Gene Expression and Evolution in the Smalltooth Sawfish, *Pristis pectinata*

Taiya M. Jarva[1], Nicole M. Phillips[1], Cory Von Eiff[1], Gregg R. Poulakis[2], Gavin Naylor[3], Kevin A. Feldheim[4], Alex S. Flynt[1]

[1]School of Biological, Environmental, and Earth Sciences. The University of Southern Mississippi, 118 College Drive, Hattiesburg, MS, 39401

[2]Charlotte Harbor Field Laboratory, Fish and Wildlife Research Institute, Florida Fish and Wildlife Conservation Commission, 585 Prineville St., Port Charlotte, FL 33954

[3] Florida Program for Shark Research, University of Florida, Gainesville, FL 32611

[4]Pritzker Laboratory for Molecular Systematics and Evolution, The Field Museum, 1400 S. Lake Shore Drive, Chicago, IL 60605

Summary

Sawfishes (Pristidae) are large, highly threatened rays named for their tooth-studded rostrum, which is used for prey sensing and capture. Of all five species, the smalltooth sawfish, *Pristis pectinata*, has experienced the greatest decline in range, currently found in only ~20% of its historic range. To better understand the genetic underpinnings of these taxonomically and morphologically unique animals, we collected transcriptomic data from several tissue types, mapped them to the recently completed reference genome and contrasted the patterns observed with comparable data from other elasmobranchs. Evidence of positive selection was detected in 79 genes in *P. pectinata*, several of which are involved in growth factor/receptor tyrosine kinase signaling and specification of organ symmetry, suggesting a role in morphogenesis. Data acquired also allow for examination of the molecular components of *P. pectinata* electrosensory

systems, which are highly developed in sawfishes and have likely been influential in their evolutionary success.

As meso- to apex-level predators, chondrichthyans (sharks, rays, and chimaeras) play essential roles in the health of marine ecosystems by increasing biodiversity, buffering against invasive species, decreasing transmission of diseases, and mitigating the effects of climate change[1,2] However, many chondrichthyan populations are in decline, primarily due to overfishing, with more than one third estimated to be threatened with extinction[3]. Sawfishes belong to one of the most threatened families, with all five species assessed as Endangered or Critically Endangered on the International Union for the Conservation of Nature (IUCN) Red List of Threatened Species[4]. Sawfishes are notable for their tooth-studded rostrum, which is used to detect, acquire, and manipulate prey[5,6]. This rostrum also makes them especially susceptible to entanglement in fishing gear, which has precipitated a global decline in their numbers over the last century.[4,7,8]

Bycatch in fisheries and habitat degradation continue to pose the greatest threats to sawfishes, including in ‘stronghold’ locations such as the United States and Australia (NMFS 2009)[9]. Fisher education and trade bans on sawfish and their parts have been used to mitigate sawfish declines[10,8]. While such strategies are essential, additional preventative initiatives are needed as sawfish populations continue to decline globally[11]. Recent research aimed at supporting development of deterrent technology found that sawfish will react to electric field stimuli, but behavioral responses were not consistent and were considered insufficient to avoid fisheries gear entanglement[12].

The electrosensing abilities of sawfishes are currently poorly understood, and little is known about the underlying the molecular mechanisms of electroreception and environmental biosensing that underpin their behaviors. The aims of this study were to collect transcriptome sequences from different tissues of the Critically Endangered smalltooth sawfish, *Pristis pectinata*, map the data to the recently completed high resolution genome assembly, and compare patterns of gene expression and sequence evolution with those derived from other elasmobranchs to identify unique components of genomic architecture. Of all sawfishes, *P. pectinata* has experienced the most severe decline in range and is present in less than 20% of its former range in the Atlantic Ocean[7,8]. Viable populations are currently restricted to Florida in the U.S. and western portions of The Bahamas and, like all sawfishes, reducing fisheries interactions is a top conservation priority[4]. The gene set collected in this study provides insight into evolutionary patterns and the mechanisms of electrosensing in *P. pectinata* and serve as critical first steps for future work.

High quality *Pristis pectinata* gene set

In collaboration with the Vertebrate Genome Project, GN produced a high-quality reference genome assembly for an adult female *P. pectinata* that had died at the Ripley's Aquarium in Myrtle Beach, South Carolina (GCA_009764475.1). The assembly was based on 60x PACBIO long read sequencing, BioNano, Hi-C and Illumina short read data (scaffold N50 101.7M; Contig N50=17M; 99.61% of the data assigned to 48 chromosomes. Genome size 2.27Gb; http://vertebrategenomesproject.org). However,

while the assembly had high contiguity, there were limited EST data associated with the original genome assembly[13]. As a result, approximately 40% of expected orthologs were absent from the predicted gene dataset (Fig. 1E)[14]. To address this issue and maximize utility of the *P*. *pectinata* genome, RNA was sequenced from tissues collected from a juvenile female (828 mm stretch total length) collected by GP under ESA Permit No. 21043. Datasets were generated from brain, kidney, liver, ovary, and skin tissues fixed in RNAlater.

Genome annotation facilitated by RNA-seq alignments yielded substantially more transcripts (159,014), than the 19,597 genes predicted from the genome alone. RNA-seq data were also assembled into a *de novo* transcriptome of 265,427 transcripts with 31.9% being >500 bp (Fig. 1A-B). Substantial improvement was seen in the contribution of the *de novo* transcriptome and RNA-seq guided annotation, with approximately 2000 genes exclusive to the *de novo* transcriptome. Less redundancy was seen in the dataset after intersecting *de novo* transcripts with predicted genes from the genome, with 9,233 transcripts shared between datasets (Fig. 1D). Combining the reference genome assembly predictions, RNA-seq assisted annotations, and a *de novo* transcriptome resulted in a greatly enhanced gene set representation (Fig. 1E). In the *de novo* transcriptome alone, less than 10% of expected orthologs were absent. In the combined gene set, the percentage of missing genes was reduced to only 6.6% (Sup. Tbl. 2).

To further validate the *P*. *pectinata* gene set, gene ontology (GO) terms were assigned and compared to the transcriptome of the little skate, *Leucoraja erinacea*[15] (Fig. 1C). Distributions of high-level GO terms were similar between *P. pectinata* and *L. erinacea,* suggesting gene content in the *de novo* transcriptome represents what is observed in related taxa. However, differences were noted, such as a higher percentage of genes involved in antioxidant activity and DNA binding in *P. pectinata* versus a higher percentage of genes in *L. erinacea* related to signaling, response to stimulus, metabolic process, and catalytic activity. This may reflect a difference in juvenile and adult tissues sampled for *P. pectinata* relative to the embryonic tissue-derived *L. erinacea* transcriptome. The near-complete *P. pectinata* gene set enables characterization of unique genetics in this species, which was not possible with predicted annotations offered by genome sequence approaches alone.

Unique genetic features of *Pristis pectinata*

To identify positively selected genes (PSGs) in *P. pectinata*, the combined dataset from *P. pectinata* and transcriptomes of four other fish species were assigned to orthologous gene groups with Orthofinder[16]. The species included were the Australian ghostshark, *Callorhinchus milii*, chain catshark, *Scyliorhinus retifer*, and *L. erinacea*, with the Indonesian coelacanth, *Latimeria menadoensis*, serving as an outgroup (Fig. 2A). *Pristis pectinata* had the second highest percentage of genes (~51%) assigned to an orthogroup and the highest percentage of species-specific genes, likely due to the substantially greater completeness of the assembly relative to the other species (Sup. Fig. 1). The resulting 3,116 genes were tested for branch-specific episodic selection

using aBSREL, revealing 79 PSGs in *P. pectinata*[17] (Supplementary Methods, Sup. Tbl. 3).

To cluster *P. pectinata* PSGs into groups which may share functional relationships or similar selection pressures, omega values from aBSREL, percent of sites under selection, gene expression rank relative to *S. retifer* orthologs, and the difference in Grand Average of Hydropathy (GRAVY) between *P. pectinata* and *L. erinacea* orthologs were collected for Principal Components Analysis (Sup. File 1). K-means clustering identified three groups containing 49, 25, and 5 genes (Fig. 2B). Generally, changes in GRAVY and omega values were highly correlated, with GRAVY values being the major contribution to variance (Sup. Fig. 2A). Between tissues, the highest correlation was seen between skin, kidney, and brain, while liver and ovary showed little correlation. Gene ontology enrichment analysis by TopGO of PSGs in Cluster 1 versus all genes analyzed showed enrichment in functions such as multicellular organism development, animal organ morphogenesis, and anatomical structure morphogenesis[18] (Fig. 2C). In comparison, Cluster 2 genes were related to regulation of response to biotic stimulus, regulation of defense response, and proteasomal protein catabolic process (Sup. Fig. 2B). Interestingly, Cluster 1 contained multiple genes implicated in developmental processes such as fibroblast growth factor/receptor tyrosine kinase (FGF/RTK) and mitogen activated protein kinase (MAPK) signaling, suggesting that *P. pectinata* PSGs share functional or evolutionary similarities. The altered FGF and MAPK signaling that may result from these diverged genes implies possible relevance to *P. pectinata*-specific morphogenesis.

