PROCEEDINGS B

royalsocietypublishing.org/journal/rspb

Research

Cite this article: Heras J, Chakraborty M, Emerson JJ, German DP. 2020 Genomic and biochemical evidence of dietary adaptation in a marine herbivorous fish. *Proc. R. Soc. B* **287**: 20192327.
http://dx.doi.org/10.1098/rspb.2019.2327

Received: 3 October 2019
Accepted: 26 January 2020

Subject Category:
Genetics and genomics

Subject Areas:
evolution, genomics, physiology

Keywords:
digestive physiology, amylase, genome, transcriptome, gene expression, carboxyl ester lipase

Author for correspondence:
Joseph Heras
e-mail: herasj01@gmail.com

Electronic supplementary material is available online at https://doi.org/10.6084/m9.figshare.c.4838775.

THE ROYAL SOCIETY PUBLISHING

Genomic and biochemical evidence of dietary adaptation in a marine herbivorous fish

Joseph Heras, Mahul Chakraborty, J. J. Emerson and Donovan P. German

Department of Ecology & Evolutionary Biology, University of California, Irvine, CA 92697-2525, USA

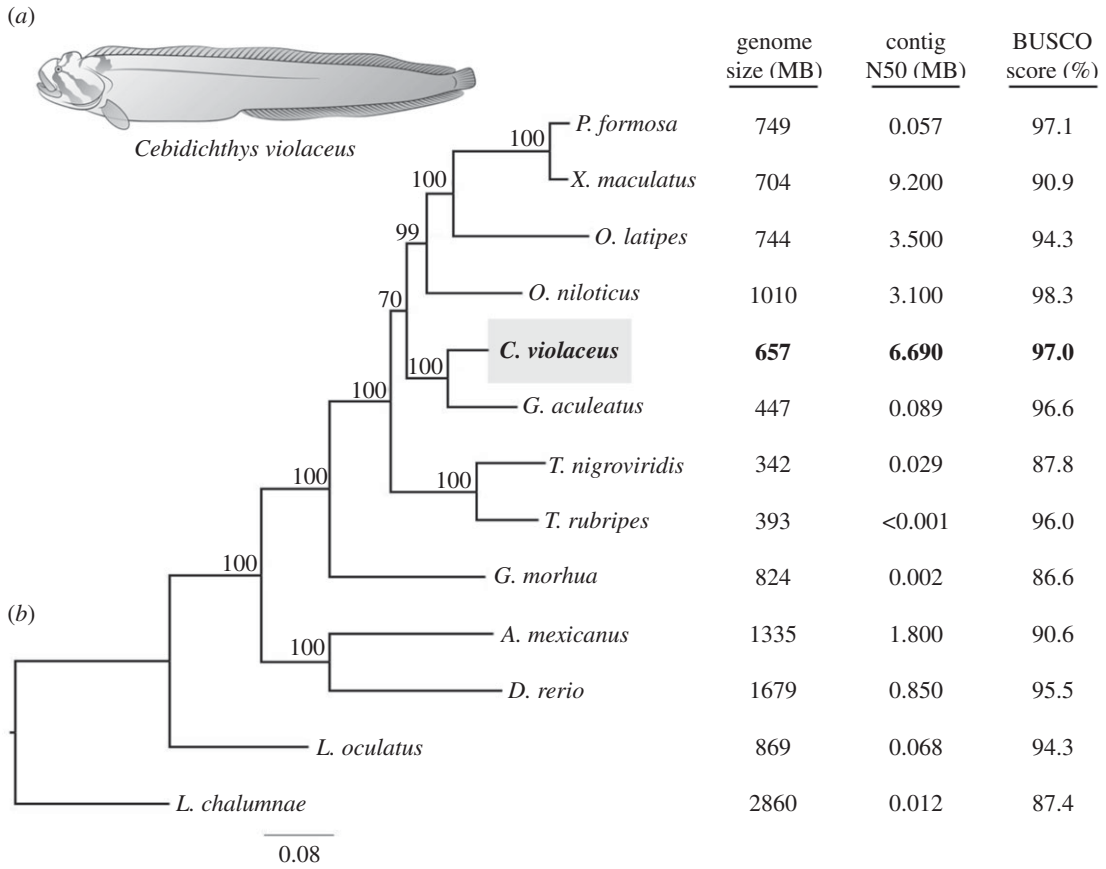
JH, 0000-0002-6462-9313

Adopting a new diet is a significant evolutionary change, and can profoundly affect an animal's physiology, biochemistry, ecology and genome. To study this evolutionary transition, we investigated the physiology and genomics of digestion of a derived herbivorous fish, *Cebidichthys violaceus*. We sequenced and assembled its genome (N50 = 6.7 Mb) and digestive transcriptome, and revealed the molecular changes related to digestive enzymes (carbohydrases, proteases and lipases), finding abundant evidence of molecular adaptation. Specifically, two gene families experienced expansion in copy number and adaptive amino acid substitutions: amylase and carboxyl ester lipase (*cel*), which are involved in the digestion of carbohydrates and lipids, respectively. Both show elevated levels of gene expression and increased enzyme activity. Because carbohydrates are abundant in the prickleback's diet and lipids are rare, these findings suggest that such dietary specialization involves both exploiting abundant resources and scavenging rare ones, especially essential nutrients, like essential fatty acids.

1. Introduction

Populations exposed to new environments often experience strong natural selection [1]. Thus, comparing closely related species has been effective in pinpointing changes that drive adaptation [2]. The comparative method can link genetic variation to molecular phenotypes that in turn change animal performance, revealing for example adaptation to new abiotic factors [3,4], changes in diet [5,6] and exposure to pollution [7]. Animal digestion is an ideal model phenotype because it is central to fitness, is understood in many species at genetic, molecular, biochemical and physiological levels and is variable throughout animal evolution [8–11]. Untangling the genetic basis of digestion is contingent on quality genomic resources [12], which have traditionally been lacking in non-model species [13,14]. Advances in genome technology have improved genomic quality, stimulating the production of genome assemblies and catalysing genetic discoveries in non-model organisms [15,16].

The digestive process requires digestive enzymes [8]. In the field of nutritional physiology, two main hypotheses are invoked to explain digestive enzyme activities in relation to ingested substrate concentrations in the animal's diet (electronic supplementary material, figure S1). The adaptive modulation hypothesis [17] suggests a positive correlation between digestive enzyme activities and ingested quantity of the substrate for those enzymes. Based on economic principles, this hypothesis is well supported in the literature for carbohydrases, as herbivorous and omnivorous animals tend to have elevated amylase activities in their guts [11,18–23], and achieve these activities largely via gene duplications [22]. The nutrient balancing hypothesis [24] suggests that there can be elevated expression of enzymes towards limiting dietary resources to ensure acquisition of essential nutrients, like essential fatty acids. Indeed, elevated carboxyl ester lipase activities are observed in fishes consuming low-lipid, high-fibre foods [11,25,26]. In this investigation, we sequenced the genome of the herbivorous fish *Cebidichthys violaceus*, which revealed extensive genetic variation and adaptive amino acid



	genome size (MB)	contig N50 (MB)	BUSCO score (%)
P. formosa	749	0.057	97.1
X. maculatus	704	9.200	90.9
O. latipes	744	3.500	94.3
O. niloticus	1010	3.100	98.3
C. violaceus	**657**	**6.690**	**97.0**
G. aculeatus	447	0.089	96.6
T. nigroviridis	342	0.029	87.8
T. rubripes	393	<0.001	96.0
G. morhua	824	0.002	86.6
A. mexicanus	1335	1.800	90.6
D. rerio	1679	0.850	95.5
L. oculatus	869	0.068	94.3
L. chalumnae	2860	0.012	87.4

Figure 1. Phylogenetic relationships from 13 fishes with sequenced genomes including *Cebidichthys violaceus*. (*a*) Illustration of *Cebidichthys violaceus*. (*b*) A maximum-likelihood (ML) tree constructed with 1000 bootstrap replicates in PHYML v3.1 using alignments of concatenated 30 protein coding genes from 13 fish species. *Cebidichthys violaceus* is highlighted in the grey box. Genomes from the following 12 species were acquired from ENSEMBL (release 91): *Poecilia formosa*, *Xiphophorus maculatus*, *Oryzias latipes*, *Oreochromis niloticus*, *Gasterosteus aculeatus*, *Tetraodon nigroviridis*, *Takifugu rubripes*, *Gadus morhua*, *Astyanax mexicanus*, *Danio rerio*, *Lepisosteus oculatus* and *Latimeria chalumnae*. Genome size and contig N50, and BUSCO v3 complete genes identified out of 2586 BUSCO groups are represented for all taxa.

variation for amylase and carboxyl ester lipase, suggesting multiple mechanisms underlying the novel derived dietary physiology in *C. violaceus*, and simultaneously supporting the adaptive modulation and nutrient balancing hypotheses in the same organism.

