope	60	24362	0		
LINEs:	8905	4191047	0.71		
CRE/SLACS	0	0	0		
L2/CR1/Rex	4980	2049678	0.35		
R1/LOA/Jockey	0	0	0		
R2/R4/NeSL	85	24828	0		
RTE/Bov-B	3316	1768649	0.3		
L1/CIN4	336	205561	0.03		
LTR elements:	825	925987	0.16		
BEL/Pao	20	36895	0.01		
Ty1/Copia	22	20106	0		
Gypsy/DIRS1	703	846315	0.14		
Retroviral	79	22637	0		
DNA transposons	10562	2582768	0.44		
hobo-Activator	4384	730619	0.12		
Tc1-IS630-Pogo	4812	1690957	0.29		
En-Spm	0	0	0		
MuDR-IS905	0	0	0		
PiggyBac	352	30854	0.01		
Tourist/Harbinger	178	26872	0		
Other (Mirage, P-element, Transib)	0	0	0		
Rolling-circles	0	0	0		
Unclassified:	203	21708	0		
Total interspersed repeats:		7945165	1.34		
Small RNA:	2430	215618	0.04		
Satellites:	4	1112	0		
Simple repeats:	468887	27486227	4.64		
Low complexity:	37947	2434577	0.41		

Supplemental Table S5: Annotation of the 30 loci used for the phylogeny in figure 1

Ortholog cluster ID	Annotation	Alignment Length
5	amyloid beta precursor protein (cytoplasmic tail) binding protein 2 (APPBP2)	1767
12	anaphase promoting complex subunit 7 (ANAPC7)	1806
25	aminoadipate-semialdehyde dehydrogenase (AASDH)	3860
39	tyrosyl-tRNA synthetase (YARS)	1634
123	PRP18 pre-mRNA processing factor 18 homolog (prpf18)	1083
131	sphingomyelin phosphodiesterase 2, neutral membrane (neutral sphingomyelinase) (SMPD2)	1456
141	UbiA prenyltransferase domain containing 1 (UBIAD1)	1086
163	dual oxidase 1-like	5184
187	phosphatidylinositol glycan anchor biosynthesis class M (pigm)	1296
191	methyltransferase like 9 (METTL9)	1194
198	chromosome 22 C6orf62 homolog (c22h6orf62)	701
214	uncharacterized protein F13E9.13, mitochondrial-like	875
218	mediator complex subunit 7 (MED7)	831
220	transmembrane protein 98 (TMEM98)	687
223	prolactin regulatory element binding (preb)	1520
230	Shwachman-Bodian-Diamond syndrome (SBDS)	783
235	ATP synthase, H+ transporting, mitochondrial F1 complex, O subunit (ATP5O)	639
236	sodium channel modifier 1 (SCNM1)	906
241	cysteine dioxygenase type 1 (CDO1)	787
242	GINS complex subunit 2 (Psf2 homolog) (GINS2)	688
252	mediator complex subunit 22 (MED22)	686
269	HD domain containing 3 (HDDC3)	550
271	mago homolog, exon junction complex subunit (MAGOH)	447
281	optic atrophy 3 (autosomal recessive, with chorea and spastic paraplegia) (OPA3)	495
288	splicing factor 3b subunit 6 (sf3b6)	378
290	polymerase (RNA) II (DNA directed) polypeptide F (POLR2F)	503
301	mitochondrial ribosomal protein S17 (mrps17)	489
311	C1D nuclear receptor corepressor (c1d)	484
315	replication protein A3, 14kDa (RPA3)	369
320	NADH:ubiquinone oxidoreductase subunit S5 (ndufs5)	324

Supplemental Table S6: Pairwise Comparison of Orthologs of *Cebidichthys violaceus* and fish genomes deposited on Ensembl

	Latimeria chalumnae	*Lepisosteus oculatus*	*Danio rerio*	*Astyanax mexicanus*	*Gadus morhua*	*Takifugu rubripes*	*Tetraodon nigroviridis*	*Gasterosteus aculeatus*	*Cebidichthys violaceus*	*Xiphophorus maculatus*	*Poecilia formosa*	*Oryzias latipes*	*Oreochromis niloticus*
Latimeria chalumnae													
Lepisosteus oculatus	12563												
Danio rerio	12374	7237											
Astyanax mexicanus	11795	13247	16214										
Gadus morhua	9052	110	1594	12385									
Takifugu rubripes	11346	6281	14060	2246	13671								
Tetraodon nigroviridis	10745	11808	13207	1399	12851	15335							
Gasterosteus aculeatus	4717	12652	12624	13279	6725	15429	2245						
Cebidichthys violaceus	9472	10483	5357	11000	7886	11805	10965	12022					
Xiphophorus maculatus	11730	13157	14824	13864	14081	2059	14318	15867	12249				
Poecilia formosa	12271	13926	4201	14610	14478	6240	14865	8343	13241	18391			
Oryzias latipes	10795	11868	3796	12418	13040	14465	5278	14674	11024	14610	15367		
Oreochromis niloticus	11858	4274	7487	7576	14102	15732	14502	15897	3913	5828	17602	14761	

