

Genome and gene evolution of seahorse species revealed by the chromosome-level genome of *Hippocampus abdominalis*

Libin He¹ | Xin Long² | Jianfei Qi¹ | Zongji Wang^{2,3} | Zhen Huang⁴ | Shuiqing Wu¹ | Xingtang Zhang⁵ | Huiyu Luo¹ | Xinxin Chen⁶ | Jinbo Lin⁶ | Qihua Yang¹ | Shiyu Huang⁷ | Qi Zhou^{2,8,9}  | Leyun Zheng¹

¹Fisheries Research Institute of Fujian, Xiamen, China

²MOE Laboratory of Biosystems Homeostasis and Protection and Zhejiang Provincial Key Laboratory for Cancer Molecular Cell Biology, Life Sciences Institute, Zhejiang University, Hangzhou, China

³Institute of Animal Sex and Development, Zhejiang Wanli University, Ningbo, China

⁴Fujian Key Laboratory of Sustainable Utilization of Featured Marine Bioresources, College of Life Science, Fujian Normal University, Fuzhou, China

⁵Shenzhen Branch, Guangdong Laboratory for Lingnan Modern Agriculture, Genome Analysis Laboratory of the Ministry of Agriculture, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen, China

⁶Xiamen Xiaodeng Aquatic Science and Technology Company Limited, Xiamen, China

⁷Fisheries College of Jimei University, Xiamen, China

⁸Department of Neuroscience and Developmental Biology, University of Vienna, Vienna, Austria

⁹Center for Reproductive Medicine, the Second Affiliated Hospital School of Medicine School of Medicine, Zhejiang University, Zhejiang University, Hangzhou, China

Correspondence

Shiyu Huang and Leyun Zheng, Fisheries Research Institute of Fujian, Xiamen, China.

Email: hys@fmu.edu.cn and 981668817@qq.com

Qi Zhou, MOE Laboratory of Biosystems Homeostasis and Protection and Zhejiang Provincial Key Laboratory for Cancer Molecular Cell Biology, Life Sciences Institute, Zhejiang University, Hangzhou, China.

Email: zhouqi1982@zju.edu.cn

Funding information

National Natural Science Foundation of China, Grant/Award Number: 32170415, 32061130208

Abstract

Seahorses belong to the teleost family Syngnathidae that evolved a distinct body plan and unique male pregnancy compared to other teleosts. As a classic model for studying evolution of viviparity and sexual selection of teleosts, seahorse species still lack a publicly available high-quality reference genome. Here, we generated the genome assembly of the big-belly seahorse, *Hippocampus abdominalis* with long-read and Hi-C technologies. We managed to place over 99% of the total length of 444.7 Mb of assembled genome into 21 linkage groups with almost no gaps. We reconstructed a phylogenomic tree with the big-belly seahorse genome and other representative Syngnathidae and teleost species. We also reconstructed the historical population dynamics of four representative Syngnathidae species. We found the gene families that underwent expansion or contraction in the Syngnathidae ancestor were enriched for immune-related or ion transporter gene ontology terms. Many of these genes were also reported to show a dynamic expression pattern during the pregnancy stages of *H. abdominalis*. We also identified putative positively selected genes in the Syngnathidae ancestor or in *H. abdominalis*, whose mouse mutants are enriched for abnormal craniofacial and limb morphological phenotypes. Overall, our study

He, Long, Qi and Wang contributed equally to this study.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2021 The Authors. *Molecular Ecology Resources* published by John Wiley & Sons Ltd.

provides an important genome resource for evolutionary and developmental studies of seahorse species, and candidate genes for future experimental works.

KEYWORDS

chromosome-level genome, Hi-C, male brood pouch, Seahorse

1 | INTRODUCTION

The great morphological and behavioural diversity of teleosts may be best exemplified by Syngnathidae species, that is, seahorses, pipefish and seadragons. All syngnathids have evolved tube-shaped mouth without teeth, an elongated, and specifically in seahorse an upright body plan without scales and pelvic fins but bony plates; and most intriguingly, the unique male pregnancy reproductive behaviour that is realised by male brood pouch (Leysen et al., 2011; Teske & Beheregaray, 2009; Wilson & Orr, 2011). The male brood pouch is inferred to develop from abdominal epithelium, in which females deposit the eggs for being fertilized and nurtured by the males until the free-swimming offspring are hatched. It is hypothesized to be functionally similar to the uterus of amniotes. Indeed, a recent study provided evidence that the pouch is critical for sustaining the lipid mass of embryos (Skalkos et al., 2020). Another study characterizing the transcriptomes of developing pouches of the big-belly seahorse (*Hippocampus abdominalis*) identified many up-regulated genes that are also upregulated during mammalian pregnancy (Whittington et al., 2015). Coincided with the evolution of male pregnancy, syngnathids seem to have undergone reconstructions of their immune gene repertoire, with major histocompatibility complex (MHC) genes having either become lost or undergone positive selection among different syngnathids (Roth et al., 2020). This reflects the trade-off between immunological tolerance and embryo rejection that has been reported in other viviparity vertebrates (Roth et al., 2020). Overall, these studies highlighted the role of syngnathids as a prominent evolutionary and developmental biology model for understanding the evolution of viviparity and novel morphological traits, as well as the mechanisms of sexual selection (Paczolt & Jones, 2010). In addition, seahorses are more vulnerable than other teleosts in the shallow water areas to overfishing for ornamental trade or traditional medicine, because of their long period of parental care, low mobility and population density (Foster & Vincent, 2004). Eighteen (about 40%) of the seahorse species are listed as International Union for Conservation of Nature (IUCN) Red list of vulnerable or endangered species. Therefore, seahorse and some other syngnathids are also considered flagship species for marine conservation (Vincent et al., 2011).

The development of genomic techniques has greatly facilitated the studies into seahorse and other syngnathids as a noncanonical teleost model, but a publicly available high-quality chromosome-level genome is lacking. A total of four Syngnathidae genomes (Lin et al., 2016, 2017; Small et al., 2016; Zhang et al., 2020) have been published using Illumina technologies, all of which are in a draft

genome form with a great number of assembly gaps still remaining within or between scaffold sequences. These data provided important insights into the global dispersal route, or identified the key genes responsible for some novel traits (e.g., bony spines and pelvic fin loss) of seahorses (Lin et al., 2016, 2017). Nevertheless, regulatory sequences of many responsible conserved developmental genes are probably located on unanchored scaffolds, or embedded in highly repetitive regions (McEwen et al., 2009; Vaughn et al., 2012). This necessitates a highly continuous reference genome of seahorse species. Here we are motivated to use the third-generation sequencing and high-throughput chromatin capture (Hi-C) technologies to produce a de novo assembled genome of *H. abdominalis*, as the first publicly available chromosomal reference genome of seahorse species. This is a widely farmed species and can be reared in the lab, thus is much more promising than many other threatened or endangered syngnathids for future evodevo studies. Comparative analyses with other representative Syngnathidae species reconstructed their phylogeny, population history, and identified new candidate genes underlying the morphological and immunological changes that distinguish the seahorses from other non-Syngnathidae species.

2 | MATERIALS AND METHODS

2.1 | Genome sequencing and assembly

The big-belly seahorse samples used in this project were obtained from a cultured population in Fujian (China), which has been approved by the Animal Ethical and Welfare Committee (Approval no. IACUC-2019-031) of the Animal Ethical and Welfare Committee of Fisheries Research Institute of Fujian. Before tissue collection, all the seahorses were cultured in indoor breeding tanks (150 L), and the seawater was from the sea and adjusted to have a salinity of 31‰ and temperature 16–18°C. The seahorses were fed three times per day (08:00 h, 12:00 h and 16:00 h) with artemia. All the seahorses were sacrificed after being anaesthetized for 30 minutes by MS-222 solution at a concentration of 50 mg/l, which is approved by FDA (Food and Drug Administration, USA). Male individuals were used for DNA extraction. We used DNeasy Mini Kit (Qiagen) to isolate genomic DNAs according to the manufacturer's instructions. The extracted DNAs were subject to size-selection in the BluePippin system and further used for PacBio long reads sequencing. We constructed SMRTbell™ libraries with 30–50 kb sizes following the official protocol downloaded from PacBio website. The resultant libraries were sequenced using PacBio Sequel platform. The PacBio long reads were first corrected using Canu 1.7

(Koren et al., 2017), with parameter `corOutCoverage = 80`. The corrected reads were assembled by three widely-used PacBio assemblers, Canu (Koren et al., 2017), wtdbg2 and SMARTdenovo (<https://github.com/ruanjue/smartdenovo>) (Ruan & Li, 2020). The quality of these assemblies was evaluated based on a series of parameters, including contig N50, assembled genome size as well as BUSCO completeness (Simão et al., 2015). The assembly from SMARTdenovo was finally chosen for further analysis, showing a contig N50 of 8.4 Mb, an assembled genome size of 445 Mb and a BUSCO completeness of 94%. To minimize the proportion of nucleotide base-pair errors in PacBio sequencing, 109.5x Illumina short reads were recruited for polishing using the Pilon program with default parameters (Walker et al., 2014).