We generated a physiological genomics dataset for this non-model species, including a highly contiguous and complete genome, the transcriptomes of digestive and hepatic tissues, and digestive enzyme activity levels. The *C. violaceus* genome had a N50, a measure of genome assembly quality, of 6.7 MB, placing it among the most contiguous teleost assemblies (figure 1). The ecological and evolutionary positions of *C. violaceus* makes these resources important for unravelling the acquisition of herbivory. Members of the family Stichaeidae, including *C. violaceus*, have independently invaded the intertidal zone multiple times, and in two cases began consuming significant amounts of algae (electronic supplementary material, figure S2) [9], a diet low in protein and lipids and rich in fibrous cell walls and soluble carbohydrates [10,11].

Herbivory is poorly represented among high quality teleost genomes (figure 1; electronic supplementary material, table S1). Because teleosts are so speciose, they represent a large number of independent acquisitions of herbivory, even though only 5% of teleosts are considered nominally herbivorous (cf. 25% for mammals [27]). Among herbivorous fishes, most do not specialize on algal thalli like *C. violaceus* does [28]. Moreover, *C. violaceus* digests algae with the aid of microbial symbionts in their hindguts, as evidenced by elevated levels of short chain fatty acids, or SCFAs, in this gut region (electronic supplementary material, figure S2; [11]). This microbial symbiosis is analogous to other highly specialized vertebrate herbivores like lagomorphs or rodents. Thus *C. violaceus* offers a unique opportunity to study extremes of dietary specialization and provide a link between the genome and the digestive physiology of this organism.

As a resource for genomic annotation, and to better understand their metabolism, we assembled the transcriptomes of nine tissues from *C. violaceus*: gill, heart, spleen, pyloric caeca, proximal intestine, middle intestine, brain, gonad (testes) and liver. We examined the *C. violaceus* genome for gene copy number of candidate genes that encode for digestive enzymes. We also used these genes along with orthologous sequences from the transcriptomes of other stichaeid species in order to estimate episodic diversifying selection. Lastly, we compared syntenic regions of the *C. violaceus* genome to other teleosts to gain a better understanding of the evolution of digestive enzymes. Overall, this study provides a detailed understanding of the evolutionary processes of dietary specialization that has occurred within this group of marine fishes and can lead to hypothesis formation regarding the evolution of dietary specialization in general.

2. Methods

For each section, more detailed methods can be found in the Supplemental Methods in the electronic supplementary material (see electronic supplementary material).

(a) Collection and preparation

One individual of *Cebidichthys violaceus* (156 mm standard length) was collected in May 2015 from San Simeon, California (35.6525° N, 121.2417° W). The individual was euthanized in MS-222 (1 g l^{-1}), dissected to remove internal organs, decapitated and preserved in liquid nitrogen. All handling of fish from capture to euthanization was conducted under approved protocol 2011-2989 of the Institutional Animal Care and Use Committee (IACUC) at the University of California, Irvine. We used 1.21 g of skin and muscle tissue to extract genomic DNA using a Genomic DNA and RNA purification kit (Macherey-Nagel, Düren, Germany). After extraction, the DNA samples were sheared and separated into large DNA molecules. We used Pacific Biosciences (PacBio) and Illumina platforms for sequencing. For PacBio sequencing, genomic DNA was sized selected with BluePippin with a 15 kb size cut-off, and 40 SMRT cells were sequenced with the PacBio RS II. In addition, from the same gDNA extraction, a multiplex gDNA-Seq Illumina sequencing library was prepared from size selected fragments which ranged from 500 to 700 bp and sequenced on two lanes that resulted in short reads (100 bp paired-end) on an Illumina HiSeq 2500. All genomic sequencing was completed at the University of California, Irvine (UCI) Genomics High-Throughput Facility (GHTF).

(b) Illumina and Pacbio hybrid assembly, genome size estimation and quality assessment

We implemented multiple bioinformatic assembly programs to generate our final assembly of the *C. violaceus* genome (electronic supplementary material, figure S3 and tables S2–S4). Sequence data generated from two lanes of Illumina HiSeq 2500 were assembled through PLATANUS v. 1.2.1 [29]. Contigs assembled from PLATANUS and reads from 40 SMRT cells of PacBio sequencing were assembled with a hybrid assembler DBG2OLC v. 1.0 [30]. Without Illumina sequence reads, we also conducted a PacBio reads only assembly with FALCON v. 0.3.0 (https://github.com/PacificBiosciences/FALCON). We then used the outputs from FALCON and DBG2OLC as input for QUICKMERGE v. 1.0 [31], a metassembler and assembly gap filler developed for long molecule-based assemblies. We then polished the assembly with two rounds of QUIVER [32] and then used PILON v. 1.16 [33] to make final improvements. N25, N50 and N75 was estimated with a perl script (Joseph Fass—http://bioinformatics.ucdavis.edu) from our final *C. violaceus* genome assembly. We processed the final genome assembly through REPEATMASKER v. 4.0.6 [34] and we used BUSCO v. 3 [35] to estimate the completeness of our genome assembly with the Vertebrata and Actinopterygii gene set.

See Supplemental Methods in the electronic supplementary material for information on transcript assembly, annotation and heatmap generation of all tissues and genes associated with diet.

(c) Identification of candidate genes, copy number and estimation of positive selection

We selected the following candidate genes involved in the digestion of dietary carbohydrates, proteins, and lipids to identify gene copy number and estimate positive diversifying selection: the carbohydrase amylase; the proteolytic enzymes aminopeptidase a, aminopeptidase b, aminopeptidase Ey, aminopeptidase N, aminopeptidase Ey-like, chymotrypsin A, chymotrypsin B, chymotrypsin-like, and trypsin; the lipolytic enzymes phospholipase B1, group XIIB secretory phospholipase A2-like protein, carboxyl ester lipase and carboxyl ester lipase-like enzyme.

To evaluate amylase gene copies, we used previously published variants of *C. violaceus* amylase (*amy2a* and *amy2b* [28]) deposited on NCBI (KT920438 and KT920439) to search our assembled genome using both MUMMER v. 3.23 [36] and BLAST [37]. We then used FASCUT, a perl script that is part of the FAST Analysis of Sequences Toolbox [38] to trim the contig that contained amylase loci and neighbouring genes. These fragments were viewed with AUGUSTUS v. 3.2.3 [39]; GENOMICUS v. 96.01 (http://www.dyogen.ens.fr/genomicus/) and ENSEMBL v. 97 were used to visualize syntenic regions of the candidate gene in *C. violaceus*, as well as *Danio rerio*, *Oryzias latipes* and *Gasterosteus aculeatus*. We then obtained the amylase sequences from multiple stichaeid species representing dietary diversity [22] (electronic supplementary material, figure S2), including: *Anoplarchus purpurescens* (carnivore), *Dictyosoma burgeri* (carnivore), *Phytichthys chirus* (omnivore), *Xiphister atropurpureus* (omnivore) and *X. mucosus* (herbivore [22,25]). The *C. violaceus* amylase sequences and orthologous sequences from the five other prickleback species were aligned in MEGA v. 7.0.26 [40] with MUSCLE (default parameters with codons [41]). Selection was estimated using branch-site models and using adaptive branch site REL (aBSREL), a branch-site model that infers the optimal number of ω (nonsynonymous/synonymous rate ratio) classes for each branch, testing whether a proportion of sites have evolved under positive selection. Next, a mixed effects model of evolution (MEME) was used to test the individual sites subject to episodic positive or diversifying selection, and a signatures of recombination genetic algorithm for recombination detection (GARD) was used as part of the DATAMONKEY v. 2.0 web application [42].

