LEAFY AND WEEDY SEADRAGON GENOMES CONNECT GENIC AND REPETITIVE DNA FEATURES TO THE EXTRAVAGANT BIOLOGY OF SYNGNATHID FISHES

Running Title: The genomes of two seadragon species

AUTHORS

Clayton M. Small[1,2=#], Hope M. Healey[1], Mark C. Currey[1], Emily A. Beck[1,2], Julian Catchen[3], Angela S. P. Lin[4], William A. Cresko[1,2#], and Susan Bassham[1=#]

[1] Institute of Ecology and Evolution, University of Oregon, Eugene, OR 97403
[2] Presidential Initiative in Data Science, University of Oregon, Eugene, OR 97403
[3] Department of Evolution, Ecology and Behavior, University of Illinois at Urbana-Champaign, Urbana, IL 61801
[4] Knight Campus for Accelerating Scientific Impact, University of Oregon, Eugene, OR 97403
[=]Contributed equally

ORCID IDs

C.M.S.	0000-0003-1615-7590
H.M.H.	0000-0001-9978-1553
M.C.C.	0000-0003-3688-7208
E.A.B.	0000-0001-9810-5094
J.C.	0000-0002-4798-660X
A.S.P.L.	0000-0001-5349-927X
W.A.C.	0000-0002-3496-8074
S.B.	0000-0002-7309-2095

[#]Corresponding Authors

Susie Bassham, William A. Cresko and Clayton M. Small

Author Contributions: S.B. and C.M.S. conceived of the project, and along with W.A.C. planned the work and provided general oversight of the project. S.B. and C.M.S. generated the seadragon sequencing data, and C.M.S. performed the genome assemblies. M.C.C. led the genome annotation and assisted C.M.S. with annotation and analysis of TEs. J.C. and S.B. performed the conserved synteny analysis. C.M.S. performed the global gene family evolution analyses. In-depth analyses of FGF/FGFR gene family evolution were performed by S.B., H.M.H. E.A.B. S.B. and A.S.P.L. prepared the weedy seadragon specimen for scanning, A.S.P.L performed Zeiss Xradia scans and initial image processing, and H.M.H. led image processing for figure composites. H.M.H. also performed the analysis of microRNAs, with initial data generation and processing performed by S.B and C.M.S. M.C.C. produced all fish illustrations for the figures. S.B. and C.M.S. took the lead in writing the manuscript, and all authors made writing contributions. The work was performed in the laboratory of W.A.C. at the University of Oregon with funding to W.A.C.

Keywords: genome sequencing, novel traits, evolution, miRNA, transposable elements, molecular evolution, fibroblast growth factors, syngnathids

Abstract

Seadragons are a remarkable lineage of teleost fishes, and they are members of the family Syngnathidae renowned for having evolved male pregnancy. Comprising three known species, seadragons are widely recognized and admired for their fantastical body forms and coloration, and their specific habitat requirements have made them flagship representatives for marine conservation and natural history interests. Until recently, a gap has been the lack of significant genomic resources for seadragons. We have produced gene-annotated, chromosome-scale genome models for the leafy and weedy seadragon to advance investigations into evolutionary innovation and elaboration of morphological traits in seadragons as well as their pipefish and seahorse relatives. We identified several interesting features specific to seadragon genomes, including divergent non-coding regions near a developmental gene important for integumentary outgrowth, a high genome-wide density of repetitive DNA, and recent expansions of transposable elements and a vesicular trafficking gene family. Surprisingly, comparative analyses leveraging the seadragon genomes and additional syngnathid and outgroup genomes revealed striking, syngnathid-specific losses in the family of fibroblast growth factors (FGFs), which likely involve re-organization of highly conserved gene regulatory networks in ways that have not previously been documented in natural populations. The resources presented here serve as important tools for future evolutionary studies of developmental processes in syngnathids and will be a key resource for conservation studies of the extravagant seadragons and their relatives.

INTRODUCTION

Seadragons are phenotypic outliers in an already exceptional clade of teleost fishes (Family Syngnathidae) that also includes seahorses and pipefishes. For this reason, seadragons are often a colorful, flagship group in discussions of adaptation and evolutionary innovation. Seadragons demonstrate strikingly derived characters compared to their pipefish and seahorse relatives, including “leafy appendages,” extreme curvature of the spine (kyphosis and lordosis), especially elongated craniofacial bones, and large body size (Figure 1)(Dawson, 1985; Stiller et al., 2015). Substantial differences in these traits exist even among the three known extant species: *Phycodurus eques* (leafy seadragon), *Phyllopteryx taeniolatus* (weedy, or common, seadragon), and the recently described *Phyllopteryx dewysea* (ruby seadragon) (Stiller et al., 2015).

In addition to being a focus for evolutionary studies, seadragons are of significant cultural and conservation interest (Connolly et al., 2002; Martin-Smith & Vincent, 2006; Stiller et al., 2017). Presumed adaptations for crypsis including the leafy appendages, unique body plan, and elaborate skin coloration contribute to the status of seadragons as distinguished and valued cultural symbols for the people of Australia, where seadragon species are endemic. Because seadragon distributions are specific to temperate Australian macroalgal reefs, and their population sizes are relatively small, seadragons are likely susceptible to negative human impacts including global climate change. Furthermore, recent population genomic studies documenting significant population structure (Klanten et al., 2020; Stiller et al., 2021) in these species are especially relevant to conservation decisions. The unique evolutionary innovations, cultural importance, and conservation challenges all elevate the need to better understand and conserve seadragon species.

To improve our understanding of highly derived phenotypic traits and genomic features within seadragons, as well as those that are shared but derived among the Syngnathidae, we sequenced and annotated chromosome-scale assemblies for a male leafy seadragon and a female weedy seadragon. In addition to the production of these genomic resources we carried out several comparative analyses among five syngnathid and many other teleost genomes to determine changes in genome organization and content, including a more detailed analysis of a few key gene families and regulatory elements that may be involved in the development of syngnathid phenotypic innovations. Lastly, we performed high-resolution 3D X-ray microscope scans of an adult, male weedy seadragon to more precisely view seadragon morphological innovations.

Our work reveals a number of seadragon-specific genomic features, including divergent conserved non-coding elements (CNEs) near key developmental genes, a unique microRNA gene repertoire, and expanded gene families related to immunity and vesicular trafficking. We also found that the seadragon genomes are highly repetitive for their sizes, with unique repeat abundance distributions. Because the seadragon lineage occupies a region of the syngnathid phylogeny that is relatively basal to most of the species diversity, we leveraged their phylogenetic position to identify several genomic synampomorphies of the family. These genomic features include the striking loss of several highly conserved fibroblast growth factor (FGF) genes, expansions and contractions of gene families related to immunity and potentially male pregnancy, and syngnathid-specific transposable element (TE) expansion.

With these new genome models and rich accompanying data, we grow the existing collection of high-quality genomic tools and insights for several syngnathid groups including genera *Syngnathus* (Roth et al., 2020; Small et al., 2016),

Hippocampus (Li et al., 2021; Lin et al., 2016), *Microphis* (Zhang et al., 2020), and most recently (published as of the writing of this paper) *Phyllopteryx* and *Syngnathoides* (Qu et al., 2021). We add to this quiver of syngnathid genomes useful in illuminating the evolution and development of puzzling syngnathid novelties such as male pregnancy and leaf-like appendages. These genomic resources will also support ongoing efforts to understand and conserve sensitive syngnathid populations, including phylogenetic umbrella species like the seadragons.

MATERIALS AND METHODS

Seadragon genome assemblies

We isolated high molecular weight genomic DNA from tissues of an adult male leafy seadragon (*Phycodurus eques*) and from an adult female common ("weedy") seadragon (*Phyllopteryx taeniolatus*). We then generated PacBio libraries for both species and sequenced 49.12 and 80.80 Gb, respectively. We also generated "shotgun" whole-genome sequencing (WGS) Illumina libraries for both species, sequencing 57.48 and 105.79 Gb, to estimate genome size and polish the PacBio assemblies (see Supplemental Methods for all software versions and non-default parameters). We assembled both genomes with Flye (Kolmogorov et al., 2019), using all PacBio data excluding "scraps" with default parameters and an estimated genome size of 600 Mb. We performed two rounds of polishing on the primary Flye assemblies with the tool arrow (Chin et al., 2013), using the PacBio reads and performed an additional two rounds of polishing for each genome with WGS Illumina data, using pilon (Walker et al., 2014). To organize Flye assemblies into putative chromosome-scale genome models we generated Hi-C libraries using Phase Genomics Proximo Animal kits, then

scaffolded using the 3D-DNA pipeline (Dudchenko et al., 2017) with breaking of original scaffolds disabled, followed by visualization and manual editing using Juicer and Juicebox (Dudchenko et al., 2018; Durand et al., 2016). We evaluated assembly quality and completeness using Quast (Gurevich et al., 2013) and BUSCO (Waterhouse et al., 2018).

Draft short-read genome assemblies for *Doryrhamphus excisus* and *Synchiropus splendidus*

To supplement our comparative analyses with additional syngnathid genomes and a close outgroup we isolated high molecular weight DNA and generated linked-reads assemblies for the bluestripe pipefish (*Doryrhamphus excisus*) and the Mandarin dragonet (*Synchiropus splendidus*). These assemblies were performed using 10x Genomics Chromium technology and the Supernova assembly software (Weisenfeld et al., 2017). Details for these assembly methods are as described in Stervander and Cresko (2021).

mRNA-seq

To generate mRNA-seq libraries, we extracted total RNA from tissues of the same *P. eques* and *P. taeniolatus* individuals as were used for PacBio genome sequencing. From the *P. eques* specimen, we dissected testis, leafy appendage, eye, and gill tissues. From *P. taenoiolatus* we dissected ovary, leafy appendage, eye, and liver tissue (see Supplemental Methods). We used the Roche KAPA HyperPrep Kit to generate indexed, stranded mRNA-seq libraries for Illumina sequencing to obtain 301.52 million paired-end 100 bp reads. We trimmed Illumina adaptors and low-quality regions from reads using process_shortreads from the Stacks software suite (Catchen

et al., 2013; Catchen et al., 2011) and aligned cleaned RNA-seq reads from both seadragon species to both *P. taeniolatus* and *P. eques* genome assemblies using STAR aligner (Dobin et al., 2013).

miRNA-seq

To generate small RNA-seq reads we purified a small fraction of RNA from each tissue above using Zymo DirectZol columns and generated indexed sequencing libraries using the NextFlex Small RNA-seq Kit v3. We ran BBMap (Bushnell, 2014) on the weedy and leafy seadragon genomes to align sequencing reads. We annotated miRNAs in the leafy seadragon based on sequence conservation with included species' miRNAs (Antarctic blackfin icefish (*Chaenocephalus aceratus*), platyfish (*Xiphophorus maculatus*), European perch (*Perca fluviatilis*), and threespine stickleback (*Gasterosteus aculeatus*), zebrafish (*Danio rerio*), spotted gar (*Lepisosteus oculatus*), and medaka (*Oryzias latipes*)) as described in the Prost! (Desvignes et al., 2019) manual (see Supplemental Methods). Additionally, to run Prost!, we supplied threespine stickleback noncoding sequences (Desvignes et al., 2019). We ran Prost! for the weedy seadragon by adding the leafy seadragon's annotations to the included species pool. We distinguished ohnologous miRNAs for the seadragon species using synolog (Catchen et al., 2009) synteny software, which aligned the leafy seadragon genome with the stickleback genome. Any miRNA not initially identified in the sequencing reads were searched for in the leafy seadragon genome with BLASTN and vista plots.

Genome annotation

To facilitate the analysis of comparable annotations among genome assemblies, we performed new annotations for leafy seadragon (*P. eques*), weedy seadragon (*P.*

157 *taeniolatus*), Greater pipefish (*Syngnathus acus*), bluestripe pipefish (*Doryrhamphus*
158 *excisus*), Mandarin dragonet (*Synchiropus splendidus*), and Pacific bluefin tuna (Smit et
159 al.) (*Thunnus orientalis*) (see Supplemental Methods for all public sequence data
160 accessions). Assemblies were soft-masked for repetitive elements and areas of low
161 complexity with RepeatMasker (Smit et al.) using custom repeat libraries made by
162 combining a teleost-specific library extracted from RepeatModeler2 (Flynn et al., 2020)
163 with species-specific repeat libraries produced by running RepeatModeler on each
164 species listed above. We aligned all RNA-seq data (including new and previously
165 published reads) as described above and supplied the .bam files to BRAKER2 (Bruna et
166 al., 2021) for genome annotation.
167 We evaluated the final BRAKER2 set of annotated genes for each genome using
168 InterProScan (Jones et al., 2014) and retained a final, "filtered" version of the annotation
169 if they showed InterProScan evidence other than disorder-based (i.e. MobiDB-lite). We
170 subjected remaining amino acid sequences to blastp searches against the NCBI
171 RefSeq protein database and retained those in the filtered annotation if they returned a
172 hit (e-value threshold of 0.001). For all downstream analyses of protein-coding
173 sequences based on these and other annotations, we selected the longest transcript
174 per locus.
175
176 **Repeat annotation**
177 We characterized the repetitive content of 16 teleost genome assemblies (Figure
178 S1; see Supp. Methods for accession numbers). With one exception (tiger tail seahorse
179 (*Hippocampus comes*)), we focused exclusively on assemblies produced by long-read
180 (i.e., PacBio or Oxford Nanopore) and/or linked-read (i.e., 10x Genomics) technologies.
181 All genomes were subject to a unified repeat library generation and annotation

workflow. Briefly, we identified repeats *de novo* for each assembly using 1. RepeatModeler2 (Flynn et al., 2020) and 2. TransposonPSI (Haas). We combined those repeat predictions with teleost repeats extracted from RepeatMasker (Smit et al.) libraries and all sequences from the FishTEDB database (Shao et al., 2018) http://www.fishtedb.org). These sequences (576,007 in total) were classified using the RepeatClassifier module of RepeatModeler. We also clustered repeats at 80% sequence identity using USEARCH (Edgar, 2010) as a way to group and ultimately enumerate repeats based on sequence divergence (see Supp. Methods). We ran RepeatMasker on all 16 genome assemblies using all 576,007 sequences and integrated the aforementioned RepeatClassifier and USEARCH cluster IDs into the RepeatMasker output. Finally, we used RepeatClassifier taxonomy and USEARCH cluster membership as alternative grouping mechanisms to characterize the distribution of repeats within and among genomes (e.g., using heatmaps and rank abundance plots), and to ordinate genomes in repeat space (i.e., using PCA). Regional repeat abundance distributions for target gene families were compared to those from random samples of genes across the genome using resampling-style hypothesis tests. All downstream repeat analysis and visualizations were carried out using the R statistical language (R_Core_Team, 2019).

