



RESEARCH ARTICLE

SkateBase, an elasmobranch genome project and collection of molecular resources for chondrichthyan fishes [v1; ref status: indexed, <http://f1000r.es/445>]

Jennifer Wyffels¹, Benjamin L. King², James Vincent³, Chuming Chen¹, Cathy H. Wu¹, Shawn W. Polson¹

¹Department of Computer and Information Sciences, Center for Bioinformatics and Computational Biology, University of Delaware, Newark, DE, 19711, USA

²Mount Desert Island Biological Laboratory, Salisbury Cove, ME, 04672, USA

³Vermont Genetics Network, University of Vermont, Burlington, VT, 05405, USA

v1 First published: 12 Aug 2014, 3:191 (doi: [10.12688/f1000research.4996.1](https://doi.org/10.12688/f1000research.4996.1))
 Latest published: 12 Aug 2014, 3:191 (doi: [10.12688/f1000research.4996.1](https://doi.org/10.12688/f1000research.4996.1))

Abstract

Chondrichthyan fishes are a diverse class of gnathostomes that provide a valuable perspective on fundamental characteristics shared by all jawed and limbed vertebrates. Studies of phylogeny, species diversity, population structure, conservation, and physiology are accelerated by genomic, transcriptomic and protein sequence data. These data are widely available for many sarcopterygii (coelacanth, lungfish and tetrapods) and actinopterygii (ray-finned fish including teleosts) taxa, but limited for chondrichthyan fishes. In this study, we summarize available data for chondrichthyes and describe resources for one of the largest projects to characterize one of these fish, *Leucoraja erinacea*, the little skate. SkateBase (<http://skatebase.org>) serves as the skate genome project portal linking data, research tools, and teaching resources.



This article is included in the **Elasmobranch Biology & Conservation Collection**

Open Peer Review

Referee Status:

Invited Referees

1 2

version 1
published
12 Aug 2014



1 **Kevin Feldheim**, Field Museum of Natural History USA

2 **Ed Heist**, Southern Illinois University USA

Discuss this article

Comments (0)

Corresponding author: Jennifer Wyffels (wyffels@udel.edu)

How to cite this article: Wyffels J, L. King B, Vincent J *et al.* **SkateBase, an elasmobranch genome project and collection of molecular resources for chondrichthyan fishes [v1; ref status: indexed, <http://f1000r.es/445>]** *F1000Research* 2014, **3**:191 (doi: [10.12688/f1000research.4996.1](https://doi.org/10.12688/f1000research.4996.1))

Copyright: © 2014 Wyffels J *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution Licence](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. Data associated with the article are available under the terms of the [Creative Commons Zero "No rights reserved" data waiver](#) (CC0 1.0 Public domain dedication).

Grant information: This work was supported by a re-entry career award to JTW, National Institute of General Medical Sciences (NIGMS) IDeA Networks of Biomedical Research Excellence (INBRE) 3P20GM103446-12S1. Skate genome sequencing was funded by National Institutes of Health (NIH) National Center for Research Resources (NCRR) ARRA Supplements to 5P20RR016463-12 (MDIBL), 5P20RR016472-12 (UD), and 5P20RR16462 (UVM). The North East Cyberinfrastructure Consortium (NECC) is funded by NIH NCRR grants 5P20RR016463-12 (MDIBL), 5P20RR016472-12 (UD), 5 P20 RR16462 (UVM), 5P20RR016457-11 (URI), and 5P20RR030360-03 (UNH) and NIH NIGMS grants 8 P20 GM103423-12 (MDIBL), 8P20GM103446-12 (UD), 8P20GM103449 (UVM), 8 P20 GM103430-11 (URI), and 8P20GM103506-03 (Dartmouth), and NSF Experimental Program to Stimulate Competitive Research (EPSCoR) grants EPS-0904155 (UM), EPS-081425 (UD), EPS-1101317 (UVM), EPS-1004057 (URI), and EPS-1101245 (UNH).

Competing interests: The authors declare to have no competing interests.

First published: 12 Aug 2014, **3**:191 (doi: [10.12688/f1000research.4996.1](https://doi.org/10.12688/f1000research.4996.1))

First indexed: 01 Sep 2014, **3**:191 (doi: [10.12688/f1000research.4996.1](https://doi.org/10.12688/f1000research.4996.1))

Introduction

Chondrichthyan fishes are composed of two subclasses, Holocephali and Elasmobranchii. Holocephalans are the more basal of the pair having first appeared more than 400 million years ago and include a single surviving order, Chimaeriformes, the chimaeras, with 39 extant species¹. Elasmobranchs appeared approximately 350 million years ago and include more than 1000 species of sharks, skates, and rays². Chondrichthyan fishes occupy a pivotal position at the base of the vertebrate phylogenetic tree. For research that includes an evolutionary component, representation of this diverse class affords a valuable perspective to evaluate all vertebrates.

Chondrichthyan fishes are circumglobal in distribution and occupy a wide range of ecological habitats. Their life history parameters are equally disparate but in general chondrichthyans are slow growing and late maturing fishes with an increased risk of extinction³⁻⁵. Fecundity is as few as 1 or 2 for viviparous species such as the sand tiger shark, *Carcharias taurus*⁶ and as high as 300 for the whale shark, *Rhincodon typus*⁷. They are of economic importance for fisheries as well as ecotourism. Management and assessment of stock is essential to ensure both ecotourism interests and food resources remain sustainable⁸. Management of fish populations has increasingly relied on molecular tools to investigate population structure, properly identify species, and compliance with fishing quotas⁹⁻¹³.

Elasmobranchs have been used as a model for biomedical research for more than 100 years. Elasmobranchs, like other cartilaginous fishes, exhibit many fundamental vertebrate characteristics, including a neural crest, jaws and teeth, an adaptive immune system, and a pressurized circulatory system. The skate is a powerful comparative model to study biological processes shared among jawed and limbed vertebrates such as development¹⁴⁻¹⁶, renal physiology¹⁷⁻²⁰, immunology²¹⁻²⁶, toxicology²⁷, neurobiology²⁸, and wound healing and regeneration²⁹. They are the most ancient vertebrates to possess an adaptive immune system that generates antibodies using a V(D)J combinatorial mechanism³⁰. Phylogenetically, cartilaginous fishes are the first vertebrates to possess a thymus, a central lymphoid organ that provides a microenvironment for the development of T cells³¹. The thymus shares a common organization with more derived vertebrates containing cortical and medullary regions^{32,33}.

In addition to shared physiological characteristics, the diversity of specializations between species allows investigations of evolution within a single clade. For example, elasmobranchs use a plethora of reproductive strategies that span the full range of maternal investment from placental viviparity to strict lecithotrophic oviparity. Besides sexual reproduction, captive elasmobranchs are capable of asexual parthenogenesis³⁴⁻³⁶. Of these reproductive mechanisms, the most tractable for research purposes is oviparity. Approximately 43% of chondrichthyans utilize oviparity including all Chimaeriformes, Heterodontiformes (bullhead sharks), Rajoidae (skates) and Scyliorhinidae (catsharks)³⁷. Many species can be maintained in captivity and will breed and lay eggs throughout an annual season³⁸. Artificial insemination has been reported for two oviparous species, the clearnose skate, *Raja eglanteria*³⁹, and the cloudy catshark, *Scyliorhinus torazame*⁴⁰. Additionally, sperm storage allows wild caught females to lay eggs for several years without requiring males or captive mating events⁴¹.

Leucoraja erinacea, the little skate, was chosen for a genome sequencing project to represent this clade of fishes because of their use as a biomedical model, experimental tractability, genome size, existing sequence data, and northeast regional distribution. The sequencing project is an ongoing effort of the North East Bioinformatics Collaborative (NEBC) of the North East Cyberinfrastructure Consortium (NECC), composed of the bioinformatics core facilities from Delaware, Maine, New Hampshire, Rhode Island, and Vermont funded by National Institutes of Health (NIH) Institutional Development Awards (IDeA) and/or National Science Foundation (NSF) Experimental Program to Stimulate Competitive Research (EPSCoR) programs.

Existing resources

There is a single order of holocephalans and 13 orders of elasmobranchs. The distribution of species in orders, families and genera is shown in [Figure 1](#). The batoids are composed of 4 orders, Rajiiformes, Myliobatiformes, Torpidiformes, and Rhinopristiformes, and contain 54% of extant chondrichthyan species. Sharks are broadly divided into two super orders, Galeomorphii and Squalomorphii that together account for 43% of extant chondrichthyan species. The galean sharks include 4 orders: Heterodontiformes, Orectolobiformes, Lamniiformes and Carcharhiniformes, and represent 30% of extant chondrichthyan species. Squalan sharks are composed of 4 orders: Squaliformes, Squatiniformes, Pristophoriformes, and Hexanchiformes, comprising 13% of extant chondrichthyan species. Among individual orders, Rajiiformes, the skates, have the most species (345) followed by Carcharhiniformes, the ground sharks (283) and Myliobatiformes (226)². These 'big three' orders contain 854 species, 72% of extant chondrichthyans.

Chondrichthyan conservation, management, and research all benefit from easily accessible and well-documented molecular resources. The organization of data and metadata in archival databases is critically important for efficient use of large and complex datasets. The International Nucleotide Sequence Database Collaboration (INSDC) is composed of three large public nucleotide repositories, DNA Data Bank of Japan (DDBJ), European Molecular Biology Laboratory-European Bioinformatics Institute (EMBL-EBI), and GenBank at the National Center for Biotechnology Information (NCBI). Recently, two new NCBI database projects were initiated to collect details of samples, BioSample, and project data, BioProject, and propagate the metadata to all associated database entries in an effort to expand the use of already existing and rapidly expanding molecular resources⁴². [Figure 2](#) illustrates the relationship between BioProject, BioSample and the sequence data for SkateBase. Because the BioProject and BioSample databases were established in 2012, not all existing datasets have metadata or details of the biological source to populate a BioSample and BioProject entry. When available, BioProject and BioSample hyperlinks are included for Sequence Read Archive (SRA), Expressed Sequence Tag (EST) and Genome Survey Sequence (GSS) datasets in the tables below.

[Table 1](#) is a summary of chondrichthyan sequence data in NCBI databases, UniProtKB, and the Protein Data Bank (PDB) with *L. erinacea*, *Callorhynchus milii* and *Scyliorhinus canicula*, the three species featured at SkateBase listed individually. The distribution of holocephalans and elasmobranchs in public databases

is illustrated in Figure 3. Despite the majority of species belonging to Elasmobranchii, the GenBank, UniProtKB/TrEMBL, and Gene databases are dominated by chimaera data derived from the genome sequence of the elephant shark, *C. milii*⁴³. Elasmobranch data predominates in UniProtKB/Swiss-Prot, PDB, BioProject and BioSample databases as well as the number of whole mitochondrial

genomes (WGM) in GenBank. The EST and SRA databases are nearly equally split between the two subclasses.