The assembled *C. violaceus* transcriptome was used to identify candidate protease genes. Once identified, candidate genes were BLASTed against our assembled transcriptomes where the highest bit score and per cent identity (greater than 70%) were used to identify the orthologues from assembled transcriptomes from the same stichaeid species used for amylase analyses [11]. All orthologues identified from the stichaeid species were used for the molecular evolution analyses (aBSREL, MEME and GARD), and used for multiple sequence alignments. JMODELTEST v. 2.1.0 [43] was used to test for a model of sequence evolution, and phylogenetic trees made using PHYML v. 3.1 [44]. Synteny was analysed as described for amylase.

Similar methods were followed for phospholipase B1, group XIIB secretory phospholipase A2-like protein, carboxyl ester lipase and carboxyl ester lipase-like enzyme using the annotated *C. violaceus* transcriptome.

(d) Identification of orthologues across teleost fishes and identification of syntenic regions

All *C. violaceus* transcripts predicted from AUGUSTUS were used for identifying orthologues among ensembl protein datasets of teleost fishes and a lobed finned fish (coelacanths) (figure 1; electronic supplementary material, tables S5 and S6). All *C. violaceus* transcripts were translated into protein sequences using ORFPREDICTOR [45]. From 13 fish species (figure 1; see Supplemental Methods in the electronic supplementary material), sequences with 60 amino acids or longer were used to conduct pairwise identification of orthologues by using INPARANOID v. 4.0 [46]. Nucleotide sequences of single copy orthologues were aligned using custom python scripts (https://github.com/JosephHeras/Heras_etal_C.violaceus_Genome_2019) and MUSCLE [41] with default settings and reviewed each alignment. GTR+I+G was selected as the best model of sequence evolution (JMODELTEST v. 2.1.0 [43]). Sequences were concatenated with SEQUENCEMATRIX 1.8 [47]. Maximum-likelihood (ML) phylogenetic trees were constructed with PhyML v. 3.1 [44] with the GTR+I+G model and 1000 bootstrap replicates. Syntenic regions of the *C. violaceus* genome was compared to *Gasterosteus aculeatus*, *Oryzias latipes*, *Danio rerio* and *Lepisosteus oculatus* (electronic supplementary material) based on their quality draft genomes and broad evolutionary distance with respect to *C. violaceus*.

3. Results and discussion

(a) Genome assembly, quality and size

PacBio Single Molecule Real Time sequencing generated approximately 30 Gb long reads (approx. 37X based on genome size = 792 Mb for *C. violaceus* [48]), whereas the Illumina effort generated 84.5 Gb (approx. 107X) paired end reads. The Illumina-only assembly was highly fragmented (N50 = 2760 bp), consistent with other fish genome assemblies [49] (www.ensembl.org). Using the Illumina contigs in concert with long reads yielded a more contiguous hybrid assembly (N50 = 2.21 Mb), whereas an assembly of the long reads alone yielded similar results (N50 = 2.45 Mb). Merging the hybrid assembly with the long read only assembly [31] (see Methods) yielded a highly contiguous assembly (N50 = 6.69 Mb), ranking it among the most contiguous teleost genomes, and, to our knowledge, the most contiguous among herbivorous fishes (electronic supplementary material, figure S4). The universal single copy orthologues (BUSCOs) [50] show that our assembly (97%) is comparable to or better than many model reference fish genomes (86.6–98.3%) (figure 1b). Based on an average of four k-mer size counts (k-mer: 25, 27, 29, 31), JELLYFISH estimates the genome size at 656 598 967 base pairs (standard deviation of 4 138 853 base pairs; electronic supplementary material, figure S5 and table S7), which is close to the *C. violaceus* genome size estimate (792 Mb [48]), and other fish genomes (figure 1*b*) [49]. We estimate that approximately 21.5% of the genome (128 Mb) is repetitive, with approximately 17% (100 Mb) and 4.5% (27.5 Mb) of the genome occupied by transposable elements and simple repeats, respectively. Only 7.9% of these TEs could be detected using the fugu TE database, so the majority of the TEs we detect in the *C. violaceus* genome are likely novel. We also were able to identify about 38 448 repeated loci where the repetitive sequence (period) range from 1 to 1983 and number of repeats identified were 1.8 to 14 140.8 bps (electronic supplementary material, figures S6 and S7).

The genome of *C. violaceus* is most similar to that of stickleback, *Gasterosteus aculeatus* (electronic supplementary material, figure S8) and less so with those of *Oryzias latipes*, *Danio rerio* and *Lepisosteus oculatus* (electronic supplementary material, figures S9–S11). Although the synteny of many of the comparisons are rearranged, these comparisons nevertheless show the relative completeness of our draft genome in comparison to these model systems.

(b) Transcriptomics

We constructed a transcriptomics dataset for nine tissues (electronic supplementary material, figures S12–S17) in *C. violaceus*. Focusing on the liver, pyloric caeca (which includes pancreatic tissue [9,22]), proximal intestine, mid intestine, and spleen (figure 2; electronic supplementary material, tables S8–S12), we identified genes associated with metabolism (figure 2*a*), and digestion (figure 2*a*). Like many herbivorous mammals, *C. violaceus* has an active microbial community in their hindgut that ferments dietary substrates to short chain fatty acids (SCFAs) (electronic supplementary material, figure S2) [11]. SCFAs are absorbed by the host and used to generate ATP in various tissues [8,52]. Because the two most commonly produced SCFAs acetate and propionate are largely metabolized to ketones in the liver, animals reliant on hindgut fermentation tend to have active ketotic pathways in their tissues [8,51]. Indeed, genes coding for proteins in ketone synthesis and degradation are upregulated in most tissues in *C. violaceus*, especially in the liver (figure 2*a*). There are also clear expression patterns for carbohydrases, proteases and lipases, confirming the suite of enzymes necessary to digest a range of nutrients (figure 2*b*). *Cebidichthys violaceus* appears to express four chymotrypsin genes and two trypsin genes (electronic supplementary material, figures S18–S20). This may be consistent with their herbivorous diet, as carnivores (e.g. salmon) may invest more in trypsin expression [53], whereas herbivores (e.g. grass carp) may express more chymotrypsin [54]. *Cebidichthys violaceus* efficiently digests protein from algae [10]. Fishes appear to possess five aminopeptidase genes that appear to be retained following fish whole genome duplication events (electronic supplementary material, figures S21–S24 and supplemental discussion), and signatures of selection were observed for these genes.

The pyloric caecal tissue of *C. violaceus* and other pricklebacks is recognized as 'pancreatic' because it is sheathed in acinar cells [9] and shows elevated activity levels of pancreatic enzymes [11,22]. However, this tissue only has two differentially expressed genes in comparison to the mid intestine (electronic supplementary material, figure S15), which is a highly absorptive region of the fish gut [55,56]. Although pyloric caeca have been documented as absorptive (i.e. similar function to the mid intestine [57]), the mid intestine is rarely recognized as also having pancreatic function. In fishes, pancreatic tissue can be embedded in the liver (forming a hepatopancreas) or diffuse along the proximal intestine, particularly in fishes with pyloric caeca [58]. Our transcriptomic and biochemical data suggest an absorptive function of the pyloric caeca, and that the acinar cells are distributed at least down to the mid intestine and not restricted to the pyloric caecal region in *C. violaceus* (electronic supplementary material, figures S16 and S17).