Gene family evolution

We used protein annotations from 21 teleost genomes (Figure 2; see Supp. Methods for accession numbers) to better understand gene family size evolution in seadragon and syngnathid lineages. We defined putative gene families via all-by-all blastp (Altschul et al., 1990) and clustering with mcl (Enright et al., 2002), then we conducted a series of gene family size evolution analyses using CAFE 5 (Mendes et al.,

2020). Briefly, we first fit an error model to account for artifactual (e.g., genome assembly or annotation error) family size variation, which was applied to subsequent CAFE runs. We fit a model assuming a single rate of gene family expansion/contraction (λ) to the data and identified gene families evolving especially rapidly using CAFE's internal likelihood ratio tests (LRTs).

Of those gene families showing evidence for rapid evolution, we identified subsets for which branch-specific LRTs suggested extreme λs along branches of interest: the terminal *P. eques* and *P. taeniolatus* branches, the internal branch leading to the seadragon lineage, and the internal branch leading to the syngnathid lineage. We also assigned each gene family to a KEGG orthology (KO) ID if possible, and we used this information to perform KEGG pathway overrepresentation analysis using ClusterProfiler (Yu et al., 2012), with respect to the branch-specific rapidly evolving gene families. Lastly, we fit several multiple-λ models to the data in order to test hypotheses of overall λ differences (from a global λ) for the internal seadragon branch and the internal syngnathid branch. To test these hypotheses we used CAFE to perform 100 gene family simulations, fit the aforementioned models to the simulated data sets, and compared observed likelihood ratios (LRs) from the data to the LR distributions from the simulations. For details, see the Supplementary Methods.

***fgf* and *fgfr* gene family characterization**

We collected fibroblast growth factor (FGF) and receptor (FGFR) amino acid sequences from Ensembl of several percomorph species (see Supplementary Methods). We aligned orthologs and screened the genome assemblies of syngnathids and outgroups based on Hidden Markov Model (HMM) profiles from the alignments, according to methods described in Small et al. (Small et al., 2016). In some cases, we

supplemented these sensitive searches with regional RNA-seq read alignments to correctly define exon boundaries. We conducted targeted, lineage-specific tests of positive selection using branch site-models in PAML (see Supplemental Methods) and tested for deleterious mutations using Provean (Choi & Chan, 2015).

X-ray tomography and 3D model reconstruction

We obtained a euthanized adult male weedy seadragon, fixed in neutral buffered formalin from the Tennessee Aquarium (Chattanooga, TN), and scanned it using a resolution of 54 µm voxels on a Zeiss XRadia 620 Versa X-ray microscope at the University of Oregon Knight Campus for Accelerating Scientific Impact's X-ray Imaging Core Facility. The anterior of the fish including the head and cleithrum were scanned again at 17 µm voxel resolution. Composite virtual three-dimensional reconstructions of the unstained specimen were generated using Dragonfly Pro and Dragonfly software (Object Research Systems). After completing scans of the unstained specimen, which allowed high contrast visualization of electron-dense bony structures, the fish was then dehydrated through an ethanol series, stained with an iodine-based X-ray contrast agent to enhance imaging of soft tissues, and a section of the rostral part of the stained tail that includes a pair of leafy appendages was scanned at 27 µm voxel resolution (Figure 1; Figure S2; Supplemental Methods).

RESULTS

PacBio assembly with Hi-C scaffolding yields chromosome-level genome models for the leafy and weedy seadragon

We estimated the haploid genome sizes for leafy and weedy seadragons to be 644.0 Mb and 597.3 Mb, respectively, based on k-mer frequency analysis of Illumina

WGS data (Vurture et al., 2017). From this analysis we also estimated genome-wide heterozygosity for each individual to be 0.27 and 0.33%. Polished PacBio assemblies were 664.24 and 650.38 Mb, with scaffold N50s of 19.59 and 9.90 Mb. The longest Flye scaffolds were 38.02 and 29.83 Mb, and BUSCO completeness frequencies were 95% for both genomes. After scaffolding both polished PacBio assemblies using *P. eques* Hi-C reads, we obtained 23 putative chromosome models for each genome (Figure S3), which reflect 93.22% and 96.10% of the total length for final leafy and weedy seadragon genome models. These assembly and completeness metrics, our ability to annotate 22,256 and 22,043 protein-coding genes in the respective genomes, and extensive evidence for conserved synteny between the two seadragon assemblies (Figure S4), all support that these genome models are of high quality relative to available resources for teleost fishes. These genomes will therefore serve as useful tools for a variety of current comparative genomic analyses.

Seadragon karyotypes are conserved relative to other syngnathid genomes but lack one of two chromosome fusions observed in *Syngnathus* and *Hippocampus*

Haploid chromosome number in syngnathid fishes, as assessed by karyotyping and genetic mapping, is reported to be 22 or 24 in seahorse (*Hippocampus*) species (Vitturi & Catalano, 1988; Vitturi et al., 1998) and 22 in Gulf pipefish (*Syngnathus scovelli*) (Small et al., 2016). In Gulf pipefish, the reduction in chromosome number from 24, the putative ancestral number in ray-finned fishes (Mank & Avise, 2006) to 22 likely resulted from fusion of two pairs of chromosomes orthologous to Chromosomes 1 and 24 and to Chromosomes 14 and 23 in platyfish (*X. maculatus*), a representative outgroup percomorph with 24 chromosomes (Small et al., 2016). Though a haploid number of 24 chromosomes in seahorse was reported (with some published confusion

about in which species this was observed (Vitturi & Catalano, 1988; Vitturi et al., 1998) the reference genome for tiger tail seahorse (*H. comes*) (Lin et al., 2016), provides conserved gene synteny evidence that both ancestral chromosome fusions were already present in the common ancestor of seahorses and *Syngnathus* pipefish (Figure S5). As stated, our inferred seadragon haploid chromosome number of 23 is based on the size distribution of Hi-C scaffolds (Figure S3). The seadragon lineage apparently shares only the ancestral Chromosome 1 to Chromosome 24 fusion with the seahorse+*Syngnathus* pipefish ancestor, leaving one to one orthologs of platyfish Chromosomes 14 and 23 (Figure S5).

Seadragon genomes are highly repetitive for their compact size, with large contributions from relatively recent TE expansions

A correlation between eukaryotic genome size and the proportion of a genome classifiable as TEs has led to an appreciation that TEs can be an important driver of genome size evolution (Kidwell, 2002). The relationship is particularly apparent among some of the well characterized genomes of teleost fish model species. Genome size and TE proportion share a positive, tightly linear relationship in a comparison of green spotted puffer (*Tetraodon nigroviridis*), threespine stickleback (*G. aculeatus*), medaka (*Oryzias latipes*), and zebrafish (*Brachydanio rerio*), whose genomes span a range from roughly 358 Mb to over 1.37 Gb (Gao et al., 2016). *Syngnathus* pipefish have genomes that rival green spotted puffer in genomic compactness, with assembly lengths of 307.0 Mb for the Gulf pipefish (Small et al., 2016) (*S. scovelli*), and 324.33 Mb for the greater pipefish (*S. acus*) (NCBI RefSeq Genome GCF_901709675.1). In the spectrum of vertebrate genomes, known seahorse genomes are also diminutive, estimated at 421

306 Mb for the lined seahorse (*H. erectus*) (Li et al., 2021) and 494 Mb for the tiger tail
307 seahorse (*H. comes*) (Lin et al., 2016).

308 Our final genome assembly sizes for seadragons were 666.5 Mb for leafy
309 seadragon and 652.2 Mb for weedy seadragon, which are appreciably larger than those
310 of *Syngnathus* and *Hippocampus* but still much smaller than many other teleost
311 genomes. Surprisingly, as much as 58.7% of the leafy and 57.9% of the weedy
312 seadragon genome is composed of repetitive sequences, as classified by our workflow
313 (see Methods). To assess whether seadragons or syngnathids are exceptional in their
314 repetitive DNA characteristics, we measured the proportion of repetitive sequence in
315 sixteen teleost genomes (Figure S1) using an in-common repeat reference library to
316 create a standardized basis for comparison. Species were chosen for taxonomic
317 breadth and because their genomes were assembled from long- or linked-read
318 sequence data (except tiger tail seahorse). Genomes of five syngnathids were
319 represented, including a seahorse and tail-brooding *Syngnathus* pipefish, the
320 seadragons, and a basal lineage relative to these, the abdominal-brooding bluestripe
321 pipefish (*D. excisus*).

322 We confirmed a strong, positive relationship between repeat content and
323 genome size for the 16 genomes using phylogenetic generalized least squares (PGLS)
324 regression ($t_{16,14}$=4.88; p=0.0002; b=0.00048), but seadragons are notable outliers with
325 large, positive residuals (Figure 3). Both seadragon genomes are unusually bloated with
326 repetitive DNA among teleosts for their relatively compact size, a feature either derived
327 specifically in the seadragon lineage or an ancestral lineage. This latter hypothesis is
328 not yet testable given currently available genome assemblies but can be addressed
329 with the addition of syngnathid genomes at key phylogenetic positions.

330 Repeat density was high across seadragon chromosomes, in stark contrast with
331 the greater pipefish (Figure 4), which shows repeat "hotspots" almost exclusively near
332 chromosome ends, and a close outgroup to syngnathids, common dragonet
333 (*Callionymus lyra*), which has relatively uniform, low repeat density in general (Figure
334 S6).

335 Several TE classes contribute notably to the large repeatomes of seadragons.
336 The *Tc1* family of the *Tc1/Mariner* superfamily of transposases is a major contributor to
337 repeatome composition variation among teleosts as revealed by PCA based on within-
338 repeatome relative class proportions, with seadragons, platyfish (*X. maculatus*), and
339 Northern pike (*Esox lucius*) genomes influenced heavily by abundant *Tc1* repeats
340 (Figure 3). Phylogenetic patterns of repeat abundance among the fish lineages we
341 analyzed suggest *Tc1* expansion in the syngnathid lineage, given lower *Tc1*
342 abundances in the close outgroups of common dragonet, (*C. lyra*), and Pacific bluefin
343 tuna (*T. orientalis*). Among syngnathids, *Tc1* repeats compose a disproportionately
344 large fraction of seadragon repeatomes (Figure 3). This class of "cut-and-paste" DNA
345 transposons is widespread in animals and especially common in teleost fishes, with
346 high abundance and variability among species (Gao et al., 2017; Gao et al., 2016). In
347 fact, phylogenetic evidence suggests *Tc1* transposons are still active and recently
348 expanding in some neoteleosts, such as threespine stickleback (*G. aculeatus*) (Gao et
349 al., 2017).

350 The second most abundant, classifiable TE category in the seadragon genomes
351 was the *BovB* family of non-LTR LINE retrotransposons (Figure 3; Figure S7). This TE
352 family is restricted to animal taxa, where its members are patchily distributed among
353 lineages and have been inferred via phylogenetic analysis to populate animal genomes
354 through horizontal transfer, perhaps via metazoan parasites (Ivancevic & Chuong, 2020;

355 Ivancevic et al., 2018). *BovB* density variation along seadragon chromosomes (Figure
356 S7) was largely concordant with overall repeat density patterns (Figure S6), suggesting
357 common mechanisms or constraints for the expansion of at least some TE families.
358 We also discovered an apparent expansion of *Tigger* transposases in the
359 syngnathid clade, which are members of the *pogo* superfamily closely related to
360 *Tc1/Mariner* (Gao et al., 2020). *Tigger* repeats are overrepresented in syngnathids
361 compared to outgroups but proportionally are conserved among the five syngnathid
362 genomes we analyzed (Figure 3). Unlike *Tc1* and *BovB*, *Tigger* repeat density variation
363 along seadragon and *Syngnathus* pipefish chromosomes deviates from the collective
364 pattern of repeat density (Figure S8), with the highest-density *Tigger* regions more
365 centrally located, although the potential significance of this is unclear.
366 Because the grouping and enumeration of repeats according to classification
367 alone comes with a loss of evolutionary resolution and precision, we also analyzed
368 repeats as clusters of sequences with ≥ 80% sequence identity. PCA based on this
369 cluster-wise treatment of repeats revealed that leafy and weedy seadragon genomes
370 are quite divergent in repeat space from the other teleost species, including the three
371 other syngnathids compared (Figure 5; Figure S9). In particular four repeat clusters are
372 heavily influential in this regard, one of which was not classifiable but especially
373 abundant in scaffolds unassigned to chromosomes (Cluster 20902), two of which
374 belong to the *BovB* LINEs mentioned above (Clusters 10278 and 10626), and the last
375 (Cluster 14395) belonging to the *Tc1* group of transposases mentioned above (Figure 5;
376 Supplemental File 1).
377 We also generated rank abundance distributions (RADs) for repeat clusters in
378 the 16 fish genomes, an approach from ecology used to understand community
379 evenness and major and rare constituents (Whittaker, 1965). Interestingly, leafy and

weedy seadragon RADs show markedly high repeat cluster abundances relative to the other genomes for repeat ranks 1-20 (Figure 5). Among the highest-ranking repeats for both seadragon species are the four repeat clusters mentioned above as major contributors to the distinctiveness of seadragon repeatomes. These findings, along with the observation that members of the top-ranking seadragon repeat clusters are rare in the other genomes (Supplemental File 1), suggest that the unique repetitive features of seadragon genomes are driven largely by recent expansions of *BovB* and *Tc1*, and a yet-to-be-classified cluster.

Gene family contractions in Syngnathidae involve innate immunity, and seadragon-specific expansions are associated with vesicular trafficking

We identified 290 total gene families as having expanded or contracted at a rate (λ) significantly higher than the background λ among 21 teleost species. Of these, 109 showed evidence for rapid size evolution along the leafy seadragon branch, 113 along the weedy seadragon branch, 38 along the internal seadragon branch, and 31 along the internal branch leading to syngnathids (Figure 2; Supplemental File 2). Based on 100 simulations of gene family evolution along the tree, we also inferred that λs for the internal seadragon and syngnathid branches are respectively distinct from the global λ for the tree, and likely distinct from one another (Figure S10).

Among the gene families most likely to have evolved rapidly in size in the ancestral syngnathid lineage, at least seven families related to innate immunity experienced contractions (Supplemental File 2), in contrast to multiple lines of evidence for expansion of inflammation and innate immunity gene families in teleosts relative to other vertebrates (Balla et al., 2020; Boudinot et al., 2011; Chang et al., 2021; Howe et al., 2016; Mattingsdal et al., 2018). Specifically, we found evidence for syngnathid-

specific contractions in NACHT, TRIM, and GIMAP gene families, consistent with some of the immunity and detoxification pathway gene families depleted in the genome of the Manado pipefish *(Microphis manadensis*) (Zhang et al., 2020).