Supplemental Table S7: Genome Size Estimation with Jellyfish v2.2.0 (Marçais and

Kmer Size	Peak	Genome Estimate Size
31	45	657,524,901
29	46	661,608,906
27	48	651,654,143
25	49	655,607,918
Average Genome		**656,598,967**
Standard Deviation		4,138,853

Supplemental Table S8: Transcriptomic Sequencing Information and Trinity Assembly

	TUBE ID	RIN Values	Barcode ID	Number of Reads 1st lane	Number of Reads 2nd lane	Trimmomatic	Normalized	Mapped (Overall)	Aligned Pairs	Transcripts Present
Liver	CV100	8.6	ACAGTG	7,954,337	8,119,254	14,705,062	1,330,380	71.5%	849,686	20,008
Brain	CV98	9.4	GCCAAT	12,832,744	13,197,868	25,319,484	7,389,717	83.7%	5,711,530	60,430
Heart	CV97,98,99	8.5	CAGATC	17,371,046	17,402,450	33,817,846	3,961,155	84.8%	3,129,930	35,570
Gill	CV96	8.4	CGATGT	12,365,345	11,647,444	22,903,000	4,518,995	80.5%	3,354,450	78,629
Pyloric Caeca	CV96	8.8	CTTGTA	16,098,871	18,297,873	33,400,321	4,396,658	83.3%	3,400,229	40,201
Proximal Intestine	CV97,98	9.7	TGACCA	10,420,112	9,421,382	15,949,732	2,621,189	69.7%	1,602,595	37,277
Middle Intestine	CV97,98	8.6	AGTTCC	16,528,911	16,500,293	31,957,043	4,805,496	83.8%	3,740,247	41,978
Spleen	CV97,98,99,100	8.2	ATGTCA	31,903,535	39,356,506	69,076,597	6,354,519	81.5%	4,796,120	48,270
Gonad (Testes)	CV99,100	9.7	AGTCAA	14,000,194	16,098,169	29,305,941	9,061,164	67.4%	5,640,612	60,487

Supplemental Table S9: RepeatMasker for the Genome Guided (TRINITY) Transcriptome

	Number of Elements	Length Occupied (bp)	Percentage of Sequence (%)	
Retroelements	345	30165	0.03%	
SINEs	18	1278	0.00%	
Penelope	2	92	0.00%	
LINEs	164	14962	0.01%	
L2/CR1/Rex	104	9935	0.01%	
R1/LOA/Jockey	8	789	0.00%	
R2/R4/NeSL	4	394	0.00%	
RTE/Bov-B	7	613	0.00%	
L1/CIN4	32	2704	0.00%	
LTR elements	163	13925	0.01%	
BEL/Pao	14	733	0.00%	
Ty1/Copia	0	0	0.00%	
Gypsy/DIRS1	110	9224	0.01%	
Retroviral	25	2683	0.00%	
DNA transposons	513	36648	0.03%	
hobo-Activator	180	12759	0.01%	
Tc1-IS630-Pogo	27	1952	0.00%	
PiggyBac	4	421	0.00%	
Tourist/Harbinger	23	2569	0.00%	
Other (Mirage, P-element, Transib)	0	0	0.00%	
Total interspersed repeats		75248	0.07%	
Small RNA	3	264	0.00%	
Satellites	13	948	0.00%	
Simple repeats	69	6138	0.01%	
Low complexity	10	1450	0.00%	

Supplemental Table S10: RepeatMasker for AUGUSTUS Predicted Genes

	Number of Elements	Length Occupied (bp)	Percentage of Sequence (%)	
Retroelements	451	82999	0.19%	
SINEs	13	827	0.00%	
Penelope	9	2950	0.01%	
LINEs	211	36107	0.08%	
L2/CR1/Rex	128	15756	0.04%	
R1/LOA/Jockey	10	813	0%	
R2/R4/NeSL	5	1512	0%	
RTE/Bov-B	12	1606	0%	
L1/CIN4	22	7877	0.02%	
LTR elements	227	46065	0.1%	
BEL/Pao	11	1225	0%	
Ty1/Copia	19	4233	0.01%	
Gypsy/DIRS1	102	18320	0.04%	
Retroviral	80	19852	0.04%	
DNA transposons	877	123135	0.28%	
hobo-Activator	431	65743	0.15%	
Tc1-IS630-Pogo	84	12063	0.03%	
PiggyBac	36	6698	0.02%	
Tourist/Harbinger	31	4641	0.01%	
Other (Mirage, P-element, Transib)	166	13032	0.03%	
Total interspersed repeats		240853	0.55%	
Small RNA	4	250	0%	
Satellites	33	16963	0.04%	
Simple repeats	10737	482867	1.09%	
Low complexity	3249	206240	0.47%	

Supplemental Table S11: Estimation of Total Genes from Cuffmerge and Augustus

Total Estimated Transcripts from Cuffmerge	101,922			
Trinotate Annotation	65,535			
Top BLASTX hit	26,356			
Total Estimates from Augustus (de novo)	29,525			
80% Hit Coverage from Cuffmerge Assembly Uniprot	5,199			