The chromosome-scale assembly was achieved using the High-throughput Chromatin Conformation Capture (Hi-C) technology. Muscle tissue DNAs were extracted from one single male individual and subjected to construction of Hi-C libraries. The cross-linked DNAs were digested overnight with MboI restriction enzyme and treated with biotines, which were added to the end of fragmented DNA sequences. The chimeric junctions formed by proximity ligation were enriched by extracting biotins and further physically sheared, generating DNA fragments with 500–700 bp size. We sequenced these DNA fragments on Illumina NovaSeq platform with the PE mode. We performed two-rounds of Hi-C reads mapping using the BWA program (Li & Durbin, 2009). In the first round, 3D-DNA pipeline (Dudchenko et al., 2017) was applied to detect and correct abnormal contact patterns in the SMARTdenovo initially assembled contigs. In the second round of reads mapping, the Hi-C corrected contigs were subjected to ALLHiC pipeline (Zhang et al., 2019) for partition, orientation and ordering, and finally anchored onto 21 pseudochromosomes with an anchoring rate of 99.35%. We further manually curated the Hi-C scaffolding based on the chromatin contact matrix in Juicebox (Robinson et al., 2018). In addition, genome completeness was assessed based on 3640 conserved ray-finned fish genes (`actinopterygii_odb10`) in BUSCO program with default parameters. GC isochore analysis of this newly assembled genome was conducted using the same method as previously published paper about the isochore map of human chromosomes (Costantini et al., 2007). In brief, we first partitioned the entire chromosomal sequences of this newly assembled genome into nonoverlapping 100 kb windows. Then we calculated and plotted their GC levels along chromosomes using the program `draw_chromosome_gc.pl` (<http://genomat.img.cas.cz>). Furthermore, the intervals bounded by 37.5, 42.5, 47.5 and 52.5% GC are used to indicate the isochore families. Collinearity analysis between *H. abdominalis* linkage groups and other three Syngnathidae fish were conducted by nucmer (Marçais et al., 2018). Only one to one alignments were kept and only segments with alignment length larger than 500 bp were presented.

2.2 | Repeat and gene annotation

We first customized a *de novo* repeat library of the big-belly seahorse genome using RepeatModeler (version 2.0) (Flynn et al., 2020).

Then we merged the *de novo* library and the repeat consensus library in RepBase to annotate all repetitive elements in the genome using RepeatMasker (open-4.0.9) (Tarailo-Graovac & Chen, 2009). We used the Kimura 2-parameter distance to estimate the divergence level between the individual TE copies versus their consensus sequences. Kimura distances between the genome copies and TE consensus were calculated using the RepeatMasker built-in scripts (`calcdivergencefromalign.pl`). For gene annotation, we combined resources of sequence homology, *de novo* prediction and transcriptome to build consensus gene models using the GETA pipeline (<https://github.com/chenlianfu/geta>). The reference protein sequences of fugu (GCA_901000725.2), medaka fish (GCA_002234675.1) and zebrafish (GCA_000002035.4) were downloaded from the Ensembl database to perform homology-based prediction. The GETA pipeline randomly selected 1000 homology-based genes to train Augustus (Stanke et al., 2004) for *de novo* prediction on the premasked genome sequences. We provided the pipeline with RNA-seq reads of pooled tissue samples (brain, liver, testis, ovary, heart, kidney, brood pouch, muscle and skin) to assemble the transcriptomes. We used Trimmomatic (Bolger et al., 2014) implemented in the GETA pipeline to first filter the RNA-seq reads with default parameters. The filtered reads were aligned to the reference genome with hisat2 (Kim et al., 2015). The genomic regions with expression were identified based on RNA-seq alignment. We further assembled and filtered the transcripts with “`sam2transfrag`” function in GETA pipeline, which keeps the isoform with the highest RNA-seq read depth for each gene as the representative transcript. Finally, gene models from these three methods were combined into a nonredundant gene set.

To identify the nearly complete functional gene repertoire of *Hox* genes in the investigated species, we first collected the known amino acid sequences of intact *Hox* genes from three species (medaka, fugu and zebrafish) from Ensembl (<https://asia.ensembl.org>) and NCBI (<http://www.ncbi.nlm.nih.gov/>). Then we performed a TBLastN search with the cutoff E-value of $1E-5$ against the genomic data using these query sequences (Altschul et al., 1990). Aligned sequence fragments were combined into one predicted gene using perl scripts if they belonged to the same query protein. Each candidate gene region was then extended for 5 kb from both ends to predict its open reading frame by GeneWise (Birney et al., 2004). Obtained sequences were verified as corresponding genes by BlastP searches against the NCBI nonredundant (nr) database.

2.3 | Evolutionary analyses

A phylogenetic tree of the big-belly seahorse and the other sequenced genomes (*Hippocampus comes*, *Hippocampus erectus*, *Syngnathus scovelli*, *Oryzias latipes*, *Tetraodon nigroviridis*, *Gasterosteus aculeatus* and *Danio rerio*) was constructed using the 3,700 orthologous single-copy genes (Howe et al., 2013; Jaillon et al., 2004; Jones et al., 2012; Kasahara et al., 2007; Lin et al., 2016, 2017; Small et al., 2016). PhyML (Guindon et al., 2010) was used to construct the phylogenetic tree and estimate the evolutionary rates. To estimate the

divergence times between species, for each species, 4-fold degenerate sites were extracted from each orthologous family and concatenated to one sequence for each species. The MCMCTree program implemented in the phylogenetic analysis by maximum likelihood (PAML) package (Yang, 2007) was used to estimate the species divergence time. Calibration time was obtained from the fossil record time provided by <http://www.fossilrecord.net/dateaclade/index.html>. Two calibration points were applied in this study as normal priors to constrain the age of the nodes described below. 96.9–150.9 million years (Ma) for the most recent common ancestor (TMRCA) of medaka and pufferfish; 149.85–165.2 Ma for TMRCA of zebrafish and stickleback. The neutral mutation rate was calculated with the formula $u = d/2T$ (u : neutral mutation rate, d : synonymous substitution rate per synonymous site (dS), T : divergence time) (Li, 1976). We scaled branch-specific estimates of dS based on four-fold degenerate sites with estimated divergence time as proxies for rate of mutation.

To examine the evolution of gene families in seahorse, genes from tiger-tail, lined and big-belly seahorses, medaka, green spotted pufferfish, three-spined stickleback and zebrafish were clustered into gene families by Treefam (Li et al., 2006) ($\text{min_weight} = 10$, $\text{min_density} = 0.34$, and $\text{max_size} = 500$). The family expansion or contraction analysis was performed by CAFÉ (De Bie et al., 2006). In CAFE, a random birth-and-death model was proposed to study gene gain and loss in gene families across a user-specified phylogenetic tree. A global parameter λ (lambda), which described both gene birth (λ) and death ($\mu = -\lambda$) rate across all branches in the tree for all gene families, was estimated using the maximum likelihood method. A conditional p -value was calculated for each gene family, and the families with conditional p -values lower than 0.05 were considered to have a significantly accelerated rate of expansion and contraction. Gene ontology (GO) analysis of expanded and contracted gene families were performed by Metascape (Zhou et al., 2019) using orthologues in zebrafish with default parameters.

For the PAML analysis, we first assigned the orthologous relationship for 6,735 gene groups among all studied Clupeocephala (8 species) using the reciprocal best blast hit algorithm and syntentic information. We used PRANK (Löytynoja & Goldman, 2005) to align the orthologous gene sequences, which takes phylogenetic information into account when placing a gap into the alignment. We filtered the PRANK alignments by gblocks (Talavera & Castresana, 2007) and excluded genes with high proportion of low complexity or repetitive sequences to avoid alignment errors. To identify the genes that evolve under positive selection (PSGs), we performed a likelihood ratio test (LRT) using the branch model by PAML. We first performed a LRT of the two-ratio model, which calculates the dN/dS ratio for the lineage of interest and the background lineage, against the one-ratio model assuming a uniform dN/dS ratio across all branches, so that to determine whether the focal lineage is evolving significantly faster ($p < 0.05$). To differentiate between episodes of positive selection and relaxation of purifying selection (RSGs), we performed a LRT of two-ratios model against the model that fixed the focal lineage's dN/dS ratio to be 1 ($p < 0.05$). For the identified

RSGs and PSGs, we used their mouse and zebrafish orthologues mutant phenotype information and performed enrichment analyses using MamPhEA (Weng & Liao, 2010) and modPhEA (Weng & Liao, 2017), a method that has been used for inferring the function of other non-model organisms (Lv et al., 2020; Tollis et al., 2019; Yin et al., 2016). Then we grouped the enriched MP terms by different tissue types. To further investigate the expression pattern of PSGs and RSGs with mouse phenotypes, we used published mouse transcriptome data of different tissues (Noguchi et al., 2017). GO analysis of PSGs and RSGs were performed by Metascape (Zhou et al., 2019) using orthologues in zebrafish with default parameters.

2.4 | Demographic history analysis

We investigated the demographic histories of the different synanthid species with whole-genome data sets using the pairwise sequentially Markovian coalescent (PSMC) approach (Li & Durbin, 2011), which can infer changes in effective population size (N_e) within 10,000 to 1,000,000 years. We performed PSMC with the heterozygous SNP loci produced by the GATK pipeline (McKenna et al., 2010). Parameters were set to "N30 -t5 -r5 -p 4+30*2+4+6+10" and bootstrapping (100 times) was performed for each species to determine variance in N_e estimates. We used the estimated values of generation time and mutation rate to scale the results. We applied branch-specific estimates of the synonymous substitution rate per synonymous site (dS) from our dated phylogeny as proxies for the rate of mutation. The age of sexual maturity was collected from published results (Hausmann, 2001; Vincent et al., 2011) (<https://web.archive.org/web/20111101033410/http://www.seahorsesource.com>) and then multiplied by a factor of two as a proxy for generation time.