Comparing biochemical properties such as hydrophobicity and polarity of developmental genes with the closest available taxon, *L. erinacea*, revealed functional changes in *P. pectinata* in three notable genes: *Crk-l*, an integrator of multiple signaling pathways; Annexin A1 (*Anxa1*), a modulator of FGF ligands and RAS, and FGF1 intracellular binding protein (*Fibp*), an EGF/MAPK modulator (Fig. 2D-H)[19,20]. CRK-L had decreased hydrophobicity in the Src Homology 2 (SH2) domain where tyrosine phospho-proteins like growth factor receptors bind and decreased hydrophobicity in its SH3 domain that associates with RAC1 or RAS[21] (Fig. 2D). ANXA1 had changes in all four Annexin repeat domains, which upon Ca^{2+} binding display phosopho-sites (Fig. 2E). Annexins activate MAPK signaling either through growth factors, or they can be directly phosphorylated by RTKs[22]. FIBP, which binds FGF1, showed fluctuations in both hydrophobicity and polarity in the annotated FIBP domain, though little is known about the mechanisms of action of this protein[23] (Fig. 2F). Selection in genes such as these, which are potentially involved in body patterning and growth factor signaling suggest a role related to rostral development. A prime example is *Crk-l*, which has been shown to cause craniofacial defects in mice through disruption of neural crest development or through the retinoic acid/*Tbx1* regulatory network[24–26].

Another major function of Cluster 1 genes is microtubule biology with two genes being significant (*Haus2* and *Tubgcp4*). Shared evolutionary trajectory with developmental regulators suggests a similar functionality in sawfish biology, such as through cilia-mediated signaling. One example is *Ccdc103*, which is necessary for ciliogenesis.

CCDC103 had numerous changes in polarity and hydrophobicity in its Rpap3 domain, which binds the axonemes of cilia (Figs. 2G). This is necessary for outer dynein arm attachment, and thus changes in this domain suggest functional differences which do not appear to be a result of gene duplication and subsequent divergence[27]. Lastly, a homeobox transcription factor involved in segmentation and body patterning, *HoxA5*, was found. The protein sequence had changes in polarity and hydrophobicity near the conserved site (residues 183–188) and the beginning of the DNA-binding homeobox domain, despite being truncated relative to *L. erinacea*. These changes suggest modified interactions between HOXA5 and its cofactors and/or targets (Fig. 2H).

To support whether changes in PSGs were specific to *P*. *pectinata* or shared with other sawfishes, sequencing of genomic DNA in functional domains was performed using samples from three other *Pristis* sawfishes: dwarf sawfish (*P. clavata*), green sawfish (*P. zijsron*), and largetooth sawfish (*P. pristis*). *Ccdc103* and *HoxA5* were chosen as they demonstrated clear chemical differences between *P. pectinata* and *L. erinacea* but were conserved enough in flanking regions of functional domains for amplification. Sequence alignments revealed that the ratios of non-synonymous to synonymous nucleotide substitutions in both genes were lower among sawfishes than when comparing sawfishes to *L. erinacea*. At *HoxA5* conserved sites, all five substitutions were shared among sawfishes but were non-synonymous with *L. erinacea,* with amino acid property changes at two sites (Fig. 2L, Sup. Fig. 3). A similar pattern was seen in *Ccdc103* among sawfishes. In the Parp3 domain of *Ccdc103*, 18/21 changes were non-synonymous between sawfishes compared to *L. erinacea*, and eight had an amino

184 property change (Fig. 2J, Sup. Fig. 4). Fewer changes were seen among sawfishes;
185 10/13 were non-synonymous, and seven of these included chemical changes. Given the
186 functions of PSGs and that the ratios of non-synonymous substitutions were lower
187 among *Pristis* sawfishes compared to *L. erinacea*, this supports that these changes
188 could be related to themes found in Cluster 1 genes. As conservation of the biochemical
189 changes occurred at the base of the *Pristis* branch and conserved in all species this
190 leads to further credence for a role in sawfish-specific morphological development.
191
192 **Characterization of electrosensory genes**
193 In sawfishes, the rostrum provides an exaggerated platform for placement of
194 electrosensing organs called ampullae of Lorenzini. Initial sensing occurs through
195 voltage-gated calcium channels (VGCC) produced from *Cacnα1D* in both rays and
196 sharks[28]. Following influx of Ca2+ ions, K+ efflux occurs leading to membrane
197 depolarization and neurotransmission. Different K+ channels participate in rays (BK-
198 channels, BK-α) and sharks (Shaker-type, Kv). Using the *P. pectinata* transcriptome
199 collection described above, orthologs of all channels involved were identified to
200 understand mediators of electrosensing in this species.
201
202 First, all VGCC subunit transcriptomes were assessed for expression in rostral skin (Fig.
203 3A). Six VGCC-related genes were expressed, including a *Cacnα1D* ortholog. Others
204 were channel accessory subunits (*Cacnα2δ, Cacnb2,* and *Cacnb4*), or a *Cacnα1C*-type
205 VGCC, which have not been previously implicated in electrosensing chondrichthyans. A
206 *Cacnb2* ortholog was found in the transcriptome but was not expressed in rostral skin.

Thus, as expected, the *Cacnα1D* ortholog is likely the primary mediator of electrosensing. However, *P. pectinata Cacnα1D* is missing a 91-residue segment that would interact with intracellular effectors which is present in other rays (Fig. 3B)[28]. A portion of this region is also deleted in other chondrichthyans such as *C. mili* and the whale shark, *Rhincodon typus*, indicating that it is a site of functional novelty. Together, these results suggest that substantial changes occur during the initial step in the ability of *P. pectinata* to detect the bioelectric signal of prey.

Next, sequences of eight orthologs of Ca^{2+} responsive K^{+} channel were identified and compared (Fig. 3C). Clustering showed both major groups, BK (α and b) and Shaker types, were present in the gene set. A BK-α ortholog with clear pore motif was found (*Kcnmα1*) along with three accessory BK-β subunits, one of which was designated as novel as it could not be definitively paired with an ortholog. The four Shaker-type channels also exhibited the requisite pore domain residues. Two of these channels have been observed in *L. erinacea*, but they have only been demonstrated to work in conjunction with *Cacnα1D* orthologs in ampullae of the chain catshark, *S. retifer*, though they have also been found to be enriched in ampullae of the American paddlefish, *Polyodon spathula* [29,30] [31].

All eight channel components are expressed in rostral skin, except for the *P. pectinata* β-4 ortholog (Kcnmb4) (Fig. 3D). Absence of this inhibitory subunit may contribute to the heightened electrosensing seen in sawfishes. Further differences were seen in expression levels of β-2 subunits. In *L. erinacea*, expression of β- subunits in

electrosensory cells is 1000- to 10000-fold lower than α-subunit expression, yet in *P. pectinata* tissue, a BK β-2 is expressed nearly 10-fold to BK-α, and the novel subunit nearly 2-fold[29]. Together, this suggests *P. pectinata* has altered BK-channel physiology. Another unusual observation was expression of all four shaker-type channels, which differs from *L. erinacea* where only two, Kv1.1 and Kv1.5, are found[29].

Other aspects of electrosensing machinery in *P. pectinata* showed potentially divergent biology. Several inserts and deletions were found in *Kcnmα1* that may alter its function (Sup. Fig. 5). The ion pore-containing BK α-subunits can differentially associate with β-subunits to modulate membrane repolarization to alter activation rates or affect calcium sensitivity, drawing into question the role of the novel β-subunit in the overall performance of the system[32–34]. Another difference found in *Kcnmα1* was the inclusion of exon 21 (known as exon 29 in *L. erinacea*) (Fig. 3E). Alternative splicing at exon 21 has been identified in BK channel ampullary transcripts of *L. erinacea* and in auditory hair cells of chick cochlea, though the effect on electrosensing is not known [35]. The inclusion of this exon appears to be variable among elasmobranchs, being present in *P. pectinata*, thorny skate, *Amblyraja radiata*, and *L. erinacea* adult ampullae, but absent in *S. retifer,* whitespotted bamboo shark, *Chiloscyllium plagiosum,* and *L. erinacea* embryonic sequence (Sup. Fig. 5C).

Implications of genetic changes in sawfishes

Here, fundamental changes in developmental genes were found in *P. pectinata*, which may influence morphogenesis of the rostrum. Elongated rostral structures have evolved

in chondrichthyans at least five times[36]. Sawfishes (Batoidea: Pristidae) and sawsharks (Selachii: Pristiophoridae) are the only extant families that possess toothed rostra, but similar structures are also found in the fossil record, including two species of chimaeras, Squaloraja *polyspondyla* and *Acanthorhina jaekel* (Holocephali), Sclerorhynchoidei (Rajiformes), and *Bandringa rayi*. (Elasmobranchii)[36–38]. Convergent evolution of elongated rostral structures suggests an evolutionary advantage in prey detection, likely through heightened bioelectric sensing, which may provide sensory information at night, at depth, or in murky shallow waters[39]. The cohort of genes uniquely under selection in *P. pectinata* may provide a framework for the genetic changes that underpin the amorphogenetic origins of saw-like rostral structures. Primarily, this appears to be through reshaping the signaling environment, as was observed in this study, with changes in FGF/MAPK mediators. Simultaneously, positive selection was also seen in *HoxA5*, a transcription factor that establishes regions along the dorsoventral axis. However, in vertebrate development, cluster 5 Hox genes are typically expressed in the somites which eventually become the upper thorax and organs, and thus changes in this gene in *P. pectinata* may be related to differences in pectoral or synarcual morphology, and not related to the rostrum.