Chondrichthyan genomes

Currently there are multiple efforts to sequence an elasmobranch genome in various stages of completion (Table 2); however, only

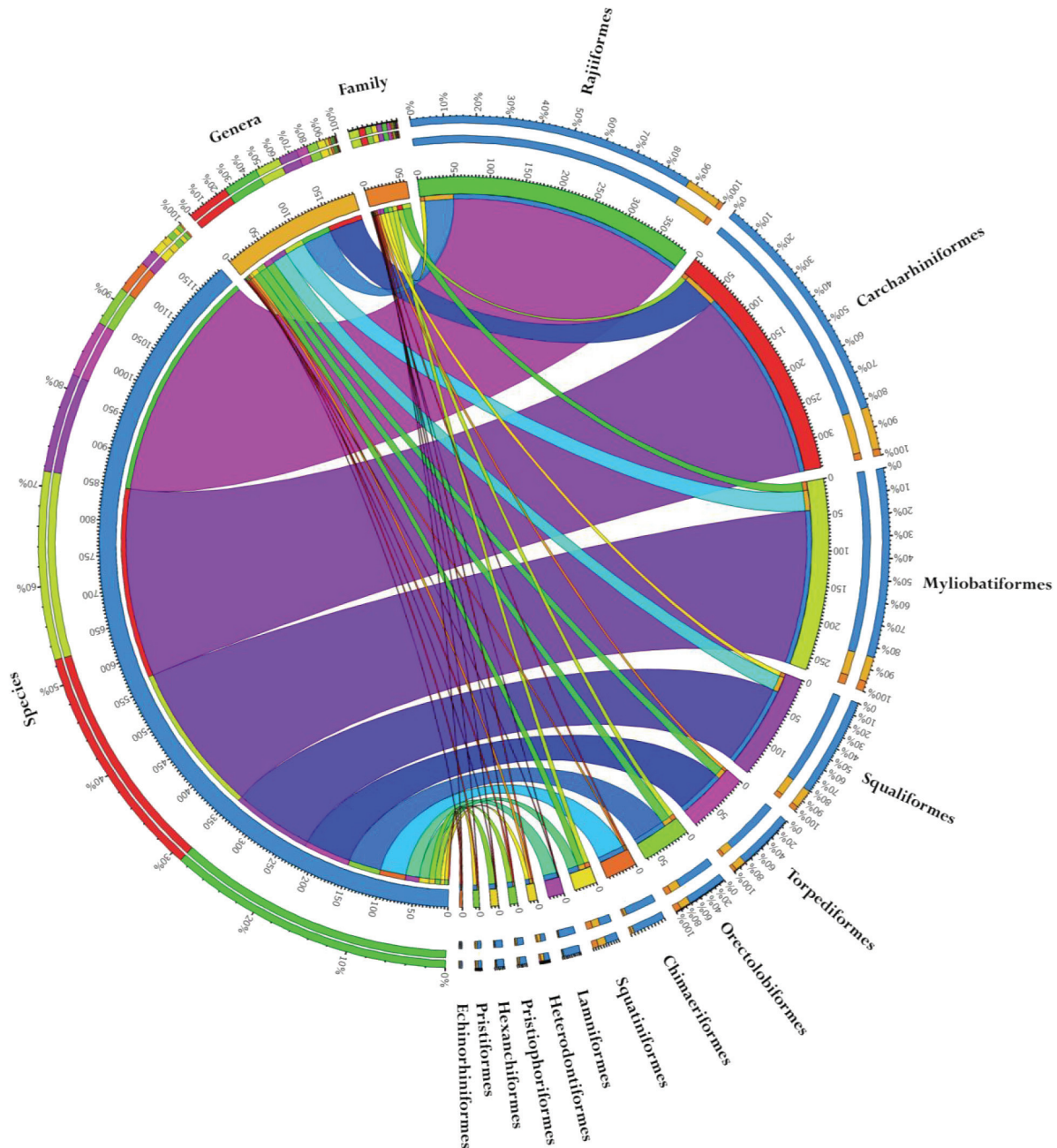


Figure 1. Species distribution within chondrichthyan orders. There is a single order of Holocephalans, Chimaeriformes, and 13 orders of elasmobranchs. The distribution of chondrichthyan species in each of the 14 orders is shown relative to the total number of species, genera and families for the clade. The batoids are composed of 4 orders, Rajiiformes, Myliobatiformes, Torpidiformes, and Rhinopristiformes, and contain 54% of extant chondrichthyan species. Sharks are broadly divided into two super orders, Galeomorphii and Squalomorphii that together include the remaining 9 orders and 43% of extant chondrichthyan species.

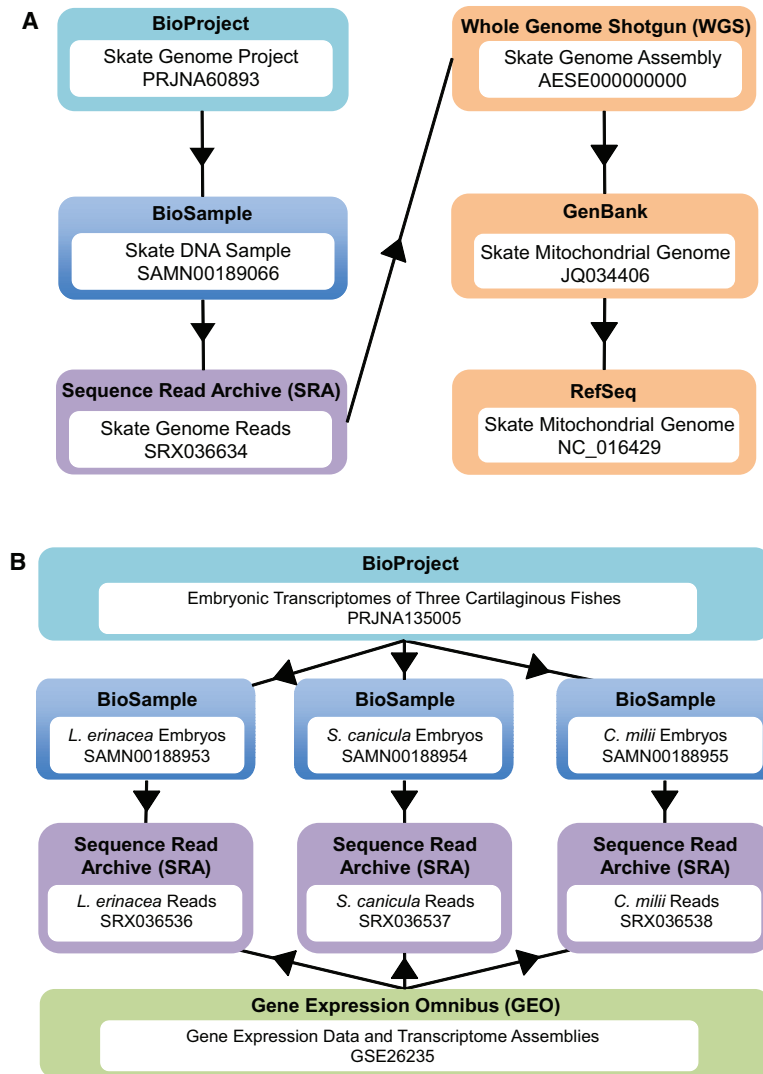


Figure 2. Representation of SkateBase data within the The National Center for Biotechnology Information (NCBI) databases. A. The little skate genome project is represented as a BioProject entry that connects all samples and data thematically. A BioSample record describes the DNA sample that was used for genome sequencing that was generated from a single stage 32 skate embryo. The Sequence Read Archive (SRA) catalogs the unassembled Illumina genome sequence data. The Whole Genome Shotgun (WGS) database contains the contiguous sequences from shotgun sequencing projects. The assembled and annotated mitochondrial genome was deposited in GenBank and subsequently included in the NCBI Reference Sequence Database (RefSeq). **B.** The project to characterize the embryonic transcriptomes of *L. erinacea*, *C. milii* and *S. canicula* is represented in a BioProject entry. Three BioSample entries, one for each species, lead to three SRA datasets. The transcriptome data is represented also in the Gene Expression Omnibus (GEO), a database of high-throughput functional genomic data derived from microarrays and next-generation sequencing technologies.

the skate genome project currently has data publically available. Efforts to sequence the whale shark are underway at the Georgia Aquarium and Emory University (personal communication, Alistair Dove, Georgia Aquarium). Genoscope leads a project to sequence the genome of another oviparous elasmobranch, the catshark, *S. canicula*. The current assembly is described in Table 2. A second version of the catshark genome with 200x coverage, including mate pair sequencing, is in progress (personal communication, Sylvie Mazan, French National Centre for Scientific Research). Among holocephalans, the genome of the elephant shark, *C. milii*, was first described in a 1.4x coverage assembly in 2006⁴⁴. With continued sequencing the assembly coverage is currently 19.25x and data has

been made available through the project website (<http://esharkgenome.imcb.a-star.edu.sg/>) and Genbank⁴³.

A powerful resource for characterizing genomes is large-insert clone libraries where each clone contains a large (~100kb) genomic region. Bacterial artificial chromosome (BAC) and P1-derived artificial chromosome (PAC) libraries are DNA constructs within a plasmid used to transform bacteria. As the bacteria grow the inserted DNA is amplified and subsequently isolated and sequenced. BACs are beneficial for genome sequencing projects because the insert size can be very large, nearly 350 kb, facilitating assembly post-sequencing. BAC/PAC libraries were built for several chondrichthyan species

Table 1. Chondrichthyan molecular sequence data in public databases.

National Center for Biotechnology Information (NCBI) databases ¹															
Taxonomy	BioProject	BioSample	Gene	GenBank				GSS lib	WGS (Mbp)	GEO ²	SRA	UniProtKB ²			
				GenBank	WMG	EST	EST lib					GSS	Swiss-Prot	TrEMBL	PDB
Chondrichthyes	16	75	21069	55810	72	192948	33	28497	5	2492.3	3	22	276	26485*	178
Holocephali	3	21	20201	39512	8	109965	6	27944	1	936.9	1	13	12	20170	0
<i>C. milii</i>	3	21	20110	39232	1	109965	6	27944	1	936.9	1	13	3	19989	0
Elasmobranchii	13	54	868	16273	64	82983	27	553	4	1555.4	2	9	264	6299	178
<i>L. erinacea</i>	3	7	13	284	1	31167	5	0	0	1555.4	1	2	6	123	0
<i>S. canicula</i>	2	8	13	645	1	1600	7	0	0	0	1	1	38	283	1

(WMG) whole mitochondrial genome, (EST) Expressed Sequence Tags, (lib) libraries (GSS) Genome Survey Sequences, (GEO) Gene Expression Omnibus, (WGS) Whole Genome Shotgun, (SRA) Sequence Read Archive, (WMG) whole mitochondrial genomes, (PDB) Protein Data Bank. * includes 16 unidentified fin entries

¹ NCBI databases accessed July 25, 2014; ² Release 2014_07 of 09-Jul-2014; ³ GEO sample accessions

Table 2. Chondrichthyan genome sequencing projects.