3 | RESULTS

3.1 | Genome assembly and annotation of the big-belly seahorse

To generate a high-quality reference genome, we produced 121.41-fold genomic coverage of PacBio reads (subread N50 length 16.3 kb), 109.5-fold Illumina reads and 119.03-fold Hi-C data from big-belly seahorse male individuals (Table S1). We tried different long-read assemblers including CANU, wtdbg2 and smartdenovo (Koren et al., 2017; Liu et al., 2021; Ruan & Li, 2020), and decided to use the contig assembly produced by smartdenovo because it outperforms CANU and wtdbg2 in terms of BUSCO completeness and/or contig continuity (Tables 1 and 2). The genome assessment showed a BUSCO completeness of 94.0% from the genome produced by smartdenovo, slightly higher than that derived from CANU (93.9%); however, the smartdenovo assembly is much more continuous than CANU assembly (contig N50: 8.4 Mb vs. 0.1 Mb). Although wtdbg2 shows an outstanding contig N50 (10.8 Mb), the BUSCO completeness is only

TABLE 1 BUSCO assessment of contig-level assemblies

Description	smartdenovo		CANU		Wtdbg2	
	Number	Perct. (%)	Number	Perct. (%)	Number	Perct. (%)
Complete BUSCOs (C)	4310	94.0	4302	93.9	3763	82.1
Complete and single-copy BUSCOs (S)	4169	90.9	2263	49.4	3669	80.0
Complete and duplicated BUSCOs (D)	141	3.1	2039	44.5	94	2.1
Fragmented BUSCOs (F)	151	3.3	111	2.4	362	7.9
Missing BUSCOs (M)	123	2.7	171	3.7	459	10.0
Total BUSCO groups searched	4584	100	4584	100	4584	100

TABLE 2 Statistics of contig assemblies

	smartdenovo	canu	Wtdbg2
No. of contigs	186	10326	308
Max length (Mb)	20.7	8.0	32.8
Assembly size (Mb)	444.7	864.2	432.5
Contig N90 (Mb)	1.9	0.04	2.0
Contig N50 (Mb)	8.4	0.1	10.8
Average (Mb)	2.4	0.08	1.4

82.1%. This contig assembly derived from PacBio reads alone produced 186 gapless contig sequences. Using 119.03-fold Hi-C data derived from muscles followed by manual curation (see Materials and Methods), these contigs were oriented and connected into 168 scaffolds with an N50 size of 21.77 Mb (Table S2). We performed the final gap filling and assembly polishing using corrected PacBio reads and cleaned Illumina data compensatory to each other, and assembled the entire genome into 444.7 Mb sequences with only 163 gaps left in the sequences. About 441.8 Mb sequences (99.3%) of the assembled genome were anchored into 21 linkage groups (Figure 1a, Table S3). Genome alignment between *H. abdominalis* and *S. scovelli* indicated that linkage group 11 in *H. abdominalis* has two homologous linkage groups in *S. scovelli*. In linkage group 11, these two homologous regions are separated by a highly repetitive region (Figure S1). The genome assembly completeness analysis identified 95.8% complete Actinopterygii BUSCO genes (94.6% complete and single copy) and only 3.0% missing BUSCOs. This indicated that the assembly covered most of the coding regions of the genome (Table S4). The big-belly seahorse genome assembly (ZJU1.0) has a dramatic 11-fold increase of continuity measured by scaffold N50 size relative to other publicly available seahorse genome assemblies (Lin et al., 2016, 2017) (Table S5). The distribution of the scaffold lengths has shifted toward much longer sequences in the big-belly seahorse assembly relative to other Syngnathidae species (Figure 1b).

Using query protein sequences of multiple teleost species, and transcriptome data of pooled seahorse tissue samples (including brain, liver, testis, ovary, heart, kidney, brood pouch, muscle and skin), we combined homology and *de novo* prediction methods and annotated a total of 22,903 consensus protein-coding genes (Table S6). The average lengths of exons and introns of these genes are 163 and 930 bp, respectively. This gene number is close to those

of other seahorse genomes. 35.63% of the big-belly seahorse genome consists of transposable elements (TEs), a higher percentage than any other Syngnathidae genomes (17.17%–24.11%) but Manado pipefish (53%) analysed so far (Lin et al., 2016, 2017; Small et al., 2016; Zhang et al., 2020). This can be explained either by the much more intact genome assembly, or lineage-specific expansion of TEs of the big-belly seahorse. Supporting the latter, we found that the big-belly seahorse genome shared one peak of TEs with a similar divergence level (Kimura 2-parameter divergence level around 20%) compared to the TE consensus sequences with tiger tailed and lined seahorses, and had a specific peak of TE divergence level that was not present compared to the reported results of other syngnathid species (Zhang et al., 2020). This probably corresponds to a very recent (Kimura divergence level <5%) TE burst only in the big-belly seahorse (Figure 1c). The expanded TEs are concentrated in DNA transposon TcMar-Tigger (3.86% of the genome vs. 1.1% of the tiger tail seahorse), LTR transposon Ngarg (3.94% of the genome vs. 2.04% of the tiger tail seahorse) and uncategorized TEs (Table S7). The chromosomes of the big-belly seahorse consist almost solely of GC-intermediate isochores L2, H1 and H2 (Figure 1d, Figure S2), which is different to those of zebrafish (*Brachydanio rerio*), medaka (*Oryzias latipes*), stickleback (*Gasterosteus aculeatus*), and pufferfish (*Tetraodon nigroviridis*) (Costantini et al., 2007). Similar to the human, the gene density is positively associated with the gradient of GC isochores in seahorses (Figure 1e). Interestingly, the local GC content of macrochromosomes shows variation similar to microchromosomes (Figure S2).

3.2 | Phylogenomics and demographic history

We compared the colinearity of our new chromosome-level genome of the big-belly seahorse to the other two scaffold-level genomes of tiger-tail and lined seahorses, and the chromosome-level genome of Gulf pipefish (Figure 2a–c). It was clear that the other two Illumina-based seahorse genomes were quite fragmented, with a total of over 100 scaffolds whose lengths are longer than 1 Mb mapped, and on average 6.6 scaffolds mapped to each linkage group of the big-belly seahorse. To provide a phylogenetic framework for all evolutionary inferences in this work, we constructed a phylogenomic tree using coding sequences of 3700 single-copy orthologous genes shared

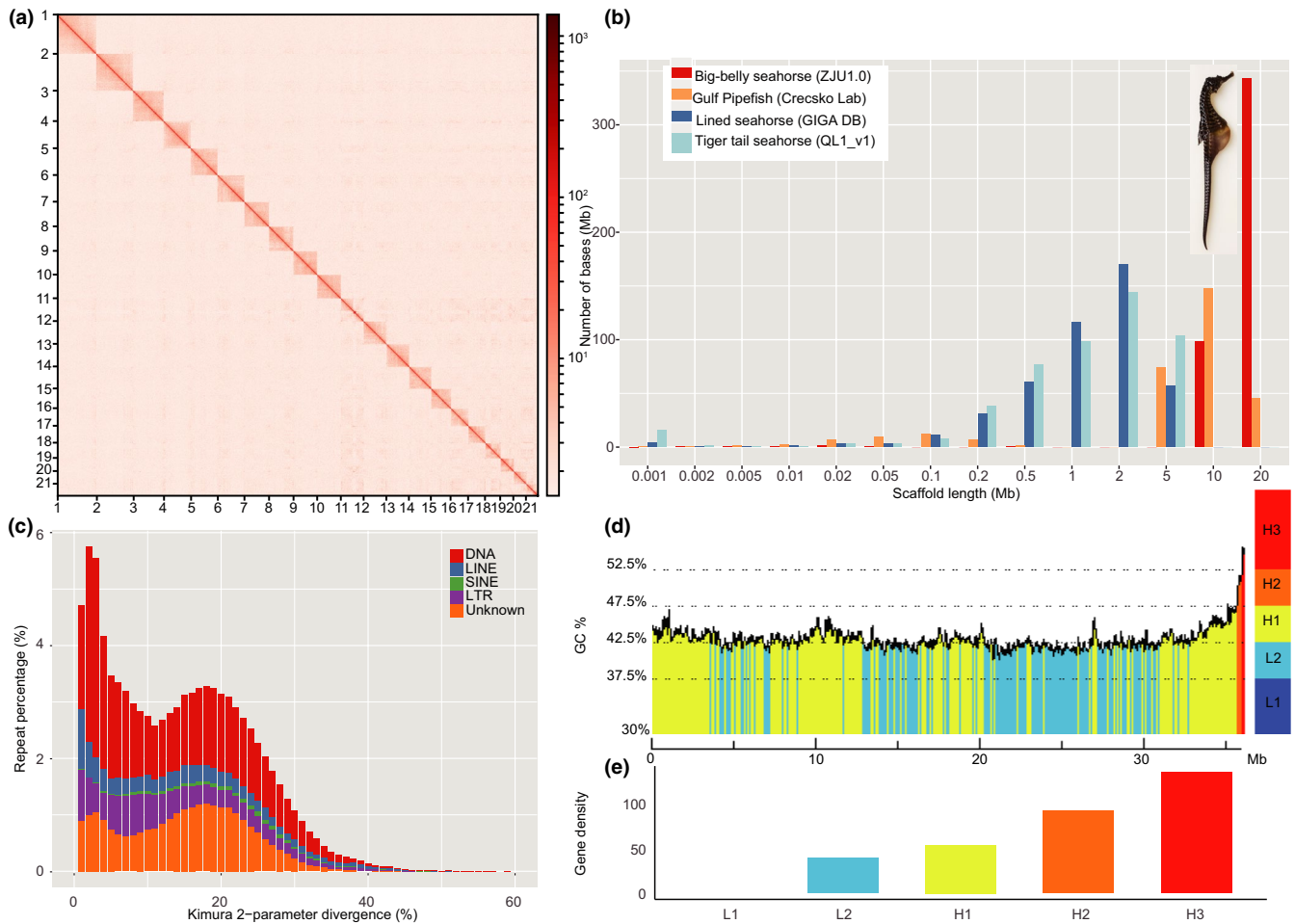


FIGURE 1 Genomic features of the big-belly seahorse, *Hippocampus abdominalis*. (a) The chromosome-level genome assembly of the big-belly seahorse was generated by combining PacBio long reads and Hi-C reads. The Hi-C contact map here shows the genome-wide all-by-all interactions of 21 chromosomes, with little off-diagonal interactions between chromosomes in this curated assembly. (b) Comparison of the scaffold length distributions between the big-belly seahorse versus the other three Syngnathidae species produced by Illumina reads. (c) The histogram shows the distributions of sequence divergence between each repeat family versus their consensus sequences. (d) Compositional overview of chromosome 1 and its GC levels. The colour-coded map shows 100 kb nonoverlapping sliding window plots. The colour code spans the spectrum of distinct GC levels, indicated by broken horizontal lines, from blue (GC-poorest isochores) to red (GC-richest isochores) (e) Gene density (gene number per Megabase) of GC isochores