Several functional categories (KEGG pathways) were overrepresented among gene families with large size changes in the seadragon lineage, including cancer, cardiomyopathy, and immunity (Supplemental File 2), primarily due to contraction events. In terms of gene family expansions along the branch leading to seadragons, however, two families with roles in vesicular trafficking - *Vacuolar Protein Sorting-Associated Protein 13B* (*vps13b*) and *Coatomer Protein Complex Subunit Beta 2* (*copb2*) – are notable. Sequences with high similarity to *copb2*, which encodes one subunit of a Golgi budding and vesicular trafficking protein complex (Waters et al., 1991), are especially abundant in seadragon genomes relative to other syngnathids and teleosts (Supplemental File 2).

Though tuna (*T. orientalis*) and platyfish (*X. latipes*), for example, have two paralogs of *copb2*, we could find evidence for only a single gene copy in the genomes of seahorse (*H. comes*) and pipefish (*S. acus*). By contrast, we detected the presence of at least nine and six copies (on the 26 longest Hi-C scaffolds) in leafy and weedy seadragon genomes (Supplemental File 3). We also found many sequences matching *copb2* on several short scaffolds, suggesting that the repetitive nature of this region prevented these from being incorporated into chromosome-scale scaffolds, or, but less likely (Figure S11), that they could be redundant artifacts. One of the *copb2* paralogs (on Pequ Hi-C scaffold 13, Ptae Hi-C scaffold 14) is the likely ortholog of platyfish *copb2* on the orthologous chromosome (Xmac 6), with the remaining copies likely expanding secondarily via an unknown mechanism.

Given the repetitive nature of these regions, we hypothesized that TE activity may have played a role in seadragon *copb2* expansion. Specifically, we first tested whether the seadragon-expanded TE classes of *BovB* and *Tc1* are overrepresented in the immediate vicinity (1-kb flanking both sides) of *copb2* copies, relative to 499 randomly resampled gene groups. Both *BovB* (4.24-fold enriched; $p < 0.002$) and *Tc1* (1.39-fold enriched; $p < 0.002$) repeat classes were significantly overrepresented in *copb2* regions (Figure S11; Supplemental File 4). We secondarily performed naïve hypothesis tests with the false discovery rate (FDR) controlled at 0.05 to identify other TE classes with potential enrichment in these regions, revealing *Rex-Babar* LINEs, and *Tigger* and *Charlie* transposons as additional candidates (Supplemental File 4).

Syngnathid fishes have lost several FGF family genes, most notably *fgf3* and *fgf4*

The FGF and FGFR gene families include well-studied ligand and receptor signaling molecules central to vertebrate craniofacial, limb, dermal appendage, sensory placode, and hindbrain development, among many other functions (reviewed in Xie et al., 2020). Because of the prominence of FGF signaling in the morphogenesis of traits that are distinctively modified in syngnathids, we explored whether FGF ligands or their receptors are exceptional in seadragons and in three other representative lineages of syngnathids: flagtail pipefishes (the most basal lineage of the four), seahorses, and *Syngnathus* pipefishes (Figure 2).

The FGF and FGFR families in seadragons and other syngnathid fishes conform, with two notable exceptions mentioned below, to a broadly typical complement of gene paralogs relative to other percomorphs. Within the Syngnathidae, there are several lineage-specific losses of FGF and FGFR genes (Table S1). We did not detect *fgf9* in any syngnathid. Seadragons and *Syngnathus* pipefish have apparently lost *fgf4-like*,

454 while tiger tail seahorse (*Hippocampus comes*) and bluestripe pipefish (*Doryrhamphus*
455 *excisus*) have retained it. *fgf17* is missing birds (Abramyan, 2015), in seadragons and
456 bluestripe pipefish but is present in tiger tail seahorse and greater pipefish. Because
457 flagtail pipefishes, the lineage to which bluestripe pipefish belongs, are basal to the
458 clade containing seadragons, seahorses, and *Syngnathus* pipefish (Longo et al., 2017),
459 *fgf17* therefore must have been lost at least twice. Evolutionary loss of *fgf17*, though
460 uncommon, is not unique to syngnathid lineages; orthologs of this gene have been
461 independently erased from other distant taxa, such as the medaka genus *Oryzias*
462 (Canestro et al., 2007). In addition, at least two clades of *fgf5* and *fgfr1b* are present in
463 seadragons, tiger tail seahorse, and blue-stripe pipefish, but these genes were not
464 detected in the two *Syngnathus* pipefish genomes (*S. acus* and *S. scovelli*).
465 The most startling gene losses from the FGF family are shared across the
466 seadragon, tiger tail seahorse, greater pipefish, and bluestripe pipefish lineages: all are
467 missing *fgf3* and *fgf4* (Figure 6; Table S1). In percomorph outgroups to the syngnathids,
468 *fgf3*, *fgf4*, and *fgf19* are clustered, with *fgf3* flanked by *ccnd1* and *lto1,* genes that are
469 also missing from seadragons, seahorses, and greater pipefish (but bluestripe pipefish
470 has retained *lto1* (Figure 6). The expected neighbors normally on the *fgf19* side of the
471 cluster (*zgc:153993* and *ano1*) are present in seadragons and both lineages of
472 pipefishes, but bluestripe pipefish has lost *fgf19*. Close outgroups to the syngnathids,
473 the razorfish (*Aeoliscus strigatus*) and the mandarinfish (*Synchiropus splendidus*) (Longo
474 et al., 2017) retain intact *fgf3/4/19* clusters, making it likely that drastic alteration to the
475 cluster is a syngnathid synapomorphy (Figure S12). In the common ancestor of lobe-
476 finned and ray-finned fishes, *fgf3*, *fgf4*, and *fgf19* were likely already clustered, and
477 remain tandemly arrayed in representatives from jawed vertebrate lineages as
478 separated as mammals and teleosts (Oulion et al., 2012). The syngnathid lineage has

479 therefore experienced a degree of change in this cluster that is perhaps unprecedented
480 throughout gnathostome evolution.

481 Exceptional absences of FGF and FGFR members in extant syngnathid
482 genomes suggest that some property of their syntenic neighborhoods is inherently
483 volatile. We tested whether the local repeat landscape flanking significant regions of
484 FGF, FGFR, and other developmental gene loss in syngnathids – the *fgf3/4/19* cluster,
485 *fgf4l*, *fgf5*, *fgf17*, *fgfr1b, eve1* (Small et al., 2016), and *tbx4* (Lin et al., 2016; Small et al.,
486 2016) - differs from that of genes in general, by using the leafy seadragon genome and
487 the close syngnathid outgroup *C. lyra* (common dragonet) as a pre-loss comparator.
488 While no individual repeat class was significantly enriched in these regions (of either
489 genome) after FDR adjustment, several show nominal evidence for enrichment and are
490 at high density in the leafy seadragon regions, including *Tc1*, *BovB*, *hAT-Ac*, and
491 additional LINEs *Rex-Babar*, and *L2* (Figure S13; Supplemental File 4). Furthermore, we
492 found support for the alternative hypothesis that *overall* repeat density in general is
493 enriched in these regions (Figure S13; Supplemental File 4), but only for the leafy
494 seadragon (Pequ: 1.45-fold enrichment; $p = 0.022$), and not the dragonet genome (Clyr:
495 1.11-fold enrichment; $p = 0.27$).
496
497 **Limited evidence for compensatory evolution in the FGF pathway**

498 FGF paralogs sometimes overlap in their expression and ability to compensate,
499 partially or fully, for loss-of-function of one another in specific developmental contexts.
500 In zebrafish, for example, knock down of either *fgf3* or *fgf10* causes relatively mild
501 effects in lateral line migration and neuromast maturation, but there are severe lateral
502 line defects when expression of both paralogs is depleted (Nechiporuk & Raible, 2008).
503 We explored whether those paralogs most recently diverged from *fgf3* and *fgf4*, or with

504 overlapping developmental roles, showed evidence for compensatory evolution prior to
505 the syngnathid radiation.

506 Syngnathid and close outgroup sequences are largely conserved for *fgf4-like*,
507 *fgf6a*, *fgf7*, *fgf8a*, *fgf8b*, *fgf10a*, *fgf10b*, *fgf19*, *fgf20a, fgf20b*, *fgf22,* and *fgf24*, and
508 exhibit no evidence of positive selection along the syngnathid branch (Table S2). We did
509 find potential evidence for positive selection in *fgf6b*, although this test result was within
510 the range of false discovery (Figure S14; Table S2). Second, although the gene was lost
511 in the seadragons and blue striped pipefish lineages, we found the strongest evidence
512 for positive selection on *fgf17*, which is retained in seahorse and *Syngnathus* pipefish
513 (Table S2).

514 Like the ligands, the syngnathid FGF receptors are largely conserved, and we
515 found no evidence of syngnathid lineage-specific positive selection for non-canonical
516 *fgfrl1a* and *fgfrl1b*, nor for canonical *fgfr1a, fgfr2*, *fgfr3*, or *fgfr4*. In contrast, *fgfr1b*
517 presented some evidence for positive selection (although not robust to false discovery),
518 displaying surprising syngnathid-specific substitutions (Table S2). These include a likely
519 deleterious substitution within the activation loop of the kinase domain, a position
520 conserved not only across vertebrate Fgfr1 proteins, but across all canonical FGFR
521 paralogy groups (Figure S15) (provean score= -6.123).

522 Although we did not find evidence for lineage-specific positive selection using
523 codon-based models, leafy and weedy seadragons share a derived six amino acid
524 deletion in a conserved region of the Fgf16 protein (Figure S16), while seahorse has
525 amino acid substitutions in this same motif. Other representative percomorphs,
526 chicken, and human Fgf16 protein sequences are identical across this region indicating
527 a high level of conservation, though no function has yet been ascribed to this protein
528 domain. Divergence of seadragon *fgf16* is not limited to the coding sequence. A

putative regulatory change is hinted at by the absence in both dragons of an approximately 240 bp conserved non-coding element (CNE) that is well preserved across percomorph fishes 5' of *fgf16* (Figure S17).

miRNA-seq data from seadragons reveal loss of conserved microRNAs

Although microRNAs (miRNAs) are important developmental regulators (Bizuayehu & Babiak, 2014), there are currently no annotations of these genes in syngnathids. In the leafy and weedy seadragons, we identified 261 and 251 miRNA genes that produce 331 and 318 unique mature miRNAs (Supplemental File 5). We found absences of numerous conserved miRNA genes including *mir10a* and *mir196b,* hox miRNAs genes that were reported missing in the Gulf pipefish genome assembly (Small et al., 2016). Six of the additional highly conserved missing miRNAs belong to two miRNA clusters from the miR-130 family. These absences include *mir130a*, *mir301a*, *mir130b*, *mir301b*, *mir130c-2*, and *mir454b*. While *mir130a, mir301a, and mir301b* are convergently missing in platyfish (*X. maculatus*) and medaka (*O. latipes*), absences of *mir130b, mir130c-2,* and *mir454b* have not yet been reported in other vertebrate species (Figure S18) (Desvignes et al., 2019; Desvignes et al., 2021; Kelley et al., 2021). In some tetrapods and teleosts, *mir301a* is located in the first intron of *ska2*. The seadragons, Gulf and greater pipefishes, and tiger tail seahorse are missing *ska2* as well as *mir301a* and *mir130a.* Throughout vertebrates, *mir130b* and *mir301b* (and, in teleosts and Coelacanth, additionally *mir130c-2* and *mir454b*) are located in the intergenic region between *sdf2l1* and *top3b*. Though these adjacent protein coding genes and the immediate syntenic neighborhood are conserved in seadragons, Gulf and greater pipefishes, and tiger tail seahorse, this cluster of microRNAs is not.

DISCUSSION

Exploring the seadragon genomes in a comparative phylogenetic context has lifted a veil on the evolution of seadragon-specific traits and has also revealed intriguing evolutionary facets of this unusual vertebrate family, the Syngnathidae, as a whole. We found that both leafy and weedy seadragon genomes stand out among their relatives in having a surprisingly large contingent of repetitive DNA. This pattern appears to have been driven largely by recent expansions of *BovB* and *Tc1*, and a yet-to-be-classified cluster of repetitive sequences specific to the seadragon lineage. One possible explanation for the large difference we observed between seadragon and *Syngnathus* pipefish repeatome size and genomic distribution could be a difference in historical effective population size (N_e). In this case, negative selection to remove rapidly expanding repeats from a population would be less effective in the face of strong genetic drift (small N_e) (Lynch, 2007; Lynch & Conery, 2003). Differences in k-mer-based heterozygosity estimates from individual WGS data are at least consistent with this idea. A heterozygosity estimate based on the Gulf pipefish data (Small et al., 2016) is roughly three times the seadragon estimates (1.01% versus 0.27% and 0.33% for leafy and weedy seadragons, respectively).

The explanation presented above for TE expansion assumes that deleterious effects are common, likely through interruption of coding regions or promoters, or by regional silencing via chromatin changes (Liu et al., 2018). However, TEs can also contribute new genes or gene regulatory sequences when a host genome co-opts ("domesticates") these exogenous genetic elements (Bejerano et al., 2006). Some classes of TE provide, for example, binding sites for master transcriptional regulators NANOG and OCT4 throughout the mouse and human genomes but in largely non-overlapping sets of loci between the taxa, potentially restructuring, in lineage-specific

579 ways, the transcriptional networks for developmental pluripotency (Kunarso et al.,
580 2010). TEs and repetitive DNA in general can fuel gene family expansion or contraction
581 by precipitating unequal crossover events; Hahn et al. (2007) suggest an “explosion” of
582 TEs in the primate lineage could be linked to its accelerated gene content evolution in
583 primates. Repeatome expansion, such as what we here observe in seadragons, could
584 have had disruptive impacts on gene regulatory networks and gene content, subject to
585 subsequent evolution via negative *or* positive selection.