(c) Physiological genomics of digestive enzymes

Digestive enzyme activity levels reveal what substrates are readily digested in an animal's digestive tract, highlighting the enzyme genes that are potential targets of selection for efficient digestion [8]. Within the field of nutritional physiology, the adaptive modulation hypothesis predicts a match between the amount of an ingested substrate (e.g. starch) and digestive enzyme expression and activity levels to digest such substrates (e.g. amylase) based on economic principles [8]. Pancreatic amylase activity is elevated in the guts of herbivores and omnivores in comparison to carnivores (especially in prickleback fishes [22,59]), matching the higher intake of soluble carbohydrates in these animals [11,21–23,60]. *Cebidichthys violaceus* has three tandem pancreatic amylase genes: two copies of *amy2a* and one copy of *amy2b* (figure 3). The three *C. violaceus amy* genes in tandem differs from other pricklebacks (and most other fishes for which genetic data are available), which tend to have one or two identical copies of *amy2* (figure 3) [22]. The two *amy2a* copies in *C. violaceus* are supported by three spanning reads, emphasizing the correct assembly of the *amy2a* tandem duplicates (electronic supplementary material, figure S25). Each amylase gene is preceded by a 4.3 Kb DNA element encoding a transposase (figure 3; electronic supplementary material, figure S26), hinting at a role of this transposable element (TE) in gene duplications in this region [61,62]. Additionally, the *amy2b* gene has a

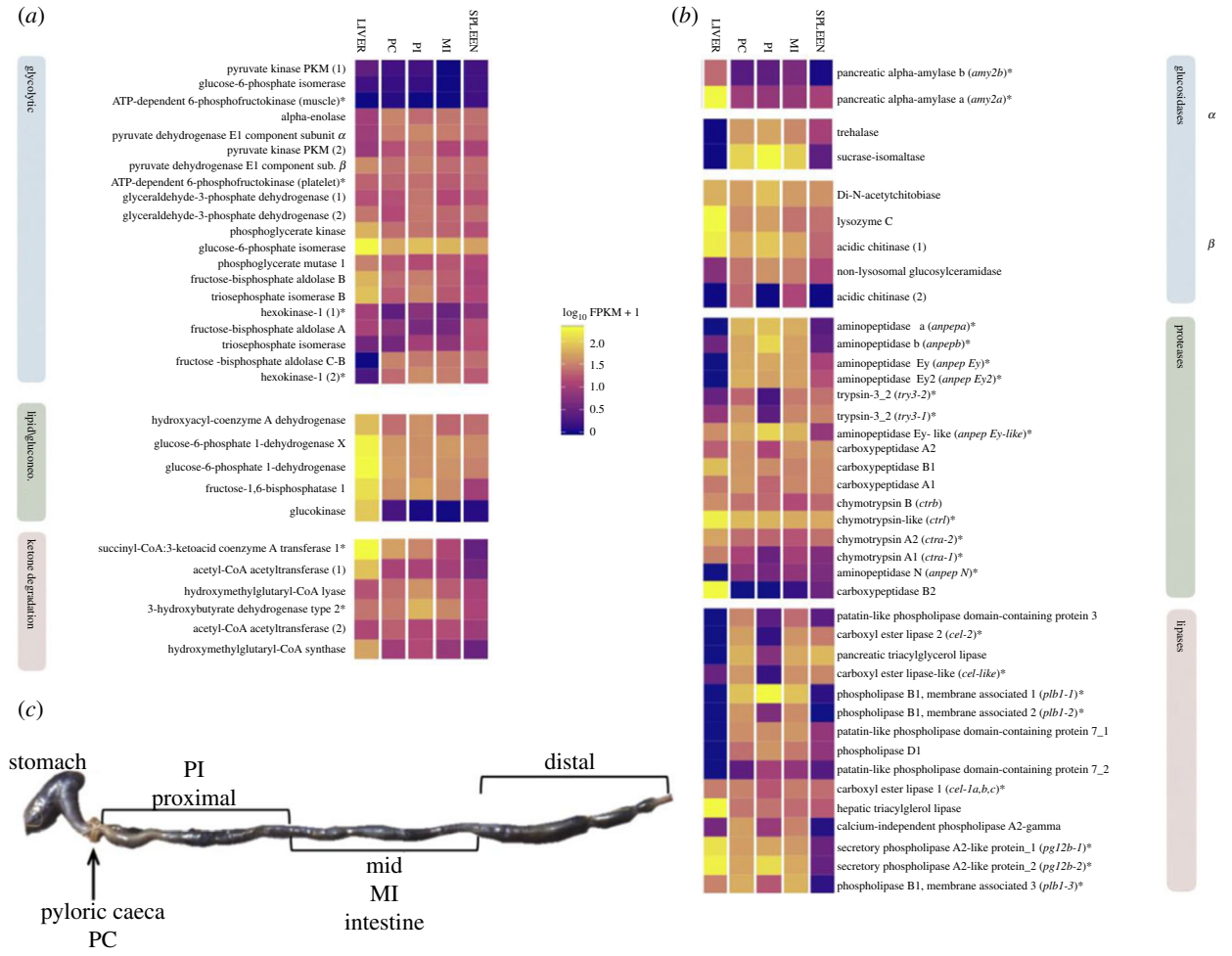


Figure 2. Gene expression profiles in tissues of *Cebidichthys violaceus*. We used brain, gill, gonads (testes), heart, liver, pyloric caeca (PC), proximal intestine (PI), middle intestine (MI) and spleen tissues from *C. violaceus* to represent the transcriptome. Only gene expression profiles of liver, PC, PI, MI and spleen are shown. Low to high expression is shown on a gradient scale from violet (darker) to yellow (lighter) respectively. Unit of expression is measured as fragments per kilobase of transcript per Million mapped reads (FPKM). (*a*) Three heatmaps were generated for transcripts representing glycolytic pathways (blue box), transcripts for enzymes associated with lipid synthesis/gluconeogenesis [51] (green box), and ketone degradation pathway transcripts (red/pink box). (*b*) Three heatmaps of digestive enzymes were generated which include carbohydrases (blue boxes), proteases (green box) and lipases (red/pink box). (*c*) A diagram of the *C. violaceus* gut, which includes the stomach, pyloric caeca, proximal intestine, middle intestine and distal intestine. Asterisk shows candidate genes for identifying positive selection. (Online version in colour.)

2025 bp long interspersed nuclear element (LINE) inside the 2nd intron of the gene. All three copies of amylase gene copies possess an approximately 470 bp fragment of a long terminal repeat retrotransposon (electronic supplementary material, figure S26). Insertion of the TEs proximal to the transcription start site and within first introns could modulate the expression of the amylase gene copies because both of these regions are typically enriched with cis-regulatory elements [62]. When testing all 11 branches for the seven prickleback taxa, only one branch was under episodic diversifying selection (*C. violaceus*, *amy2b*; figure 3*c*). The AMY2A and AMY2B proteins have different isoelectric points (7.86 versus 8.62 [22]), which hints that they may be active in different parts of the gut, and the transcriptomic data show that *amy2a* is expressed at a fairly constant level throughout the proximal GI tract (including the pyloric caeca), whereas *amy2b* is expressed mostly in the mid intestine (figure 2).

Amylolytic activities in the guts of *C. violaceus* are similar to those in the two species of *Xiphister* (figure 3*a*), yet the *Xiphister* taxa have two copies of *amy2a* and no amy2b [22], and *C. violaceus* and *X. mucosus* digest algal starch with similar efficiencies [10]. Thus, the phenotype of elevated amylase activity can be achieved via increased gene copy number leading to increased expression of the genes or by increased expression of fewer genes with similar performance outcomes at the whole animal level [10,22].

Herbivores consume a food that is simultaneously low in lipid [63], has more plant-derived galactolipids and betaine lipids than animal material [64,65], is high in fibre [66], and fibre binds to lipid, impeding its digestion [67]. Thus, carboxyl ester lipase (*cel*) [68] represents an important digestive enzyme gene for herbivorous fishes because lipids (especially essential fatty acids) are crucial for survival. In fishes, CEL is the most important lipolytic enzyme due to its broad specificity [68], and ability to hydrolyse galactolipids [69,70] (in fact, fishes lack a pancreatic lipase–colipase system like mammals have, further emphasizing CEL as a key lipase [68–70]). In pricklebacks, lipolytic activities are elevated in herbivores [11,25]. Moreover, *Danio rerio* fed a high-fibre, low-lipid diet, analogous to a herbivorous diet in the laboratory, had elevated lipase activities in their guts [26]. Thus, it appears that herbivorous pricklebacks, and other herbivorous fishes [11,25,26] invest in lipase expression to ensure lipid digestion from their algal diet, consistent with the Nutrient Balancing Hypothesis, under which animals invest in the synthesis of digestive enzymes to acquire limiting nutrients [24].