between four Syngnathidae species and four outgroup teleost species (Figure S3), with the maximum likelihood method (Guindon et al., 2010) (Figure 2d). The resulting phylogeny has a 100% full bootstrapping support for all the studied nodes. It places the big-belly seahorse as the sister species to tiger-tail and lined seahorses, consistent with the reported result that the big-belly seahorse is the earliest branching species among *Hippocampus* species (Li, Olave, et al., 2021; Wang et al., 2019; Wilson & Rouse, 2010). We estimated that seahorses diverged from the gulf pipefish about 45.3 (29.9–84.1) million years ago (MYA), and syngnathid fishes and medaka diverged 134.0 (111.7–158.6) MYA assuming a molecular clock. These results are consistent with the oldest known syngnathid fossil (48–50 MYA) (Bellwood, 1996; Teske & Beheregaray, 2009). However, the estimated divergence time between the big-belly seahorse and the ancestor of tiger-tail and lined seahorses is different from that of recently published results (11.4 vs. 23.1 million years) (Li, Olave,

et al., 2021). This is probably because of different reconstruction methods (maximum likelihood vs. neighbour-joining trees), calibration points and a greater number of loci used in this study (3700 vs. 100) (Materials and Methods). Since we used many more loci and also maximum likelihood methods for phylogenetic inference, our results might have more support than that previously reported.

Previous comparative analysis of *H. comes* revealed its higher evolutionary rate than other studied teleosts (Lin et al., 2016). Here, we showed that this elevation of genome-wide substitution rate is concentrated in the ancestor of Syngnathidae species; all studied seahorse species, however, showed a low genome-wide substitution rate compared to each other, or to other teleosts, after they diverged from their most recent common ancestor (Figure 2d). This suggested that most morphological and reproductive behavioural changes shared by seahorse and pipefish are probably caused by the rapid genome evolution in their common ancestor. While the following

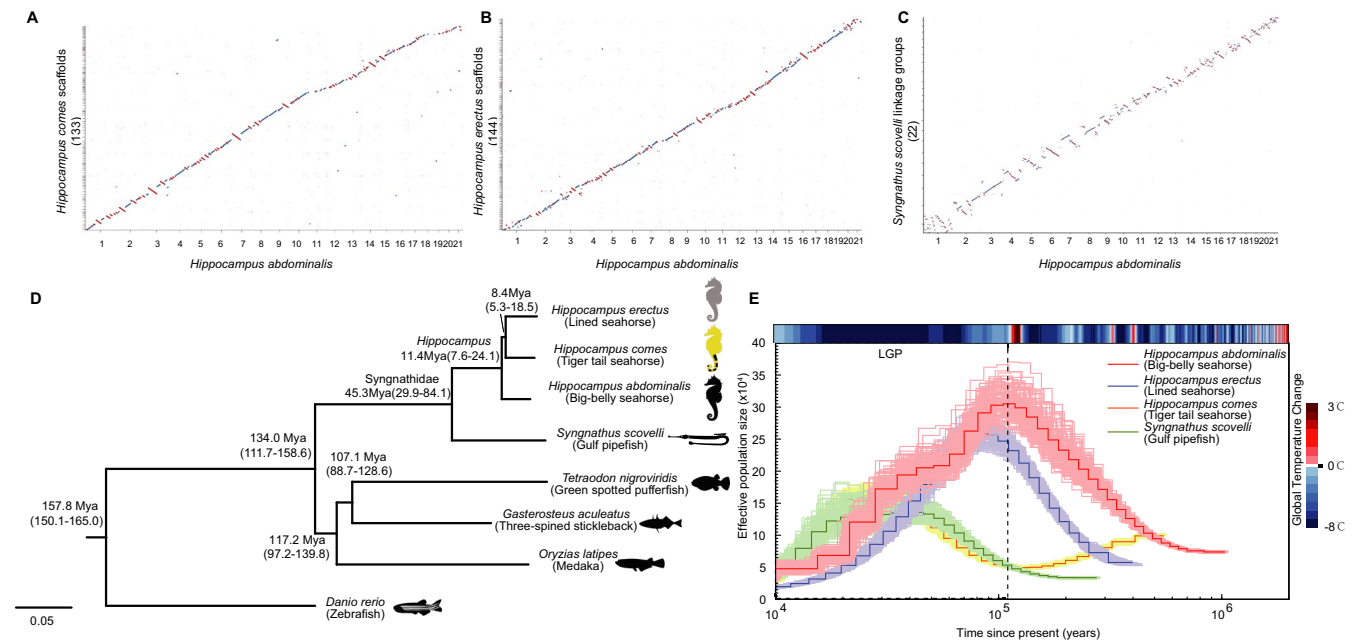


FIGURE 2 Phylogeny and demographic history of the representative Syngnathidae species. (a–c) Collinearity analysis between *H. abdominalis* vs. *H. comes*, *H. erectus* and *S. scovelli*. We performed the collinearity analysis between *H. abdominalis* linkage groups and the scaffolds longer than 1 Mb in *H. comes* and *H. erectus* and linkage groups in *S. scovelli*. Blue segments represent alignments to the positive strand of *H. abdominalis*. Red segments represent alignments to the negative strand of *H. abdominalis*. Numbers in the parentheses are the number of scaffolds longer than 1Mb in *H. comes* and *H. erectus*, and the number of linkage groups in *S. scovelli*. (d) Maximum likelihood tree reconstructed using single-copy orthologous genes. Branch lengths are scaled to the specific substitution rates estimated by PhyML. The numbers on each branch indicate the estimated divergence time (in million years), and 95% highest posterior densities. All phylogenetic nodes have full bootstrap support. (e) PSMC-inferred trajectories of four syngnathid species. Coloured bold line of each species is the population size dynamics inferred from PSMC analyses, with the lighter, thinner lines indicating variations in population size derived from 100 bootstraps. Top rectangles show respective time periods with global temperature changes. LGP: last glacial period

genomic deceleration among seahorses may be associated with their much reduced metabolic rate and sedentary lifestyle compared to other teleosts. Systematic studies in the future of more seahorse genomes would provide a more detailed picture on the change of substitution rate across different seahorse species.

We estimated the neutral mutation rate of the big-belly seahorse to be 6.97×10^{-9} substitutions/site/year based on four-fold degenerate (4D) sites (Table S8). This estimate and allele frequencies estimated from the published genomic data (Table S9) (Li, Olave, et al., 2021; Lin et al., 2017; Small et al., 2016) were used to reconstruct the historical fluctuations of population size (N_e) of studied Syngnathidae species, including one pipefish and three seahorse species, with the pairwise sequentially Markovian coalescent (PSMC) (Li & Durbin, 2011) method. We found large variations of N_e among the studied species, whose mean values over time (Table S8) ranged from 135,984–169,503 in lined and big-belly seahorses, to approximately 82,964–92,222 in gulf pipefish and tiger-tail seahorse from approximately 10 million to 10 thousand years ago. Specifically, the N_e of big-belly and lined seahorses reached their maximum sizes around 100,000 years ago, when there was a high global seawater level (Miller et al., 2005) associated with the warm climate. This historical expansion of N_e has also been reported in many other seahorse (Li, Olave, et al., 2021) and teleost (Li, Bian, et al., 2021; Xu et al.,

2019) (e.g., medaka and half-smooth tongue sole) species. While all the studied species exhibited a decline of N_e during the last glacial period (LGP) between 110,000 and 12,000 years ago (Figure 2e). Such a decline is probably associated with the considerable loss of shallow water habitat caused by the reduction of sea level and global temperature (Miller et al., 2005). Two species (lined and tiger-tail seahorses) in our study were classified as vulnerable on the IUCN Red List of Threatened Species (<http://www.iucnredlist.org/>) (Table S8). Our results showed that these vulnerable and least concerned species have experienced a population size reduction predating the more recent declines impacted by human activities.