A consequence of potentially altered FGF/MAPK signaling in *P. pectinata* is divergent cellular response to environmental stressors such as changes in salinity, temperature, and dissolved oxygen. In Florida, juveniles have an affinity for warm (>30°C), brackish waters (18–30) with high dissolved oxygen levels (>6 mg/L).[40] Cold water temperatures (<8–12°C) are known to alter habitat use and cause mortality[40,41], while the

276 physiological impacts of hypoxia and low salinities are unknown. As *P. pectinata* begins
277 to recover, populations should re-establish in historical habitats. While there is evidence
278 that such re-expansions have begun in the northern Gulf of Mexico and in the southern
279 Indian River Lagoon, Florida[42] (G. R. Poulakis unpublished data), water quality issues
280 and fluctuating environmental conditions, caused in part by freshwater diversions that
281 modify the hydrology of estuaries and cause algal blooms and hypoxic conditions, may
282 pose a risk to sawfish health and viability[42].
283
284 The genes involved in this environmental signaling can be affected by marine pollutants
285 including polycyclic aromatic hydrocarbons (PAHs), heavy metals, bisphenol-A's
286 (BPAs), brominated flame retardants (BFRs), and polychlorinated benzodioxins (PCBs).
287 For example, *Crk-l* expression is affected by BPAs, BFRs, and PCBs, and *Anxa1* is
288 upregulated in response to PAH exposure in sea turtles as a response to increased
289 production of reactive oxygen species[43]. Cadmium, PCBs, and other persistent organic
290 pollutants can bioaccumulate in elasmobranch tissues, and may disrupt critical
291 processes such as metabolism, immune function, and reproduction, though pollutant
292 levels in sawfish tissues have not been reported[44–46]. Future research should identify
293 contaminant levels in sawfish tissues and document sources of these pollutants in
294 habitats where *P. pectinata* reside and may be re-establishing. Characterizing direct
295 physiological effects of pollutants in sawfishes via *ex situ* experiments is not feasible;
296 however, a potential circumvention of this issue could be to identify biomarkers in *P.*
297 *pectinata* which are correlated with toxicological risks[47].
298

A top conservation priority for *P. pectinata*, and all sawfishes, is to reduce injuries and mortalities in fisheries, especially trawl fisheries. Understanding the basis of sawfish electrosensing at a molecular level may support the development of more effective deterrent technologies that exploit this sensing modality. In future studies, molecular-level responses to stimuli, such as electric fields, should be studied to refine optimal experimental conditions, and used in parallel with aquarium trials to elicit avoidance responses[12]. Analysis of electrosensory genes revealed numerous differences in functional domains and expression of channels and subunits which were previously not implicated in electrosensing in chondrichthyans. As basal batoids, sawfish may possess unique electrosensing biology that leverages both types of K+ channels found in sharks and other rays. Considering ray-type electrosensing alone, elevated expression of multiple previously uncharacterized subunits suggests *P. pectinata* may have substantially altered BK-channel physiology. Altogether, these results suggest unique physiology of ion channels and subunits in *P. pectinata,* and their identification could allow tailored bycatch-reduction technology. Developing this technology for trawls is a priority as shrimp trawls have been identified as still having the highest bycatch risk for sawfishes in the 'stronghold' nations of the U.S. and Australia[9].

In addition to providing an annotation of the publicly available genome, this study highlights the potential value of genomic approaches to conservation efforts, particularly through the identification and characterization of electrosensory genes and subunits and genes that could be used to reduce bycatch. Global sawfish recovery requires aggressive conservation planning that uses novel methodological approaches to

support modern, high-tech solutions. Results from this work suggest that the impacts of pollutants need to be more deeply investigated and that physiological responses may differ between sawfishes and existing model organisms. These data will also support the ability to build experimental systems to test channel behavior in a controlled *in vitro* setting that can be used to assess and quantify the performance of sawfish electrosensing to facilitate development of behavior-modifying technology. Together, these data and insights provide the foundation to support key future research, with the goal of supporting global recovery of imperiled sawfishes.

Supplementary methods

Sample collection

RNAs were collected from the brain, liver, kidneys, ovary, and skin tissues of a juvenile female *P. pectinata*. Tissues were preserved in RNAlater and stored at -80°C until use. Samples from each tissue were homogenized and extracted with the TRIzol method. Concentrations and purities of RNA extracts were measured by nanodrop and bioanalyzer 2100 (Sup. Tbl. 1).

Transcriptome sequencing and assembly

The transcriptome was assembled using Trinity on the Magnolia High Performance Computing (HPC) cluster[48]. Blast2GO was used to assign gene ontology terms to transcripts, and WeGO was used for comparison between *P. pectinata* and *L. erinacea*[49,50]. Completeness was assessed with BUSCO in transcriptome mode against the most recent vertebrate lineage, odb10. To annotate the genome with coding

sequences and to capture a more complete gene set for positive selection analysis, the published *P. pectinata* genome from an adult female (GCA_009764475.2) was used in addition to the assembled transcriptome.

Genome annotation for positive selection analysis geneset

Gene content exclusive to the transcriptome was isolated using STAR aligner with default parameters to align reads from all tissue samples to the genome[51]. Coverage over the genome was assessed with BEDtools and reads with low coverage (< 5 reads) were then aligned to the assembled transcriptome using Hisat2 with the –dta-cufflinks option enabled[52,53] Cufflinks was used to map RNA reads to coordinates in the genome, which were then intersected, excluding overlapping transcripts, with Augustus-predicted gene sequence coordinates for the genome using BEDTools Intersect[54]. All BAM and SAM file conversion and sorting was performed with SAMtools[55]. Unique gene sequences from the transcriptome and genome were concatenated into one dataset for further analysis and BUSCO annotation output was used to remove redundant sequences and assess completeness. To obtain transcript expression, each tissue library was individually mapped to the constructed transcriptome using Bowtie2, and SAMtools Idxstats was used to quantify mapped reads[56].

Positive selection in transcriptome

Open reading frames were predicted from all input transcriptomes using TransDecoder with default parameters[57]. All coding sequences were concatenated into one dataset and Orthofinder was executed with default parameters to cluster orthologous gene

groups[16] Sequences in each gene family were annotated using eggNOG with Diamond mode enabled[58]. Annotations were also used to split gene families into paralogous groups using custom scripts. For positive selection, only groups that retained at least one sequence from each taxon were retained for analysis. Transcripts were discarded if their length was more than 100 amino acids shorter or longer than the average length of the gene, and only genes with Pal2Nal alignments longer than 20 amino acids were kept, excluding trees with insufficient branch lengths for analysis. ABSREL, which identifies episodic selection in individual branches, was used to analyze each group of orthologous genes[17]. Mafft v7.475 was used for protein alignment, FastTree 2.1.10 for tree construction, and Pal2Nal v14 with the –nogap option to provide gap-free codon-based nucleotide alignments[59–61]. Output JSON files were parsed with custom scripts using the JsonLite package[62]. To examine potential effects of substitutions in genes of interest under selection, chemical properties were compared between protein sequences from *P. pectinata* and *L. erinacea* using Expasy ProtScale[63]. The Kyte & Doolittle (1982) scale was used for hydrophobicity and the Zimmerman scale for polarity. After manual gap correction, scale values at each residue for *P. pectinata* were subtracted from *L. erinacea* values and the change plotted by residue. Domains of each protein were obtained from Interpro Scan. Analysis of PSGs of interest using DAMBE found no significant substitution saturation in any species from any alignment, indicating that the species are not too diverged to obtain meaningful positive selection results[64].

For PCA analysis, grand average values of hydropathicity (GRAVY) for each *P. pectinata* and *L. erinacea* protein were obtained from Expasy ProtParam. If there were

two sequences in an orthogroup for *L. erinacea*, the closest aligning sequence was selected for each *P*. *pectinata* gene from a multiple sequence alignment, and the difference was taken between the values. Expression of homologous genes under selection in *P. pectinata* was compared to similar tissues from *S. retifer* by aligning publicly available single-end RNAseq libraries from brain, liver, kidney, ovary, and skin to the longest *S. retifer* homolog sequence (Sup. Tbl. 4). Alignment was performed with Hisat2 and quantified with SAMtools Idxstats[52,55]. Alignment statistics for each tissue type can be found in Sup. Table 5. Expression values for each gene were normalized to TPM and subtracted from *P. pectinata* values. Principal component analysis was performed using the factoextra R package and visualized with ggpubr[65] (Supplementary File 1).

Conservation of changes in sawfishes

To determine whether signals of positive selection in genes of interest for *P. pectinata* were also present in other sawfish species, nucleotide changes in the three other *Pristis* sawfishes were examined for HoxA5 and Ccdc103. Primers were designed to amplify divergent homologous functional domains between *P. pectinata* and *L. erinacea*. Primer sequences used were: *HoxA5* forward: 5'GACTTATGTGCAGTTTTCGCATCCA 3'*;* *HoxA5* reverse: 3'AACTACCTCCTCAAATTC 5'; *Ccdc103* forward: 5'CTGCTGCTCAGGAAATCCAC 3'; *Ccdc103* reverse: 3'AGCGGAGTTTAGCCGTGACTG 5'.

DNA samples from *P. clavata*, *P. zijsron*, and *P. pristis* were mixed with Phire Hot Start II DNA polymerase (ThermoFisher), dNTPs, deionized water, and primers for either *Ccdc103* or *HoxA5*, followed by PCR amplification using a Mastercycler Pro and electrophoresis apparatus. The reactions were amplified for 35 cycles at 98°C for 30s/5s, 53°C/58.5°C for 15s, and 72°C/72°C for 1min/1 min for *HoxA5* and *Ccdc103* respectively. The agarose gel bands were purified with a GeneJET Plasmid MiniPrep (ThermoFisher) and sent to Eurofins Genomics for sequencing. ApE (RRID:SCR_014266) was used to translate DNA sequences into the appropriate reading frame, and Clustal Omega was used to align sequences to *P. pectinata* and *L. erinacea*[66]. BLAT from the UCSC genome browser and the *P. pectinata* genome were used to verify amplification of the correct target sequence. Alignments were manually examined between species to identify conserved non-synonymous changes among sawfishes that were not found in *L. erinacea*.