586 Using global and targeted approaches we explored expansion and contraction
587 of gene families in seadragons and their relatives, and possible connections of gene
588 content changes to observed TE distributions. Perhaps the most obvious trend we
589 observed for the syngnathid lineage in general was contraction of particular immunity
590 and detoxification pathway gene families, some of which have previously been
591 described (Lin et al., 2016; Zhang et al., 2020). GTPase of the immunity-associated
592 protein genes (GIMAP), for example, was among the contracted families in seadragons
593 that was also detected in the seahorse genome (Lin et al., 2016), and our more
594 comprehensive analysis supports that the GIMAP contraction occurred prior to the
595 radiation of syngnathids. This family exists as a single eight-member cluster in
596 mammals (Krucken et al., 2004). GIMAP genes have multiplied in other teleost lineages,
597 numbering up to nearly 190 genes. Balla et al. (2020) found zebrafish GIMAP genes
598 respond to pathogenic viral exposure and suggest the gene expansion could have been
599 evolutionarily favored by the relatively long period that hatchlings must rely on innate
600 immunity before the development of a functional adaptive immune system. Male-
601 brooded syngnathid embryos might enjoy a luxury not available to free-spawned
602 progeny like those of zebrafish, namely pathogen climate control afforded by the
603 paternal immune system before their own adaptive immunity develops. The male’s

604 immune system strikes a balance of defending against foreign agents but without
605 rejecting his own brood. The syngnathid contraction of the GIMAP family is also
606 interesting, therefore, in light of the possibility that *gimap4*, which does persist in the
607 lineage, could promote immunologic tolerance to embryos in brood pouch tissues (Roth
608 et al., 2020). Roth et al. (2020) based this assertion on the observation that *gimap4* is
609 up-regulated in *Syngnathus* pregnancy tissues, where it could contribute to local
610 suppression of the lymphocyte population (Roth et al., 2020; Schnell et al., 2006).
611 We detected seadragon-specific copy number expansion of a coatamer
612 complex gene *copb2*. Variants of the *BovB* and *Tc1* transposable elements were
613 enriched surrounding the supernumerary gene copies, suggesting a TE-driven
614 mechanism for the expansion. Copb2 forms part of a protein complex involved in
615 retrograde vesicle budding from the Golgi apparatus and secretion of macromolecule
616 cargo, such as collagen, which is critical for bone and connective tissue development.
617 Mice and fish developing with deficits in *copb2* have delayed bone mineralization and
618 low bone density, as well as defects in type II collagen trafficking and secretion (Marom
619 et al., 2021). Zebrafish mutants of *copb2* develop mispatterned, kinked notochords with
620 poorly formed vacuoles and a disorganized perinotochordal basement membrane,
621 which is secreted by notochord sheath cells (Coutinho et al., 2004). In teleosts, the
622 notochord, particularly its sheath, plays an instructive role in patterning the vertebrae
623 (Gray et al., 2014; Peskin et al., 2020).
624 Given the elaborated bony exoskeleton in seadragons, their stiff bodies with
625 connective tissue-dense leafy ornaments, and their kinked axial skeletons with varied
626 and regionalized vertebral forms (Figure 1), the proliferation of *copb2* gene sequences
627 in their genomes ignites curiosity about the possible evolutionary developmental
628 consequences of this expansion for the seadragons' unique exo- and endoskeletons.

629 Kinked vertebral columns in the guppy (*Poecelia reticulata*) mutant *curveback* are
630 characterized by wedge-shaped vertebrae (Gorman et al., 2007). Similar to the guppy
631 mutant and to the developmental malformation of vertebrae in human Scheuermann's
632 kyphosis sufferers, vertebrae are keystone-shaped in the weedy seadragon at locations
633 of spinal curvature (Figure 1; Figure S2).

634 One of the most surprising gene family reductions we uncovered is shared by all
635 of the syngnathid lineages we explored; it is the loss of *fgf3* and *fgf4*. Loss of these two
636 fibroblast growth factor genes in syngnathids is striking because their orthologs in other
637 vertebrates have long been hypothesized to play nearly indispensable pleiotropic
638 developmental roles in the pharyngeal arches, teeth, brain, cranial placodes, epidermal
639 appendages, limbs, and the segmental axis (Anderson et al., 2020; Boulet et al., 2004;
640 Cooper et al., 2017; Cooper et al., 2018; Crump et al., 2004; David et al., 2002;
641 Jackman et al., 2013; Leger & Brand, 2002; Lu et al., 2006; Maves et al., 2002; Miyake
642 & Itoh, 2013; Nechiporuk & Raible, 2008; Prykhozhij & Neumann, 2008; Reuter et al.,
643 2019; Walshe et al., 2002; Wang et al., 2007). It is reasonable to weigh whether losing
644 these two multifunctional signaling ligands could have had significant and broad
645 consequences to both deeply conserved developmental pathways and their
646 morphological readouts. Another possibility is that these developmental pathways had
647 diverged neutrally, or through changes in other pathway members, from anciently
648 conserved functions along the syngnathid lineage, permitting genes that had once been
649 critical to become expendable. It is nevertheless valuable to note that syngnathid fishes
650 share peculiarities in many features from the constellation of vertebrate traits that *fgf3*
651 and *fgf4* are known to help pattern.

652 Dermal integuments of syngnathid fishes are bony plates, and in several
653 syngnathid lineages including the seadragons these have been elaborated to

magnificence, sometimes independently. A subset of these plates in seadragons bear blunted struts of bone that end in fleshy paddle-shaped ornaments, the “leaves” and “weeds” of the species featured in this report (Figure 1). Another percomorph clade, the pufferfishes, are adorned with bony spines likely evolved from elasmoid scales. Shono et al., (Shono et al., 2019), showed that *fgf3* (and *fgf1a*) are expressed in developing pufferfish dermal spines. Given this and a trove of other evidence for an FGF signaling role in the development and diversification of scales, spines, and denticles in ray-finned and chondrichthyan fishes (Albertson et al., 2018; Aman et al., 2018; Cooper et al., 2017; Cooper et al., 2018; Daane et al., 2016; Kim et al., 2019; Shono et al., 2019), absence of *fgf3* and *fgf4* in the often heavily armored, elaborately spined syngnathids clearly suggests that derived mechanisms for integumentary bone development are at play in this lineage.

Syngnathids have evolved elongated faces with an unusual hyoid apparatus integral to specialized suction feeding (Figure 1) (Leysen et al., 2010), and they are toothless. These craniofacial features arise developmentally from complex interactions between - at least - endoderm, mesoderm, and neural crest components. Both *fgf3* and *fgf4* are expressed in the pharyngeal arches, which form skeletal elements of the jaw, hyoid, and, in fish, the gill supports (David et al., 2002; Herzog et al., 2004; Niswander & Martin, 1992) *fgf3* plays an essential role in craniofacial development. When *fgf3* is knocked out in zebrafish, mutants die within 7-9 days; craniofacial defects (particularly the full loss of cartilage in gill arches) are predicted to be the cause of their early mortality (Herzog et al., 2004). Additionally, when *fgf3* expression is disrupted in the mesendoderm in developing zebrafish, an “inverted” backward directed ceratohyal cartilage is formed (David et al., 2002). It is tempting to speculate that evolutionary loss of these genes could have led to altered craniofacial architecture of the elongated

679 syngnathid face, either directly or through the effects of genetic compensation
680 percolating through this signaling pathway.

681 Syngnathid toothlessness is also particularly interesting given *fgf3* and *fgf4*
682 expression in zebrafish dental epithelium and their suspected roles in tooth
683 morphogenesis (Jackman et al., 2004). Gilbert et al. (2019) propose a model for
684 patterning of dentition in ray-finned fishes in which a tooth primordium acts as an
685 organizer that induces the development of subsequent teeth, likely via secretion of Fgf3
686 and Fgf4. Other tooth developmental genes are known to be lost or reduced in copy
687 number in syngnathids, including *eve1* (Small et al., 2016); see above) and P/Q-rich
688 SCPP enamel/enameloid matrix genes (Lin et al., 2016; Qu et al., 2021; Zhang et al.,
689 2020). These losses imply an erosion of tooth development pathways spanning
690 induction to mineralization, with our discovery of *fgf3/4* loss enlarging the pool of
691 candidate causative genes for edentulism in syngnathids.

692 The syngnathid central nervous system features its own peculiarities. Bennedetti
693 (1991) described greater pipefish (*S. acus*) and long-snouted seahorse (*Hippocampus*
694 *guttulatus*) to have highly modified or no discernable Mauthner neurons, the large,
695 rhombomere 4 (r4) reticulospinal neurons critical for the rapid “C-start” escape
696 response in many fishes and some amphibians (reviewed in Korn & Faber, 2005).
697 Syngnathids are reputed also to lack mechanosensory lateral line neuromasts, from
698 which the Mauthner cells receive synaptic inputs (Benedetti, 1991; Korn & Faber, 1975).
699 In zebrafish, joint impairment of *fgf3* and *fgf8* impacts segmental identity of
700 rhombomeres 3 and 5 and their reticulospinal neurons (Maves et al., 2002; Walshe et
701 al., 2002). Furthermore, depletion of *fgf3* and *fgf10* reduced the number of zebrafish
702 posterior lateral line neuromasts and inhibited migration of the placode down the length
703 of the body (Nechiporuk & Raible, 2008). These observations present a compelling case

for future interrogation of a link between *fgf3* loss and derived hindbrain and sensory development in syngnathids.

Though it is known that paralogous FGF ligands can compensate for one another in some experimental contexts (Maves et al., 2002; Walshe et al., 2002), in general, we did not find sweeping evidence for adaptive evolution of paralogous FGF proteins or their receptors in the wake of syngnathid *fgf3* and *fgf4* gene losses. Correlated changes could instead have included evolution of non-coding, regulatory sequences of FGF/FGFR genes or changes to other gene families that interact with FGF signaling. Syngnathids, we found, have lost six deeply conserved miRNAs in the miR-130 family. Biological implications for these missing miRNA genes are uncertain, though it is possible that their losses could relate to derived syngnathid-specific traits and gene pathway changes. For instance, angiogenesis and tissue remodeling are critical components in syngnathid male pregnancy tissues (Stolting & Wilson, 2007), and mir130a is connected to vascular repatterning in mammals and teleosts (Chen & Gorski, 2008; Singh et al., 2017) mir130 genes are also known to affect FGF signaling. In chicken, knockdowns of *mir130* lead to increased expression of *fgf8* and FGF pathway member, *Erk 1/2* (Bobbs et al., 2012; Lopez-Sanchez et al., 2015). There is also a suspected regulatory relationship between FGF receptors and these miRNAs; in human differentiating stem cells, nuclear located FGFR1, nFGFR1, binds to the promoters of *mir301b and mir130b* (Stachowiak & Stachowiak, 2016; Terranova et al., 2015). Loss of miRNAs that regulate and are potentially regulated by FGFs and FGFRs could indicate a further extension of the derived restructuring of FGF signaling pathways in syngnathids.

Evidence for positive selection as a compensatory consequence of having lost *fgf3* and *fgf4* in a syngnathid ancestor proved to be scarce. *fgf16*, however, provides a

729 possible example of a separate evolutionary scenario: derived functional change in the
730 seadragon lineage, at both the protein and cis-regulatory levels. Leafy and weedy
731 seadragon fgf16 proteins share an unusual deletion in a deeply conserved motif, and
732 the seadragon gene also appears to have lost a 5' non-coding sequence that is
733 otherwise conserved among syngnathids and distantly related percomorphs. In
734 zebrafish, this gene is necessary for outgrowth of the pectoral fin, upstream of *fgf4* and
735 *fgf8* (Nomura et al., 2006), and we show it is expressed similarly in fin margins in a
736 representative percomorph, threespine stickleback (Figure S19) (*Gasterosteus*
737 *aculeatus*). The eponymous "leaves" of seadragons that are fleshy extensions borne on
738 bony supports extending from the dermal plates, are apparently stiffened by a core of
739 collagenous tissue rather than ossified structures such as fin rays (Figure 1). Homology
740 of the leafy ornaments with fins might pertain only at the level of shared genetic
741 pathways, such as in the case of a cooption or re-deployment of FGF-signaling to elicit
742 proliferation and outgrowth of these superficially fin-like structures. A role in scale or
743 dermal plate patterning for *fgf16* is not known from teleosts, though in birds the
744 orthologous gene, potentially acting with *fgfr1*, can suppress both *shh* expression and
745 downy feather elongation when locally overexpressed (Chen et al., 2016). The
746 seadragon-specific changes in a known AER and integument-patterning gene are
747 intriguing in the context of this morphological oddity, the leafy "appendages". The
748 seahorse fgf16 protein, we found, bears substitutions in the same amino acid motif
749 deleted in seadragons. The fact that two lineages that have evolved elaborate bony and
750 fleshy ornaments (seadragons and seahorses) both show divergence in a deeply
751 conserved motif of this gene is tantalizing and warrants further comparative work.
752

753 **SUMMARY**

754 The genome models of the leafy and weedy seadragon provide a first step
755 towards unlocking how the fantastical forms of these animals and their close relatives
756 have evolved. Even with our preliminary analyses we have identified likely recent
757 expansions of transposable elements, with hypothesized involvement in significant
758 changes to gene families of known consequence, the most surprising being the loss of
759 fgf3 and fgf4 in the sygnathid lineage. Repetitive DNA, including that generated by the
760 activity of TEs, can precipitate genomic rearrangements such as deletions, duplications,
761 and inversions (reviewed in Bourque et al., 2018; Schrader & Schmitz, 2019). It is
762 therefore likely that historical and possibly continuing TE activity in genomes of the
763 syngnathid lineage has resulted in alterations to genome structure and content, with
764 extreme consequences for developmental genetic programs that are otherwise deeply
765 conserved across vertebrates.
766

767 ## ACKNOWLEDGEMENTS

768 We are truly grateful for the dedicated efforts of L. Matsushige and the seadragon
769 husbandry staff of the Birch Aquarium at Scripps, who preserved *P. eques* samples
770 crucial for this work. We also thank the Tennessee Aquarium, and particularly Aquarist
771 K. Hurt for generous sharing of *P. taeniolatus* samples. We are indebted to M.
772 Weitzman and D. Turnbull from the UO GC3F for critical assistance with library
773 preparation and sequencing. We would like to acknowledge staff at Carl Zeiss
774 Microscopy, LLC, particularly J. Mancuso, A. Browning, K. Skinner, and R. White for
775 collaborating to scan the Gulf pipefish in Figure S2. We are grateful for T. Desvignes for
776 guidance on the miRNA annotations. We especially thank C. Kimmel for his helpful
777 comments on the manuscript. This work was funded by National Institutes of Health
778 grants RR032670 and P50GM098911 (to WAC), National Science Foundation grant
779 OPP-2015301 (to WAC, SB, and CMS), and Oregon Research Excellence Funds (to
780 WAC). HMH was supported by the Genetics Training Program (NIH T32GM007413) at
781 UO. Purchase of the Zeiss Xradia 620 Versa at the University of Oregon was made
782 possible with funding support from the M.J. Murdock Charitable Trust Grant SR-
783 201812008 (to WAC).
784

785 ## CONFLICTS OF INTEREST

786 The authors declare no conflicts of interest.

Data Accessibility: All raw sequencing data associated with this study will be available upon peer-reviewed publication via the NCBI Sequencing Read Archive (SRA), under BioProjects PRJNA765699 and PRJNA765702. Annotation and summary files not in the Supplementary Information will be available upon peer-reviewed publication in the Dryad repository.