3.3 | Evolution of seahorse gene families

Turnover of genes or gene family numbers is one of the major mechanisms underlying the phenotypic divergence of related species. There are 10,491 homologous gene groups shared by four species (big-belly, tiger-tail and lined seahorse and gulf pipefish) (Figure 3a). Besides, 1118 gene families were specific to seahorses, and 268 gene families were specifically found only in the big-belly seahorse. Interestingly, the MHC protein complex genes, including *B2M*, *H2-Ab1* and *HLA-DPB1*, are enriched ($p = 8.53 \times 10^{-5}$,

Chi-squared test) among the seahorse-specific gene families, and *B2M* has been reported to play an important role during the evolution of male pregnancy and showing copy number variation

among Syngnathiformes species (Table S10) (Roth et al., 2020). Several seahorse-specific genes related to protein and ion transport, including *SELENBP1*, *ATP2C1*, and *AQP3B*, are also reported

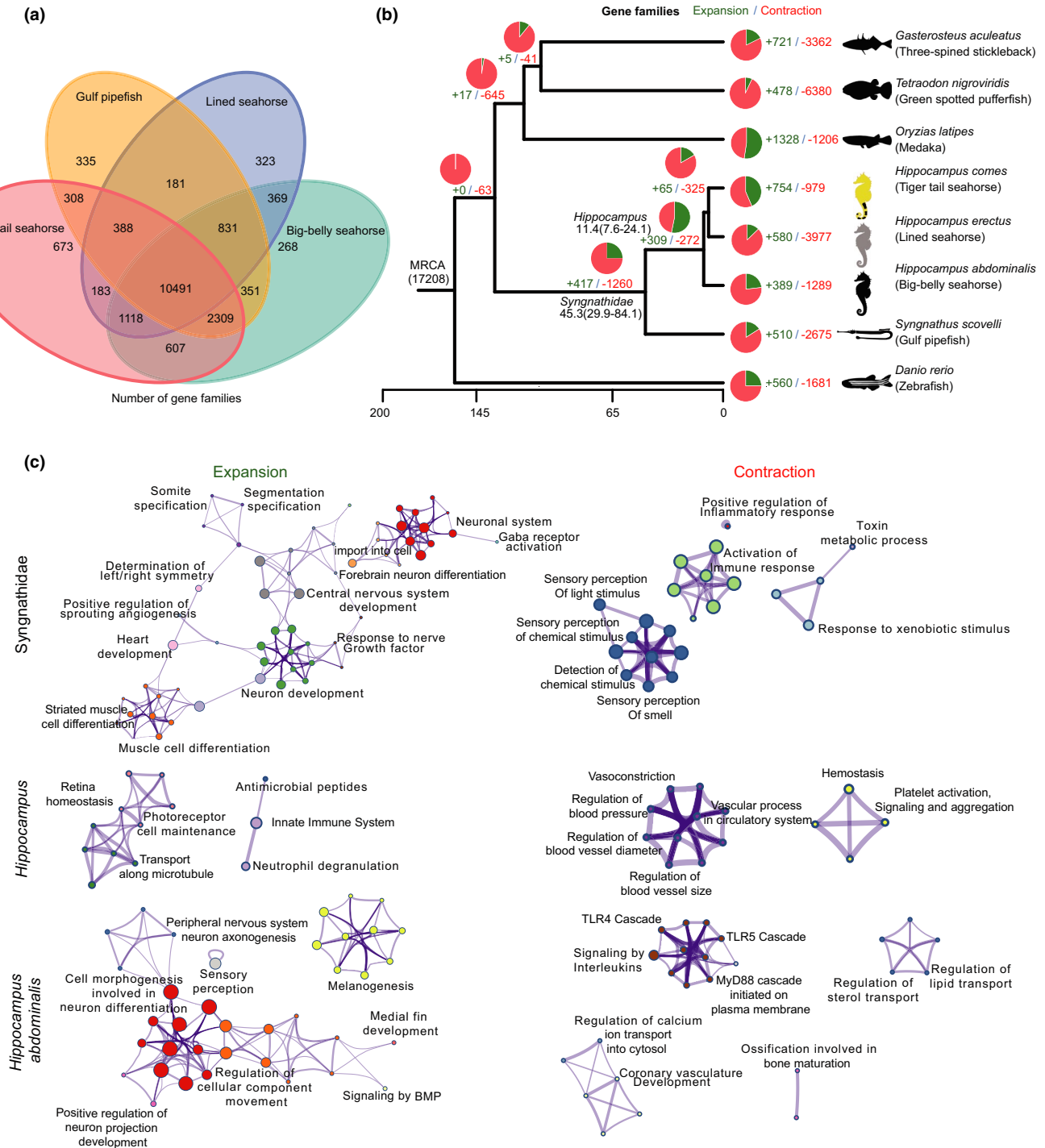


FIGURE 3 Evolution of gene families of the big-belly seahorse. (a) Venn diagram showing the specific and shared gene families of the four Syngnathidae species. (b) The inferred numbers of expanded (green) or contracted (red) gene families at each phylogenetic node of Syngnathidae species. We labelled the divergence time on each node. MRCA: most recent common ancestor. Animal icons are made by Freepik from www.flaticon.com, from <https://thenounproject.com> and from <http://phylopic.org>. (c) Enriched GO terms of expanded and contracted gene families identified by Metascape (Zhou et al., 2019). Nodes are coloured by their identities. Each node represents one enriched term. Within each branch, the size of nodes represents the percentage of input genes belonging to each GO term. Terms with Kappa scores (Cohen, 1960) >0.3 are connected by edges, the thicker, the higher similarity

to be upregulated during pregnancy (Table S11) (Whittington et al., 2015).

To systematically identify the gene gain and loss events across Syngnathidae species, we next mapped the expansions and contractions of gene families onto their phylogenetic tree (Figure 3b). Compared to other teleosts, there were 1260 gene families that have undergone contraction in the ancestor of Syngnathidae. These families are enriched at various immune-related gene ontology (GO) terms such as regulation of immune response, T cell receptor signaling pathway, positive regulation of inflammatory response etc. Contraction of specific innate immunity Toll-like receptor (TLR) pathway gene families, *TLR2*, 4 and 5 cascades were also found specifically in *H. abdominalis* (Table S12). Some gene families (e.g., *CASPA*) of the “innate immune systems” or “antimicrobial peptides” GO term are instead significantly expanded in the Hippocampus ancestor. These results are consistent with the reported complex turnovers of immune-related gene families possibly associated with development of male pouch among syngnathids (Figure 3c, Tables S11–S13). Inflammatory response related gene *TLR18*, which has undergone expansion in the ancestor of Syngnathidae, is also reported to be upregulated during pregnancy of *H. abdominalis* (Whittington et al., 2015). *NSMAF*, a gene related to leucocyte chemotaxis during immune response, has undergone copy number expansion in the ancestor of Syngnathidae (Boeckle et al., 2012) and was also reported to be upregulated during pregnancy of broadnosed pipefish (Roth et al., 2020). Interestingly, several previously uncharacterized GO terms related to vascularization, for example, “vasoconstriction”, “regulation of blood vessel size”, “hemostasis” are significantly enriched among the contracted gene families in the Hippocampus ancestor. There might be adaptive loss in these gene families that is associated with the development of more sophisticated male brood pouch of seahorse relative to that of pipefish, when vascularization is a critical step. It is worth comparing the expression pattern of genes (e.g., endothelin receptors, Table S12) of such GO terms in future between the developing pouches of seahorse and pipefish to further elucidate their functional roles.

On the other hand, gene families of ion transport, for example, “calcium ion import”, “potassium ion transport”, “ammonium ion metabolic process”, “iron ion transport” etc., and those of the vascularization process, for example, “aorta development” and “artery development” are significantly expanded in the ancestor of Syngnathidae or Hippocampus, or specifically in *H. abdominalis* (Table S12). There are several interesting cases of ion transporter or metabolic genes that underwent significant family contraction or expansion. Their expression level has also been reported to be dynamically upregulated or downregulated in the male brood pouch of *H. abdominalis* during the pregnancy stages from conception to parturition (Whittington et al., 2015). These gene families include *SLC6a11a*, a sodium ion transmembrane transporter, *SLC25a14*, an anion transporter, *CETP*, which is involved in lipid transfer, and *CHIA.4* that is involved in the carbohydrate metabolic process, etc (Tables S11–S12). This suggested that some genes involved in nutrient exchange and osmoregulation in the male brood pouch, or the vascularization

process during the pouch development of seahorse have undergone adaptive expansion or contraction of family members.

3.4 | Rapidly evolving genes possibly related to the seahorse specific traits

We then focused on the genes that have a one-to-one orthologous relationship with other teleosts. We first examined the *Hox* gene cluster, which is responsible for body axis patterning and appendage development (Ferrier & Holland, 2001). We annotated a total of 44 *Hox* and 2 *Evx* genes organized in seven clusters (*HoxAa–HoxDb*) of the big-belly seahorse (Figure S4). Consistent with previously reported results (Hoegg et al., 2007; Small et al., 2016), the entire *HoxCb* cluster, *HoxB10a*, *HoxB8b*, *HoxD13a* were absent in the genomes of seahorse and other percomorphs (stickleback, pufferfish, medaka, etc.). By parsimony, we inferred that *HoxA7a* and *HoxC3a* were convergently lost in the ancestors of Syngnathidae and Tetraodontidae. Since the *Hox7* mutants in mice show rib defects (Chen et al., 1998), the loss of *Hox7* genes may account for the fact that both syngnathids and pufferfish do not have ribs (Chen et al., 1998). *HoxD11b* was convergently lost in *H. abdominalis* (it is present in the tiger tail seahorse genome: Lin et al., 2016), medaka and zebrafish. Its function remains to be elucidated in future by for example, the knockout experiments in sticklebacks or fugu that still possess the gene. *EvxBa* or *eve1* (a paralogue of *evenskipped* gene family) was specifically lost in the ancestor of Syngnathidae, which was previously suggested to be associated with the teeth loss of pipefish and seahorse (Small et al., 2016). Several *Hox* genes are also found to have undergone rapid evolution. *HoxB3a*, *HoxC5a*, *HoxC4a* and *HoxD3a* are identified as Hippocampus RSGs, and *HoxD12a*, *HoxC4a* and *HoxC9a* are identified as *H. abdominalis* RSGs (Table S14).