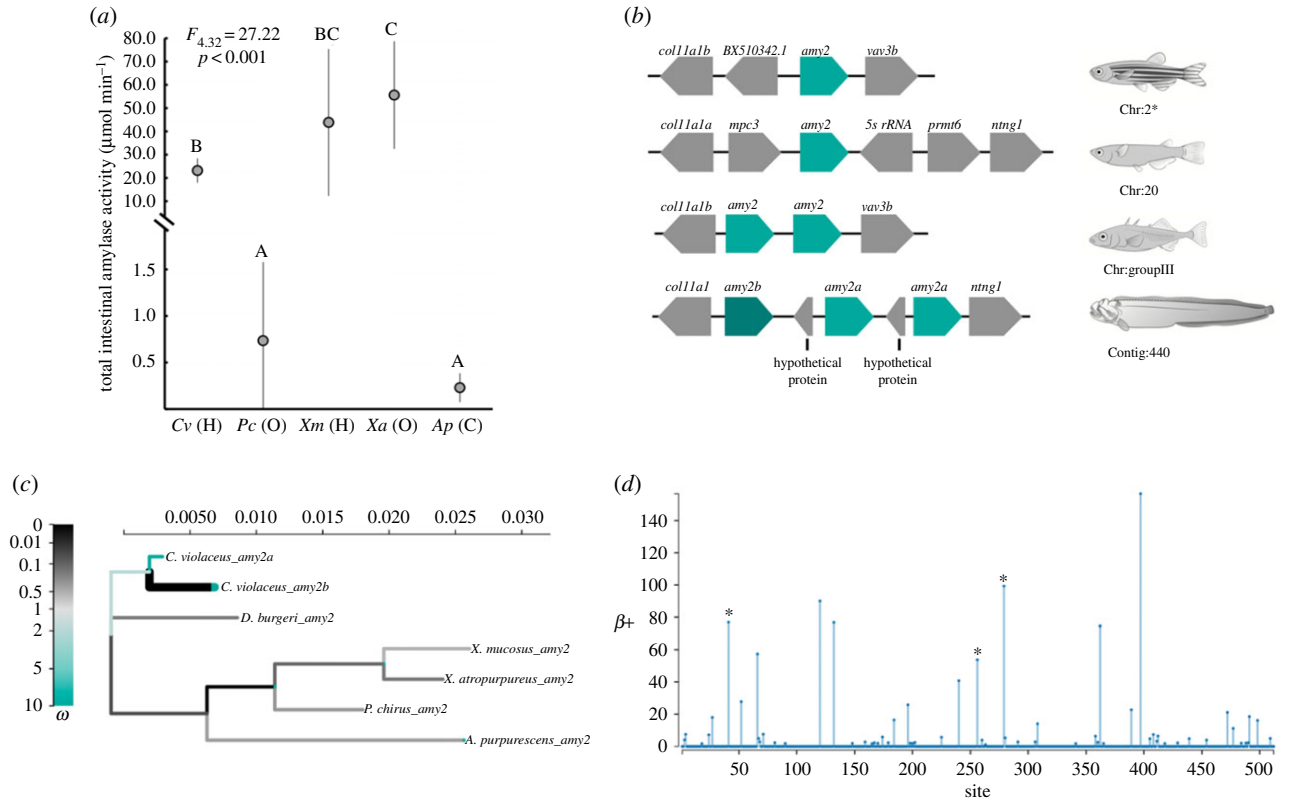


Figure 3. Enzyme activity, gene copy number and molecular evolution of amylase (a carbohydrase). (*a*) Total gut standardized amylase activity for *Cebidichthys violaceus* (*Cv*) and other stichaeid species: *Phytichthys chirus* (*Pc*), *Xiphister mucosus* (*Xm*), *Xiphister atropurpureus* (*Xa*) and *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* [11]. Interspecific comparisons were made for amylase with ANOVA, where circles that share a letter are not significantly different. (*b*) Synteny map for amylase genes from *Danio rerio*, *Oryzias latipes*, *Gasterosteus aculeatus* and *C. violaceus*. Asterisk denotes that *D. rerio* also has additional amylase loci present on chromosome 17. (*c*) An adaptive branch-site random effects likelihood (aBSREL) test for episodic diversification phylogenetic tree constructed for amylase genes from *C. violaceus* (*amy2a* and *amy2b*) and other stichaeid species. ω is the ratio of nonsynonymous to synonymous substitutions. The colour 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. (*d*) The output of a mixed effects model of evolution (MEME) to detect episodic positive/diversifying selection at sites. $\beta+$ is the non-synonymous substitution rate at a site for the positive/neutral evolution throughout the sequence of the gene. Asterisks indicate that the positive/diversifying site is statistically significant: $^{**}p < 0.01$, $^{*}p < 0.05$. (Online version in colour.)

In terms of lipid composition, algae have mostly galactolipids, betaine lipids, and phosphatidylcholine in their tissues, whereas animal material mostly contains triacylglycerides; each contains phospholipids in membranes [67,68]. CEL would be the main fish lipase to hydrolyse galactolipids into free fatty acids [65,69,70]. In mammals, omnivores consuming plant material (i.e. more galactolipids) have elevated lipolytic activities towards galactolipids, whereas carnivores do not [71]. This is also true in insects [72]. Galactolipid assimilation has been confirmed in *Danio rerio* [73]. Here, we see an expansion in *cel* gene copy number in an animal consuming more galactolipids (figure 4), whereas we did not observe any expansion in gene copy numbers of phospholipases or triacylglycerol lipase (figure 2*b*; electronic supplementary material, figures S27 and S28).

In the *C. violaceus* genome, we identified four tandem copies of *cel* on contig 445 (three to *cel-1* and one *cel-2*; figure 4; electronic supplementary material, figure S29), and a *cel-like* locus on contig 138 (electronic supplementary material, figure S30). Each of the individual CEL proteins contains the requisite bile salt binding motif [70] (electronic supplementary material, figure S31). Two of the *cel-1* genes (*cel-1b* and *cel-1c*) are more similar to each other in intron/exon arrangements than either are *cel-1a* or *cel-2*, suggesting that *cel-1b* and *cel-1c* copies are more recent gene duplications. Interestingly, *cel-1a* and *cel-2* possess two different LINEs in their first and last introns, respectively, contributing to the structural diversification of the two gene copies (electronic supplementary material, figure S29). We see positive episodic selection for *cel-1* (figure 4*e*; electronic supplementary material, table S13).

The *cel-1* gene is expressed across tissues, except for brain, whereas *cel-2* shows primarily gut expression (figure 2*b*). The protein sequences of the *cel-1* copies are similar (pairwise distance of 0.014–0.054; Poisson corrected). Whereas *C. violaceus* has four copies of *cel* (*cel-1* and *cel-2)* in their genome, other fishes for which the genome has been sequenced, and other stichaeids (based on transcriptomes of their relevant tissues), appear to have two *cel* genes (figure 4*b*). *cel-2* shows evidence of recombination (figure 4); there is no evidence of positive diversifying selection. Phylogenetic analyses of *cel* loci show that *cel-1* and *cel-2* are represented in *G. aculeatus*, whereas *O. latipes* and *D. rerio* do not have a divergent *cel-2* locus (electronic supplementary material, figure S30 and table S14).