FIGURES

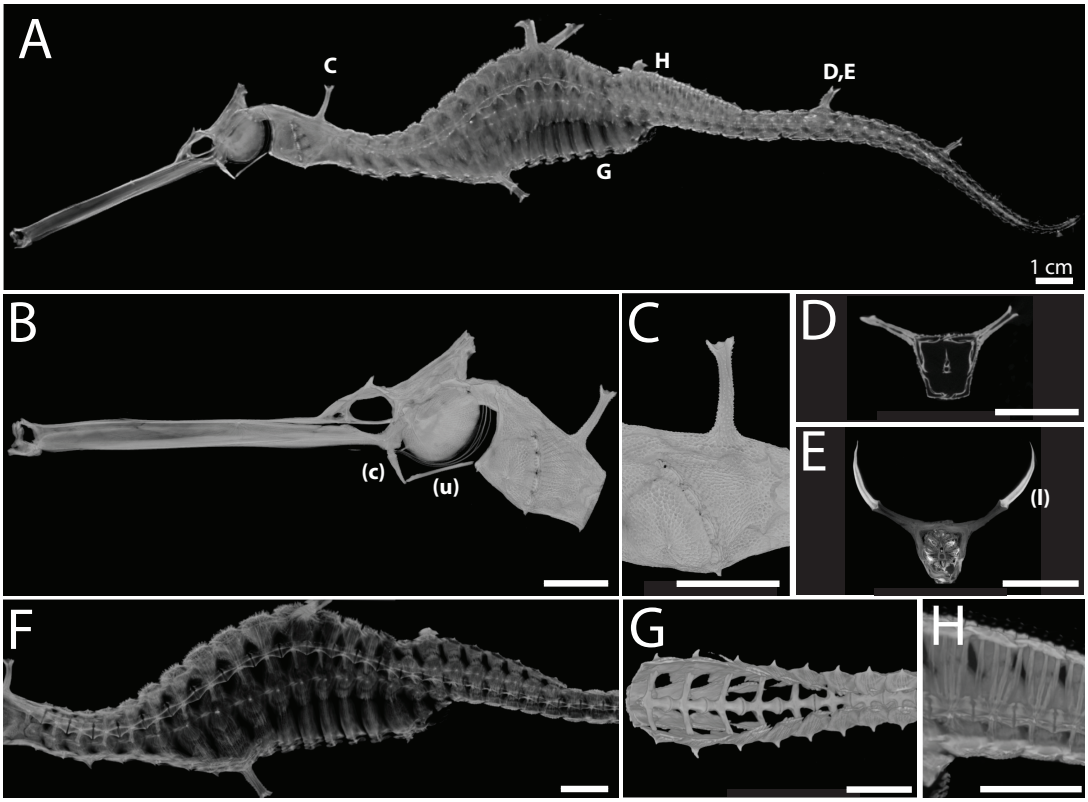


Figure 1. The anatomy of the weedy seadragon includes remarkably elongated facial features terminating in toothless, upturned jaws, an unusual hyoid apparatus specialized for suction feeding, a bony exoskeleton with elaborate spines that support fleshy "leaves", and a sinusoidal spine of rib-less vertebrae that vary in shape and size. A) a lateral view of the skeleton of *Phyllopterix taeniolatus* reconstructed by X-ray microscopy. B) detail of the head (the ceratohyal, c, and urohyal, u, of the hyoid apparatus are noted). C) Detail of the pectoral region (lateral view) showing a dorsal, unpaired "leafy" appendage support surrounded by other dermal plates with much shorter spines. D) optical cross section of the tail through a pair of leafy appendage spines. E) optical cross section through the same appendages as in D) but with a contrast agent that reveals the fleshy leaves, seen edge on (denoted by l). F) lateral view shows keystone-shaped vertebrae at curvatures - both kyphosis and lordosis - of the

spine. G) ventral view of the rib-less abdominal vertebrae. H) a lateral detail of the specialized vertebrae beneath the propulsive dorsal fin.

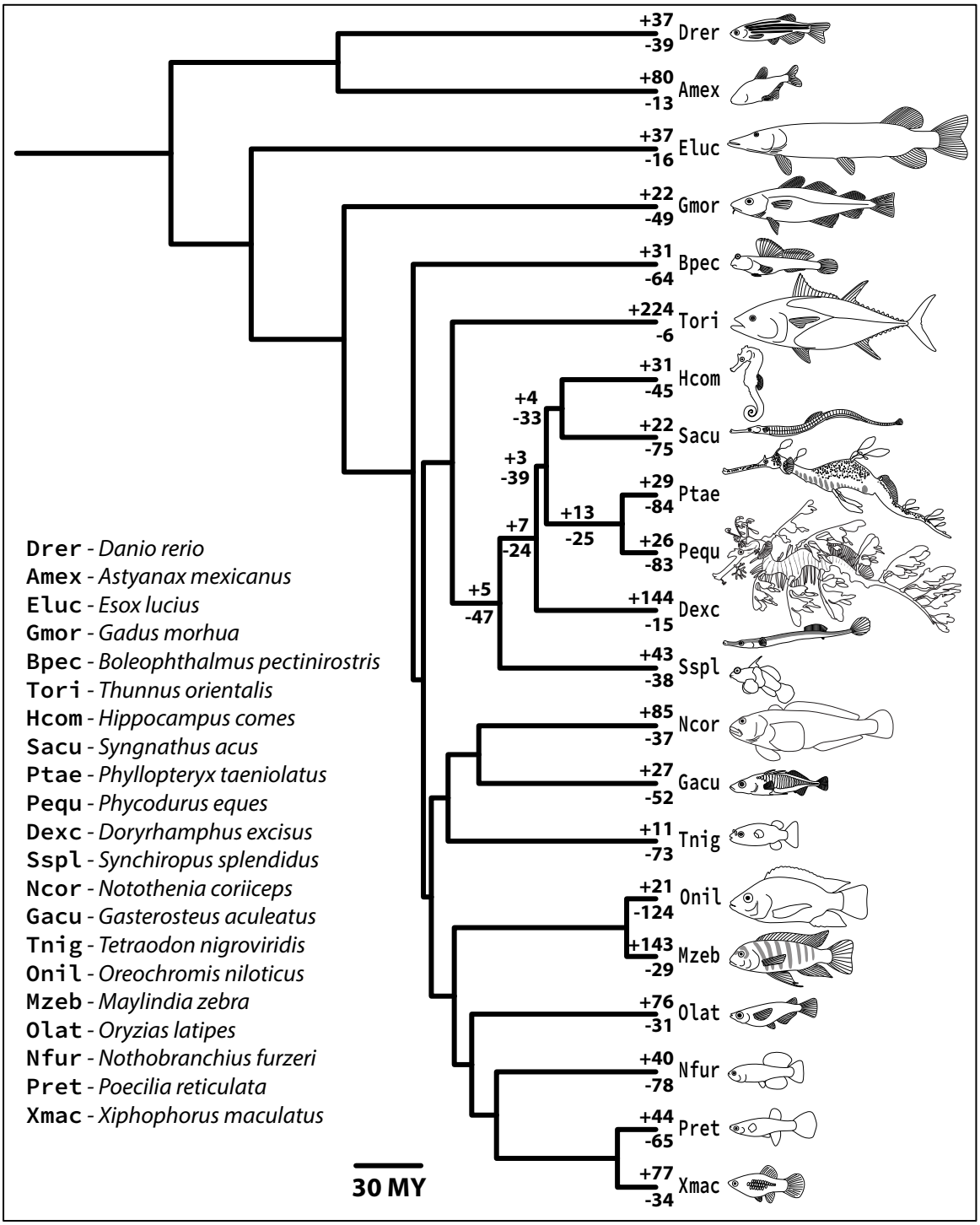


Figure 2. Genomes of 21 teleost species provide phylogenetic context for gene family evolution in seadragons and syngnathid fishes. Represented are evolutionary

relationships among a sample of morphologically diverse teleosts, according to a time-calibrated phylogenetic tree adapted from Rabosky et al. (2018). We used published and newly generated protein-coding gene annotations for the species pictured here to understand putative gene family expansions and contractions in lineages of interest, specifically the seadragons. The number of gene families with statistical evidence for expansion (top values) and contraction (bottom values) along all terminal branches and internal syngnathiform branches are shown. Note that the four-letter species symbols in the legend are used throughout this article.

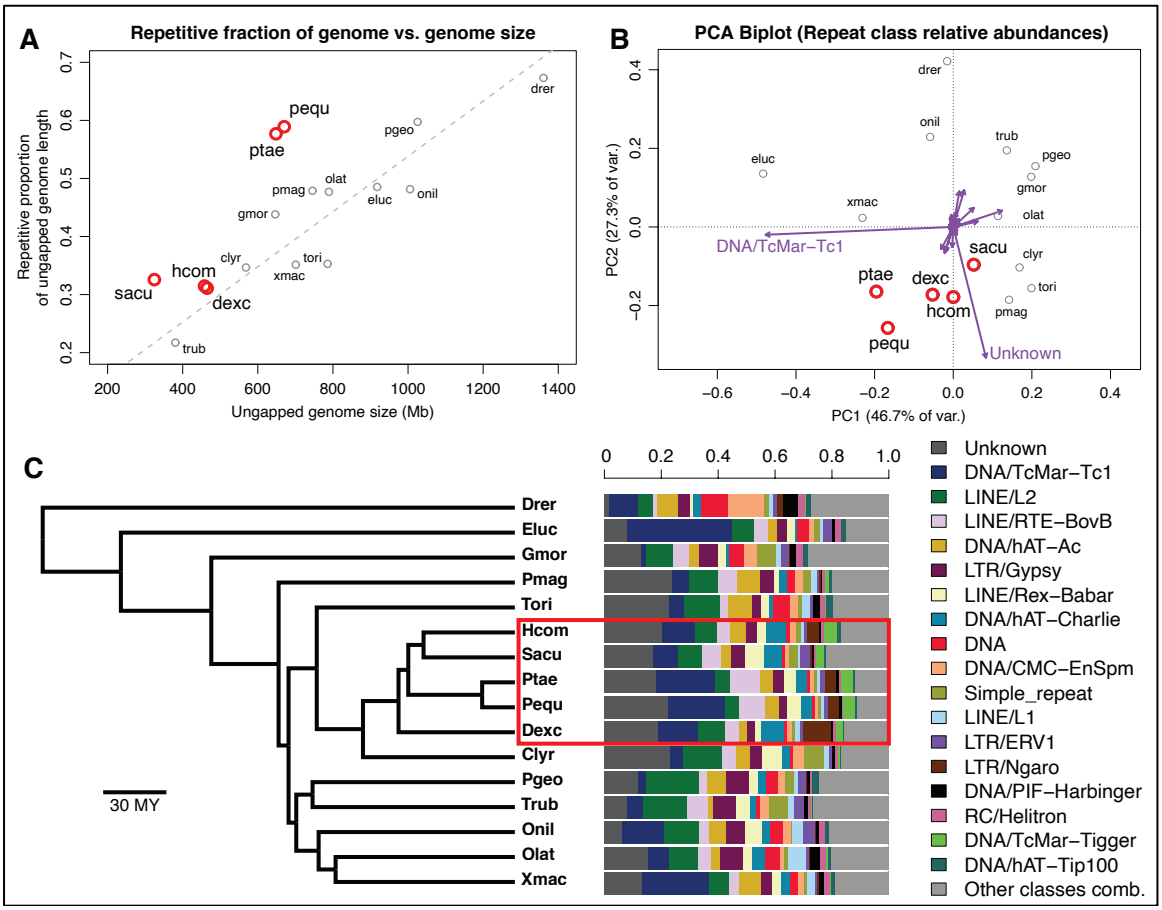
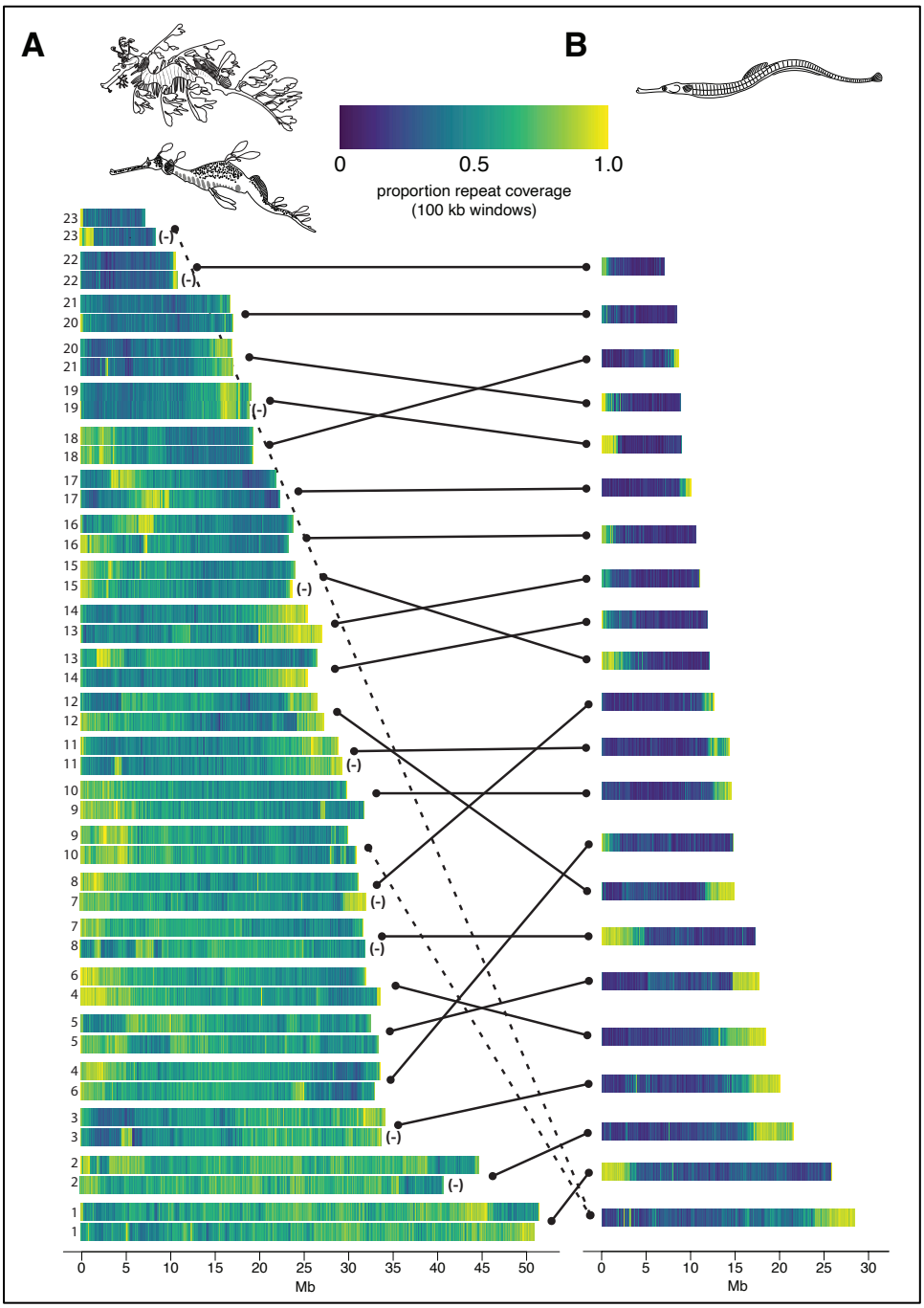