	Website	Genome size (Gb)	Coverage	Contigs	N50 (bp)	Platform	Facility	Genbank	Data ¹	BioProject	BioSample	Date
Holocephali												
<i>Callorhynchus milii</i>	esharkgenome	0.910	19.25x	21,203	1466	Sanger & 454	IMCB	AAVX020000000	244 M	PRJNA236996	SAMIN00000800	20-Dec-13*
Elasmobranchii												
<i>Leucoraja erinacea</i>	skatebase.org	3.42	26x	2,62,365	665	Illumina PE	NECC	AESE010000000	105 G	PRJNA60893	SAMN00189066	22-Dec-11
<i>Scyliorhinus canicula</i>	-	3.5	32x	3,449,662	1,292	Illumina PE	Genoscope-CEA	-	-	-	-	-
<i>Rhincodon typus</i>	-	3.44(est.)	35x			Illumina & 454	Emory University & Georgia Aquarium	-	-	PRJNA255419	SAMN02918461 SAMN02918462	16-Jul-14

¹ (M) Mega or (G) Giga base pairs; (PE) paired end; (est) estimated; (ICMB) Institute of Molecular and Cell Biology, A*STAR, (NECC) North East Cyberinfrastructure Consortium
* replaced original sequence data GenBank AAVX000000000.1 (1.4x coverage) released 20-DEC-2006

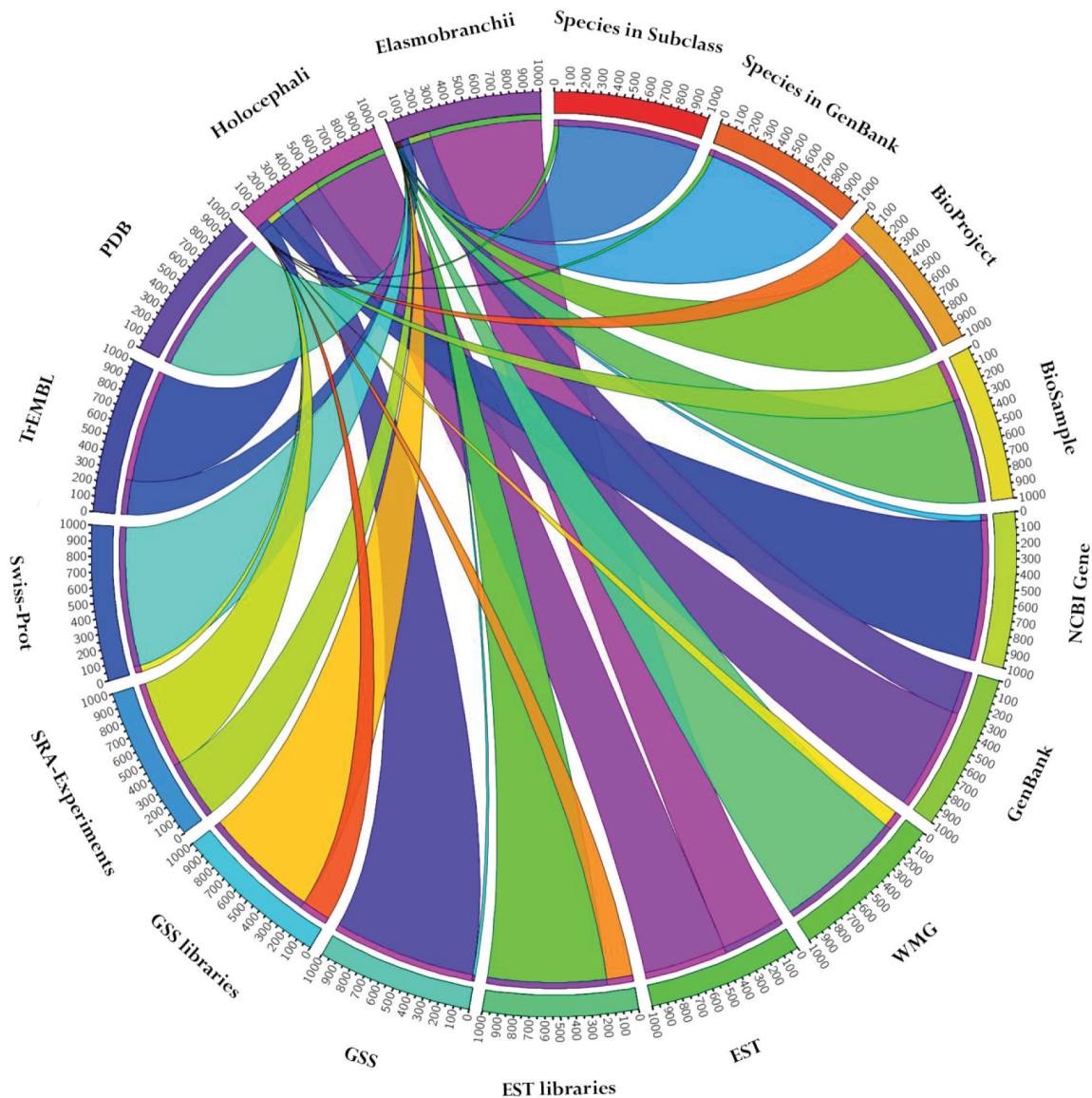


Figure 3. Holocephalan and elasmobranch resources in public nucleotide and protein databases. The distribution of data for Holocephali (chimaeras) and elasmobranchii (sharks and rays) subclasses of chondrichthyan fishes does not always reflect their species distribution. The number of species represented in GenBank is representative of the actual species distribution but the amount of data in GenBank is not. Holocephalan data forms the majority of the NCBI Gene, GenBank, Genome Survey Sequence (GSS) and UniProt TrEMBL databases. The number of Short Reach Archive (SRA) experiments and EST sequences in nearly equal for each subclass and the remaining databases are primarily populated by elasmobranch data.

including the nurse shark, *Ginglymostoma cirratum*⁴⁵; elephant shark, *C. milli*⁴⁶; little skate, *L. erinacea*⁴⁷; horn shark, *Heterodontus francisci*⁴⁸; dogfish shark, *Squalus acanthias*^{49,50}; and catshark, *S. canicula*⁵¹. These libraries were used to successfully characterize a handful of genomic regions such as little skate *HoxA* cluster^{47,52}, horn shark *Hox A* and *D* clusters⁵³, catshark *HoxA*, *B* and *D* clusters^{51,52}, *C. milii* *HoxA-D* clusters⁵⁴, immunoglobulin receptor IgW C regions³⁰ and neurohypophysial gene loci⁴⁶.

RNA databases

Transcriptome sequencing seeks to characterize all genes expressed in a tissue or set of tissues in a sample. Technologies to identify

the complete RNA transcript sequence have developed from studies of a small number of transcripts to comprehensive characterizations. The application of large-scale cDNA cloning of Expressed Sequence Tags (ESTs) gave initial characterizations of 5-prime and/or 3-prime ends of transcripts in several elasmobranchs including *L. erinacea* and *S. acanthias* (Table 3). EST sequence data are available in the EST divisions of the GenBank, EMBL and DDBJ databases that make up the INSDC. cDNA clones and their sequences from these EST projects have enabled the complete characterization of the full-length cDNA sequence of several genes. In the last five years, high-throughput RNA sequencing (RNA-Seq) has been applied to comprehensively examine the complete

sequence of transcripts in tissues of cartilaginous fishes. Among the most valuable RNA-Seq datasets are those from whole late-stage embryos following organogenesis. Our project has generated these datasets for *L. erinacea*, *S. canicula* and *C. milii*⁵². Public RNA-Seq data sets can be found in the NCBI Gene Expression Omnibus and Short Read Archive (SRA) databases or the EBI ArrayExpress and European Nucleic Acid (ENA) archives (Table 3 and Table 4).

Mitochondrial genomes

Individual mitochondrial genes such as cytochrome c oxidase subunit I (CO1 or COX1) and NADH-ubiquinone oxidoreductase chain 2 (NADH2 or MT-ND2) have been used extensively to construct molecular phylogenies⁵⁵⁻⁵⁷. The Fish barcode of life (FISH-BOL) a working group of the International Barcode of Life Project (iBOL), has CO1 barcodes for 54% of elasmobranchs and 62% of holocephalans (<http://www.fishbol.org>, accessed July 24, 2014). Recently, whole mitochondrial sequences are increasingly popular for their increased granularity when resolving branches of phylogenetic trees¹. Whole mitochondrial genome sequences currently are available for 72 species of sharks, skates, rays and chimaeras. These sequences are accessible in the GenBank, EMBL and DDBJ databases summarized in Table 5⁵⁸.

Chondrichthyan Tree of Life

Currently, molecular data for cartilaginous fishes is being collected as part of the Chondrichthyan Tree of Life project (<http://sharkrays.org>). The project website currently includes 5 elements: 1) an interactive phylogenetic tree⁵⁵; 2) scientific illustrations of specimens; 3) range information for all extant species; 4) interactive comparative anatomy through segmented CT scan data; and 5) DNA sequence for 1265 single copy orthologous genes⁵⁹. Project data will be available in public databases as well as through the project website once collection and analysis is complete (personal communication, Gavin Naylor, Medical University of South Carolina).

Protein databases

Given the improved technologies to characterize full-length transcripts using RNA-Seq, there are increasingly more protein sequence data for chondrichthyans. The UniProt Consortium, consists of groups from the European Bioinformatics Institute (EBI), the Swiss Institute of Bioinformatics (SIB) and the Protein Information Resource (PIR). The consortium maintains the UniProt Knowledgebase (UniProtKB), a comprehensive and standardized catalogue of protein sequences and functional annotation knowledgebase⁶⁰. Proteins with UniProtKB accessions are first automatically annotated, unreviewed UniProtKB/TrEMBL entries that progress to UniProtKB/Swiss-Prot entries following curator review. Among Chondrichthyes, there are 12 UniProtKB/Swiss-Prot and 20,170 UniProtKB/TrEMBL entries for holocephalans and 264 UniProtKB/Swiss-Prot and 6,299 UniProtKB/TrEMBL entries for elasmobranchs in Release 2014_07 of 09-Jul-2014 of the knowledgebase (Table 1). An unidentified fin sample accounts for 16 UniProtKB/TrEMBL entries that are not included in either Holocephali or Elasmobranchii. PDB, an archive of protein macromolecular structural data, has 178 entries for Chondrichthyes, all elasmobranchs⁶¹. Of these, 76% are derived from 2 species from a single family, Torpediniformes, the electric rays, and in total only 10 species are represented in PDB.

The distribution of data in NCBI databases, PDB, and UniProtKB for chondrichthyan orders is shown in Figure 4. When order Chimaeriformes is included (Figure 4A) the distributions are disproportionate due to the large volume of annotated sequence data from the elephant shark genome. The distributions are repeated exclusively for elasmobranchs. To understand if the data distribution is representative of the number of species in each order, a species distribution is included in each chart. A cladogram (Figure 4B) is linked to the chart legend and illustrates the phylogeny between chondrichthyan orders.