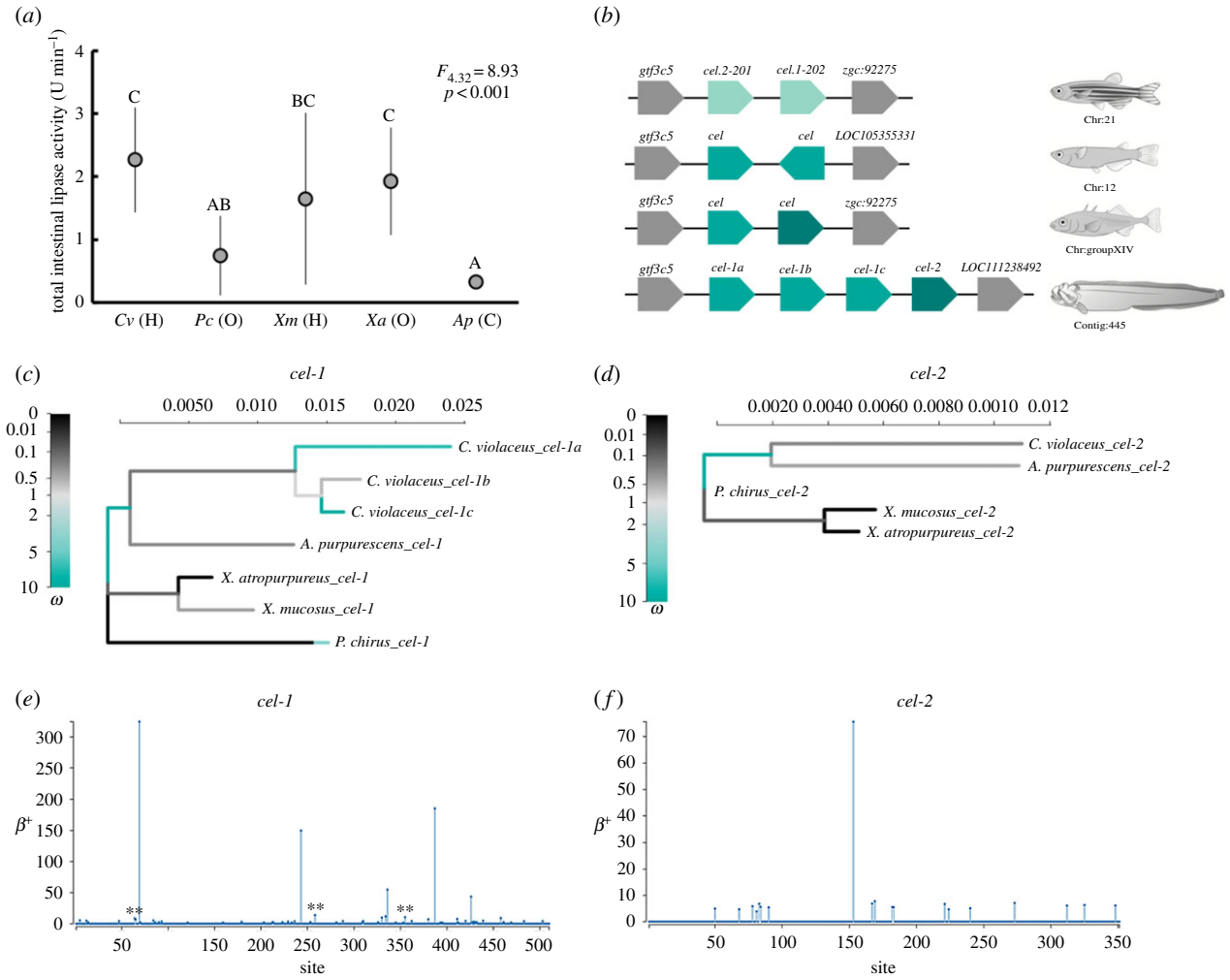


Figure 4. Enzyme activity, gene copy number and molecular evolution of carboxyl ester lipase (*cel*). (*a–f*) As described in figure 3, but for *cel-1* and *cel-2*. (Online version in colour.)

Cebidichthys violaceus digests algal lipid with consistent efficiency across a range of lipid concentrations, whereas *X. mucosus* shows decreasing lipid digestibility with decreasing dietary lipid content [10], suggesting that CEL diversity may affect lipid digestibility. Close examination of the CEL function in fishes with different diets is clearly warranted, especially using a pH-stat method, which allows for the examination of the hydrolysis of specific types of lipids [71,72].

Genomic scans of humans with lipid-rich diets (Nganasans and Yakuts), shows that they have experienced selection on lipases and proteins involved in lipid metabolism [6], consistent with the adaptive modulation hypothesis [17]. Although lipids compose a proportionally small part of algal mass (less than 10%), galactolipids are a major component of these lipids [64,65]. Thus, on a gross scale, elevated lipolytic activity in the guts of herbivorous fishes aligns with the nutrient balancing hypothesis [24], but on a more detailed level, an enzyme that hydrolyses a common lipid (galactolipids) in algae showing expansion in copy number in a herbivore also agrees with the adaptive modulation hypothesis. Nevertheless, elevated lipolytic activity in the guts of *D. rerio* fed a low-lipid, high-fibre diet in the laboratory [26] lends further support to the nutrient balancing hypothesis [24].

In conclusion, we produced a highly contiguous fish genome, and coupled with a rich literature on the nutritional physiology of *C. violaceus* and other stichaeid fishes, we were able to analyse our genomic data in a nutritional physiology context. Our results show that both, the adaptive modulation hypothesis [17], and the nutrient balancing hypothesis [24], can have genetic underpinnings within the same organism. This powerful physiological genomics approach will provide a model for nutritional physiological research. There is strong interest in using more plant-based aquaculture feeds, including plant lipid. The diversity of *cel* genes in *C. violaceus* may provide utility in genetically modified aquaculture fishes. Finally, given that *C. violaceus* is commonly found in Marine Protected Areas on the west coast of the United States, and is targeted for aquaculture in northern California where it is a delicacy, our data will also have application for conservation and better culturing techniques for this species.

Ethics. All handling of fish from capture to euthanization was conducted under approved protocol no. 2011-2989 of the Institutional Animal Care and Use Committee (IACUC) at the University of California, Irvine.

Data accessibility. All sequence data was deposited to NCBI's GenBank under the bioproject ID no. PRJNA384078. Included under this bioproject are the Illumina genomic and tissue transcriptomic objects under SAMN06857690–SAMN0687699, 40 SMRT cells of PacBio sequencing under SAMN06857690, and the final genome assembly (NJBE00000000).

Competing interests. We declare we have no competing interests.

Funding. This project was supported by the National Science Foundation grant no. IOS-1355224 (to D.P.G.). The work was supported in part by US National Institutes of Health (NIH) grant R01GM123303 (to J.E.E.).

Acknowledgements. We thank H. Yip and Q. B. Nguyen-Phuc for assistance in sample collection. D. Canestro for providing assistance and facilities at the Kenneth S. Norris Rancho Marino Reserve (Cambria, CA). We thank M. Oakes, V. Ciobanu, S. A. Chung, D. Yu and Y. Kanomata at the UC Irvine Genomics High-Throughput Facility. A. Long, J. Baldwin-Brown and K. Thorton for RNA-Seq assembly suggestions. The comparative physiology group at UCI for feedback on the content of this manuscript. We thank N. Nirale for assistance on NCBI's GenBank. We thank A. Dingeldein, S. David and M. Tan for illustrations.