Electrosensory genes

Full coding sequences for electrosensing genes were downloaded from NCBI and used for manual positive selection analysis. Analysis included the VGCC (*Cacnα1D*) and several β-subunits, potassium-activated (BK) channel α subunit, and several Shaker (Kv) channels which have been implicated as major ion channels involved in electroreception in sharks and rays[28] Transcripts for *P. pectinata* were identified with BLASTP using *L. erinacea* sequences from NCBI (acc. AJP74816.1, KY355736.1) as query and a protein database was constructed from the transcriptome by Hmmer2GO[67,68]. BK*β* and Shaker-type channels were identified using BLAST and

confirmed using Interpro Scan[69]. Multiple sequence alignments were performed using ClustalW with default parameters and 1000 bootstraps and visualized using the GGMsa package in R[70]. The phylogenetic tree was visualized with the GGTree package[71]. GenBank accessions for the elasmobranch sequences most similar to the uncharacterized BK*β* subunit were: XP_041.47886.1 , XP_032892500.1, GCB66272.1, XP_038668782.1, and XP_043564107.1 (accessed 9/26/2022).

Data availability statement

Raw sequence data used to assemble the transcriptome can be found under NCBI accession PRJNA864825.

Author contributions

NP, KF, and AF administered the project. GN contributed the high coverage reference genome assembly. AF, TJ, and NP designed and executed experiments. TJ and AF analyzed and interpreted data. CVE performed PCRs for *Ccdc103* and *HoxA5*. GP acquired and maintained the endangered species collection permit and collected the tissue samples from *Pristis pectinata*. TJ, AF, NP, GP, and GN wrote the manuscript and/or contributed to final edits.

Acknowledgments

Funding for this project was provided by NOAA Awards NA16NMF4720062 (field work) and NA18NMF0080237 (laboratory processing), start-up funds from the University of Florida to GN for the long read sequencing and assembly of the reference genome. The

authors acknowledge HPC at The University of Southern Mississippi supported by the National Science Foundation under the Major Research Instrumentation (MRI) program via Grant #ACI 1626217. Thanks to David Morgan, Jeff Whitty, and the Western Australia Department of Fisheries for providing tissue samples of *P. clavata*, *P. pristis, and P. zijsron*.

References

1. Ferretti, F., Worm, B., Britten, G. L., Heithaus, M. R. & Lotze, H. K. Patterns and ecosystem consequences of shark declines in the ocean. *Ecol Lett* **13**, 1055–1071 (2010).
2. Ritchie, E. G. *et al.* Ecosystem restoration with teeth: what role for predators? *Trends Ecol Evol* **27**, 265–271 (2012).
3. Dulvy, N. K. *et al.* Overfishing drives over one-third of all sharks and rays toward a global extinction crisis. *Current Biology* **31**, 4773-4787.e8 (2021).
4. Carlson, J. K., Wiley, T. & Smith, K. *The IUCN Red List of Threatened Species*. http://www.iucnredlist.org/details/18175/0. (2013).
5. Wueringer, B. E. Electroreception in Elasmobranchs: Sawfish as a Case Study. *Brain Behav Evol* **80**, 97–107 (2012).
6. Poulakis, G. *et al.* Sympatric elasmobranchs and fecal samples provide insight into the trophic ecology of the smalltooth sawfish. *Endanger Species Res* **32**, 491–506 (2017).
7. Dulvy, N. K. *et al.* Ghosts of the coast: global extinction risk and conservation of sawfishes. *Aquat Conserv* **26**, 134–153 (2016).
8. Brame, A. *et al.* Biology, ecology, and status of the smalltooth sawfish Pristis pectinata in the USA. *Endanger Species Res* **39**, 9–23 (2019).
9. Graham, J. *et al.* Commercial fishery bycatch risk for large juvenile and adult smalltooth sawfish (<scp> *Pristis pectinata* </scp>) in Florida waters. *Aquat Conserv* **32**, 401–416 (2022).
10. Wiley, T. & Brame, A. *5-Year Review: Summary and Evaluation of United States Distinct Population Segment of Smalltooth Sawfish.* https://repository.library.noaa.gov/view/noaa/19253 (2018).
11. Yan, H. F. *et al.* Overfishing and habitat loss drive range contraction of iconic marine fishes to near extinction. *Sci Adv* **7**, (2021).
12. Abrantes, K. *et al.* Potential of electric fields to reduce bycatch of highly threatened sawfishes. *Endanger Species Res* **46**, 121–135 (2021).
13. Salzberg, S. L. Next-generation genome annotation: we still struggle to get it right. *Genome Biol* **20**, 92 (2019).
14. Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. v. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015).

15. King, B. L., Gillis, J. A., Carlisle, H. R. & Dahn, R. D. A Natural Deletion of the *HoxC* Cluster in Elasmobranch Fishes. *Science (1979)* **334**, 1517–1517 (2011).
16. Emms, D. M. & Kelly, S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol* **16**, 157 (2015).
17. Smith, M. D. *et al.* Less Is More: An Adaptive Branch-Site Random Effects Model for Efficient Detection of Episodic Diversifying Selection. *Mol Biol Evol* **32**, 1342–1353 (2015).
18. Alexa A & Rahnenfuhrer J. topGO: Enrichment Analysis for Gene Ontology. Preprint at (2022).
19. Katoh, M. & Katoh, M. Cross-talk of WNT and FGF signaling pathways at GSK3ß to regulate ß-catenin and SNAIL signaling cascades. *Cancer Biol Ther* **5**, 1059–1064 (2006).
20. Balasubramanian, R. & Zhang, X. Mechanisms of FGF gradient formation during embryogenesis. *Semin Cell Dev Biol* **53**, 94–100 (2016).
21. Antoku, S. & Mayer, B. J. Distinct roles for Crk adaptor isoforms in actin reorganization induced by extracellular signals. *J Cell Sci* **122**, 4228–4238 (2009).
22. Babbin, B. A. *et al.* Annexin I Regulates SKCO-15 Cell Invasion by Signaling through Formyl Peptide Receptors. *Journal of Biological Chemistry* **281**, 19588–19599 (2006).
23. Thauvin-Robinet, C. *et al.* Homozygous FIBP nonsense variant responsible of syndromic overgrowth, with overgrowth, macrocephaly, retinal coloboma and learning disabilities. *Clin Genet* **89**, e1–e4 (2016).
24. Newbern, J. *et al.* Mouse and human phenotypes indicate a critical conserved role for ERK2 signaling in neural crest development. *Proceedings of the National Academy of Sciences* **105**, 17115–17120 (2008).
25. Guris, D. L., Duester, G., Papaioannou, V. E. & Imamoto, A. Dose-Dependent Interaction of Tbx1 and Crkl and Locally Aberrant RA Signaling in a Model of del22q11 Syndrome. *Dev Cell* **10**, 81–92 (2006).
26. Yutzey, K. E. DiGeorge Syndrome, Tbx1, and Retinoic Acid Signaling Come Full Circle. *Circ Res* **106**, 630–632 (2010).
27. King, S. M. & Patel-King, R. S. The Oligomeric Outer Dynein Arm Assembly Factor CCDC103 Is Tightly Integrated within the Ciliary Axoneme and Exhibits Periodic Binding to Microtubules. *Journal of Biological Chemistry* **290**, 7388–7401 (2015).
28. Bellono, Nicholas. Leitch, Duncan. Julius, David. Molecular Basis of Ancestral Vertebrate Electroreception. *Nature* **543**, 391–396 (2017).
29. Clusin, William T., Wu, Ting-Hsuan, Shi, Ling-Fang. Kao, P. N. Further Studies of Ion Channels in the Electroreceptor of the Skate Through Deep Sequencing, Cloning and Cross Species Comparisons. *Gene* **718**, (2019).
30. Bellono, Nicholas. Leitch, Duncan. Julius, David. Molecular tuning of electroreception in sharks and skates. *Nature* **558**, 122–126 (2018).
31. Modrell, Melinda S., Lyne, Mike. Carr, Adrian R. Zakon, Harold H. Buckley, David. Campbell, Alexander S. Davis, Marcus C. Micklem, Gos. Baker, C. VH. Insights Into Electrosensory Organ Development, Physiology and Evolution From A Lateral Line-Enriched Transcriptome. *Elife* **6**, (2017).
32. Gonzalez-Perez, Vivian. Lingle, C. J. Regulation of BK Channels by Beta and Gamma Subunits. *Annu Rev Physiol* **81**, 113–137 (2019).