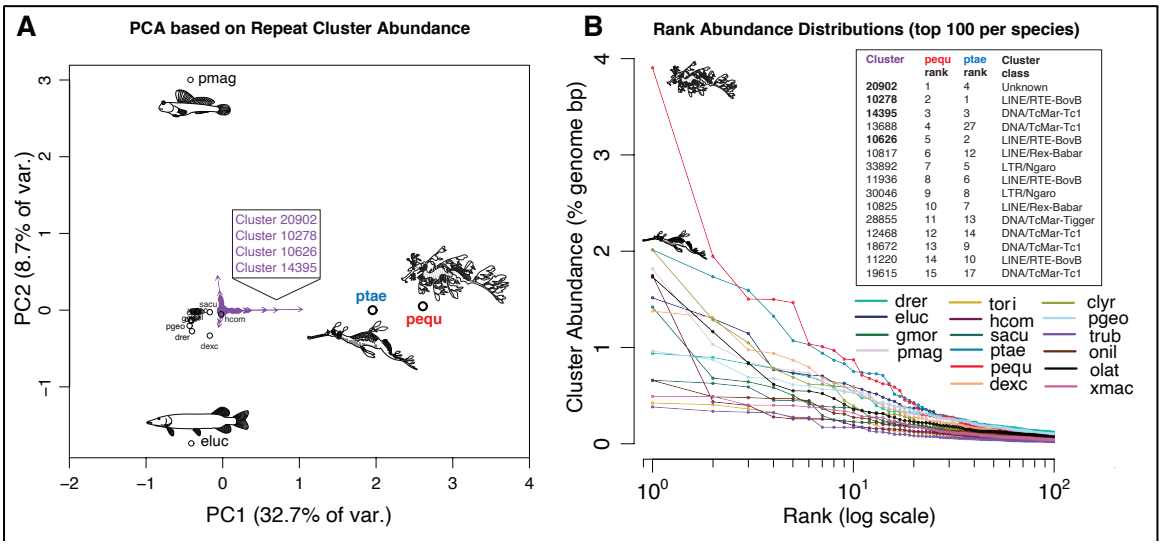
Figure 3. Seadragon genomes demonstrate a large fraction of repetitive DNA relative to other teleost fishes, characterized by substantial contributions from TcMar-Tc1 transposons and unclassified constituents. A) Scatterplot showing the strong, positive relationship between genome assembly size (x-axis) and the proportion of the genome annotated as repetitive for 16 recent teleost genome assemblies. Note that seadragon genomes (pequ and ptae) are especially repetitive (~60%) given their relatively small size (~650 Mb). Dashed line shows a phylogenetic generalized least squares (PGLS) regression fit, and syngnathid genomes are in red. B) A Principal Components Analysis (PCA) biplot shows similarity of the 16 genomes based on relative frequencies of repeat classes. TcMar-Tc1 and “Unknown” repeat classes load especially heavily on PC1 and PC2, respectively, as indicated by vectors (purple arrows) drawn in the space, and these contribute to the distinctiveness of syngnathid genomes (in red). C) Barplots

847 showing the relative abundances of repeat classes across the 16 genomes, ordered
848 (from left to right, and top to bottom in the legend) from highest to lowest mean relative
849 abundance. Phylogenetic relationships among the 16 species are presented as a time-
850 calibrated tree from Rabosky et al. (2018) and the syngnathid clade is indicated by a red
851 rectangle.
852
853
854



855
856
857 **Figure 4.** Leafy and weedy seadragon chromosomes are densely and non-uniformly
858 populated by repetitive DNA. A) Orthologous pairs of the 23 seadragon chromosomes,

ordered ascending from shortest to longest leafy seadragon sequence. For each pair the leafy seadragon chromosome is on top, and the weedy seadragon ortholog is below. Cases in which the orientation of the weedy seadragon chromosome has been reversed to align with the leafy ortholog are denoted by "(-)". B) Chromosome models of the greater pipefish (*Syngnathus acus*), also ascending from shortest to longest. Lines connect seadragon and pipefish chromosomes with large regions of orthology, as inferred via conserved synteny analysis. Dashed lines reflect the fusion of two ancestral syngnathid chromosomes that is derived in the lineage leading to *Syngnathus* and *Hippocampus*. Overall repeat basepair occupation of 100 kb windows (expressed as a proportion) is presented as a heatmap.



Cluster	pequ rank	ptae rank	Cluster class
20902	1	4	Unknown
10278	2	1	LINE/RTE-BovB
14395	3	3	DNA/TcMar-Tc1
13688	4	27	DNA/TcMar-Tc1
10626	5	2	LINE/RTE-BovB
10817	6	12	LINE/Rex-Babar
33892	7	5	LTR/Ngaro
11936	8	6	LINE/RTE-BovB
30046	9	8	LTR/Ngaro
10825	10	7	LINE/Rex-Babar
28855	11	13	DNA/TcMar-Tigger
12468	12	14	DNA/TcMar-Tc1
18672	13	9	DNA/TcMar-Tc1
11220	14	10	LINE/RTE-BovB
19615	15	17	DNA/TcMar-Tc1

Figure 5. Relatively recent repeat expansions in the seadragon lineage drive the uniqueness of seadragon repeatomes. Repeat clusters defined at the 80% sequence identity level and quantified by the proportion of total genome length they occupy were used to conduct principal components analysis (PCA). A) A bi-plot of the first two PCA axes, showing extreme separation of seadragons from the other 14 species in repeatome space, particularly along the leading axis of variation (x-axis). Purple arrows represent individual repeat clusters, and how strongly (and in what direction) they influence the position of genomes in this repeatome space. Four repeat clusters (shown in a box) with especially large loadings on PC1 are strongly associated with seadragon repeatome uniqueness. B) Rank abundance distributions for the top 100 repeat clusters in each of the 16 species included in the repeat analysis. The top 20 clusters for each seadragon species are consistently elevated in abundance, relative to corresponding ranks in the other 14 fish genomes. Shown in a box are the top 15 leafy seadragon clusters, their ranks in weedy seadragon, and the repeat class to which they likely belong. Note that the top 3 (and 5th) clusters in the list correspond to the boxed clusters in panel A.

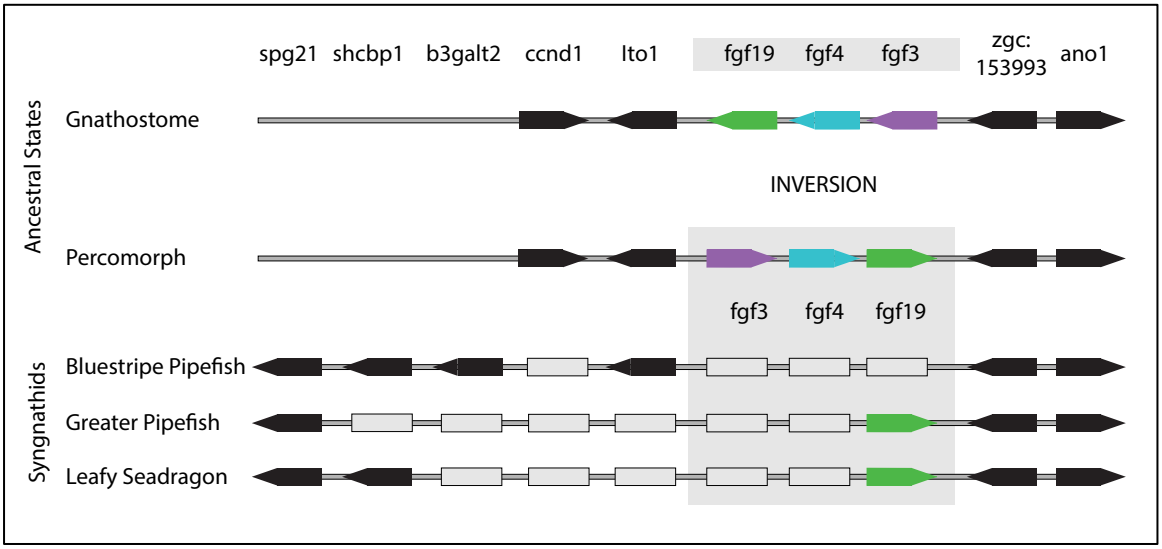


Figure 6. The *fgf3/4/19* cluster locus has experienced surprising gene losses in the syngnathid lineage. While retaining the same immediate gene neighbors, the FGF cluster became inverted in the percomorph fishes relative to outgroups like zebrafish and tetrapods. Several genes (grey rectangles) appear to have been deleted from the locus in the syngnathids, though not all of the losses are shared. *spg21*, *shcbp1*, and *b3galt2*, which neighbor the locus in the syngnathid lineage, are separated from the FGF cluster by other genes in non-syngnathid percomorphs like platyfish. Genes that appear lost from the locus cannot be found anywhere else in the genome assemblies of pipefishes, seahorses or seadragons. Arrows show gene order and orientation; grey rectangles represent apparent gene losses.

REFERENCES

Abramyan, J. (2015). Lineage-specific loss of FGF17 within the avian orders Galliformes and Passeriformes. *Gene*, *563*(2), 180-189. https://doi.org/10.1016/j.gene.2015.03.027

Albertson, R. C., Kawasaki, K. C., Tetrault, E. R., & Powder, K. E. (2018). Genetic analyses in Lake Malawi cichlids identify new roles for Fgf signaling in scale shape variation. *Commun Biol*, *1*, 55. https://doi.org/10.1038/s42003-018-0060-4

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol*, *215*(3), 403-410. https://doi.org/10.1016/S0022-2836(05)80360-2

Aman, A. J., Fulbright, A. N., & Parichy, D. M. (2018). Wnt/beta-catenin regulates an ancient signaling network during zebrafish scale development. *Elife*, *7*. https://doi.org/10.7554/eLife.37001

Anderson, M. J., Magidson, V., Kageyama, R., & Lewandoski, M. (2020). Fgf4 maintains Hes7 levels critical for normal somite segmentation clock function. *Elife*, *9*. https://doi.org/10.7554/eLife.55608

Balla, K. M., Rice, M. C., Gagnon, J. A., & Elde, N. C. (2020). Linking virus discovery to immune responses visualized during zebrafish infections. *Curr Biol*, *30*(11), 2092-2103 e2095. https://doi.org/10.1016/j.cub.2020.04.031

Bejerano, G., Lowe, C. B., Ahituv, N., King, B., Siepel, A., Salama, S. R., Rubin, E. M., Kent, W. J., & Haussler, D. (2006). A distal enhancer and an ultraconserved exon are derived from a novel retroposon. *Nature*, *441*(7089), 87-90. https://doi.org/10.1038/nature04696

Benedetti. (1991). Mauthner neurons in syngnathid bony fish.

Bizuayehu, T. T., & Babiak, I. (2014). MicroRNA in teleost fish. *Genome Biol Evol*, *6*(8), 1911-1937. https://doi.org/10.1093/gbe/evu151

Bobbs, A. S., Saarela, A. V., Yatskievych, T. A., & Antin, P. B. (2012). Fibroblast growth factor (FGF) signaling during gastrulation negatively modulates the abundance of microRNAs that regulate proteins required for cell migration and embryo patterning. *J Biol Chem*, *287*(46), 38505-38514. https://doi.org/10.1074/jbc.M112.400598

Boudinot, P., van der Aa, L. M., Jouneau, L., Du Pasquier, L., Pontarotti, P., Briolat, V., Benmansour, A., & Levraud, J. P. (2011). Origin and evolution of TRIM proteins: new insights from the complete TRIM repertoire of zebrafish and pufferfish. *PLoS ONE*, *6*(7), e22022. https://doi.org/10.1371/journal.pone.0022022

Boulet, A. M., Moon, A. M., Arenkiel, B. R., & Capecchi, M. R. (2004). The roles of Fgf4 and Fgf8 in limb bud initiation and outgrowth. *Dev Biol*, *273*(2), 361-372. https://doi.org/10.1016/j.ydbio.2004.06.012

Bourque, G., Burns, K. H., Gehring, M., Gorbunova, V., Seluanov, A., Hammell, M., Imbeault, M., Izsvak, Z., Levin, H. L., Macfarlan, T. S., Mager, D. L., & Feschotte,

C. (2018). Ten things you should know about transposable elements. *Genome Biol*, *19*(1), 199. https://doi.org/10.1186/s13059-018-1577-z

Bruna, T., Hoff, K. J., Lomsadze, A., Stanke, M., & Borodovsky, M. (2021). BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genom Bioinform*, *3*(1), lqaa108. https://doi.org/10.1093/nargab/lqaa108

Bushnell, B. (2014). BBMap: A fast, accurate, splice-aware aligner. *sourceforge.net/projects/bbmap/*.

Canestro, C., Yokoi, H., & Postlethwait, J. H. (2007). Evolutionary developmental biology and genomics. *Nat Rev Genet*, *8*(12), 932-942. https://doi.org/10.1038/nrg2226

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: an analysis tool set for population genomics. *Mol Ecol*, *22*(11), 3124-3140. https://doi.org/10.1111/mec.12354

Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W., & Postlethwait, J. H. (2011). Stacks: building and genotyping Loci de novo from short-read sequences. *G3 (Bethesda)*, *1*(3), 171-182. https://doi.org/10.1534/g3.111.000240

Catchen, J. M., Conery, J. S., & Postlethwait, J. H. (2009). Automated identification of conserved synteny after whole-genome duplication. *Genome Res*, *19*(8), 1497-1505. https://doi.org/10.1101/gr.090480.108

Chang, M. X., Xiong, F., Wu, X. M., & Hu, Y. W. (2021). The expanding and function of NLRC3 or NLRC3-like in teleost fish: Recent advances and novel insights. *Dev Comp Immunol*, *114*, 103859. https://doi.org/10.1016/j.dci.2020.103859

Chen, C. K., Ng, C. S., Wu, S. M., Chen, J. J., Cheng, P. L., Wu, P., Lu, M. Y., Chen, D. R., Chuong, C. M., Cheng, H. C., Ting, C. T., & Li, W. H. (2016). Regulatory differences in natal down development between altricial zebra finch and precocial chicken. *Molecular Biology and Evolution*, *33*(8), 2030-2043. https://doi.org/10.1093/molbev/msw085

Chen, Y., & Gorski, D. H. (2008). Regulation of angiogenesis through a microRNA (miR-130a) that down-regulates antiangiogenic homeobox genes GAX and HOXA5. *Blood*, *111*(3), 1217-1226. https://doi.org/10.1182/blood-2007-07-104133

Chin, C. S., Alexander, D. H., Marks, P., Klammer, A. A., Drake, J., Heiner, C., Clum, A., Copeland, A., Huddleston, J., Eichler, E. E., Turner, S. W., & Korlach, J. (2013). Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat Methods*, *10*(6), 563-569. https://doi.org/10.1038/nmeth.2474

Choi, Y., & Chan, A. P. (2015). PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics*, *31*(16), 2745-2747. https://doi.org/10.1093/bioinformatics/btv195

Connolly, R., Melville, A., & Keesing, J. (2002). Abundance, movement and individual identification of leafy seadragons, *Phycodurus eques* (Pisces: Syngnathidae). *Marine and Freshwater Research*, *53*, 777-780.