References

1. Herrel A, Huyghe K, Vanhooydonck B, Backeljau T, Breugelmans K, Grbac I, Van Damme R, Irschick DJ. 2008 Rapid large-scale evolutionary divergence in morphology and performance associated with exploitation of a different dietary resource. *Proc. Natl Acad. Sci. USA* **105**, 4792–4795. (doi:10.1073/pnas.0711998105)
2. Lamichhaney S *et al*. 2015 Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature* **518**, 371–375. (doi:10.1038/nature14181)
3. Protas ME, Hersey C, Kochanek D, Zhou Y, Wilkens H, Jeffery WR, Zon LI, Borowsky R, Tabin CJ. 2005 Genetic analysis of cavefish reveals molecular convergence in the evolution of albinism. *Nat. Genet.* **38**, 107–111. (doi:10.1038/ng1700)
4. Chakraborty M, Fry JD. 2015 Parallel functional changes in independent testis-specific duplicates of *Aldehyde dehydrogenase* in *Drosophila*. *Mol. Biol. Evol.* **32**, 1029–1038. (doi:10.1093/molbev/msu407)
5. Harris SE, Munshi-South J. 2017 Signatures of positive selection and local adaptation to urbanization in white-footed mice (*Peromyscus leucopus*). *Mol. Ecol.* **26**, 6336–6350. (doi:10.1111/mec.14369)
6. Hsieh P, Hallmark B, Watkins J, Karafet TM, Osipova LP, Gutenkunst RN, Hammer MF. 2017 Exome sequencing provides evidence of polygenic adaptation to a fat-rich animal diet in indigenous Siberian populations. *Mol. Biol. Evol.* **34**, 2913–2926. (doi:10.1093/molbev/msx226)
7. Vega-Retter C, Rojas-Hernandez N, Vila I, Espejo R, Loyola DE, Copaja S, Briones M, Nolte AW, Véliz D. 2018 Differential gene expression revealed with RNA-Seq and parallel genotype selection of the ornithine decarboxylase gene in fish inhabiting polluted areas. *Sci. Rep.* **8**, 1–3. (doi:10.1038/s41598-018-23182-z)
8. Karasov WH, del Rio CM. 2007 *Physiological ecology: how animals process energy, nutrients, and toxins*. Princeton, NJ: Princeton University Press.
9. Kim KH, Horn MH, Sosa AE, German DP. 2013 Sequence and expression of an α-amylase gene in four related species of prickleback fishes (Teleostei: Stichaeidae): ontogenetic, dietary, and species-level effects. *J. Comp. Physiol. B* **184**, 221–234. (doi:10.1007/s00360-013-0780-1)
10. Horn MH, Neighbors MA, Murray SN. 1986 Herbivore responses to a seasonally fluctuating food supply: growth potential of two temperate intertidal fishes based on the protein and energy assimilated from their macroalgal diets. *J. Exp. Mar. Biol. Ecol.* **103**, 217–234. (doi:10.1016/0022-0981(86)90142-5)
11. German DP, Sung A, Jhaveri P, Agnihotri R. 2015 More than one way to be an herbivore: convergent evolution of herbivory using different digestive strategies in prickleback fishes (Stichaeidae). *Zoology* **118**, 161–170. (doi:10.1016/j.zool.2014.12.002)
12. Wang Z, Du J, Lam S, Mathavan S, Matsudaira P, Gong Z. 2010 Morphological and molecular evidence for functional organization along the rostrocaudal axis of the adult zebrafish intestine. *BMC Genomics* **11**, 392. (doi:10.1186/1471-2164-11-392)
13. De Santis C, Bartie KL, Olsen RE, Taggart JB, Tocher DR. 2015 Nutrigenomic profiling of transcriptional processes affected in liver and distal intestine in response to a soybean meal-induced nutritional stress in Atlantic salmon (*Salmo salar*). *Comp. Biochem. Physiol. Part D Genomics Proteomics* **15**, 1–11. (doi:10.1016/j.cbd.2015.04.001)
14. Lie KK, Tørresen OK, Solbakken MH, Rønnestad I, Tooming-Klunderud A, Nederbragt AJ, Jentoft S, Sæle Ø. 2018 Loss of stomach, loss of appetite? Sequencing of the ballan wrasse (*Labrus bergylta*) genome and intestinal transcriptomic profiling illuminate the evolution of loss of stomach function in fish. *BMC Genomics* **19**, 186. (doi:10.1186/s12864-018-4570-8)
15. Wang D *et al*. 2019 Whole genome sequencing of the giant grouper (*Epinephelus lanceolatus*) and high-throughput screening of putative antimicrobial peptide genes. *Mar. Drugs* **17**, 503. (doi:10.3390/md17090503)
16. Lehmann R *et al*. 2018 Finding Nemo's Genes: a chromosome-scale reference assembly of the genome of the orange clownfish *Amphiprion percula*. *Mol. Ecol. Resour.* **19**, 570–585. (doi:10.1111/1755-0998.12939)
17. Karasov WH. 1992 Tests of the adaptive modulation hypothesis for dietary control of intestinal nutrient transport. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **263**, R496–R502. (doi:10.1152/ajpregu.1992.263.3.r496)
18. Perry GH *et al*. 2007 Diet and the evolution of human amylase gene copy number variation. *Nat. Genet.* **39**, 1256–1260. (doi:10.1038/ng2123)
19. German DP, Neuberger DT, Callahan MN, Lizardo NR, Evans DH. 2010 Feast to famine: the effects of food quality and quantity on the gut structure and function of a detritivorous catfish (Teleostei: Loricariidae). *Comp. Biochem. Physiol. A Mol. Integr. Physiol.* **155**, 281–293. (doi:10.1016/j.cbpa.2009.10.018)
20. Kohl KD, Weiss RB, Dale C, Dearing MD. 2011 Diversity and novelty of the gut microbial community of an herbivorous rodent (*Neotoma bryanti*). *Symbiosis* **54**, 47–54. (doi:10.1007/s13199-011-0125-3)
21. 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)
22. German DP, Foti DM, Heras J, Amerkhanian H, Lockwood BL. 2016 Elevated gene copy number does not always explain elevated amylase activities in fishes. *Physiol. Biochem. Zool.* **89**, 277–293. (doi:10.1086/687288)
23. Boehlke C, Zierau O, Hannig C. 2015 Salivary amylase—the enzyme of unspecialized euryphagous animals. *Arch. Oral Biol.* **60**, 1162–1176. (doi:10.1016/j.archoralbio.2015.05.008)
24. Clissold FJ, Tedder BJ, Conigrave AD, Simpson SJ. 2010 The gastrointestinal tract as a nutrient-balancing organ. *Proc. R. Soc. B* **277**, 1751–1759. (doi:10.1098/rspb.2009.2045)
25. German DP, Horn MH, Gawlicka A. 2004 Digestive enzyme activities in herbivorous and carnivorous prickleback fishes (Teleostei: Stichaeidae): ontogenetic, dietary, and phylogenetic effects. *Physiol. Biochem. Zool.* **77**, 789–804. (doi:10.1086/422228)
26. 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*). *J. Comp. Physiol. B* **188**, 237–253. (doi:10.1007/s00360-017-1122-5)
27. Choat JH, Clements KD. 1998 Vertebrate herbivores in marine and terrestrial environments: a nutritional ecology perspective. *Annu. Rev. Ecol. Syst.* **29**, 375–403. (doi:10.1146/annurev.ecolsys.29.1.375)
28. Horn MH. 1989 Biology of marine herbivorous fishes. *Oceanogr. Mar. Biol. Ann. Rev.* **27**, 167–272.
29. Kajitani R *et al*. 2014 Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun short reads. *Genome Res.* **24**, 1384–1395. (doi:10.1101/gr.170720.113)
30. 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**, 31900. (doi:10.1038/srep31900)
31. Chakraborty M, Baldwin-Brown JG, Long AD, Emerson JJ. 2016 Contiguous and accurate de novo assembly of metazoan genomes with modest long read coverage. *Nucleic Acids Res.* **44**, e147. (doi:10.1093/nar/gkw654)
32. Chin C-S *et al*. 2013 Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat. Methods* **10**, 563–569. (doi:10.1038/nmeth.2474)