Next, we examined the evolutionary rates of seahorse genes in the context of Syngnathidae phylogeny. We hypothesized that the genes associated with seahorse specific traits have probably undergone rapid evolution as a result of either adaptive evolution, or relaxed selective constraints. The evolutionary rate was characterized by the lineage-specific ratios (ω) of nonsynonymous to synonymous substitution rates. We identified a total of 1503 genes that have probably undergone positive selection (positively selected genes, PSG) and 1873 with relaxed selective constraints (genes with relaxed selective constraints, RSG) at different branches, using a likelihood model and the conserved lineage-specific tests (Table S14). Assuming a similar function between orthologues, we then examined the expression patterns, GO terms and the mutant phenotypes of these genes' mouse and zebrafish orthologues to elucidate whether they are associated with the seahorse-specific traits (Tables S15–S20) (Weng & Liao, 2017). Many orthologues of the identified PSGs or RSGs among different Syngnathidae lineages showed an expected tissue-biased expression pattern in mouse that is related to its function (Figure S5). For example, some Hippocampus specific RSGs with inferred T cell related phenotypes based on their mouse

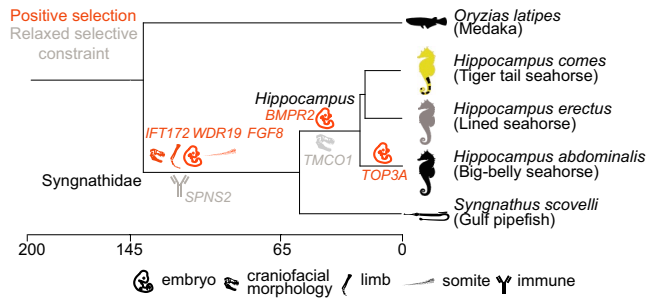


FIGURE 4 Rapidly evolving genes of seahorse species. Phylogenetic distribution of enriched mutant phenotypes (MP) of mouse orthologues of seahorse genes. Each MP term is shown by an organ icon, and significantly enriched for genes undergoing positive selection (PSG, red) or relaxed selective constraints (RSG, grey) inferred by lineage-specific PAML analyses. Organ icons are made from <https://thenounproject.com>. Gene examples that have undergone putative positive selection or relaxed selective constraints were labelled onto each branch

orthologue mutant phenotypes show a biased expression in mouse thymus.

Interestingly, we found the mutant phenotypes belonging to the terms of “abnormal craniofacial development”, “abnormal hindlimb morphology” and “abnormal mouth morphology” are significantly (Bonferroni corrected $p < 0.05$, Fisher’s exact test) enriched among PSGs in the Syngnathidae ancestor (Figure 4). The mouse mutant of one such pleiotropic PSG *IFT172* shows an abnormal morphology or shortened size of snout, limbs and spinal cord (Friedland-Little et al., 2011; Howard et al., 2010), and its zebrafish mutant shows a curved body shape (Lunt et al., 2009), which resembles that of seahorses. There are many more PSGs or RSGs whose mutant tail phenotypes are influenced in *Hippocampus* ancestor, or in *H. abdominalis* than those in the Syngnathidae ancestor (48 vs. 1 genes), which might be related to the specific trait of prehensile tail of the seahorses versus pipefish. Many of these gene mouse mutants (e.g., *SPRED1*, *KNTC1* and *MKX*, Table S17) show curly, wavy or kinked tails. The mutant phenotypes of embryonic lethality between implantation and placentaion or prenatal lethality, etc are enriched among the PSGs specifically in the *H. abdominalis* lineage. By contrast, RSGs are enriched for immune-related phenotypic terms such as increased NK T cell number in the ancestor of Syngnathidae.

4 | DISCUSSION

We produced the first publicly available chromosome-level genome of seahorse to facilitate the evolutionary and developmental studies of Syngnathidae species. Syngnathidae species have developed a unique body plan among teleosts, including different shapes of male brood pouch, elongated tube-like mouth, lack of ribs and pelvic fins and so on. The responsible genetic changes are expected to have occurred at the ancestor of Syngnathidae or specifically at the ancestor of seahorses (e.g., the upright posture) after they diverged from

other teleosts. Using the new genome of *H. abdominalis*, we first reconstructed the phylogeny and demographic history of syngnathids. We then identified many candidate genes that underwent gene family expansion or contraction, or episodes of positive selection or relaxed selective constraints. Consistent with the previous findings, seahorse and pipefish species both have significant modifications of immune gene repertoires, but in different gene families (Roth et al., 2020). This was hypothesized to be associated with the different levels of male pregnancy and parent-offspring conflict between the two groups of syngnathids. As seahorses have a more specialized and fully enclosed brood pouch compared to that of pipefish. In addition to the reported changes involving MHC pathway genes (Roth et al., 2020), we here found several TLR cascade genes have changed their copy numbers, and some (e.g., *TLR18*) also changes its expression during the pregnancy stages in the brood pouch, but in different lineages of Syngnathidae. We highlight those immune genes like *TLR18* that show both sequence changes and dynamic pouch expression patterns are more likely to be targets of positive selection in response to evolution of male pregnancy. We also identified many candidate PSG/RSG genes whose mouse orthologue mutants show craniofacial, limb or tail modifications. Future functional validation of these genes in other teleost model species, for example, zebrafish, are expected to shed light on the mechanisms underlying the acquisition of novel morphological traits in both seahorses and other vertebrates.

ACKNOWLEDGEMENTS

Animal icons are made by Freepik from www.flaticon.com, from <https://thenounproject.com> (including “limb” icon by designer468, “Thymus” icon by IcoLabs, “Testis” icon by Becris, “Ovary” icon by Dairy Free Design and “Kidney” icon by Alexander Blagochevsky) and from <http://phylopic.org>. This work was supported by the National Natural Science Foundation of China (32170415, 32061130208), Natural Science Foundation of Zhejiang Province (LD19C190001), European Research Council Starting Grant (grant agreement 677696) to Q. Z.; the Fund of Development of Special Projects in Marine and Fishery from Xiamen city (18CZY008HJ03), and Fujian Provincial Fund for Oceanology and Fishery (2019HYJG08) to L.Z..

AUTHOR CONTRIBUTIONS

Qi Zhou, Leyun Zheng, Shiyu Huang, and Libin He conceived the study. Libin He, Zhen Huang, Jianfei Qi and Shuiqing Wu collected the samples and data. Zongji Wang, Huiyu Luo, Xingtang Zhang, Jianfei Qi, Huiyu Luo, XinXin Chen, Jinbo Lin, Qihua Yang, Shiyu Huang, Qi Zhou and Leyun Zheng performed the analyses. Zongji Wang, Xin Long, Libin He, Qi Zhou and Leyun Zheng interpreted the results and wrote the manuscript. All authors read and approved the final manuscript.

DATA AVAILABILITY STATEMENT

All the genomic reads and the assembled genome and annotation generated in this study have been deposited on NCBI

Short Reads Archive (<https://www.ncbi.nlm.nih.gov/Traces/study/?acc=PRJNA713516>) under the BioProject Accession Number PRJNA713516, and in Genome Warehouse (GWH) database (<https://bigd.big.ac.cn/gwh/>) under the accession number PRJCA004882.