SkateBase

SkateBase (<http://skatebase.org>) is the public portal for the little skate genome project and is a valuable collection of data and learning resources. The NEBC little skate genome project team hosted three week-long workshops and a mitochondrial genome annotation jamboree with the goal of using the project data to develop a bioinformatics aware workforce and foster collaborative and distributed big data research. The lecture materials and worked annotation examples are included at SkateBase for educational use (<http://skatebase.org/workshops>). The project vitae contains an overview and timeline of the genome project effort along with key personnel, project related publications and presentations, the curation team, and citation information for researchers utilizing the resource in their publication. A Gene Table currently represents manually curated genes derived from workshops and curriculum with extensive annotation evidence. The number of gene entries will continue to grow through usage and expansion of the SkateBase educational modules. Plans to update the annotation interface to enable community annotation by domain experts is planned for the future.

SkateBase provides links to web resources with chondrichthyan data including the Chondrichthyan Tree of Life, Elephant Shark Genome Project (<http://esharkgenome.imcb.a-star.edu.sg>), the first described genome for a chimaera⁴³, and Vertebrate TimeCapsule, (<http://transcriptome.cdb.riken.go.jp/vtcap>), a project that aims to develop a gene database to represent evolution and development for vertebrates and currently includes transcriptome data for a hagfish (*Eptatretus burger*), shark (*S. torazame*) and birchir (*Polypterus senegalus*)⁶². SkateBase data is linked locally as well as from NCBI in the Gene Expression Omnibus (GSE26235), GenBank (AESE01000000) and Sequence Read Archive (SRA026856) to ensure convenient and easy access. A link to the American Elasmobranch Society (<http://www.elasmo.org>), a non-profit organization with the mission of advancing the scientific study of living and fossil sharks, skates, rays, and chimaeras and promoting education, conservation, and wise utilization of natural resources, connects domain scientists to the little skate genome project.

SkateBase data includes embryonic transcriptomes for three chondrichthyan species, a chimaera, *C. milii*, a shark, *S. canicula* and the little skate, *L. erinacea* as well as the first draft of the little skate genome. The assembled skate genome sequence gave a single high-coverage contiguous sequence that represented the entire length of the mitochondrial genome. The mitochondrial genome was subsequently annotated as part of a Jamboree in 2011⁶³. The annotated sequence is represented by the NCBI Reference Sequence (RefSeq) project, accession NC_016429, and provides extensive information for each gene.

Table 3. National Center for Biotechnology Information (NCBI) Expressed Sequence Tags (EST) and Genome Survey Sequences (GSS) databases (release 130101): Chondrichthyan sequence data.

BioSample	BioSample Description	Library ID	Organism	Sample age/sex	Sample type	ESTs	Facility ¹	Date
Holocephali								
Chimaeriformes								
182978	Whole-genome shotgun library of the elephant shark (aka elephant fish)	GSS: LIBGSS_009694	<i>Callorhynchus milii</i>	-	testis	27944	IMCB	2004
1000678	Elephant shark full-length cDNA library from testis	EST: LIBEST_027873	<i>Callorhynchus milii</i>	-	testis	29234	IMCB	2012
1000677	Elephant shark full-length cDNA library from spleen	EST: LIBEST_027872	<i>Callorhynchus milii</i>	-	spleen	16664	IMCB	2012
1000676	Elephant shark full-length cDNA library from liver	EST: LIBEST_027871	<i>Callorhynchus milii</i>	-	liver	16573	IMCB	2012
1000675	Elephant shark full-length cDNA library from kidney	EST: LIBEST_027870	<i>Callorhynchus milii</i>	-	kidney	19246	IMCB	2012
1000674	Elephant shark full-length cDNA library from intestine	EST: LIBEST_027869	<i>Callorhynchus milii</i>	-	intestine	12146	IMCB	2012
1000673	Elephant shark full-length cDNA library from gills	EST: LIBEST_027868	<i>Callorhynchus milii</i>	-	gills	16012	IMCB	2012
Elasmobranchii: Batoids (rays and skates)								
Torpediformes								
158311	Torpedo marmorata electric organ	EST: LIBEST_003755	<i>Torpedo marmorata</i>	-	electric organ	8	CNRS	2000
158310	Torpedo marmorata electric lobe	EST: LIBEST_003754	<i>Torpedo marmorata</i>	-	electric lobe	26	CNRS	2000
157461	pFL61-TEL	EST: LIBEST_002905	<i>Torpedo marmorata</i>	-	electric lobe	2	CNRS	2000
157406	pFL61-EL	EST: LIBEST_002849	<i>Torpedo marmorata</i>	-	electric lobe	5	CNRS	2000
154382	Torpedo californica electric organ	EST: LIBEST_020696	<i>Torpedo californica</i>	-	electric organ	10185	Children's National Medical Center, USA	2006
Rajiformes								
175126	Little Skate Multiple Tissues, Normalized	EST: LIBEST_015890	<i>Leucoraja erinacea</i>	adult	mixed ^a	5698	MDIBL	2004
176484	Little Skate Liver, Normalized	EST: LIBEST_017626	<i>Leucoraja erinacea</i>	adult	liver	6016	MDIBL	2005
165533	Little Skate embryo cell line 1 (LEE-1): 5' sequences	EST: LIBEST_022984	<i>Leucoraja erinacea</i>	embryonic cell line	stage 28	4825	MDIBL	2006
154366	Little skate embryo tissues: 5' sequences	EST: LIBEST_020422	<i>Leucoraja erinacea</i>	embryo	stage 19, 20, 25	5600	MDIBL	2006
166469	Skate Multiple Tissues, Normalized	EST: LIBEST_023576	<i>Leucoraja erinacea</i>	adult	mixed ^a	9028	MDIBL	2008
Elasmobranchii: Selachii (sharks)								
Carcharhiniformes								
168576	Dogfish testis - round spermatids zone (SSH)	EST: LIBEST_025578	<i>Scyliorhinus canicula</i>	adult	testis	20	Caen University	2009
168575	Dogfish testis - spermatogonia zone (SSH)	EST: LIBEST_025577	<i>Scyliorhinus canicula</i>	adult	testis	12	Caen University	2010
222714	Scyliorhinus canicula juvenile library	EST: LIBEST_026904	<i>Scyliorhinus canicula</i>	juvenile	5 days post-hatch	56	enoscope-CEA	2011
222713	Scyliorhinus canicula embryonic, stages 9–15 library	EST: LIBEST_026903	<i>Scyliorhinus canicula</i>	embryo	stages 9–15	628	Genoscope-CEA	2011
222712	Scyliorhinus canicula embryonic, stages 19–25 library	EST: LIBEST_026902	<i>Scyliorhinus canicula</i>	embryo	stages 19–25	772	Genoscope-CEA	2011
222711	Scyliorhinus canicula embryonic, stages 19–24 library	EST: LIBEST_026901	<i>Scyliorhinus canicula</i>	embryo	stages 19–24	33	Genoscope-CEA	2011
222710	Scyliorhinus canicula adult brain library	EST: LIBEST_026900	<i>Scyliorhinus canicula</i>	adult	brain	79	Genoscope-CEA	2011

BioSample	BioSample Description	Library ID	Organism	Sample age/sex	Sample type	ESTs	Facility ¹	Date
699400	cloudy catshark embryo cDNA library	EST: LIBEST_027410	<i>Scyliorhinus torazame</i>	embryo	stage 31	2942	RIKEN	2011
Orectolobiformes								
183175	GC_Ba	GSS: LIBGSS_009945	<i>Ginglymostoma cirratum</i>	adult	red blood cells	178	University of Arizona	2005
184343	shark whole genome shotgun library 2	GSS: LIBGSS_011249	<i>Chiloscyllium plagiosum</i>	female	ventral fin	177	Tgen	2008
184342	shark whole genome shotgun library 1	GSS: LIBGSS_011248	<i>Chiloscyllium plagiosum</i>	female	ventral fin	194	Tgen	2008
166749	Shark liver regeneration	EST: LIBEST_023789	<i>Chiloscyllium plagiosum</i>	adult	liver	2103	BGI	2008
176026	cDNA library of Shark hepatic regeneration tissues	EST: LIBEST_017019	<i>Chiloscyllium plagiosum</i>	none	Hour 24 after 2/3 partial hepatectomy	17	CPU	2005
254067	Toll like receptor ligand induced Spleen	EST: LIBEST_027180	<i>Chiloscyllium griseum</i>	male	spleen	1051	MVC	2011
254066	Spleen of <i>Chiloscyllium griseum</i>	EST: LIBEST_027179	<i>Chiloscyllium griseum</i>	male	spleen	1000	MVC	2011
1797282	Suppressive subtractive hybridization library from peptidoglycan induced spleen of the shark	EST: LIBEST_028031	<i>Chiloscyllium griseum</i>	male	spleen	315	MVC	2012
Squaliformes								
175664	Dogfish Shark Multiple Tissues, Normalized	EST: LIBEST_016552	<i>Squalus acanthias</i>	adult	mixed ^b	15078	MDIBL	2004
176998	Dogfish Shark Embryo-derived Cell Line SAE, Normalized	EST: LIBEST_018195	<i>Squalus acanthias</i>	embryonic cell line	embryo with external yolk sac	5824	MDIBL	2005
154362	Spiny dogfish shark rectal gland EST library	EST: LIBEST_020417	<i>Squalus acanthias</i>	-	rectal gland	5085	MDIBL	2006
150616	Dogfish Shark Rectal Gland, Normalized	EST: LIBEST_020023	<i>Squalus acanthias</i>	adult	rectal gland	6575	MDIBL	2006
Hexanchiformes								
178140	Hexanchus griseus DNA (Hunter C)	GSS: LIBGSS_003277	<i>Hexanchus griseus</i>	-	-	4	HGMP-RC	2001

¹ (ICMB) Institute of Molecular and Cell Biology, A*STAR, (HGMP-RC) Human Genome Mapping Project Resource Centre, Hinxton, (Tgen) Translational Genomics Research Institute AZ, USA, (CNRS) National Center for Scientific Research, France, (MDIBL) Mount Desert Island Biological Laboratory, (CPU) China Pharmaceutical University, (MVC) Madras Veterinary College, TANJUVAS, (BGI) Beijing Genomics Institute (SSH) Suppressive subtractive hybridization; (mixed^a) liver, kidney, brain, testis, ovary, gill, heart, spleen, rectal gland; (mixed^b) rectal gland, kidney, brain, testis, ovary, gill, intestine, heart, spleen

Table 4. National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) database: Chondrichthyan sequence data.