991 Cooper, R. L., Martin, K. J., Rasch, L. J., & Fraser, G. J. (2017). Developing an ancient
992 epithelial appendage: FGF signalling regulates early tail denticle formation in
993 sharks. *Evodevo*, *8*, 8. https://doi.org/10.1186/s13227-017-0071-0

994 Cooper, R. L., Thiery, A. P., Fletcher, A. G., Delbarre, D. J., Rasch, L. J., & Fraser, G. J.
995 (2018). An ancient Turing-like patterning mechanism regulates skin denticle
996 development in sharks. *Sci Adv*, *4*(11), eaau5484.
997 https://doi.org/10.1126/sciadv.aau5484

998 Coutinho, P., Parsons, M. J., Thomas, K. A., Hirst, E. M., Saude, L., Campos, I.,
999 Williams, P. H., & Stemple, D. L. (2004). Differential requirements for COPI
1000 transport during vertebrate early development. *Dev Cell*, *7*(4), 547-558.
1001 https://doi.org/10.1016/j.devcel.2004.07.020

1002 Crump, J. G., Maves, L., Lawson, N. D., Weinstein, B. M., & Kimmel, C. B. (2004). An
1003 essential role for Fgfs in endodermal pouch formation influences later craniofacial
1004 skeletal patterning. *Development*, *131*(22), 5703-5716.
1005 https://doi.org/10.1242/dev.01444

1006 Daane, J. M., Rohner, N., Konstantinidis, P., Djuranovic, S., & Harris, M. P. (2016).
1007 Parallelism and epistasis in skeletal evolution identified through use of
1008 phylogenomic mapping strategies. *Molecular Biology and Evolution*, *33*(1), 162-
1009 173. https://doi.org/10.1093/molbev/msv208

1010 David, N. B., Saint-Etienne, L., Tsang, M., Schilling, T. F., & Rosa, F. M. (2002).
1011 Requirement for endoderm and FGF3 in ventral head skeleton formation.
1012 *Development*, *129*(19), 4457-4468. https://doi.org/10.1242/dev.129.19.4457

1013 Dawson, C. E. (1985). *Indo-Pacific pipefishes (Red Sea to the Americas)*. Gulf Coast
1014 Research Laboratory.

1015 Desvignes, T., Batzel, P., Sydes, J., Eames, B. F., & Postlethwait, J. H. (2019). miRNA
1016 analysis with Prost! reveals evolutionary conservation of organ-enriched
1017 expression and post-transcriptional modifications in three-spined stickleback and
1018 zebrafish. *Sci Rep*, *9*(1), 3913. https://doi.org/10.1038/s41598-019-40361-8

1019 Desvignes, T., Sydes, J., Montfort, J., Bobe, J., & Postlethwait, J. H. (2021). Evolution
1020 after whole-genome duplication: teleost microRNAs. *Molecular Biology and*
1021 *Evolution*, *38*(8), 3308-3331. https://doi.org/10.1093/molbev/msab105

1022 Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P.,
1023 Chaisson, M., & Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq
1024 aligner. *Bioinformatics*, *29*(1), 15-21.
1025 https://doi.org/10.1093/bioinformatics/bts635

1026 Dudchenko, O., Batra, S. S., Omer, A. D., Nyquist, S. K., Hoeger, M., Durand, N. C.,
1027 Shamim, M. S., Machol, I., Lander, E. S., Aiden, A. P., & Aiden, E. L. (2017). De
1028 novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-
1029 length scaffolds. *Science*, *356*(6333), 92-95.
1030 https://doi.org/10.1126/science.aal3327

1031 Dudchenko, O., Shamim, M. S., Batra, S. S., Durand, N. C., Musial, N. T., Mostofa, R.,
1032 Pham, M., Glenn St Hilaire, B., Yao, W., Stamenova, E., Hoeger, M., Nyquist, S.
1033 K., Korchina, V., Pletch, K., Flanagan, J. P., Tomaszewicz, A., McAloose, D.,
1034 Pérez Estrada, C., Novak, B. J., Omer, A. D., & Aiden, E. L. (2018). The

Juicebox Assembly Tools module facilitates *de novo* assembly of mammalian genomes with chromosome-length scaffolds for under $1000. *bioRxiv*. <Go to WoS>://WOS:000255412600014

Durand, N. C., Robinson, J. T., Shamim, M. S., Machol, I., Mesirov, J. P., Lander, E. S., & Aiden, E. L. (2016). Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. *Cell Syst*, *3*(1), 99-101. https://doi.org/10.1016/j.cels.2015.07.012

Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, *26*(19), 2460-2461. https://doi.org/10.1093/bioinformatics/btq461

Enright, A. J., Van Dongen, S., & Ouzounis, C. A. (2002). An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res*, *30*(7), 1575-1584. https://doi.org/10.1093/nar/30.7.1575

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. *Proc Natl Acad Sci U S A*, *117*(17), 9451-9457. https://doi.org/10.1073/pnas.1921046117

Gao, B., Chen, W., Shen, D., Wang, S., Chen, C., Zhang, L., Wang, W., Wang, X., & Song, C. (2017). Characterization of autonomous families of Tc1/mariner transposons in neoteleost genomes. *Mar Genomics*, *34*, 67-77. https://doi.org/10.1016/j.margen.2017.05.003

Gao, B., Shen, D., Xue, S., Chen, C., Cui, H., & Song, C. (2016). The contribution of transposable elements to size variations between four teleost genomes. *Mob DNA*, *7*, 4. https://doi.org/10.1186/s13100-016-0059-7

Gao, B., Wang, Y., Diaby, M., Zong, W., Shen, D., Wang, S., Chen, C., Wang, X., & Song, C. (2020). Evolution of pogo, a separate superfamily of IS630-Tc1-mariner transposons, revealing recurrent domestication events in vertebrates. *Mob DNA*, *11*, 25. https://doi.org/10.1186/s13100-020-00220-0

Gibert, Y., Samarut, E., Ellis, M. K., Jackman, W. R., & Laudet, V. (2019). The first formed tooth serves as a signalling centre to induce the formation of the dental row in zebrafish. *Proc Biol Sci*, *286*(1904), 20190401. https://doi.org/10.1098/rspb.2019.0401

Gorman, K. F., Tredwell, S. J., & Breden, F. (2007). The mutant guppy syndrome curveback as a model for human heritable spinal curvature. *Spine (Phila Pa 1976)*, *32*(7), 735-741. https://doi.org/10.1097/01.brs.0000259081.40354.e2

Gray, R. S., Wilm, T. P., Smith, J., Bagnat, M., Dale, R. M., Topczewski, J., Johnson, S. L., & Solnica-Krezel, L. (2014). Loss of col8a1a function during zebrafish embryogenesis results in congenital vertebral malformations. *Dev Biol*, *386*(1), 72-85. https://doi.org/10.1016/j.ydbio.2013.11.028

Gurevich, A., Saveliev, V., Vyahhi, N., & Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. *Bioinformatics*, *29*(8), 1072-1075. https://doi.org/10.1093/bioinformatics/btt086

Haas, B. J. TransposonPSI. *http://transposonpsi.sourceforge.net*.

Hahn, M. W., Demuth, J. P., & Han, S. G. (2007). Accelerated rate of gene gain and loss in primates. *Genetics*, *177*(3), 1941-1949. https://doi.org/10.1534/genetics.107.080077

Herzog, W., Sonntag, C., von der Hardt, S., Roehl, H. H., Varga, Z. M., & Hammerschmidt, M. (2004). Fgf3 signaling from the ventral diencephalon is required for early specification and subsequent survival of the zebrafish adenohypophysis. *Development*, *131*(15), 3681-3692. https://doi.org/10.1242/dev.01235

Howe, K., Schiffer, P. H., Zielinski, J., Wiehe, T., Laird, G. K., Marioni, J. C., Soylemez, O., Kondrashov, F., & Leptin, M. (2016). Structure and evolutionary history of a large family of NLR proteins in the zebrafish. *Open Biol*, *6*(4), 160009. https://doi.org/10.1098/rsob.160009

Ivancevic, A., & Chuong, E. B. (2020). Transposable elements teach T cells new tricks. *Proc Natl Acad Sci U S A*, *117*(17), 9145-9147. https://doi.org/10.1073/pnas.2004493117

Ivancevic, A. M., Kortschak, R. D., Bertozzi, T., & Adelson, D. L. (2018). Horizontal transfer of BovB and L1 retrotransposons in eukaryotes. *Genome Biol*, *19*(1), 85. https://doi.org/10.1186/s13059-018-1456-7

Jackman, W. R., Davies, S. H., Lyons, D. B., Stauder, C. K., Denton-Schneider, B. R., Jowdry, A., Aigler, S. R., Vogel, S. A., & Stock, D. W. (2013). Manipulation of Fgf and Bmp signaling in teleost fishes suggests potential pathways for the evolutionary origin of multicuspid teeth. *Evol Dev*, *15*(2), 107-118. https://doi.org/10.1111/ede.12021

Jackman, W. R., Draper, B. W., & Stock, D. W. (2004). Fgf signaling is required for zebrafish tooth development. *Dev Biol*, *274*(1), 139-157. https://doi.org/10.1016/j.ydbio.2004.07.003

Jones, P., Binns, D., Chang, H. Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A. F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S. Y., Lopez, R., & Hunter, S. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics*, *30*(9), 1236-1240. https://doi.org/10.1093/bioinformatics/btu031

Kelley, J. L., Desvignes, T., McGowan, K. L., Perez, M., Rodriguez, L. A., Brown, A. P., Culumber, Z., & Tobler, M. (2021). microRNA expression variation as a potential molecular mechanism contributing to adaptation to hydrogen sulphide. *J Evol Biol*, *34*(6), 977-988. https://doi.org/10.1111/jeb.13727

Kidwell, M. G. (2002). Transposable elements and the evolution of genome size in eukaryotes. *Genetica*, *115*(1), 49-63. https://doi.org/10.1023/a:1016072014259

Kim, D. I., Kai, W., Hosoya, S., Sato, M., Nozawa, A., Kuroyanagi, M., Jo, Y., Tasumi, S., Suetake, H., Suzuki, Y., & Kikuchi, K. (2019). The genetic basis of scale-loss phenotype in the rapid radiation of Takifugu fishes. *Genes (Basel)*, *10*(12). https://doi.org/10.3390/genes10121027

Klanten, O. S., Gaither, M. R., Greaves, S., Mills, K., O'Keeffe, K., Turnbull, J., McKinnon, R., & Booth, D. J. (2020). Genomic and morphological evidence of distinct populations in the endemic common (weedy) seadragon *Phyllopteryx*

taeniolatus (Syngnathidae) along the east coast of Australia. *PLoS ONE*, *15*(12), e0243446. https://doi.org/10.1371/journal.pone.0243446

Kolmogorov, M., Yuan, J., Lin, Y., & Pevzner, P. A. (2019). Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol*, *37*(5), 540-546. https://doi.org/10.1038/s41587-019-0072-8

Korn, H., & Faber, D. S. (1975). Inputs from the posterior lateral line nerves upon the goldfish Mauthner cell. I. Properties and synaptic localization of the excitatory component. *Brain Research*, *96*(2), 342-348. https://doi.org/10.1016/0006-8993(75)90745-3

Korn, H., & Faber, D. S. (2005). The Mauthner cell half a century later: a neurobiological model for decision-making? *Neuron*, *47*(1), 13-28. https://doi.org/10.1016/j.neuron.2005.05.019

Krucken, J., Schroetel, R. M., Muller, I. U., Saidani, N., Marinovski, P., Benten, W. P., Stamm, O., & Wunderlich, F. (2004). Comparative analysis of the human gimap gene cluster encoding a novel GTPase family. *Gene*, *341*, 291-304. https://doi.org/10.1016/j.gene.2004.07.005

Kunarso, G., Chia, N. Y., Jeyakani, J., Hwang, C., Lu, X., Chan, Y. S., Ng, H. H., & Bourque, G. (2010). Transposable elements have rewired the core regulatory network of human embryonic stem cells. *Nat Genet*, *42*(7), 631-634. https://doi.org/10.1038/ng.600

Leger, S., & Brand, M. (2002). Fgf8 and Fgf3 are required for zebrafish ear placode induction, maintenance and inner ear patterning. *Mech Dev*, *119*(1), 91-108. https://doi.org/10.1016/s0925-4773(02)00343-x

Leysen, H., Jouk, P., Brunain, M., Christiaens, J., & Adriaens, D. (2010). Cranial architecture of tube-snouted gasterosteiformes (*Syngnathus rostellatus* and *Hippocampus capensis*). *J Morphol*, *271*(3), 255-270. https://doi.org/10.1002/jmor.10795

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., Zhang, H., Qu, M., Liu, S., Chen, Z., Zhong, J., Zhang, H., Meng, L., Wang, K., Yin, J., Huang, L., Venkatesh, B., Meyer, A., Lu, X., & Lin, Q. (2021). Genome sequences reveal global dispersal routes and suggest convergent genetic adaptations in seahorse evolution. *Nat Commun*, *12*(1), 1094. https://doi.org/10.1038/s41467-021-21379-x

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., Min, J., Zhang, C., Qiu, Y., Bai, J., He, W., Bian, C., Zhang, X., Shan, D., Qu, H., Sun, Y., Gao, Q., Huang, L., Shi, Q., Meyer, A., & Venkatesh, B. (2016). The seahorse genome and the evolution of its specialized morphology. *Nature*, *540*(7633), 395-399. https://doi.org/10.1038/nature20595

Liu, N., Lee, C. H., Swigut, T., Grow, E., Gu, B., Bassik, M. C., & Wysocka, J. (2018). Selective silencing of euchromatic L1s revealed by genome-wide screens for L1 regulators. *Nature*, *553*(7687), 228-232. https://doi.org/10.1038/nature25179

Longo, S. J., Faircloth, B. C., Meyer, A., Westneat, M. W., Alfaro, M. E., & Wainwright, P. C. (2017). Phylogenomic analysis of a rapid radiation of misfit fishes (Syngnathiformes) using ultraconserved elements. *Mol Phylogenet Evol*, *113*, 33-48. https://doi.org/10.1016/j.ympev.2017.05.002

Lopez-Sanchez, C., Franco, D., Bonet, F., Garcia-Lopez, V., Aranega, A., & Garcia-Martinez, V. (2015). Negative Fgf8-Bmp2 feed-back is regulated by miR-130 during early cardiac specification. *Dev Biol*, *406*(1), 63-73. https://doi.org/10.1016/j.ydbio.2015.07.007

Lu, P., Minowada, G., & Martin, G. R. (2006). Increasing Fgf4 expression in the mouse limb bud causes polysyndactyly and rescues the skeletal defects that result from loss of Fgf8 function. *Development*, *133*(1), 33-42. https://doi.org/10.1242/dev.02172

Lynch, M. (2007). *The Origins of Genome Architecture*. Oxford University Press.