ORCID

Qi Zhou  <https://orcid.org/0000-0002-7419-2047>

REFERENCES

- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, 215(3), 403–410. [https://doi.org/10.1016/s0022-2836\(05\)80360-2](https://doi.org/10.1016/s0022-2836(05)80360-2)
- Bellwood, D. R. (1996). The Eocene fishes of Monte Bolca: The earliest coral reef fish assemblage. *Coral Reefs*, 15(1), 11–19. <https://doi.org/10.1007/bf01626074>
- Birney, E., Clamp, M., & Durbin, R. (2004). GeneWise and Genomewise. *Genome Research*, 14(5), 988–995. <https://doi.org/10.1101/gr.1865504>
- Boecker, A., Sieger, D., Neacsu, C. D., Kashkar, H., & Krönke, M. (2012). Factor associated with neutral sphingomyelinase activity mediates navigational capacity of leukocytes responding to wounds and infection: Live imaging studies in zebrafish larvae. *The Journal of Immunology*, 189(4), 1559–1566. <https://doi.org/10.4049/jimmu.nol.1102207>
- Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>
- Chen, F., Greer, J., & Capecchi, M. R. (1998). Analysis of Hoxa7/Hoxb7 mutants suggests periodicity in the generation of the different sets of vertebrae. *Mechanisms of Development*, 77(1), 49–57. [https://doi.org/10.1016/s0925-4773\(98\)00126-9](https://doi.org/10.1016/s0925-4773(98)00126-9)
- Cohen, J. (1960). A coefficient of agreement of nominal scales. *Educational and Psychological Measurement*, 20(1), 37–46. <https://doi.org/10.1177/001316446002000104>
- Costantini, M., Auletta, F., & Bernardi, G. (2007). Isochore patterns and gene distributions in fish genomes. *Genomics*, 90(3), 364–371. <https://doi.org/10.1016/j.ygeno.2007.05.006>
- De Bie, T., Cristianini, N., Demuth, J. P., & Hahn, M. W. (2006). CAFE: A computational tool for the study of gene family evolution. *Bioinformatics*, 22(10), 1269–1271. <https://doi.org/10.1093/bioinformatics/btl097>
- Dudchenko, O., Batra, S. S., Omer, A. D., Nyquist, S. K., Hoeger, M., Durand, N. C., & Aiden, E. L. (2017). De novo assembly of the genome using Hi-C yields chromosome-length scaffolds. *Science*, 356(6333), 92–95. <https://doi.org/10.1126/science.aal3327>
- Ferrier, D. E., & Holland, P. W. (2001). Ancient origin of the Hox gene cluster. *Nature Reviews Genetics*, 2(1), 33–38. <https://doi.org/10.1038/35047605>
- Flynn, J. M., Hubley, R., Goubert, C., Rosen, J., Clark, A. G., Feschotte, C., & Smit, A. F. (2020). RepeatModeler2 for automated genomic discovery of transposable element families. *Proceedings of the National Academy of Sciences of the United States of America*, 117(17), 9451–9457. <https://doi.org/10.1073/pnas.1921046117>
- Foster, S. J., & Vincent, A. C. J. (2004). Life history and ecology of seahorses: Implications for conservation and management. *Journal of Fish Biology*, 65(1), 1–61. <https://doi.org/10.1111/j.0022-1112.2004.00429.x>
- Friedland-Little, J. M., Hoffmann, A. D., Ocbina, P. J. R., Peterson, M. A., Bosman, J. D., Chen, Y., Cheng, S. Y., Anderson, K. V., & Moskowitz, I. P. (2011). A novel murine allele of intraflagellar transport protein 172 causes a syndrome including VACTERL-like features with hydrocephalus. *Human Molecular Genetics*, 20(19), 3725–3737. <https://doi.org/10.1093/hmg/ddr241>
- 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. *Systematic Biology*, 59(3), 307–321. <https://doi.org/10.1093/sysbio/syq010>
- Hausmann, J. S. (2001). *The relationship between sex roles, sex ratios, and sexual dimorphisms in the mating system of the Gulf Pipefish, Syngnathus Scovelli* (Ph.D. Thesis).
- Hoegg, S., Boore, J. L., Kuehl, J. V., & Meyer, A. (2007). Comparative phylogenomic analyses of teleost fish Hox gene clusters: lessons from the cichlid fish *Astatotilapia burtoni*. *BMC Genomics*, 8(1), 317. <http://dx.doi.org/10.1186/1471-2164-8-317>
- Howard, P. W., Howard, T. L., & Maurer, R. A. (2010). Generation of mice with a conditional allele for Ift172. *Transgenic Research*, 19(1), 121–126. <https://doi.org/10.1007/s11248-009-9292-x>
- Howe, K., Clark, M. D., Torroja, C. F., Torrance, J., Berthelot, C., Muffato, M., Collins, J. E., Humphray, S., McLaren, K., Matthews, L., McLaren, S., Sealy, I., Caccamo, M., Churcher, C., Scott, C., Barrett, J. C., Koch, R., Rauch, G.-J., White, S., ... Stemple, D. L. (2013). The zebrafish reference genome sequence and its relationship to the human genome. *Nature*, 496(7446), 498–503. <https://doi.org/10.1038/nature12111>
- Jaillon, O., Aury, J.-M., Brunet, F., Petit, J.-L., Stange-Thomann, N., Mauceli, E., Bouneau, L., Fischer, C., Ozouf-Costaz, C., Bernot, A., Nicaud, S., Jaffe, D., Fisher, S., Lutfalla, G., Dossat, C., Segurens, B., Dasilva, C., Salanoubat, M., Levy, M., ... Roest Crolius, H. (2004). Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype. *Nature*, 431(7011), 946–957. <https://doi.org/10.1038/nature03025>
- Jones, F. C., Grabherr, M. G., Chan, Y. F., Russell, P., Mauceli, E., Johnson, J., Swofford, R., Pirun, M., Zody, M. C., White, S., Birney, E., Searle, S., Schmutz, J., Grimwood, J., Dickson, M. C., Myers, R. M., Miller, C. T., Summers, B. R., Knecht, A. K., ... Kingsley, D. M. (2012). The genomic basis of adaptive evolution in threespine sticklebacks. *Nature*, 484(7392), 55–61. <https://doi.org/10.1038/nature10944>
- Kasahara, M., Naruse, K., Sasaki, S., Nakatani, Y., Qu, W., Ahsan, B., Yamada, T., Nagayasu, Y., Doi, K., Kasai, Y., Jindo, T., Kobayashi, D., Shimada, A., Toyoda, A., Kuroki, Y., Fujiyama, A., Sasaki, T., Shimizu, A., Asakawa, S., ... Kohara, Y. (2007). The medaka draft genome and insights into vertebrate genome evolution. *Nature*, 447(7145), 714–719. <https://doi.org/10.1038/nature05846>
- Kim, D., Langmead, B., & Salzberg, S. L. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nature Methods*, 12(4), 357–360. <https://doi.org/10.1038/nmeth.3317>
- Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H., & Phillippy, A. M. (2017). Canu: Scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Research*, 27(5), 722–736. <https://doi.org/10.1101/gr.215087.116>
- Leysen, H., Christiaens, J., De Kegel, B., Boone, M. N., Van Hoorebeke, L., & Adriaens, D. (2011). Musculoskeletal structure of the feeding system and implications of snout elongation in *Hippocampus reidi* and *Dunckerocampus dactylophorus*. *Journal of Fish Biology*, 78(6), 1799–1823. <https://doi.org/10.1111/j.1095-8649.2011.02957.x>
- Li, C., Olave, M., Hou, Y., Qin, G., Schneider, R. F., Gao, Z., Tu, X., Wang, X., Qi, F., Nater, A., Kautt, A. F., Wan, S., Zhang, Y., Liu, Y., Zhang, H., Zhang, B. O., Zhang, H., Qu, M., Liu, S., ... Lin, Q. (2021). Genome sequences reveal global dispersal routes and suggest convergent genetic adaptations in seahorse evolution. *Nature Communications*, 12(1), 1094. <https://doi.org/10.1038/s41467-021-21379-x>
- Li, H., Coghlan, A., Ruan, J., Coin, L. J., Hériché, J.-K., Osmotherly, L., & Durbin, R. (2006). TreeFam: A curated database of phylogenetic