BioProject	BioSample	SRA description	SRA	Organism	Age	Sample type	Platform ¹	Data ²	Facility ³	Date
Holocephali										
PRJNA18361	SAMIN00000800	454 sequencing of <i>Callorhynchus milii</i> genomic fragment library	SRX001870*	<i>Callorhynchus milii</i>	adult	testis	LS454	244.9 M	IMCB	2008
PRJNA135005	SAMIN00188955	GSM643959: <i>Callorhynchus milii</i> pooled Stage 32 embryos	SRX036538	<i>Callorhynchus milii</i>	embryos	stage 32	Illumina SE	3.3 G	MDIBL	2011
PRJNA168475	SAMIN02699939	Illumina sequencing of elephant shark thymus RNA	SRX220387	<i>Callorhynchus milii</i>	-	thymus	Illumina PE	9.7 G	IMCB	2013

BioProject	BioSample	SRA description	SRA	Organism	Age	Sample type	Platform ¹	Data ²	Facility ³	Date
PRJNA168475	SAMN02699938	Illumina sequencing of elephant shark testis RNA	SRX154861	<i>Callorhinchus milii</i>	-	testis	Illumina PE	7.3 G	IMCB	2013
PRJNA168475	SAMN02699937	Illumina sequencing of elephant shark spleen RNA	SRX154860	<i>Callorhinchus milii</i>	-	spleen	Illumina PE	6.3 G	IMCB	2013
PRJNA168475	SAMN02699936	Illumina sequencing of elephant shark ovary RNA	SRX154859	<i>Callorhinchus milii</i>	-	ovary	Illumina PE	7.9 G	IMCB	2013
PRJNA168475	SAMN02699935	Illumina sequencing of elephant shark liver RNA	SRX154858	<i>Callorhinchus milii</i>	-	liver	Illumina PE	16.7 G	IMCB	2013
PRJNA168475	SAMN02699934	Illumina sequencing of elephant shark muscle RNA	SRX154857	<i>Callorhinchus milii</i>	-	muscle	Illumina PE	11.1 G	IMCB	2013
PRJNA168475	SAMN02699933	Illumina sequencing of elephant shark kidney RNA	SRX154856	<i>Callorhinchus milii</i>	-	kidney	Illumina PE	9 G	IMCB	2013
PRJNA168475	SAMN02699932	Illumina sequencing of elephant shark intestine RNA	SRX154855	<i>Callorhinchus milii</i>	-	intestine	Illumina PE	11.2 G	IMCB	2013
PRJNA168475	SAMN02699931	Illumina sequencing of elephant shark heart RNA	SRX154854	<i>Callorhinchus milii</i>	-	heart	Illumina PE	6.9 G	IMCB	2013
PRJNA168475	SAMN02699930	Illumina sequencing of elephant shark gills RNA	SRX154852	<i>Callorhinchus milii</i>	-	gills	Illumina PE	5.4 G	IMCB	2013
PRJNA168475	SAMN02699929	Illumina sequencing of elephant shark brain RNA	SRX154851	<i>Callorhinchus milii</i>	-	brain	Illumina PE	10.5 G	IMCB	2013
Elasmobranchii										
PRJNA60893	SAMN00189066	Initial Characterization of Leucoraja erinacea Genome Using 500bp Paired-End Sequencing	SRX036634*	<i>Leucoraja erinacea</i>	embryo	stage 32	Illumina PE	105 G	NECC	2011
PRJNA135005	SAMN00188953	GSM643957: Leucoraja erinacea pooled Stage 20–29 embryos	SRX036536	<i>Leucoraja erinacea</i>	embryos	stage 20–29	Illumina SE	3.8 G	MDIBL	2011
PRJNA135005	SAMN00188954	GSM643958: Scyliorhinus canicula pooled Stage 24–30 embryos	SRX036537	<i>Scyliorhinus canicula</i>	embryos	stage 24–30	Illumina SE	3.9 G	MDIBL	2011
PRJDA61447	SAMD00003843	Torazame EST	DRX0000491	<i>Scyliorhinus torazame</i>	embryos	stage 23–31	LS454	43.6 M	RIKEN	2011
PRJNA177971	SAMN01915239	<i>Carcharodon carcharias</i> cDNA Illumina sequence reads	SRX228421	<i>Carcharodon carcharias</i>	juvenile	heart	Illumina SE	7.9 G	Cornell	2013
PRJNA177971	SAMN01915239	<i>Carcharodon carcharias</i> heart transcriptome	SRX228332	<i>Carcharodon carcharias</i>	juvenile	heart	LS454	408.4 M	Cornell	2013
PRJNA183979	SAMN01831510	Illumina sequencing of Nurse Shark thymus transcripts	SRX219866	<i>Ginglymostoma cirratum</i>	-	thymus	Illumina PE	12 G	IMCB	2013
PRJNA183979	SAMN01831509	Illumina sequencing of Nurse Shark spleen transcripts	SRX219865	<i>Ginglymostoma cirratum</i>	-	spleen	Illumina PE	11.2 G	IMCB	2013
PRJNA240112	SAMN02673223	<i>Neotrygon kuhlii</i> barb venom gland transcriptome	SRX481088	<i>Neotrygon kuhlii</i>	-	barb venom gland	Illumina PE	84.3 M	LSTM	2014

* genomic data; ¹ (SE) single end or (PE) paired end; ² (M) Mega or (G) Giga base pairs

³ (MDIBL) Mount Desert Island Biological Laboratory, (ICMB) Institute of Molecular and Cell Biology, A*STAR, (LSTM) Liverpool School of Tropical Medicine, (NECC) North East Cyberinfrastructure Consortium

Table 5. Whole mitochondrial sequences for chondrichthyan fishes.

Accessions						
BioProject	NCBI Ref_seq	GenBank	Organism	bp	*G+C	Date
<u>Holocephali</u>						
Chimaeriformes						
PRJNA50265	NC_014281.1	HM147135.1	<i>Callorhynchus callorynchus</i>	16758	34	21-Oct-10
PRJNA50271	NC_014284.1	HM147136.1	<i>Callorhynchus capensis</i>	16760	34.1	21-Oct-10
PRJNA50273	NC_014285.1	HM147137.1	<i>Callorhynchus milii</i>	16769	33.7	21-Oct-10
PRJNA11978	NC_003136.1	AJ310140.1	<i>Chimaera monstrosa</i>	18580	38.6	14-Nov-06
PRJNA50279	NC_014288.1	HM147138	<i>Chimaera fulva</i>	21336	38.2	19-Oct-10
PRJNA50287	NC_014292.1	HM147140.1	<i>Harriotta raleighana</i>	18024	42.5	19-Oct-10
PRJNA50283	NC_014290.1	HM147139.1	<i>Hydrolagus lemures</i>	21233	39.4	19-Oct-10
PRJNA50289	NC_014293.1	HM147141.1	<i>Rhinochimaera pacifica</i>	24889	41.6	19-Oct-10
<u>Elasmobranchii: Batoids (rays and skates)</u>						
Myliobatiformes						
PRJNA247653	NC_024102.1	KJ617038.1	<i>Gymnura poecilura</i>	17874	45.1	7-May-14
PRJNA239601	NC_023525.1	KF751650.1	<i>Himantura granulata</i>	17657	39.1	25-Feb-14
PRJNA229016	NC_022837.1	KF482070.1	<i>Aetobatus flagellum</i>	20201	40.9	3-Nov-13
PRJNA198706	NC_021132.1	KC526959.1	<i>Dasyatis akajei</i>	17658	40.4	10-Mar-14
PRJNA190131	NC_020352.2	KC196067.2 KC633222.1	<i>Dasyatis bennetti</i> <i>Dasyatis bennetti</i>	17668 17717	40.2 40.1	22-Jul-13 20-Feb-14
PRJNA182669	NC_019643.1	JX524174.1	<i>Dasyatis zugei</i>	18264	36.6	24-May-13
PRJNA15549	NC_007230.1	AY597334.1	<i>Plesiobatis daviesi</i>	17514	41.9	20-Mar-07
PRJNA232219	NC_023116	KF709642.1	<i>Potamotrygon motoro</i>	17448	43.3	14-Jan-14
PRJNA177278	NC_018784.1	JX392983.1	<i>Mobula japanica</i>	18880	37.4	18-Jan-13
PRJNA212605	NC_021767.1	KC992792.1	<i>Neotrygon kuhlii</i>	18039	39.5	17-Jul-13
PRJNA182647	NC_019641.1	JX827260.1	<i>Taeniura meyeni</i>	17638	41.6	8-Nov-13
Rajiformes						
PRJNA239623	NC_023505.2	KF318309.2	<i>Dipturus kwangtungensis</i>	16912	41.6	13-Mar-14
PRJNA81399	NC_016429.1	JQ034406.1	<i>Leucoraja erinacea</i>	16724	40.3	28-Nov-11
PRJNA13984	NC_007173.1	AY525783.1	<i>Okamejei kenojei</i>	16972	42.4	15-Jun-05
PRJNA11877	NC_000893.1	AF106038.1	<i>Amblyraja radiata</i>	16783	40.3	22-Apr-09
PRJNA214406	NC_021964	KC914434.1	<i>Raja rhina</i>	16910	41.4	11-Sep-13
PRJNA214407	NC_021963.1	KC914433.1	<i>Hongoeo koreana</i>	16905	42.2	11-Sep-13
PRJNA244226	NC_023944.1	KF648508.1	<i>Zearaja chilensis</i>	16909	41.1	1-May-14
Rhinopristiformes						
PRJNA228994	NC_022821.1	KF381507.1	<i>Pristis clavata</i>	16804	39.8	13-Nov-13
PRJNA229000	NC_022841.1	KF534708.1	<i>Rhinobatos hynnicephalus</i>	16776	40.3	13-Nov-13
PRJNA244205	NC_023951.1	KJ140136.1	<i>Rhinobatos schlegelii</i>	16780	39.6	6-Apr-14
<u>Elasmonbranchii: Selachii (sharks)</u>						
Carcharhiniformes						
PRJNA246074	NC_024055.1	KF728380.1	<i>Carcharhinus acronotus</i>	16719	38.4	29-Apr-14
PRJNA244183	NC_023948.1	KF956523.1	<i>Carcharhinus amblyrhynchoides</i>	16705	38.2	6-Apr-14
PRJNA239607	NC_023522.1	KF646785.1	<i>Carcharhinus leucas</i>	16704	37.4	25-Feb-14
PRJNA252486	NC_024284.1	KJ720818.1	<i>Carcharhinus melanopterus</i>	16706	38.6	7-Jun-14
PRJNA193929	NC_020611.1	KC470543.1	<i>Carcharhinus obscurus</i>	16706	38.6	8-Nov-13
PRJNA239626	NC_023521.1	KF612341.1	<i>Carcharhinus sorrah</i>	16707	38.9	25-Feb-14