Lynch, M., & Conery, J. S. (2003). The origins of genome complexity. *Science*, *302*(5649), 1401-1404. https://doi.org/10.1126/science.1089370

Mank, J. E., & Avise, J. C. (2006). Phylogenetic conservation of chromosome numbers in Actinopterygiian fishes. *Genetica*, *127*(1-3), 321-327. https://doi.org/10.1007/s10709-005-5248-0

Marom, R., Burrage, L. C., Venditti, R., Clement, A., Blanco-Sanchez, B., Jain, M., Scott, D. A., Rosenfeld, J. A., Sutton, V. R., Shinawi, M., Mirzaa, G., DeVile, C., Roberts, R., Calder, A. D., Allgrove, J., Grafe, I., Lanza, D. G., Li, X., Joeng, K. S., Lee, Y. C., Song, I. W., Sliepka, J. M., Batkovskyte, D., Washington, M., Dawson, B. C., Jin, Z., Jiang, M. M., Chen, S., Chen, Y., Tran, A. A., Emrick, L. T., Murdock, D. R., Hanchard, N. A., Zapata, G. E., Mehta, N. R., Weis, M. A., Scott, A. A., Tremp, B. A., Phillips, J. B., Wegner, J., Taylor-Miller, T., Gibbs, R. A., Muzny, D. M., Jhangiani, S. N., Hicks, J., Stottmann, R. W., Dickinson, M. E., Seavitt, J. R., Heaney, J. D., Eyre, D. R., Undiagnosed Diseases, N., Westerfield, M., De Matteis, M. A., & Lee, B. (2021). COPB2 loss of function causes a coatopathy with osteoporosis and developmental delay. *Am J Hum Genet*, *108*(9), 1710-1724. https://doi.org/10.1016/j.ajhg.2021.08.002

Martin-Smith, K. M., & Vincent, A. C. J. (2006). Exploitation and trade of Australian seahorses, pipehorses, sea dragons and pipefishes (Family Syngnathidae). *Oryx*, *40*(2), 141-151. https://doi.org/10.1017/s003060530600010x

Mattingsdal, M., Jentoft, S., Torresen, O. K., Knutsen, H., Hansen, M. M., Robalo, J. I., Zagrodzka, Z., Andre, C., & Gonzalez, E. B. (2018). A continuous genome assembly of the corkwing wrasse (*Symphodus melops*). *Genomics*, *110*(6), 399-403. https://doi.org/10.1016/j.ygeno.2018.04.009

Maves, L., Jackman, W., & Kimmel, C. B. (2002). FGF3 and FGF8 mediate a rhombomere 4 signaling activity in the zebrafish hindbrain. *Development*, *129*(16), 3825-3837. https://www.ncbi.nlm.nih.gov/pubmed/12135921

Mendes, F. K., Vanderpool, D., Fulton, B., & Hahn, M. W. (2020). CAFE 5 models variation in evolutionary rates among gene families. *Bioinformatics*. https://doi.org/10.1093/bioinformatics/btaa1022

Miyake, A., & Itoh, N. (2013). Fgf22 regulated by Fgf3/Fgf8 signaling is required for zebrafish midbrain development. *Biol Open*, *2*(5), 515-524. https://doi.org/10.1242/bio.20134226

Nechiporuk, A., & Raible, D. W. (2008). FGF-dependent mechanosensory organ patterning in zebrafish. *Science*, *320*(5884), 1774-1777. https://doi.org/10.1126/science.1156547

Niswander, L., & Martin, G. R. (1992). Fgf-4 expression during gastrulation, myogenesis, limb and tooth development in the mouse. *Development*, *114*(3), 755-768. https://www.ncbi.nlm.nih.gov/pubmed/1618140

Nomura, R., Kamei, E., Hotta, Y., Konishi, M., Miyake, A., & Itoh, N. (2006). Fgf16 is essential for pectoral fin bud formation in zebrafish. *Biochem Biophys Res Commun*, *347*(1), 340-346. https://doi.org/10.1016/j.bbrc.2006.06.108

Oulion, S., Bertrand, S., & Escriva, H. (2012). Evolution of the FGF Gene Family. *Int J Evol Biol*, *2012*, 298147. https://doi.org/10.1155/2012/298147

Peskin, B., Henke, K., Cumplido, N., Treaster, S., Harris, M. P., Bagnat, M., & Arratia, G. (2020). Notochordal signals establish phylogenetic identity of the teleost spine. *Curr Biol*, *30*(14), 2805-2814 e2803. https://doi.org/10.1016/j.cub.2020.05.037

Prykhozhij, S. V., & Neumann, C. J. (2008). Distinct roles of Shh and Fgf signaling in regulating cell proliferation during zebrafish pectoral fin development. *BMC Dev Biol*, *8*, 91. https://doi.org/10.1186/1471-213X-8-91

Qu, M., Liu, Y., Zhang, Y., Wan, S., Ravi, V., Qin, G., Jiang, H., Wang, X., Zhang, H., Zhang, B., Gao, Z., Huysseune, A., Zhang, Z., Zhang, H., Chen, Z., Yu, H., Wu, Y., Tang, L., Li, C., Zhong, J., Ma, L., Wang, F., Zheng, H., Yin, J., Witten, P. E., Meyer, A., Venkatesh, B., & Lin, Q. (2021). Seadragon genome analysis provides insights into its phenotype and sex determination locus. *Sci Adv*, *7*(34). https://doi.org/10.1126/sciadv.abg5196

R_Core_Team. (2019). A language and environment for statistical computing. R Foundation for Statistical Computing,

. *Vienna, Austria. URL https://www.R-project.org/*.

Rabosky, D. L., Chang, J., Title, P. O., Cowman, P. F., Sallan, L., Friedman, M., Kaschner, K., Garilao, C., Near, T. J., Coll, M., & Alfaro, M. E. (2018). An inverse latitudinal gradient in speciation rate for marine fishes. *Nature*, *559*(7714), 392-395. https://doi.org/10.1038/s41586-018-0273-1

Reuter, I., Jackels, J., Kneitz, S., Kuper, J., Lesch, K. P., & Lillesaar, C. (2019). Fgf3 is crucial for the generation of monoaminergic cerebrospinal fluid contacting cells in zebrafish. *Biol Open*, *8*(6). https://doi.org/10.1242/bio.040683

Roth, O., Solbakken, M. H., Torresen, O. K., Bayer, T., Matschiner, M., Baalsrud, H. T., Hoff, S. N. K., Brieuc, 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. *Proc Natl Acad Sci U S A*, *117*(17), 9431-9439. https://doi.org/10.1073/pnas.1916251117

Schnell, S., Demolliere, C., van den Berk, P., & Jacobs, H. (2006). Gimap4 accelerates T-cell death. *Blood*, *108*(2), 591-599. https://doi.org/10.1182/blood-2005-11-4616

1251 Schrader, L., & Schmitz, J. (2019). The impact of transposable elements in adaptive
1252 evolution. *Mol Ecol*, *28*(6), 1537-1549. https://doi.org/10.1111/mec.14794

1253 Shao, F., Wang, J., Xu, H., & Peng, Z. (2018). FishTEDB: a collective database of
1254 transposable elements identified in the complete genomes of fish. *Database*
1255 *(Oxford)*, *2018*. https://doi.org/10.1093/database/bax106

1256 Shono, T., Thiery, A. P., Cooper, R. L., Kurokawa, D., Britz, R., Okabe, M., & Fraser, G.
1257 J. (2019). Evolution and developmental diversity of skin spines in pufferfishes.
1258 *iScience*, *19*, 1248-1259. https://doi.org/10.1016/j.isci.2019.06.003

1259 Singh, B. N., Tahara, N., Kawakami, Y., Das, S., Koyano-Nakagawa, N., Gong, W.,
1260 Garry, M. G., & Garry, D. J. (2017). Etv2-miR-130a-Jarid2 cascade regulates
1261 vascular patterning during embryogenesis. *PLoS ONE*, *12*(12), e0189010.
1262 https://doi.org/10.1371/journal.pone.0189010

1263 Small, C. M., Bassham, S., Catchen, J., Amores, A., Fuiten, A. M., Brown, R. S., Jones,
1264 A. G., & Cresko, W. A. (2016). The genome of the Gulf pipefish enables
1265 understanding of evolutionary innovations. *Genome Biol*, *17*(1), 258.
1266 https://doi.org/10.1186/s13059-016-1126-6

1267 Smit, A. F. A., Hubley, R., & Green, P. RepeatMasker at http://repeatmasker.org.

1268 Stachowiak, M. K., & Stachowiak, E. K. (2016). Evidence-based theory for integrated
1269 genome regulation of ontogeny—an unprecedented role of nuclear FGFR1
1270 signaling. *J Cell Physiol*, *231*(6), 1199-1218. https://doi.org/10.1002/jcp.25298

1271 Stervander, M., & Cresko, W. A. (2021). A highly contiguous nuclear genome assembly
1272 of the mandarinfish *Synchiropus splendidus* (Syngnathiformes:

1273 Callionymidae). *G3 (Bethesda)*, *in press*. https://doi.org/10.1093/g3journal/jkab306

1274 Stiller, J., da Fonseca, R. R., Alfaro, M. E., Faircloth, B. C., Wilson, N. G., & Rouse, G.
1275 W. (2021). Using ultraconserved elements to track the influence of sea-level
1276 change on leafy seadragon populations. *Mol Ecol*, *30*(6), 1364-1380.
1277 https://doi.org/10.1111/mec.15744

1278 Stiller, J., Wilson, N. G., Donnellan, S., & Rouse, G. W. (2017). The leafy seadragon,
1279 *Phycodurus eques*, a flagship species with low but structured genetic variability.
1280 *J Hered*, *108*(2), 152-162. https://doi.org/10.1093/jhered/esw075

1281 Stiller, J., Wilson, N. G., & Rouse, G. W. (2015). A spectacular new species of
1282 seadragon (Syngnathidae). *R Soc Open Sci*, *2*(2), 140458.
1283 https://doi.org/10.1098/rsos.140458

1284 Stolting, K. N., & Wilson, A. B. (2007). Male pregnancy in seahorses and pipefish:
1285 beyond the mammalian model. *BioEssays*, *29*(9), 884-896.
1286 https://doi.org/10.1002/bies.20626

1287 Terranova, C., Narla, S. T., Lee, Y. W., Bard, J., Parikh, A., Stachowiak, E. K.,
1288 Tzanakakis, E. S., Buck, M. J., Birkaya, B., & Stachowiak, M. K. (2015). Global
1289 developmental gene programing involves a nuclear form of fibroblast growth
1290 factor receptor-1 (FGFR1). *PLoS ONE*, *10*(4), e0123380.
1291 https://doi.org/10.1371/journal.pone.0123380

Vitturi, R., & Catalano, E. (1988). Karyotypes in two species of the genus *Hippocampus* (Pisces: Syngnatiformes). *Marine Biology*, *99*(1), 119-121. https://doi.org/10.1007/bf00644985

Vitturi, R., Libertini, A., Campolmi, M., Calderazzo, F., & Mazzola, A. (1998). Conventional karyotype, nucleolar organizer regions and genome size in five Mediterranean species of Syngnathidae (Pisces, Syngnathiformes). *Journal of Fish Biology*, *52*(4), 677-687. https://doi.org/10.1111/j.1095-8649.1998.tb00812.x

Vurture, G. W., Sedlazeck, F. J., Nattestad, M., Underwood, C. J., Fang, H., Gurtowski, J., & Schatz, M. C. (2017). GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics*, *33*(14), 2202-2204. https://doi.org/10.1093/bioinformatics/btx153

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

Walshe, J., Maroon, H., McGonnell, I. M., Dickson, C., & Mason, I. (2002). Establishment of hindbrain segmental identity requires signaling by FGF3 and FGF8. *Curr Biol*, *12*(13), 1117-1123. https://doi.org/10.1016/s0960-9822(02)00899-0

Wang, X. P., Suomalainen, M., Felszeghy, S., Zelarayan, L. C., Alonso, M. T., Plikus, M. V., Maas, R. L., Chuong, C. M., Schimmang, T., & Thesleff, I. (2007). An integrated gene regulatory network controls stem cell proliferation in teeth. *PLoS Biol*, *5*(6), e159. https://doi.org/10.1371/journal.pbio.0050159

Waterhouse, R. M., Seppey, M., Simao, F. A., Manni, M., Ioannidis, P., Klioutchnikov, G., Kriventseva, E. V., & Zdobnov, E. M. (2018). BUSCO Applications from quality assessments to gene prediction and phylogenomics. *Molecular Biology and Evolution*, *35*(3), 543-548. https://doi.org/10.1093/molbev/msx319

Waters, M. G., Serafini, T., & Rothman, J. E. (1991). 'Coatomer': a cytosolic protein complex containing subunits of non-clathrin-coated Golgi transport vesicles. *Nature*, *349*(6306), 248-251. https://doi.org/10.1038/349248a0

Weisenfeld, N. I., Kumar, V., Shah, P., Church, D. M., & Jaffe, D. B. (2017). Direct determination of diploid genome sequences. *Genome Res*, *27*(5), 757-767. https://doi.org/10.1101/gr.214874.116

Whittaker, R. H. (1965). Dominance and diversity in land plant communities: numerical relations of species express the importance of competition in community function and evolution. *Science*, *147*(3655), 250-260. https://doi.org/10.1126/science.147.3655.250

Xie, Y., Su, N., Yang, J., Tan, Q., Huang, S., Jin, M., Ni, Z., Zhang, B., Zhang, D., Luo, F., Chen, H., Sun, X., Feng, J. Q., Qi, H., & Chen, L. (2020). FGF/FGFR signaling in health and disease. *Signal Transduct Target Ther*, *5*(1), 181. https://doi.org/10.1038/s41392-020-00222-7

1334 Yu, G., Wang, L. G., Han, Y., & He, Q. Y. (2012). clusterProfiler: an R package for
1335 comparing biological themes among gene clusters. *OMICS*, *16*(5), 284-287.
1336 https://doi.org/10.1089/omi.2011.0118

1337 Zhang, Y.-H., Ravi, V., Qin, G., Dai, H., Zhang, H.-X., Han, F.-M., Wang, X., Liu, Y.-H.,
1338 Yin, J.-P., Huang, L.-M., Venkatesh, B., & Lin, Q. (2020). Comparative genomics
1339 reveal shared genomic changes in syngnathid fishes and signatures of genetic
1340 convergence with placental mammals. *National Science Review*, *7*(6), 964-977.
1341 https://doi.org/10.1093/nsr/nwaa002

1342