33. Li, Q. & Yan, J. Modulation of BK Channel Function by Auxiliary Beta and Gamma Subunits. in 51–90 (2016). doi:10.1016/bs.irn.2016.03.015.
34. Brenner, Robert. Jegla, Tim J. Wickenden, Alan. Liu, Yi. Aldritch, R. W. Cloning and Functional Characterization of Novel Large Conductance Calcium-activated Potassium Channel β Subunits, hKCNMB3 and hKCNMB4. *Journal of Biological Chemistry* **275**, 6453–6461 (2000).
35. King, Benjamin., Fang Shi, Ling., Kao, Peter. & Clusin, William. Calcium Activated K+ Channels in The Electroreceptor of the Skate Confirmed by Cloning. Details of Subunits and Splicing. *Gene* **578**, 63–73 (2015).
36. Welten, M., Smith, M. M., Underwood, C. & Johanson, Z. Evolutionary origins and development of saw-teeth on the sawfish and sawshark rostrum (Elasmobranchii; Chondrichthyes). *R Soc Open Sci* **2**, 150189 (2015).
37. Sallan, L. C. & Coates, M. I. The long-rostrumed elasmobranch *Bandringa* Zangerl, 1969, and taphonomy within a Carboniferous shark nursery. *J Vertebr Paleontol* **34**, 22–33 (2014).
38. GREENFIELD, T. Corrections to the nomenclature of sawskates (Rajiformes, Sclerorhynchoidei). *Bionomina* **22**, (2021).
39. Wueringer BE, Jnr LS, Kajiura SM, Tibbetts IR, Hart NS, C. S. Electric Field Detection in Sawfish and Shovelnose Rays. *PLoS One* **7**, (2012).
40. Poulakis, G. R., Stevens, P. W., Timmers, A. A., Wiley, T. R. & Simpfendorfer, C. A. Abiotic affinities and spatiotemporal distribution of the endangered smalltooth sawfish, Pristis pectinata, in a south-western Florida nursery. *Mar Freshw Res* **62**, 1165 (2011).
41. Feldheim, K., Fields, A., Chapman, D., Scharer, R. & Poulakis, G. Insights into reproduction and behavior of the smalltooth sawfish Pristis pectinata. *Endanger Species Res* **34**, 463–471 (2017).
42. Lehman, R. N. *et al.* Environmental DNA evidence of the Critically Endangered smalltooth sawfish, Pristis pectinata, in historically occupied US waters. *Aquat Conserv* **32**, 42–54 (2022).
43. Cocci, P., Mosconi, G. & Palermo, F. A. Gene expression profiles of putative biomarkers in juvenile loggerhead sea turtles (Caretta caretta) exposed to polycyclic aromatic hydrocarbons. *Environmental Pollution* **246**, 99–106 (2019).
44. Hamers, T. *et al.* In Vitro Profiling of the Endocrine-Disrupting Potency of Brominated Flame Retardants. *Toxicological Sciences* **92**, 157–173 (2006).
45. Martins, M. F., Costa, P. G. & Bianchini, A. Maternal transfer of polycyclic aromatic hydrocarbons in an endangered elasmobranch, the Brazilian guitarfish. *Chemosphere* **263**, 128275 (2021).
46. Tiktak, G. P. *et al.* Are concentrations of pollutants in sharks, rays and skates (Elasmobranchii) a cause for concern? A systematic review. *Mar Pollut Bull* **160**, 111701 (2020).
47. Cullen, J. A., Marshall, C. D. & Hala, D. Integration of multi-tissue PAH and PCB burdens with biomarker activity in three coastal shark species from the northwestern Gulf of Mexico. *Science of The Total Environment* **650**, 1158–1172 (2019).
48. Grabherr, M. G. *et al.* Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* **29**, 644–652 (2011).
49. Ye, J. *et al.* WEGO 2.0: a web tool for analyzing and plotting GO annotations, 2018 update. *Nucleic Acids Res* **46**, W71–W75 (2018).

50. Gotz, S. *et al.* High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res* **36**, 3420–3435 (2008).
51. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
52. Kim, D., Langmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods* **12**, 357–360 (2015).
53. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
54. Trapnell, C. *et al.* Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* **28**, 511–515 (2010).
55. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
56. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357–359 (2012).
57. Hass, B. J. TransDecoder. Preprint at https://github.com/TransDecoder/TransDecoder.
58. Huerta-Cepas, J. *et al.* Fast Genome-Wide Functional Annotation through Orthology Assignment by eggNOG-Mapper. *Mol Biol Evol* **34**, 2115–2122 (2017).
59. Katoh, K. & Standley, D. M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol Biol Evol* **30**, 772–780 (2013).
60. Suyama, M., Torrents, D. & Bork, P. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res* **34**, W609–W612 (2006).
61. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree: Computing Large Minimum Evolution Trees with Profiles instead of a Distance Matrix. *Mol Biol Evol* **26**, 1641–1650 (2009).
62. Ooms, J. The jsonlite Package: A Practical and Consistent Mapping Between JSON Data and R Objects. (2014).
63. Gasteiger, E. *et al.* Protein Identification and Analysis Tools on the ExPASy Server. in *The Proteomics Protocols Handbook* 571–607 (Humana Press, 2005). doi:10.1385/1-59259-890-0:571.
64. Xia, X. DAMBE7: New and Improved Tools for Data Analysis in Molecular Biology and Evolution. *Mol Biol Evol* **35**, 1550–1552 (2018).
65. Kassambra, A. ggpubr: 'ggplot2' Based Publication Ready Plots. Preprint at (2020).
66. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* **7**, 539 (2011).
67. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J Mol Biol* **215**, 403–410 (1990).
68. Staton, E. Hmmer2GO. Preprint at (2014).
69. Jones, P. *et al.* InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**, 1236–1240 (2014).
70. Zhou, L. *et al.* ggmsa: a visual exploration tool for multiple sequence alignment and associated data. *Brief Bioinform* **23**, (2022).
71. Yu, G., Smith, D. K., Zhu, H., Guan, Y. & Lam, T. T. ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol Evol* **8**, 28–36 (2017).

634

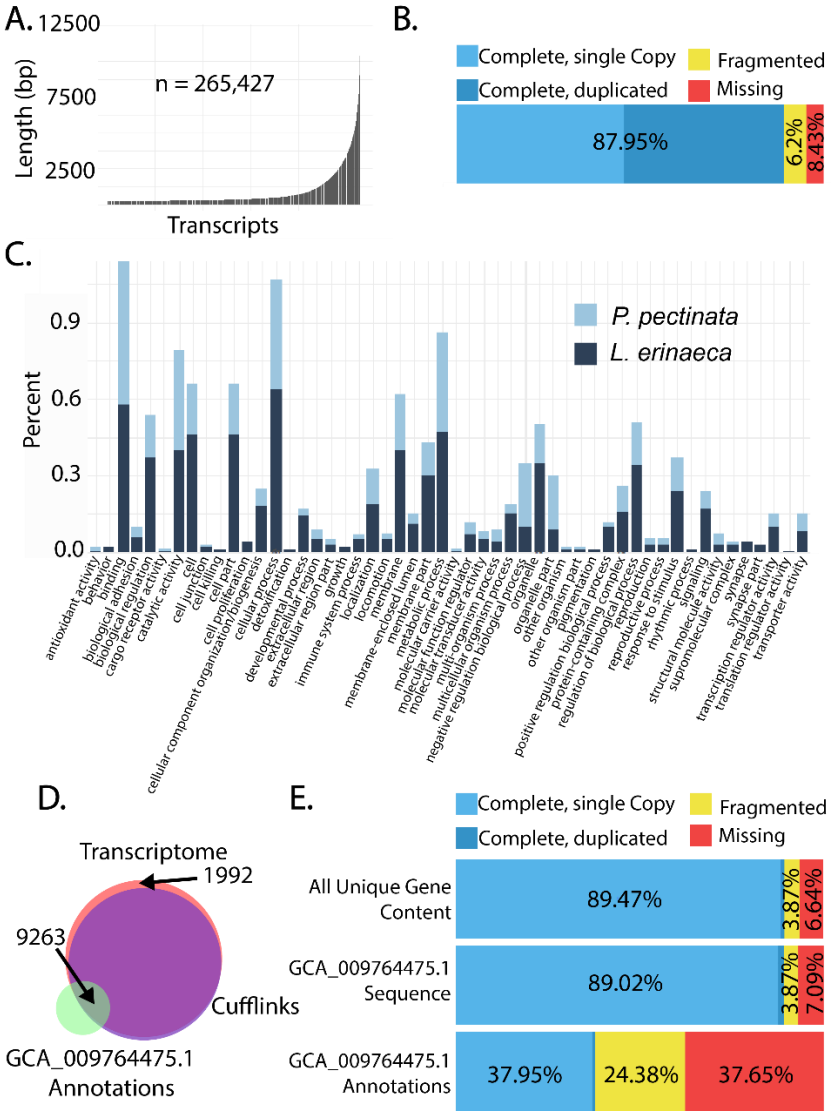


Figure 1. Establishment of a near complete gene set for the smalltooth sawfish, *Pristis pectinata*. A) Log length distribution of transcripts in assembled transcriptome with total number of transcripts (n = 265,427). **B)** Outcome of Benchmarking Universal Single Copy Ortholog (BUSCO) analysis to assess completeness of assembled transcriptome. **C)** WeGO (Ye et al., 2018) gene ontology analysis of *P. pectinata* and little skate, *Leucoraja erinacea,* transcriptomes. Genes are classified according to biological process, molecular function, or cellular component and plotted by percent of genes related to meaningful biological activities. *P. pectinata* has a higher percentage of genes related to DNA binding, negative regulation of biological processes, organelle part, and antioxidant activity, while *L. erinacea* has more genes related to cellular processes, response to stimulus, signaling, and catalytic activity. **D)** Venn diagram (Hulsen et al., 2008) showing overlap of the number of transcripts provided with the public genome (GCA_009764475.1, green), the Cufflinks assisted annotation of the public genome (purple), and transcripts in the de novo assembled transcriptome (pink). **E)** BUSCO assessment for all unique gene content from combined transcriptome and genome recourses, and annotations associated with GCA_009764475.1.