Accessions						
BioProject	NCBI Ref_seq	GenBank	Organism	bp	*G+C	Date
PRJNA217222	NC_022193.1	KF111728.1	<i>Galeocerdo cuvier</i>	16703	36.9	31-Oct-13
PRJNA236275	NC_023361.1	KF646786.1	<i>Glyphis garricki</i>	16702	39.2	13-Jan-14
PRJNA212606	NC_021768.2	KF006312.2	<i>Glyphis glyphis</i>	16701	39	25-Jul-14
PRJNA239588	NC_023527.1	KF889325.1	<i>Mustelus griseus</i>	16754	39	25-Feb-14
PRJNA11875	NC_000890.1	AB015962.1	<i>Mustelus manazo</i>	16707	38.3	8-Apr-00
PRJNA228986	NC_022819.1	KF356249.1	<i>Prionace glauca</i>	16705	37.5	13-Nov-13
PRJNA226181	NC_022735.1	AB560493.1	<i>Pseudotriakis microdon</i>	16700	36.4	29-Oct-13
PRJNA168394	NC_018052.1	JQ693102.1	<i>Scoliodon macrorhynchus</i>	16693	37	31-Mar-14
PRJNA11849	NC_001950.1	Y16067.1	<i>Scyliorhinus canicula</i>	16697	38	18-Apr-05
PRJNA226138	NC_022679.1	JX827259.1	<i>Sphyrna lewini</i>	16726	39.5	8-Nov-13
Orectolobiformes						
PRJNA163947	NC_017882.1	JQ434458.1	<i>Chiloscyllium griseum</i>	16755	36.1	6-Mar-12
PRJNA37667	NC_012570.1	JX162601.1	<i>Chiloscyllium plagiosum</i>	16725	37.4	25-Jul-12
PRJNA81281	NC_016686.1	JQ082337.1	<i>Chiloscyllium punctatum</i>	16703	36.8	31-Mar-14
PRJNA217221	NC_022148.1	KF111729.1	<i>Orectolobus japonicus</i>	16706	37.3	19-Sep-13
PRJNA238093	NC_023455.1	KF679782.1 KC633221	<i>Rhincodon typus</i> <i>Rhincodon typus</i>	16875 16928	37.1 37.1	19-Mar-14 31-Mar-14
Lamniformes						
PRJNA239610	NC_023520.1	KF569943.1	<i>Carcharias taurus</i>	16773	39.5	5-Feb-14
PRJNA221185	NC_022415.1	KC914387.1	<i>Carcharodon carcharias</i>	16744	40.8	31-Oct-13
PRJNA232870	NC_023266.1	KF597303.1	<i>Cetorhinus maximus</i>	16670	40.6	14-Jan-14
PRJNA226140	NC_022691.1	KF361861.1	<i>Isurus oxyrinchus</i>	16701	43.2	28-Sep-13
PRJNA247657	NC_024101.1	KJ616742.1	<i>Isurus paucus</i>	16704	43.8	7-May-14
PRJNA252473	NC_024269.1	KF962053.1	<i>Lamna ditropis</i>	16699	41.8	30-May-14
PRJNA207613	NC_021442.1	KC702506.1	<i>Megachasma pelagios</i>	16694	36.7	13-May-13
PRJNA33525	NC_011825.1	EU528659.1	<i>Mitsukurina owstoni</i>	17743	38.8	29-Dec-08
PRJNA228992	NC_022822.1	KF412639.1	<i>Alopias pelagicus</i>	16692	38.6	18-Dec-13
PRJNA207614	NC_021443.1	KC757415.1	<i>Alopias superciliosus</i>	16719	39.3	26-Jun-13
Heterodontiformes						
PRJNA11979	NC_003137.1	AJ310141.1	<i>Heterodontus francisci</i>	16708	39.9	14-Nov-06
PRJNA209901	NC_021615.1	KC845548.1	<i>Heterodontus zebra</i>	16720	40	18-Jun-13
Squaliformes						
PRJNA246067	NC_024059.1	KJ128289.1	<i>Cirrhigaleus australis</i>	16543	38.8	29-Apr-14
PRJNA226141	NC_022734.1	AB560492.1	<i>Somniosus pacificus</i>	16730	39.3	29-Oct-13
PRJNA11856	NC_002012.1	Y18134.1	<i>Squalus acanthias</i>	16738	38.8	18-Apr-05
Squatiniiformes						
PRJNA252467	NC_024276.1	KJ619663.1	<i>Squatina japonica</i>	16689	37.9	4-Jun-14
Pristiophoriformes						
PRJNA247682	NC_024110.1	AB721306.1	<i>Pristiophorus japonicus</i>	18430	44.5	10-May-14
Hexanchiformes						
PRJNA226134	NC_022732.1	AB560490.1	<i>Hexanchus griseus</i>	17223	36.3	29-Oct-13
PRJNA226149	NC_022733.1	AB560491.1	<i>Hexanchus nakamurai</i>	18605	36.3	29-Oct-13
PRJNA226155	NC_022730.1	AB560488.1	<i>Heptranchias perlo</i>	18909	35.9	29-Oct-13
PRJNA226147	NC_022729.1	AB560487.1	<i>Chlamydoselachus anguineus</i>	17314	35	29-Oct-13
PRJNA226123	NC_022731.1	AB560489.1	<i>Notorynchus cepedianus</i>	16990	38.2	29-Oct-13

*Metazoan Mitochondrial Genomes Accessible dataset Metamiga (<http://amiga.cbmeg.unicamp.br/>)

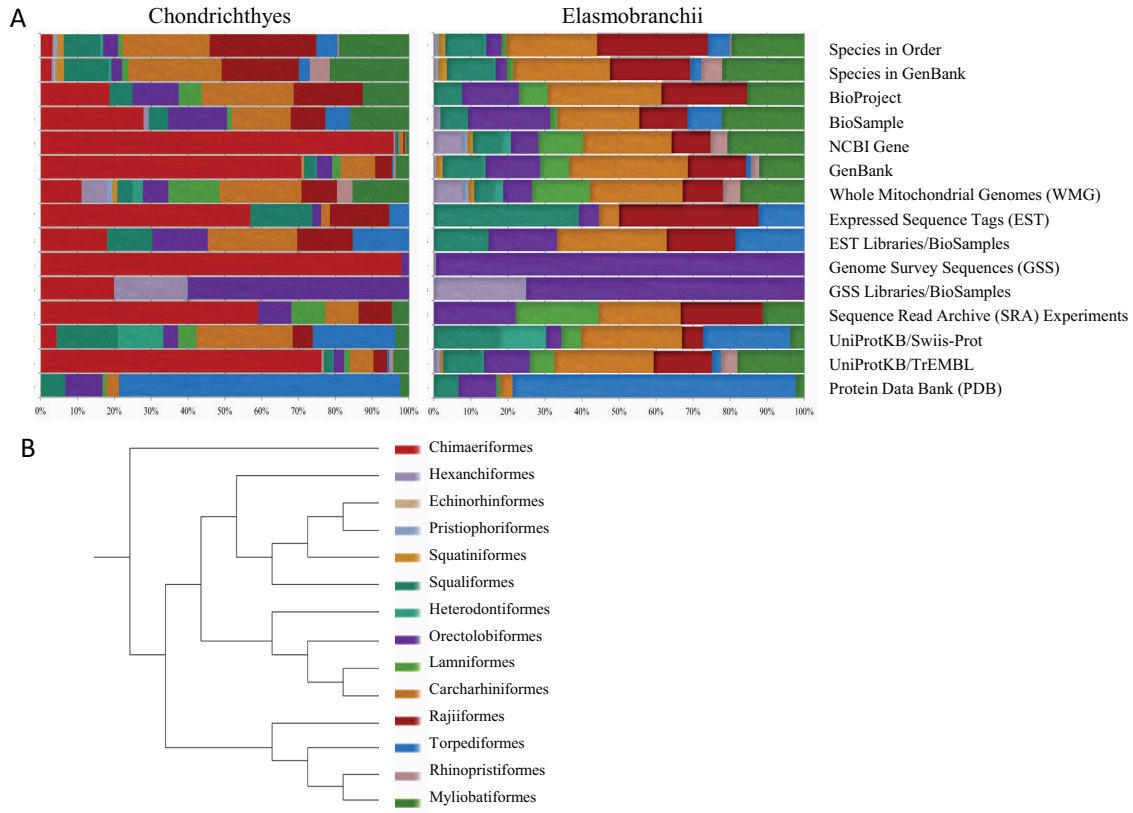


Figure 4. A survey of public data and phylogeny for chondrichthyan orders. A. The 14 orders of chondrichthyan fish and their relative distribution in public nucleotide and protein databases for Chondrichthyes and Elasmobranchii are shown individually. The species distribution for each Order and GenBank are similar indicating sequence data has been collected for a broad range of chondrichthyans. For Chondrichthyes, the elephant shark genome project data contributes the majority of the data in NCBI Gene, GenBank, Genome Survey Sequence (GSS), and the Short Reach Archive (SRA) databases. The NCBI GSS, GSS libraries, and Protein Data Bank (PDB) are the least diverse with representation of 1–6 of the 14 Orders. The color of each Order as represented in the bar chart is included in the cladogram key with left to right in the bar chart corresponding with top to bottom in the cladogram. **B.** A cladogram of Chondrichthyes illustrates the phylogeny relationship between the 14 Orders. The color code associated with each Order appears consecutively in the bar chart.

Whole embryos were used to build the transcriptome libraries available at SkateBase³⁵. Two *C. milii* embryos, stage 32, were combined and used to build a chimaera library. The transcriptome library for *S. canicula* was assembled from six pooled embryos, stages 24–30. The embryonic skate transcriptome library was assembled using six pooled embryos ranging in stage from 20–29. This combination of stages encompasses a large portion of the developmental period for these fishes and represents a catalog of genes important for organogenesis of all or part of every physiological system. Early developmental events are similar for nearly all elasmobranchs regardless of reproductive mode or adult body form enabling the data to be useful for more than just the specific species from where it was derived⁶⁴. Since all three embryonic transcriptomes contain a similar stage embryo direct comparison for temporal expression patterns is possible. Skatebase includes tools for data investigation, SkateBLAST, a sequence retrieval tool, Skate Contig Lookup, and genome browsers for three skate whole mitochondrial sequences, *L. erinacea*, the thorny skate, *Amblyraja radiata*, and, the ocellate spot skate, *Okamejei kenoei*. Skatebase contains resources that can be used for teaching and research purposes. As an example, two use cases follow, one for sequence or homology based research and the other for education.

SkateBLAST

A common task for researchers is searching for genes of interest in a genome or transcriptome. Knowledge of the gene sequence at the DNA or RNA level is needed for many different types of studies, including phylogenetic analysis or designing primers for quantitative PCR gene expression studies. Here we describe the major steps necessary to identify relevant sequences for a gene of interest using the BLAST sequence similarity tool at SkateBase. SkateBase features a web interface to BLAST, named SkateBLAST, that builds upon the ViroBLAST package version 2.2⁶⁵, with custom modifications allowing parallel cluster-based execution of queries and enhanced display of results. The overall workflow consists of a) entering a query sequence and selecting the database to search; b) evaluating the alignments returned; c) retrieving the sequence from one of the SkateBLAST databases; and d) checking to make sure that the retrieved sequence aligns best to the query sequence. The following description provides a brief tutorial on the overall workflow while describing tools at SkateBase.