33. Walker BJ *et al*. 2014 Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE* **9**, e112963. (doi:10.1371/journal.pone.0112963)
34. Smit AF. 2004 Repeat-Masker Open-3.0. See http://www.repeatmasker.org.
35. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015 BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212. (doi:10.1093/bioinformatics/btv351)
36. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL. 2004 Versatile and open software for comparing large genomes. *Genome Biol.* **5**, R12. (doi:10.1186/gb-2004-5-2-r12)
37. Altschul S. 1997 Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402. (doi:10.1093/nar/25.17.3389)
38. Lawrence TJ, Kauffman KT, Amrine KCH, Carper DL, Lee RS, Becich PJ, Canales CJ, Ardell DH. 2015 FAST: FAST Analysis of Sequences Toolbox. *Front. Genet.* **6**, 172. (doi:10.3389/fgene.2015.00172)
39. Stanke M, Morgenstern B. 2005 AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Res.* **33**, W465–W467. (doi:10.1093/nar/gki458)
40. Kumar S, Stecher G, Tamura K. 2016 MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **33**, 1870–1874. (doi:10.1093/molbev/msw054)
41. Edgar RC. 2004 MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* **5**, 113. (doi:10.1186/1471-2105-5-113)
42. Weaver S, Shank SD, Spielman SJ, Li M, Muse SV, Kosakovsky Pond SL. 2018 Datamonkey 2.0: a modern web application for characterizing selective and other evolutionary processes. *Mol. Biol. Evol.* **35**, 773–777. (doi:10.1093/molbev/msx335)
43. Posada D. 2008 jModelTest: phylogenetic model averaging. *Mol. Biol. Evol.* **25**, 1253–1256. (doi:10.1093/molbev/msn083)
44. Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010 New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321. (doi:10.1093/sysbio/syq010)
45. Min XJ, Butler G, Storms R, Tsang A. 2005 OrfPredictor: predicting protein-coding regions in EST-derived sequences. *Nucleic Acids Res.* **33**, W677–W680. (doi:10.1093/nar/gki394)
46. O'Brien KP. 2004 Inparanoid: a comprehensive database of eukaryotic orthologs. *Nucleic Acids Res.* **33**, D476–D480. (doi:10.1093/nar/gki107)
47. Vaidya G, Lohman DJ, Meier R. 2011 SequenceMatrix: concatenation software for the fast assembly of multi-gene datasets with character set and codon information. *Cladistics* **27**, 171–180. (doi:10.1111/j.1096-0031.2010.00329.x)
48. Hinegardner R, Rosen DE. 1972 Cellular DNA content and the evolution of teleostean fishes. *Am. Nat.* **106**, 621–644. (doi:10.1086/282801)
49. Zerbino DR *et al*. 2017 Ensembl 2018. *Nucleic Acids Res.* **46**, D754–D761. (doi:10.1093/nar/gkx1098)
50. Waterhouse RM, Seppey M, Simão FA, Manni M, Ioannidis P, Klioutchnikov G, Kriventseva EV, Zdobnov EM. 2017 BUSCO applications from quality assessments to gene prediction and phylogenomics. *Mol. Biol. Evol.* **35**, 543–548. (doi:10.1093/molbev/msx319)
51. Willmott ME, Clements KD, Wells RMG. 2005 The influence of diet and gastrointestinal fermentation on key enzymes of substrate utilization in marine teleost fishes. *J. Exp. Mar. Biol. Ecol.* **317**, 97–108. (doi:10.1016/j.jembe.2004.11.008)
52. Bergman EN. 1990 Energy contributions of volatile fatty acids from the gastrointestinal tract in various species. *Physiol. Rev.* **70**, 567–590. (doi:10.1152/physrev.1990.70.2.567)
53. Rungruangsak-Torrissen K, Moss R, Andresen LH, Berg A, Waagbø R. 2006 Different expressions of trypsin and chymotrypsin in relation to growth in Atlantic salmon (*Salmo salar* L.). *Fish Physiol. Biochem.* **32**, 7–23. (doi:10.1007/s10695-005-0630-5)
54. Gioda CR, Pretto A, Freitas CD, Leitemperger J, Loro VL, Lazzari R, Lissner LA, Baldisserotto B, Salbego J. 2017 Different feeding habits influence the activity of digestive enzymes in freshwater fish. *Ciênc. Rural* **47**, 1–7. (doi:10.1590/0103-8478cr20160113)
55. Buddington RK, Chen JW, Diamond J. 1987 Genetic and phenotypic adaptation of intestinal nutrient transport to diet in fish. *J. Physiol.* **393**, 261–281. (doi:10.1113/jphysiol.1987.sp016823)
56. Tengjaroenkul B, Smith BJ, Caceci T, Smith SA. 2000 Distribution of intestinal enzyme activities along the intestinal tract of cultured Nile tilapia, *Oreochromis niloticus* L. *Aquaculture* **182**, 317–327. (doi:10.1016/s0044-8486(99)00270-7)
57. Buddington RK, Diamond JM. 1987 Pyloric ceca of fish: a 'new' absorptive organ. *Am. J. Physiol. Gastrointest. Liver Physiol.* **252**, G65–G76. (doi:10.1152/ajpgi.1987.252.1.g65)
58. Clements KD, Raubenheimer D. 2006 Feeding and nutrition. In *The physiology of fishes* (ed. DH Evans), pp. 47–82. Boca Raton, FL: CRC Press.
59. Chan AS, Horn MH, Dickson KA, Gawlicka A. 2004 Digestive enzyme activities in carnivores and herbivores: comparisons among four closely related prickleback fishes (Teleostei: Stichaeidae) from a California rocky intertidal habitat. *J. Fish Biol.* **65**, 848–858. (doi:10.1111/j.0022-1112.2004.00495.x)
60. Horn MH, Gawlicka AK, German DP, Logothetis EA, Cavanagh JW, Boyle KS. 2006 Structure and function of the stomachless digestive system in three related species of New World silverside fishes (Atherinopsidae) representing herbivory, omnivory, and carnivory. *Mar. Biol.* **149**, 1237–1245. (doi:10.1007/s00227-006-0281-9)
61. Feschotte C, Pritham EJ. 2007 DNA transposons and the evolution of eukaryotic genomes. *Annu. Rev. Genet.* **41**, 331–368. (doi:10.1146/annurev.genet.40.110405.090448)
62. Pantzartzi CN, Pergner J, Kozmik Z. 2018 The role of transposable elements in functional evolution of amphioxus genome: the case of opsin gene family. *Sci. Rep.* **8**, 1. (doi:10.1038/s41598-018-20683-9)
63. Neighbors MA, Horn MH. 1991 Nutritional quality of macrophytes eaten and not eaten by two temperatezone herbivorous fishes: a multivariate comparison. *Mar. Biol.* **108**, 471–476. (doi:10.1007/bf01313657)
64. Kato M, Sakai M, Adachi K, Ikemoto H, Sano H. 1996 Distribution of betaine lipids in marine algae. *Phytochemistry* **42**, 1341–1345. (doi:10.1016/0031-9422(96)00115-x)
65. Li-Beisson Y, Thelen JJ, Fedosejevs E, Harwood JL. 2019 The lipid biochemistry of eukaryotic algae. *Prog. Lipid Res.* **74**, 31–68. (doi:10.1016/j.plipres.2019.01.003)
66. Painter TJ. 1983 Algal polysaccharides. In *The polysaccharides* (ed. GO Aspinall), pp. 195–285. Amsterdam, The Netherlands: Elsevier.
67. German JB, Xu R, Walzem R, Kinsella JE, Knuckles B, Nakamura M, Yokoyama WH. 1996 Effect of dietary fats and barley fiber on total cholesterol and lipoprotein cholesterol distribution in plasma of hamsters. *Nutr. Res.* **16**, 1239–1249. (doi:10.1016/0271-5317(96)00127-3)
68. Murray HM, Gallant JW, Perez-Casanova JC, Johnson SC, Douglas SE. 2003 Ontogeny of lipase expression in winter flounder. *J. Fish Biol.* **62**, 816–833. (doi:10.1046/j.1095-8649.2003.00067.x)
69. Olsen RE, Ringø, E. 1997 Lipid digestibility in fish: a review. *Recent Res. Dev. Lipid Res.* **1**, 199–264.
70. Sæle Ø, Nordgreen A, Olsvik PA, Hamre K. 2010 Characterization and expression of digestive neutral lipases during ontogeny of Atlantic cod (*Gadus morhua*). *Comp. Biochem. Physiol. A Mol. Integr. Physiol.* **157**, 252–259. (doi:10.1016/j.cbpa.2010.07.003)
71. Amara S, Barouh N, Lecomte J, Lafont D, Robert S, Villeneuve P, De Caro A, Carrière F. 2010 Lipolysis of natural long chain and synthetic medium chain galactolipids by pancreatic lipase-related protein 2. *Biochim. Biophys. Acta Mol. Cell Biol. Lipids* **1801**, 508–516. (doi:10.1016/j.bbalip.2010.01.003)
72. Christeller JT, Amara S, Carrière F. 2011 Galactolipase, phospholipase and triacylglycerol lipase activities in the midgut of six species of lepidopteran larvae feeding on different lipid diets. *J. Insect Physiol.* **57**, 1232–1239. (doi:10.1016/j.jinsphys.2011.05.012)
73. Gedi MA, Magee KJ, Darwish R, Eakpetch P, Young I, Gray DA. 2019 Impact of the partial replacement of fish meal with a chloroplast rich fraction on the growth and selected nutrient profile of zebrafish (*Danio rerio*). *Food Funct.* **10**, 733–745. (doi:10.1039/c8fo02109k)