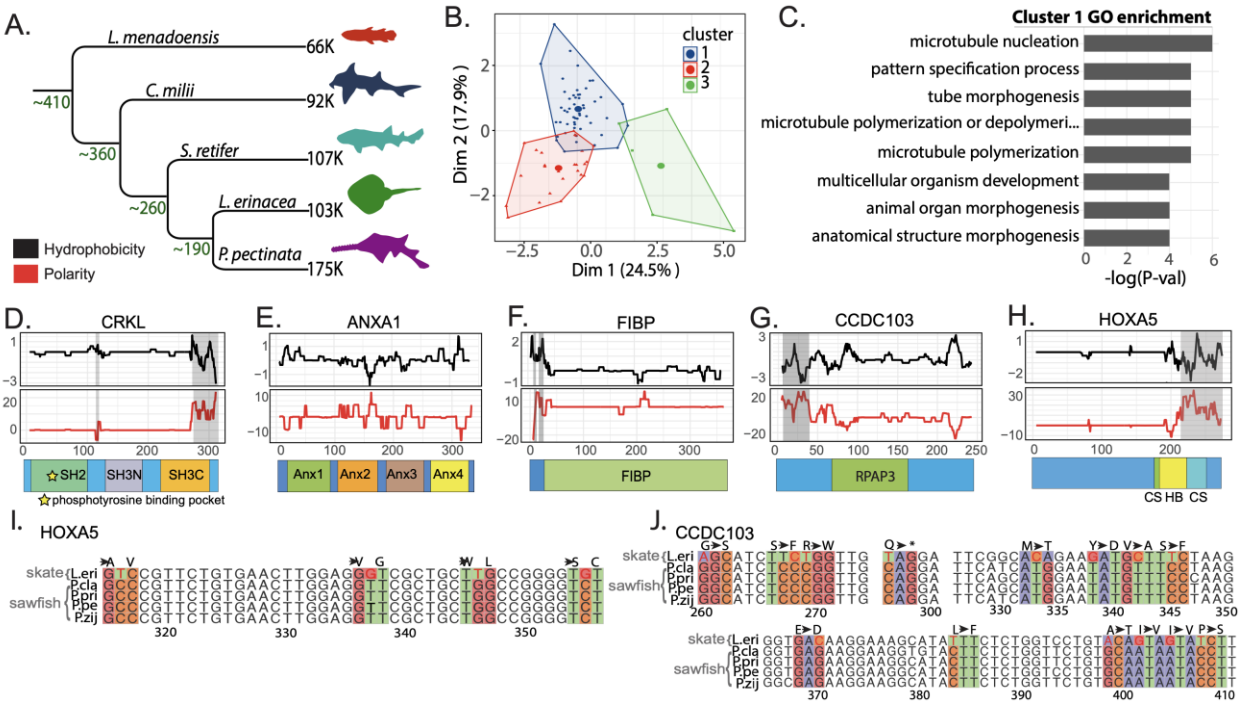


Figure 2. Smalltooth sawfish, *Pristis pectinata*, genes under positive selection. A) Phylogeny of taxa used in positive selection analysis with approximate divergence times shown at each node and number of transcripts in each transcriptome noted at each terminal branch. Species included: *Latimeria menadoensis*, *Callorhinchus milii*, *Scyliorhinus retifer*, *Leucoraja erinacea*, and *P. pectinata*. **B)** Results of k-means clustering analysis of selected genes using omega value, percent of sites, change in GRAVY value relative to little skate, *L. erinacea*, and change in expression relative to *S. retifer*. **C)** Top 8 most enriched gene ontology (GO) terms related to biological processes from genes which were grouped into Cluster 1 by PCA analysis plotted by -log(p-value). **D-H)** Changes in hydrophobicity and polarity of protein sequence relative to *L. erinacea.* Sequence alignment gaps are shown by shaded regions. Functional protein domains retrieved from InterproScan and literature are shown as colored boxes below each plot. **I)** Sequence alignment of PCR-amplified region of interest from HoxA5 between *Pristis* sawfishes and *L. erinacea* with codons of interest colored by nucleotide. **J)** Alignment of region of interest from amplification of CCDC103.Alignments are colored by nucleotide.

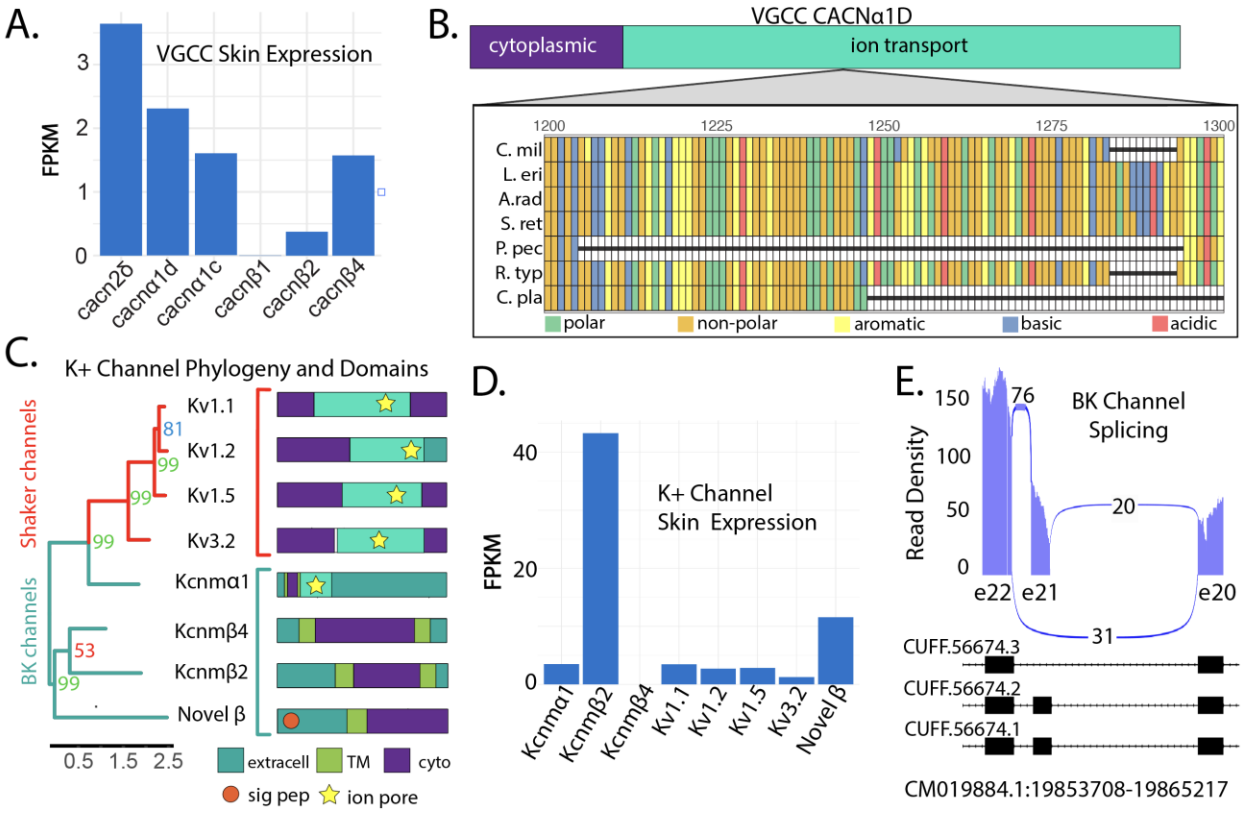
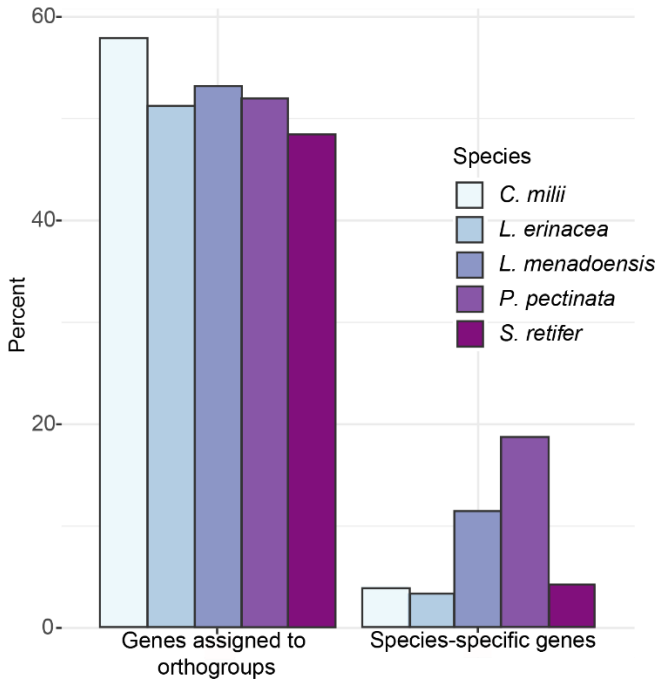
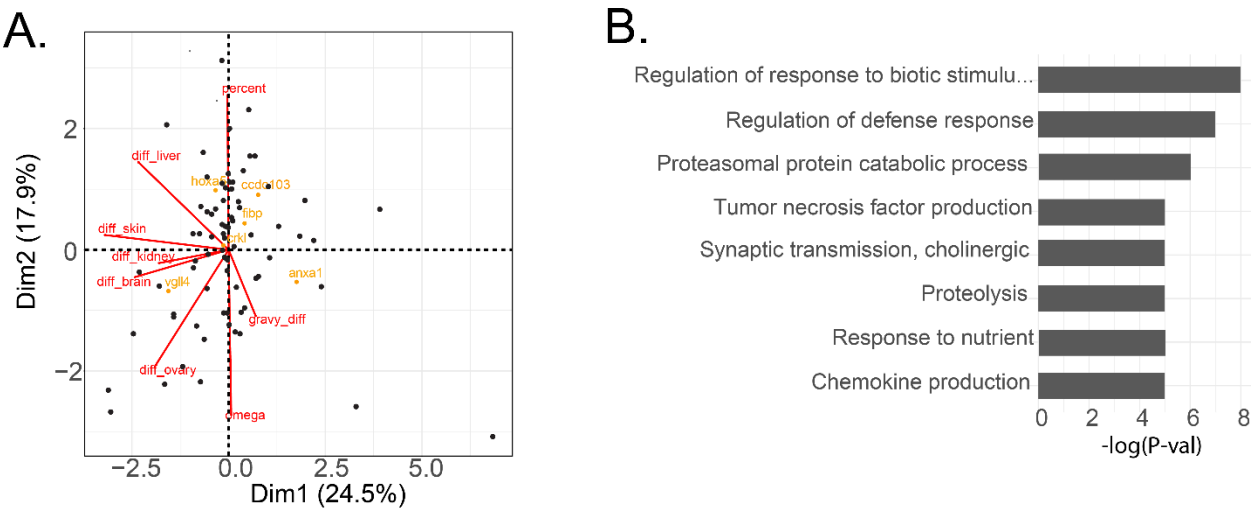