Figure 5 demonstrates the use of SkateBLAST to find expressed sequences for the gene, suppressor of cytokine signaling 6 (*SOCS6*). *SOCS6* is a E3 ubiquitin ligase that interacts with c-KIT to suppress

A

SkateBase

North East Cyberinfrastructure Consortium

Home Workshops Tools Downloads & Links Project Vitae Gene Table Curator Access

SkateBLAST

BLAST Basic Setup - Enter required parameters below

BLAST Algorithm [?]
tblastn - Translated Nucleotide Database

Query Sequence
Enter query sequences here in Fasta format

Subject Database [?]
Genome Assemblies
Little Skate Genomic Contigs Build 2
Transcriptome Assemblies
Little Skate Transcriptome Contigs Build 2
Small-Spotted Catshark Transcriptome Contigs
Elephant Shark Transcriptome Contigs
Little Skate Mitochondrial Genome
Thorny Skate Mitochondrial Genome
Ocellate Spot Skate Mitochondrial Genome
Mitochondrial Genes
Little Skate Mitochondrial Features
Thorny Skate Mitochondrial Features
Ocellate Spot Skate Mitochondrial Features

Advanced Options - Set optional BLAST+ parameters below

Expect threshold: 10
Word size: 11
Max target sequences: 50
Match/Mismatch scores: Z-3
Gap costs: Existence: 5, Extension: 2

Filter: Low complexity regions
Mask: Mask for lookup table only
 Mask for lower case letters
 Perform ungapped alignment

Alignment output format: pairwise

BLAST Reset

Adapted from ViroBLAST © 2005-2010 University of Washington. All rights reserved. Terms of Service

B

SkateBase

North East Cyberinfrastructure Consortium

Home Workshops Tools Downloads & Links Project Vitae Gene Table Curator Access

SkateBLAST

Raw BLAST Results

Filter BLAST results by bit score rank:
Show: All for each query sequence

Filter BLAST results by cutoff criterion:
Similarity Cutoff: 95 %
Bit Score Cutoff: 100

Download BLAST Results

Examine and Download Full BLAST Report
Raw BLAST Output Report

Retrieve and download subject sequences in FASTA format:
Check here to download all sequences...
OR select particular sequences of interest below

Query	Subject	Bit Score	Identity (Query Lem)	Similarity	E-value
sp O14544 SOCS6_HUMAN	gn SkateBase S-transcriptB2-ctg15542	755	403/539 (53%)	75	0.0
sp O14544 SOCS6_HUMAN	gn SkateBase S-transcriptB2-ctg17287	140	90/167 (53%)	54	8e-36
sp O14544 SOCS6_HUMAN	gn SkateBase S-transcriptB2-ctg11052	61.6	67/194 (33%)	35	5e-16
sp O14544 SOCS6_HUMAN	gn SkateBase S-transcriptB2-ctg21448	55.8	39/112 (33%)	35	5e-08
sp O14544 SOCS6_HUMAN	gn SkateBase S-transcriptB2-ctg31771	21.25	60/212 (33%)	28	1e-05
sp O14544 SOCS6_HUMAN	gn SkateBase S-transcriptB2-ctg12549	47.4	33/109 (33%)	30	1e-05
sp O14544 SOCS6_HUMAN	gn SkateBase S-transcriptB2-ctg12890	48.5	38/118 (33%)	32	1e-05
sp O14544 SOCS6_HUMAN	gn SkateBase S-transcriptB2-ctg94112	43.1	17/37 (33%)	46	2e-04
sp O14544 SOCS6_HUMAN	gn SkateBase S-transcriptB2-ctg10822	77.54	37/115 (33%)	32	2e-04
sp O14544 SOCS6_HUMAN	gn SkateBase S-transcriptB2-ctg82536	35.0	22/78 (33%)	28	0.033
sp O14544 SOCS6_HUMAN	gn SkateBase S-transcriptB2-ctg5563	35.0	17/36 (33%)	47	0.042
sp O14544 SOCS6_HUMAN	gn SkateBase S-transcriptB2-ctg17412	7.29	27/80 (33%)	34	0.10
sp O14544 SOCS6_HUMAN	gn SkateBase S-transcriptB2-ctg10554	9.07	14/35 (33%)	40	0.17
sp O14544 SOCS6_HUMAN	gn SkateBase S-transcriptB2-ctg1423	83.02	24/58 (33%)	41	0.28
sp O14544 SOCS6_HUMAN	gn SkateBase S-transcriptB2-ctg13513	28.56	18/64 (33%)	28	0.39
sp O14544 SOCS6_HUMAN	gn SkateBase S-transcriptB2-ctg14232	28.06	23/60 (33%)	38	0.41
sp O14544 SOCS6_HUMAN	gn SkateBase S-transcriptB2-ctg1414	13.34	35/100 (33%)	35	0.48
sp O14544 SOCS6_HUMAN	gn SkateBase S-transcriptB2-ctg4053	19.66	16/32 (33%)	50	0.49
sp O14544 SOCS6_HUMAN	gn SkateBase S-transcriptB2-ctg10415	3.53	25/91 (33%)	27	0.54
sp O14544 SOCS6_HUMAN	gn SkateBase S-transcriptB2-ctg14320	90.67	25/77 (33%)	32	0.64
sp O14544 SOCS6_HUMAN	gn SkateBase S-transcriptB2-ctg13733	41.45	19/46 (33%)	41	1.6
sp O14544 SOCS6_HUMAN	gn SkateBase S-transcriptB2-ctg99397	19.98	14/28 (33%)	50	2.1
sp O14544 SOCS6_HUMAN	gn SkateBase S-transcriptB2-ctg1550	24.12	26/96 (33%)	25	2.1
sp O14544 SOCS6_HUMAN	gn SkateBase S-transcriptB2-ctg32605	222.88	20/48 (33%)	42	2.7
sp O14544 SOCS6_HUMAN	gn SkateBase S-transcriptB2-ctg5087	24.98	15/38 (33%)	39	3.0
sp O14544 SOCS6_HUMAN	gn SkateBase S-transcriptB2-ctg12218	28.9	25/78 (33%)	32	3.9
sp O14544 SOCS6_HUMAN	gn SkateBase S-transcriptB2-ctg12654	18.99	20/68 (33%)	29	9.5
sp O14544 SOCS6_HUMAN	gn SkateBase S-transcriptB2-ctg7762	24.63	14/34 (33%)	41	9.8

Adapted from ViroBLAST © 2005-2010 University of Washington. All rights reserved. Terms of Service

MDIB | UNIVERSITY OF DELAWARE | Rhode Island | Vermont | National Science Foundation | MAINE

Funding provided by:
NIH National Institute of General Medical Sciences NIGMS Program
National Science Foundation EPSCoR Program

SkateBase is hosted and maintained by the University of Delaware
Center for Bioinformatics and Computational Biology
contact: info@skatebase.org

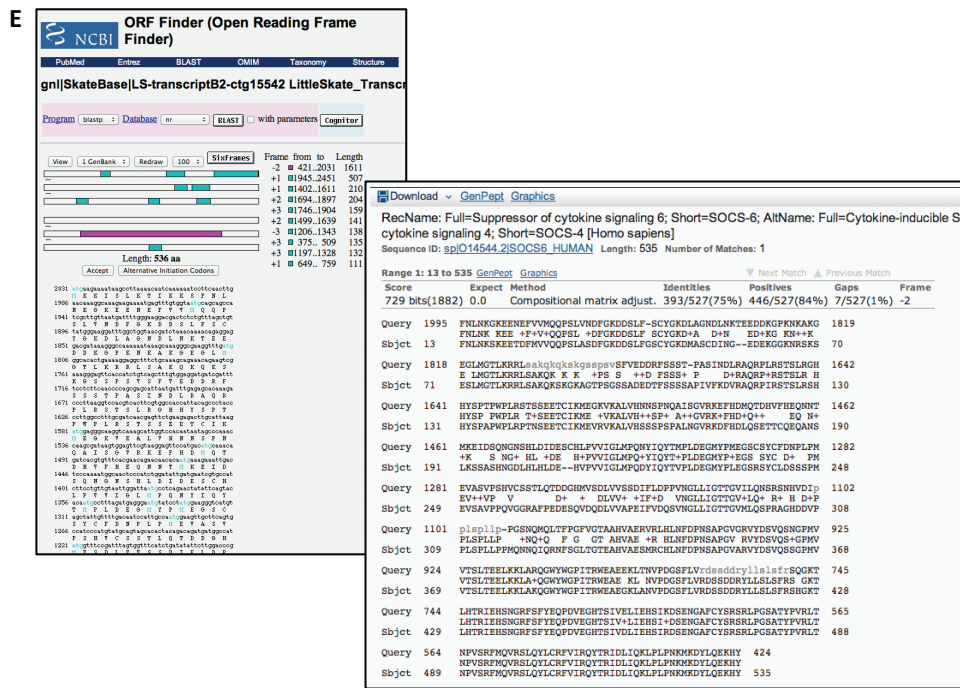


Figure 5. Example of using SkateBase and NCBI resources to find transcriptome data for SOCS6. **A.** SkateBLAST query form showing the four steps to align the UniProt sequence for human SOCS6 (O14544) against the skate embryonic transcriptome using tblastn. Step 1 is to enter the sequence in FASTA format. The second step is to choose the tblastn program that will align the query protein sequence against translated sequences in all six possible reading frames. The third step is to select the embryonic transcriptome as the sequence database to search. The fourth step is to launch the search. **B.** The complete BLAST output can be accessed by clicking the "Inspect BLAST output" link at the top of the summary report page. This is necessary to examine the sequence alignments. **C.** Four important fields in the output should be examined carefully to interpret the alignments and determine which returned alignment best represents the skate ortholog to SOCS6. First, the alignment score, E-value, alignment length and percent identity can be used to interpret the overall alignment significance. Alignment coverage with respect to the query protein sequence and the subject transcriptome sequence can be interpreted by comparing the alignment coordinates to the length of the query protein sequence and length of the transcriptome sequence. In this example, the entire query protein sequence is covered by this transcriptome sequence. **D.** The SkateBase Contig Lookup tool can be used to retrieve the transcriptome sequence found in the SOCS6 tblastn search in FASTA format. Sequences from the skate genome assembly or the skate, *S. canicula* or *C. milii* transcriptome assemblies can be retrieved using this tool. **E.** Output from the NCBI ORF Finder tool showing a 536aa ORF in the skate transcriptome contig that best represents SOCS6 (left). Alignment from blastx search of the skate transcriptome sequence (contig 15542) against human UniProt using NCBI BLAST to validate that the contig aligned best to human SOCS6 rather than another human gene.

cellular proliferation through its SH2 domain⁶⁶. The first step to identify *SOCS6* in the skate transcriptome begins with entering the protein sequence for human SOCS6 that was obtained from UniProt and searching this sequence against the skate transcriptome using the tblastn program. The next step is to evaluate the alignments to determine which transcriptome sequences best represent *SOCS6*. When interpreting the pairwise alignments from SkateBlast as in any BLAST tool, it is important to examine: a) alignment statistics; b) alignment coverage; and c) presence of protein domains that you may expect to be conserved. The alignment statistics are reported to ascertain whether you would expect the given alignment by chance or not. There are three key alignment statistics, the expectation (E)-value, percent identity and alignment length. The E-value represents the probability that you would expect an alignment with that alignment score or better by random chance, thus the lower the E-value, the better the alignment. Conversely, the greater the percent identity (percent identical sequence) and alignment length, the more similar the two sequences are assumed to be. Alignment

coverage with respect to the query or subject sequence (alignment length divided by the length of the query or subject sequence) can also be an important consideration, as low coverage suggests that important regions of one or both sequences may not be represented in the alignment. Finally, there may be particular sequence features, such as protein domains, that you would expect to find in the alignment. If those domains are missing, then it suggests that you have a partial or misleading alignment.