Supplemental Table S12: Estimation of Full Length of Transcripts from all nine transcriptomes

Hit percent coverage bin	Count in bin	>Bin below
100	2692	2692
90	1258	3950
80	1249	5199
70	1358	6557
60	1598	8155
50	1991	10146
40	2411	12557
30	2912	15469
20	2966	18435
10	1132	19567

Supplemental Table S13: Candidate Genes (Digestive Enzymes) Selection Analyses

Gene Name	Contig Location	Alignment Length (Nucleotide)	Alignment Length (AA)	% of Swissprot Annotation Hit	Gard (sites)	Absrel	Meme (sites)
Aminopeptidase A (*anpepa*)	86	2127	709	73.40%	111, 1039	No evidence	194*,412**,445**,593**
Aminopeptidase B (*anpepb*)	3	1497	499	51.34%	801	No evidence	156*
Aminopeptidase Ey (*anpep Ey*)	3	1500	500	51.71%	No recombination	No evidence	No Sites
Aminopeptidase Ey-like (*anpep Ey-like*)	78	927	358	37.02%	309	1 branch	38*
Aminopeptidase N (*anpep N*)	78	1449	483	50.00%	616	No evidence	351**
Amylase (*amy2a* and *amy2b*)	440	1536	512	100.79%	No recombination	1 branch	41*, 256*, 279*
Carboxyl Ester Lipase 1 (*cel-1a,b,c*)	445	1530	510	85.43%	No recombination	No evidence	64**, 258**, 355**
Carboxyl Ester Lipase 2 (*cel-2*)	445	1053	351	58.79%	241, 303, 975	No evidence	No sites
Carboxyl Ester Lipase-like (*cel-like*)	138	1560	520	87.10%	No recombination	No evidence	No sites
Chymotrypsin A (*ctra-1 & ctra-2*)	55	789	263	100.00%	99	No evidence	No sites
Chymotrypsin B (*ctrb*)	55	792	264	107.76%	81, 336, 581	2 branches	112**
Chymotrypsin-like (*ctrl*)	427	789	263	99.62%	No recombination	No evidence	No sites
Phospholipase B1 (*plb1-1*)	442	2697	899	60.99%	202	No evidence	66**, 97*, 289*, 438**, 800*, 821*
Phospholipase B1 (*plb1-2*)	434	2004	668	45.32%	532	1 branch	183**, 230*,476*
Phospholipase B1 (*plb1-3*)	356	582	194	13.16%	95, 124, 282	No evidence	No sites
Phospholipase B12 (*pg12b-1*)	428	606	202	103.59%	No recombination	No evidence	No sites
Phospholipase B12 (*pg12b-2*)	413	582	194	99.49%	95, 124, 282	No evidence	25**,31**,32**,33**,37**,83**,142**,146**,185**
Trypsin-3_1 (*try3-1*)	427	750	250	105.04%	21	No evidence	No sites
Trypsin-3_2 (*try3-2*)	435	732	244	102.52%	141	No evidence	91*

** is an indication that the positive/diversifying site is statistically significant with a p-value < 0.01 and * is for p-value < 0.05.

Supplementary Table S14: Ensembl IDs used for *anpep* and *cel*

Species Name	Gene Name	Ensembl ID
Danio rerio	*anpepb*	ENSDARG00000103878
Oryzias latipes	*anpepb*	ENSORLG00020014580
Gasterosteus aculeatus	*anpepb*	ENSGACG00000014140
Danio rerio	*si:ch211*	ENSDARG00000097285
Oryzias latipes	*anpep Ey*	ENSORLG00020014549
Danio rerio	*anpep*	ENSDARG00000089706
Oryzias latipes	*anpep N*	ENSORLG00000014691
Gasterosteus aculeatus	*anpep-201*	ENSGACG00000014748
Oryzias latipes	*anpep Ey-like*	ENSORLG00000029229
Gasterosteus aculeatus	*anpep-202*	ENSGACG00000014748
Danio rerio	*anpepa*	ENSDARG00000041083
Oryzias latipes	*anpepa*	ENSORLG00000019272
Gasterosteus aculeatus	*anpepa*	ENSGACG00000002363
Petrus marinus	*anpep*	ENSPMAG00000003227
Petrus marinus	*anpep*	ENSPMAG00000009142
Petrus marinus	*anpep*	ENSPMAG00000009172
Danio rerio	*cel.2-201*	ENSDARG00000029822
Danio rerio	*cel.1-202*	ENSDARG00000017490
Oryzias latipes	*cel-1a*	ENSORLG00000014439
Oryzias latipes	*cel-1b*	ENSORLG00000014464
Gasterosteus aculeatus	*cel-2*	ENSGACG00000018127
Gasterosteus aculeatus	*cel-1*	ENSGACG00000018130
Oryzias latipes	*cel-like*	ENSORLG00000016428
Eptatretus burgeri	*cel*	ENSEBUG00000006718
Ctenopharyngodon	*cel*	CI01000006 (scaffold)

*_C. idella cel_ gene was taken from the Grass Carp Genome Database (http://bioinfo.ihb.ac.cn/gcgd).