Figure 3. Smalltooth sawfish, *Pristis pectinata*, electroreception machinery. A) FPKM normalized expression of voltage gated calcium channel α and β subunits in rostral skin. **B)** Multiple sequence alignment of *Cacnα1D* depicting missing 91-residue segment in *P. pectinata* ion channel domain relative to other elasmobranchs. Species included: *Callorhinchus milii* (C. mil), *Leucoraja erinacea* (L. eri), *Amblyraja radiata* (A. rad), *Scyliorhinus retifer* (S. ret), *P. pectinata* (P. pec), *Rhinchodon typus* (R. typ), and *Chiloscyllium plagiosum* (C. pla). Alignments are colored by amino acid chemistry. **C)** Phylogenetic construction of potassium channels and subunits with bootstrap values noted (left). Sequences were aligned with ClustalW and tree was constructed by RAxML with 1000 bootstraps. Shaker channel sequences are highlighted in red, while BK channels are highlighted with light blue. Annotated protein domains of each sequence obtained using Interpro Scan (Right). EC is extracellular, TM, transmembrane, cyto is cytoplasm, and pore is ion channel pore. **D)** Normalized expression of potassium channels potentially involved electrosensing in rostral skin. Kcnm- transcripts correspond to BK channel subunits, Kv are voltage-gated Shaker channels. Novel β refers to the uncharacterized BK subunit found in transcriptome data. **E)** GGSashimi plot of alternatively spliced BK channel transcripts in all *P. pectinata* tissues. Transcriptome expression confirms that the BK channel transcript with highest expression in the rostral skin retains exon 21 (exon 29 in *L. erinacea*).

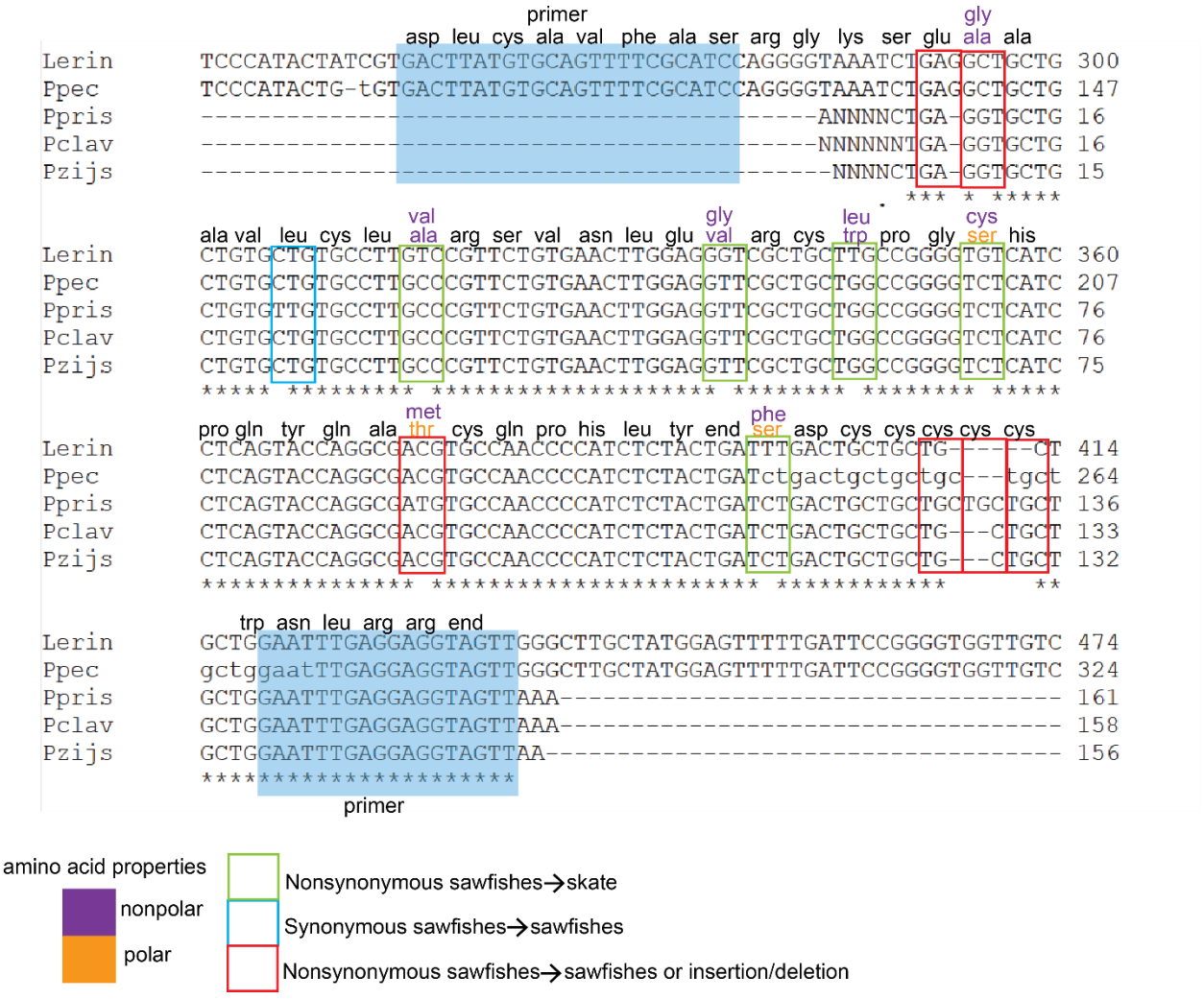


684

685 **Supplementary Figure 1.** Left: Percentage of genes assigned to orthogroups by OrthoFinder.
686 Right: Percentage of genes unique to each species. Species included: chain catshark,
687 *Scyliorhinus retifer*, Indonesian coelacanth, *Latimeria menadoensis*, Australian ghostshark,
688 *Callorhinchus milii*, smalltooth sawfish, *Pristis pectinata*, and little skate, *Leucoraja erinacea*.

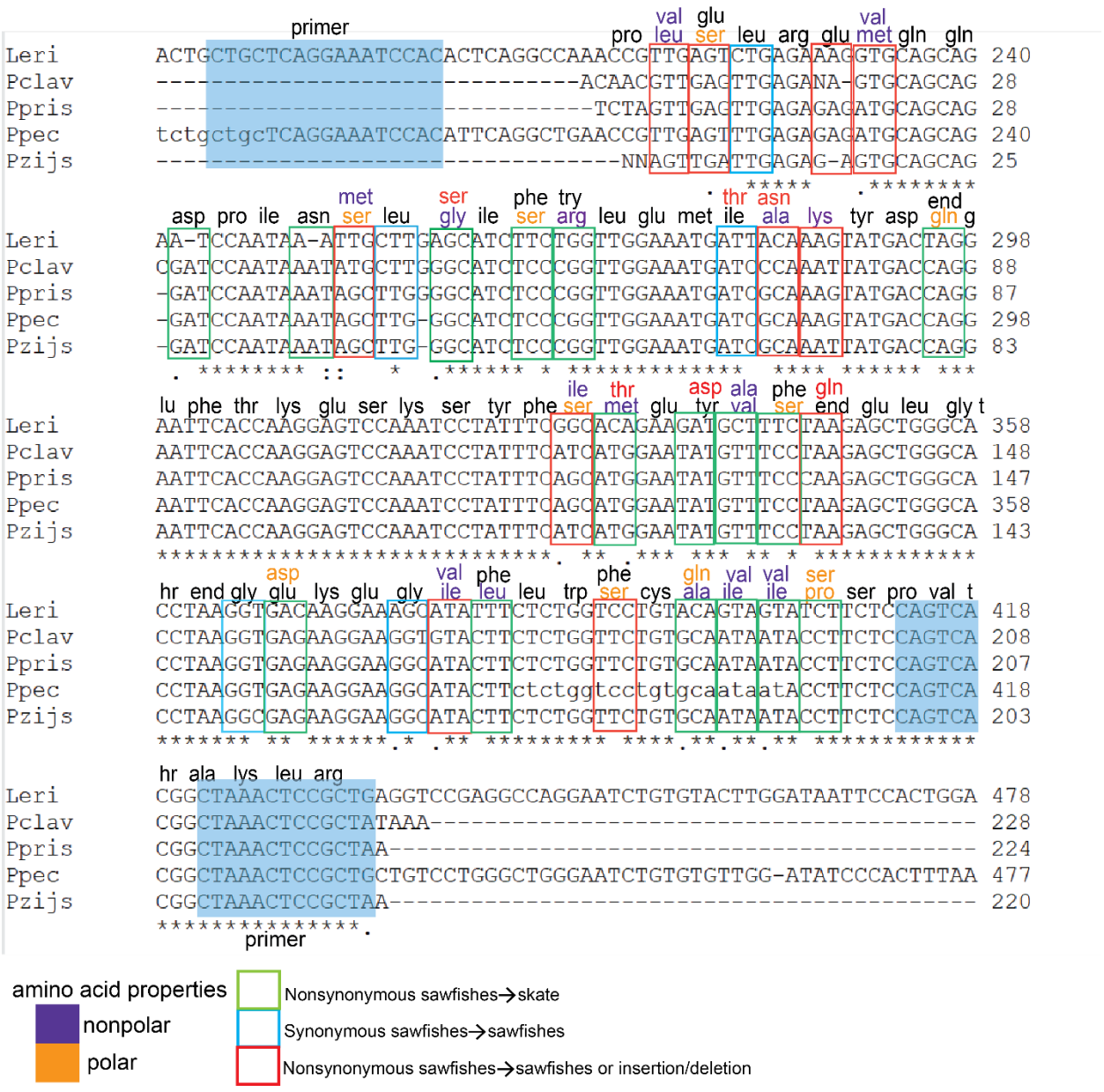


Supplementary Figure 2. Smalltooth sawfish, *Pristis pectinata*, genes under selection. A) PCA of genes under selection using omega value, percent of sites, change in GRAVY value relative to little skate, *Leucoraja erinacea*, and change in expression relative to chain catshark, *Scyliorhinus retifer*. Genes of interest from Cluster 1 are shown in yellow, variables are shown in red. **B)** Top 8 most enriched gene ontology terms related to biological processes from genes grouped into Cluster 2 by PCA analysis plotted by -log(p-value).