- trees of animal gene families. *Nucleic Acids Research*, 34, D572–580. <https://doi.org/10.1093/nar/gkj118>
- Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 25(14), 1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>
- Li, H., & Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. *Nature*, 475(7357), 493–496. <https://doi.org/10.1038/nature10231>
- Li, J., Bian, C., Yi, Y., Yu, H., You, X., & Shi, Q. (2021). Temporal dynamics of teleost populations during the Pleistocene: A report from publicly available genome data. *BMC Genomics*, 22(1), 490. <https://doi.org/10.1186/s12864-021-07816-7>
- Li, W. H. (1976). Effect of migration on genetic distance. *The American Naturalist*, 110(975), 841–847. <https://doi.org/10.1086/283106>
- Lin, Q., Fan, S., Zhang, Y., Xu, M., Zhang, H., Yang, Y., Lee, A. P., Woltering, J. M., Ravi, V., Gunter, H. M., Luo, W., Gao, Z., Lim, Z. W., Qin, G., Schneider, R. F., Wang, X., Xiong, P., Li, G., Wang, K., ... Venkatesh, B. (2016). The seahorse genome and the evolution of its specialized morphology. *Nature*, 540(7633), 395–399. <https://doi.org/10.1038/nature20595>
- Lin, Q., Qiu, Y., Gu, R., Xu, M., Li, J., Bian, C., Zhang, H., Qin, G., Zhang, Y., Luo, W., Chen, J., You, X., Fan, M., Sun, M., Xu, P., Venkatesh, B., Xu, J., Fu, H., & Shi, Q. (2017). Draft genome of the lined seahorse, *Hippocampus erectus*. *GigaScience*, 6(6), 1–6. <https://doi.org/10.1093/gigascience/gix030>
- Liu, H., Wu, S., Li, A., & Ruan, J. (2021). SMARTdenovo: a de novo assembler using long noisy reads. *Gigabyte*, 2021, 1–9. <https://doi.org/10.46471/gigabyte.15>
- Löytynoja, A., & Goldman, N. (2005). An algorithm for progressive multiple alignment of sequences with insertions. *Proceedings of the National Academy of Sciences of the United States of America*, 102(30), 10557–10562. <https://doi.org/10.1073/pnas.0409137102>
- Lunt, S. C., Haynes, T., & Perkins, B. D. (2009). Zebrafish *ift57*, *ift88*, and *ift172* intraflagellar transport mutants disrupt cilia but do not affect hedgehog signaling. *Developmental Dynamics*, 238(7), 1744–1759. <https://doi.org/10.1002/dvdy.21999>
- Lv, M., Zhang, Y., Liu, K., Li, C., Wang, J., Fan, G., Liu, X., Yang, H., Liu, C., Mahboob, S., Liu, J., & Shao, C. (2020). A chromosome-level genome assembly of the anglerfish *Lophius litulon*. *Front. Genet*, 11, 581161. <https://doi.org/10.3389/fgene.2020.581161>
- Marçais, G., Delcher, A. L., Phillippy, A. M., Coston, R., Salzberg, S. L., & Zimin, A. (2018). MUMmer4: A fast and versatile genome alignment system. *PLoS Computational Biology*, 14(1), e1005944. <https://doi.org/10.1371/journal.pcbi.1005944>
- McEwen, G. K., Goode, D. K., Parker, H. J., Woolfe, A., Callaway, H., & Elgar, G. (2009). Early evolution of conserved regulatory sequences associated with development in vertebrates. *PLoS Genetics*, 5(12), e1000762. <https://doi.org/10.1371/journal.pgen.1000762>
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, 20(9), 1297–1303. <https://doi.org/10.1101/gr.107524.110>
- Miller, K. G., Kominz, M. A., Browning, J. V., Wright, J. D., Mountain, G. S., Katz, M. E., Sugarman, P. J., Cramer, B. S., Christie-Blick, N., & Pekar, S. F. (2005). The Phanerozoic record of global sea-level change. *Science*, 310(5752), 1293–1298. <https://doi.org/10.1126/science.1116412>
- Noguchi, S., Arakawa, T., Fukuda, S., Furuno, M., Hasegawa, A., Hori, F., Ishikawa-Kato, S., Kaida, K., Kaiho, A. I., Kanamori-Katayama, M., Kawashima, T., Kojima, M., Kubosaki, A., Manabe, R.-I., Murata, M., Nagao-Sato, S., Nakazato, K., Ninomiya, N., Nishiyori-Sueki, H., ... Hayashizaki, Y. (2017). FANTOM5 CAGE profiles of human and mouse samples. *Scientific Data*, 4, 170112. <https://doi.org/10.1038/sdata.2017.112>
- Paczolt, K. A., & Jones, A. G. (2010). Post-copulatory sexual selection and sexual conflict in the evolution of male pregnancy. *Nature*, 464(7287), 401–404. <https://doi.org/10.1038/nature08861>
- Robinson, J. T., Turner, D., Durand, N. C., Thorvaldsdóttir, H., Mesirov, J. P., & Aiden, E. L. (2018). Juicebox.js provides a cloud-based visualization system for Hi-C data. *Cell Systems*, 6(2), 256–258.e251. <https://doi.org/10.1016/j.cels.2018.01.001>
- Roth, O., Solbakken, M. H., Tørresen, O. K., Bayer, T., Matschiner, M., Baalsrud, H. T., Hoff, S. N. K., Briec, M. S. O., Haase, D., Hanel, R., Reusch, T. B. H., & Jentoft, S. (2020). Evolution of male pregnancy associated with remodeling of canonical vertebrate immunity in seahorses and pipefishes. *Proceedings of the National Academy of Sciences of the United States of America*, 117(17), 9431–9439. <https://doi.org/10.1073/pnas.1916251117>
- Ruan, J., & Li, H. (2020). Fast and accurate long-read assembly with wtdbg2. *Nature Methods*, 17(2), 155–158. <https://doi.org/10.1038/s41592-019-0669-3>
- Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, 31(19), 3210–3212. <https://doi.org/10.1093/bioinformatics/btv351>
- Skalkos, Z. M. G., Van Dyke, J. U., & Whittington, C. M. (2020). Paternal nutrient provisioning during male pregnancy in the seahorse *Hippocampus abdominalis*. *Journal of Comparative Physiology B*, 190(5), 547–556. <https://doi.org/10.1007/s00360-020-01289-y>
- Small, C. M., Bassham, S., Catchen, J., Amores, A., Fuiten, A. M., Brown, R. S., Jones, A. G., & Cresko, W. A. (2016). The genome of the Gulf pipefish enables understanding of evolutionary innovations. *Genome Biology*, 17(1), 258. <https://doi.org/10.1186/s13059-016-1126-6>
- Stanke, M., Steinkamp, R., Waack, S., & Morgenstern, B. (2004). AUGUSTUS: A web server for gene finding in eukaryotes. *Nucleic Acids Research*, 32, W309–W312. <https://doi.org/10.1093/nar/gkh379>
- Talavera, G., & Castresana, J. (2007). Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Systematic Biology*, 56(4), 564–577. <https://doi.org/10.1080/10635150701472164>
- Tarailo-Graovac, M., & Chen, N. (2009). Using RepeatMasker to identify repetitive elements in genomic sequences. *Current Protocols in Bioinformatics*, 25(1), Unit 4.10. <https://doi.org/10.1002/0471250953.bi0410s25>
- Teske, P. R., & Beheregaray, L. B. (2009). Evolution of seahorses' upright posture was linked to Oligocene expansion of seagrass habitats. *Biology Letters*, 5(4), 521–523. <https://doi.org/10.1098/rsbl.2009.0152>
- Tollis, M., Robbins, J., Webb, A. E., Kuderna, L. F. K., Caulin, A. F., Garcia, J. D., Bèrubè, M., Pourmand, N., Marques-Bonet, T., O'Connell, M. J., Palsbøll, P. J., & Maley, C. C. (2019). Return to the sea, get huge, beat cancer: An analysis of cetacean genomes including an assembly for the humpback whale (*Megaptera novaeangliae*). *Molecular Biology and Evolution*, 36(8), 1746–1763. <https://doi.org/10.1093/molbev/msz099>
- Vaughn, J. N., Ellingson, S. R., Mignone, F., von Arnim, A. (2012). Known and novel post-transcriptional regulatory sequences are conserved across plant families. *RNA*, 18(3), 368–384. <https://doi.org/10.1261/rna.031179.111>
- Vincent, A. C. J., Foster, S. J., & Koldewey, H. J. (2011). Conservation and management of seahorses and other Syngnathidae. *Journal of Fish Biology*, 78(6), 1681–1724. <https://doi.org/10.1111/j.1095-8649.2011.03003.x>
- Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C. A., Zeng, Q., Wortman, J., Young, S. K., & Earl, A. M.

- (2014). Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One*, 9(11), e112963. <https://doi.org/10.1371/journal.pone.0112963>
- Wang, X., Han, X., Zhang, Y., Liu, S., & Lin, Q. (2019). Phylogenetic analysis and genetic structure of the seahorse, *Hippocampus fuscus* from the Arabian and Red Sea based on mitochondrial DNA sequences. *Mitochondrial DNA Part A*, 30(1), 165–171. <https://doi.org/10.1080/24701394.2018.1467410>
- Wang, X., Zhang, Y., Zhang, H., Qin, G., & Lin, Q. (2019). Complete mitochondrial genomes of eight seahorses and pipefishes (Syngnathiformes: Syngnathidae): Insight into the adaptive radiation of syngnathid fishes. *BMC Evolutionary Biology*, 19(1), 119. <https://doi.org/10.1186/s12862-019-1430-3>
- Weng, M.-P., & Liao, B.-Y. (2010). MamPhEA: A web tool for mammalian phenotype enrichment analysis. *Bioinformatics*, 26(17), 2212–2213. <https://doi.org/10.1093/bioinformatics/btq359>
- Weng, M.-P., & Liao, B.-Y. (2017). modPhEA: Model organism phenotype enrichment analysis of eukaryotic gene sets. *Bioinformatics*, 33(21), 3505–3507. <https://doi.org/10.1093/bioinformatics/btx426>
- Whittington, C. M., Griffith, O. W., Qi, W., Thompson, M. B., & Wilson, A. B. (2015). Seahorse brood pouch transcriptome reveals common genes associated with vertebrate pregnancy. *Molecular Biology and Evolution*, 32(12), 3114–3131. <https://doi.org/10.1093/molbev/msv177>
- Wilson, A. B., & Orr, J. W. (2011). The evolutionary origins of Syngnathidae: Pipefishes and seahorses. *Journal of Fish Biology*, 78(6), 1603–1623. <https://doi.org/10.1111/j.1095-8649.2011.02988.x>
- Wilson, N. G., & Rouse, G. W. (2010). Convergent camouflage and the non-monophyly of 'seadragons' (Syngnathidae: Teleostei): suggestions for a revised taxonomy of syngnathids. *Zoologica Scripta*, 39(6), 551–558. <https://doi.org/10.1111/j.1463-6409.2010.00449.x>
- Xu, S., Zhao, L., Xiao, S., & Gao, T. (2019). Whole genome resequencing data for three rockfish species of *Sebastes*. *Scientific Data*, 6(1), 97. <https://doi.org/10.1038/s41597-019-0100-z>
- Yang, Z. (2007). PAML 4: Phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, 24(8), 1586–1591. <https://doi.org/10.1093/molbev/msm088>
- Yin, W., Wang, Z.-J., Li, Q.-Y., Lian, J.-M., Zhou, Y., Lu, B.-Z., Jin, L.-J., Qiu, P.-X., Zhang, P., Zhu, W.-B., Wen, B. O., Huang, Y.-J., Lin, Z.-L., Qiu, B.-T., Su, X.-W., Yang, H.-M., Zhang, G.-J., Yan, G.-M., & Zhou, Q. I. (2016). Evolutionary trajectories of snake genes and genomes revealed by comparative analyses of five-pacer viper. *Nature Communications*, 7, 13107. <https://doi.org/10.1038/ncomms13107>
- Zhang, X., Zhang, S., Zhao, Q., Ming, R., & Tang, H. (2019). Assembly of allele-aware, chromosomal-scale autopolyploid genomes based on Hi-C data. *Nature Plants*, 5(8), 833–845. <https://doi.org/10.1038/s41477-019-0487-8>
- Zhang, Y.-H., Ravi, V., Qin, G., Dai, H. E., Zhang, H.-X., Han, F.-M., Wang, X., Liu, Y.-H., Yin, J.-P., Huang, L.-M., Venkatesh, B., & Lin, Q. (2020). Comparative genomics reveal shared genomic changes in syngnathid fishes and signatures of genetic convergence with placental mammals. *National Science Review*, 7(6), 964–977. <https://doi.org/10.1093/nsr/nwaa002>
- Zhou, Y., Zhou, B., Pache, L., Chang, M., Khodabakhshi, A. H., Tanaseichuk, O., Benner, C., & Chanda, S. K. (2019). Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nature Communications*, 10(1), 1523. <https://doi.org/10.1038/s41467-019-09234-6>

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: He, L., Long, X., Qi, J., Wang, Z., Huang, Z., Wu, S., Zhang, X., Luo, H., Chen, X., Lin, J., Yang, Q., Huang, S., Zhou, Q., & Zheng, L. (2021). Genome and gene evolution of seahorse species revealed by the chromosome-level genome of *Hippocampus abdominalis*. *Molecular Ecology Resources*, 00, 1–13. <https://doi.org/10.1111/1755-0998.13541>