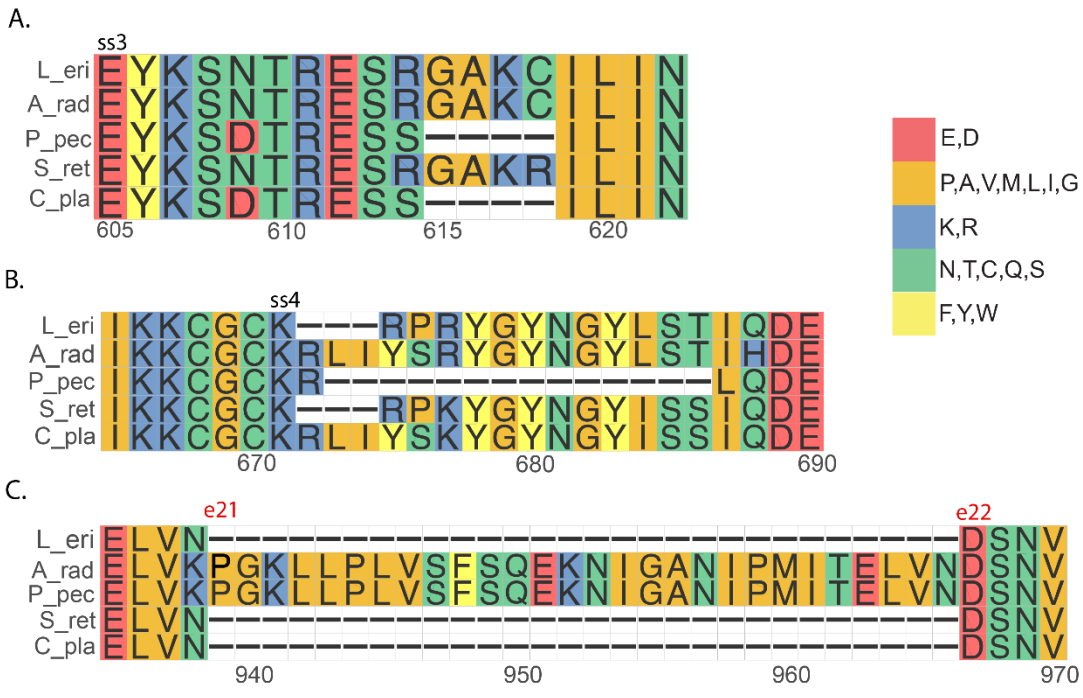
Supplementary Figure 3

Annotation of substitutions in *HoxA5* between dwarf, green, largetooth, and smalltooth sawfishes (*Pristis clavata*, *P. zijsron*, *P. pristis*, *P. pectinata*, respectively) and little skate, *Leucoraja erinacea*. Changes in amino acids are noted above aligned nucleotide sequences. Primer sequences are shaded in blue.



Supplementary Figure 4

Annotation of substitutions in *Ccdc103* between dwarf, green, largetooth, and smalltooth sawfishes (*Pristis clavata*, *P. zijsron*, *P. pristis*, *P. pectinata*, respectively) and little skate, *Leucoraja erinacea*. Changes in amino acids are noted above aligned nucleotide sequences. Primer sequences are shaded in blue.



Supplementary Figure 5

Insertions, deletions, and splice sites identified in *Pristis pectinata* (P_pec) BK alpha K^+ channel relative to *Leuroraja erinacea* (L_eri), *Amblyraja radiata* (A_rad), *Scyliorhinus retifer* (S_ret) and *Chiloscyllium plagiosum* (C_pla). Ss = splice site. e = exon. Plot colors indicate amino acid chemical properties.

Tissue	Concentration (ng/ul)	Purity (A260/A280)	RIN
Brain	174.7	1.83	6.7
Liver	962.4	2.01	6.9
Ovary	1660.7	2.03	7.0
Kidney	1150.0	2.04	7.3
Skin	273.9	1.98	7.5

Supplementary Table 1. Concentrations and purities of smalltooth sawfish, *Pristis pectinata*, RNA samples by tissue obtained by nanodrop and Bioanalyzer. For RNA, chemical purity is indicated by A260/A280 of 2.0. RNA integrity number (RIN) ranges from 1 to 10, where 10 is intact and 1 is degraded.

	Smalltooth sawfish, *Pristis pectinata*	Chain catshark *Scyliorhinus retifer*	Little skate, *Leucoraja erinacea*	Indonesian coelacanth, *Latimeria menadoensis,*	Australian ghostshark, *Callorhinchus milii*
No. transcripts in dataset	175,569	107,231	103,996	66,138	92,334
Accession	NA	GEO: GSM643958	GEO: GSM643957	GAPS01066138	GEO: GSM643959
%BUSCO complete	89.47% C 3.87% F	56.6% C 22.1% F	60.4% C 20.4% F	40.9% C 18.9% F	47.3% C 26.9% F

Supplementary Table 2. Number of transcripts, NCBI accession numbers, and percentage of complete (C) and fragmented (F) orthologs via BUSCO analysis of transcriptomes for all taxa used in positive selection analyses.

Species	Unique genes	Total genes
Smalltooth sawfish, *Pristis pectinata*	77	96
Chain catshark, *Scyliorhinus retifer*	26	38
Australian ghostshark, *Callorhinchus milii*	76	95
Little skate, *Leucoraja erinacea*	40	51
Indonesian coelacanth, *Latimeria menadoensis*	49	60

Supplementary Table 3. Number of unique genes and total number of genes found under selection in aBSREL analysis by species.

Sample	Accession code	Run
Adult 1 kidney	SAMD00098998	DRR111789
Adult 1 medulla	SAMD00098989	DRR111780
Adult 1 ovary	SAMD00098994	DRR111785
Adult 1 liver	SAMD00098997	DRR111788
Adult 3 skin	SAMD00099043	DRR111834

724 **Supplementary Table 4.** NCBI sample names, accession codes, and run identifiers for the
725 chain catshark, *Scyliorhinus retifer,* expression data used in PCA analysis.

Sample	Total reads	Unaligned	Aligned once	Multimapping	Overall rate
A1 Kidney	10208584	99.41%	0.58%	0.01%	0.59%
A1 Medulla	8936204	99.32%	0.67%	0.01%	0.68%
A1 Ovary	8892600	99.30%	0.69%	0.01%	0.70%
A1 Liver	7162306	99.61%	0.38%	0.01%	0.39%
A3 Skin	9125803	99.35%	0.63%	0.02%	0.65%

Supplementary Table 5. Number and percent of reads mapped to genes orthologous to the smalltooth sawfish, *Pristis pectinata*, PSGs per sample for the chain catshark, *Scyliorhinus retifer*, expression data used in PCA analysis.

Gene	Chrom	Start	Stop	Score	Query Acc.
hoxA1	Chr5	107566999	107568411	902	ACT65756.1
hoxA2	Chr5	107572840	107574441	1070	ACT65755.1
hoxA3	Chr5	107579173	107581805	1056	ACT65754.1
hoxA4	Chr5	107599253	107600338	706	ACT65753.1
hoxA5	Chr5	107611902	107613201	706	ACT65752.1
hoxA6	Chr5	107615218	107616103	680	ACT65751.1
hoxA7	Chr5	107624042	107625644	611	ACT65750.1
hoxA9	Chr5	107633224	107634655	610	ACT65749.1
hoxA10	Chr5	107642407	107644791	686	ACT65748.1
hoxA11	Chr5	107656057	107658036	818	ACT65747.1
hoxA13	Chr5	107669970	107671270	860	ACT65746.1
hoxB1	Chr25	3190815	3192527	718	ASS31204.1
hoxB2	Chr25	3200448	3202103	728	ASS31205.1
hoxB3	Chr25	3206913	3208894	1055	ASS31206.1
hoxB4	Chr25	3228207	3229258	652	ASS31207.1
hoxB5	Chr25	3242638	3244302	727	ASS31208.1
hoxB6	Chr25	3246731	3248076	611	ASS31209.1
hoxB7	Chr25	3254738	3256947	596	XP_032890850.1
hoxB8	Chr25	3258510	3259634	611	ASS31210.1
hoxB9	Chr25	3268829	3270311	493	ASS31211.1
hoxB10	Chr25	3278434	3284468	488	ASS31212.1
hoxB13	Chr25	3355373	3356805	793	ASS31213.1
hoxD1	Chr1	47299641	47300923	482	ASS31214.1
hoxD2	Chr1	47291801	47293231	808	ASS31215.1
hoxD3	Chr1	47286156	47288710	929	ASS31216.1
hoxD4	Chr1	47268047	47269186	623	ASS31217.1
hoxD5	Chr1	47258508	47259603	566	ASS31218.1
hoxD8	Chr1	47248080	47249022	665	ASS31219.1
hoxD9	Chr1	47240275	47241612	590	ASS31220.1
hoxD10	Chr1	47233908	47235988	715	ASS31221.1
hoxD11	Chr1	47224113	47225493	592	ASS31222.1
hoxD12	Chr1	47216240	47217755	701	ASS31223.1
hoxD13	Chr1	47209657	47210874	773	ASS31224.1

Supplementary Table 6. BLAT results of Hox genes into the smalltooth sawfish, *Pristis pectinata*, genome browser with chromosome positions, starts, stops, highest scores, and query accession numbers. Only the highest scores were included for simplicity.