Once a transcriptome sequence of interest, such as contig15542, is identified in the SkateBLAST results, you must do a reciprocal search of that sequence against a database of protein sequences to confirm that the sequence aligns best to your gene of interest. You can retrieve the full sequence directly from the BLAST tool or using the Skate Contig Lookup tool (Figure 5D): a) specify the transcriptome that you had originally searched using SkateBLAST; b) enter the sequence identification or contig number is entered into the query box; and c) select the 'GO' button. The user can copy the

returned sequence and use it for further exploration of sequence homology at NCBI or similar databases.

SkateBase classroom use case: teach concepts of gene and protein annotation

SkateBase includes valuable teaching resources derived from the project workshops on gene and protein annotation. Infrastructure for sequence annotation was developed and modules for use in teaching are available. Access to the teaching modules is through the Curator Access link from the homepage and permission is granted by request using the email link at the bottom of each page, info@SkateBase.org. Once successfully logged into the site, access to pre-computed blast results, guides and examples, annotation forms, and links to external tools helpful for sequence analysis are available. Gene annotation begins with a transcriptome contig identified through a SkateBlast search as illustrated above. The portion of the transcript that codes for protein is identified using an open reading frame or ORF finder tool. Annotation follows a workflow where complimentary sequences from the transcriptome and genome are aligned allowing annotation of both sequences using Sequence Ontology vocabulary⁶⁷. The evidence is recorded in an annotation form that records information about the annotator and sequences and includes a comment box for questions and comments between students and teachers or curators and annotators. The annotation form records the pairwise alignment of the transcriptome and genome contigs, notes concerning mismatches or gaps, as well as output from the ORF tool. The untranslated regions (UTR) at the beginning and end of each sequence, 5' UTR and 3' UTR regions, as well as the intron/exon structure for the genomic contig and CDS for the transcriptomic contig are recorded in the Gene

Annotation Form. When completing the Gene Annotation Form, the appropriate activity must be selected and can be customized to specify the user's course ID, institution or workshop title to track annotation history. Protein annotation uses the rapid annotation interface for proteins, RACE-P, developed by the PIR. A UniProt accession number is required to initiate a new annotation form. The form is composed of 6 blocks of information, protein information, gene information, a bibliography, Gene Ontology (GO), computational analysis using tools such as Pfam⁶⁸, TMHMM⁶⁹, SignalP⁷⁰, COILS⁷¹, NetPhos⁷² and EMBOSS⁷³, and protein family evidence.

Discussion

The volume of data in GenBank continues to grow exponentially, doubling nearly every 18 months. The first sequences for chondrichthyes appeared in 1983 and the overall data trend for chondrichthyes is similar to all of GenBank with three exceptions. First, the rate of increase is less than GenBank. Second, the number of sequences deposited during the first decade of the 21st century was nearly stagnant in comparison. Third, a large spike is observed in late 2012 attributed to the Elephant Shark Genome Project data (Figure 6). Molecular data is increasingly important for all aspects of research utilizing chondrichthyan fishes⁷⁴. It can be a forensic tool to understand species when fins are landed without carcass and ensure protected species and quotas are respected⁷⁵⁻⁷⁹. For migrating species molecular data serves as a surrogate to classical tagging data to understand population structure and range⁸⁰⁻⁸³. In studies of evolution, molecular data provides estimates of divergence time and supplements morphological and ecological traits as the basis for a phylogeny. The benefits and uses of molecular data for these fishes are limited

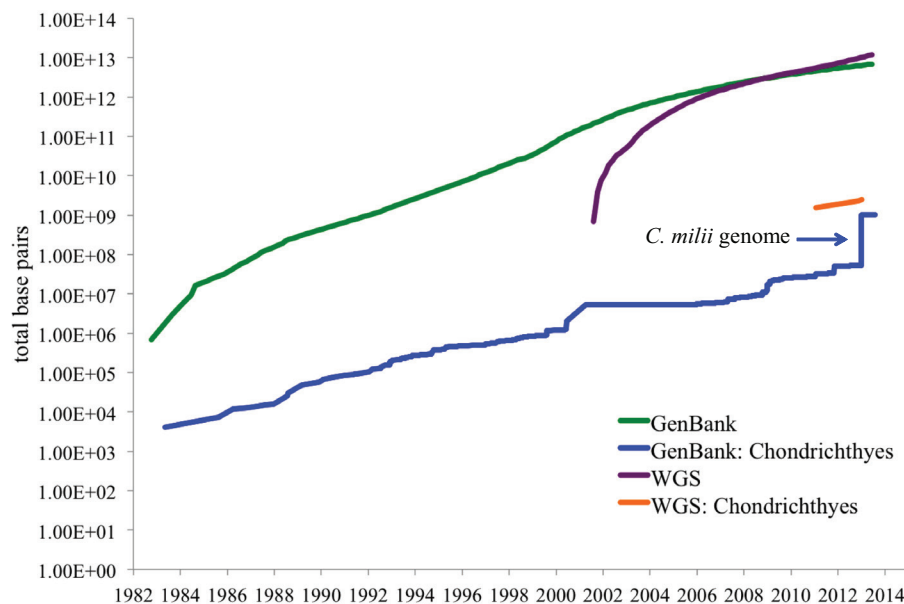


Figure 6. GenBank and WGS data trends for Chondrichthyes and all taxa. GenBank is the National Institutes of Health (NIH) genetic sequence database and together with the DNA Databank of Japan (DDBJ) and the European Molecular Biological Laboratory (EMBL) comprise the International Nucleotide Sequence Database Collaboration (INSDC). The cumulative base pair total for all taxa as well as chondrichthyan only data are given versus time for GenBank and Whole Genome Shotgun (WGS) data. The Elephant Shark Genome Project is responsible for the spike in chondrichthyan GenBank in 2011. The little skate and elephant shark genome projects are currently the only two WGS datasets (yellow line).

only by the amount of data available. SkateBase provides the only genomic data publically available for an elasmobranch in addition to embryonic transcriptomes, data tools, and educational resources.

Sequencing projects require significant funding and personnel commitments but generate a large amount of information that can be translated to knowledge by domain experts. The efficiency of this process is affected most by allowing the scientific community to access the data. The value of data sharing can be measured by the number of publications that result from its distribution. To date, 19 publications in peer-reviewed journals have used data derived from SkateBase (<http://skatebase.org/vitae>). Molecular data are the means to investigate genes and develop reagents for gene expression studies by PCR or *in situ* hybridization. Small scale sequencing efforts that generate limited or fragmented data often get deposited to hard disks and remain 'buried' and out of reach. Efforts to deposit this data at public sequence repositories are encouraged to build the foundation of data required to describe this dynamic and ancient clade of fishes. We invite investigators to contact the authors in an effort to survey the volume of private data available for potential distribution through SkateBase.

The transcriptome data at SkateBase serves as a platform to teach molecular techniques, technologies, and bioinformatics in the context of studying elasmobranchs. As next generation sequencing (NGS) technologies evolve it is important for scientists and students to understand how the sequence was generated and caveats of workflow for each data type in order to recognize errors and customize analysis algorithms. The educational materials and infrastructure at SkateBase have been used by University of Delaware, Georgetown University, MDI Biological Laboratory, University of Maine at Machias, University of Rhode Island, and most recently the Virginia Institute of Marine Science to teach gene and protein annotation concepts. We invite and look forward to continued expansion of the SkateBase educational platform as we refine the infrastructure and expand the data available for investigation through continued sequencing efforts.

Author contributions

The little skate genome project is a collaborative effort with participation from all authors. JTW and BLK prepared the first draft of the manuscript. SWP oversaw the design and implementation of

the SkateBase web portal and SkateBLAST tool. All authors were involved with the revision of the draft manuscript and have agreed to the final content.

Competing interests

The authors declare to have no competing interests.

Grant information

This work was supported by a re-entry career award to JTW, National Institute of General Medical Sciences (NIGMS) IDeA Networks of Biomedical Research Excellence (INBRE) 3P20GM103446-12S1. Skate genome sequencing was funded by National Institutes of Health (NIH) National Center for Research Resources (NCRR) ARRA Supplements to 5P20RR016463-12 (MDIBL), 5P20RR016472-12 (UD), and 5P20RR16462 (UVM). The North East Cyberinfrastructure Consortium (NECC) is funded by NIH NCRR grants 5P20RR016463-12 (MDIBL), 5P20RR016472-12 (UD), 5 P20 RR16462 (UVM), 5P20RR016457-11 (URI), and 5P20RR030360-03 (UNH) and NIH NIGMS grants 8 P20 GM103423-12 (MDIBL), 8P20GM103446-12 (UD), 8P20GM103449 (UVM), 8 P20 GM103430-11 (URI), and 8P20GM103506-03 (Dartmouth), and NSF Experimental Program to Stimulate Competitive Research (EPSCoR) grants EPS-0904155 (UM), EPS-081425 (UD), EPS-1101317 (UVM), EPS-1004057 (URI), and EPS-1101245 (UNH).

Acknowledgements

The authors thank the North East Bioinformatics Collaborative of the North East Cyberinfrastructure Consortium and Karl Steiner and Steven Stanhope of DE-INBRE, Patricia Hand of ME-INBRE, and Judith VanHouten of VT-INBRE. We thank the Delaware Biotechnology Institute (DBI) for hosting the website, database and NECC Shared Data Center. We thank Karol Miaskiewicz for systems administration support and Gang Li and Zhiwen Li for web development. We thank Bruce Kingham and the DBI Sequencing and Genotyping Center. Sylvie Mazan, Alistair Dove, Dave Ebert, Cecelia Arighi, Qinghua Wang, James Sulikowski and Eric Haenni contributed data, samples, artwork or editorial comments. We thank Gavin Naylor and the Chondrichthyan Tree of Life project team including Lindsay Marshall (illustrations), Jason Davies (database, computational work, and visualizations), Will White and Peter Last (maps and taxonomy), Shannon Corrigan and Lei Yang (gene capture data), and Callie Crawford and Thomas Fussell (CT scanning and anatomy) for providing a description of the project scope.

References

- Inoue JG, Miya M, Lam K, *et al.*: **Evolutionary origin and phylogeny of the modern holocephalans (Chondrichthyes: Chimaeriformes): a mitogenomic perspective.** *Mol Biol Evol.* 2010; 27(11): 2576–86.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Ebert DA, Ho H, White WT, *et al.*: **Introduction to the systematics and biodiversity of sharks, rays, and chimaeras (Chondrichthyes) of Taiwan.** 2013; 3752(1): 5–19.
[Publisher Full Text](#)
- Stevens J, Bonfil R, Dulvy NK, *et al.*: **The effects of fishing on sharks, rays, and chimaeras (chondrichthyans), and the implications for marine ecosystems.** *ICES J Mar Sci.* 2000; 57(3): 476–94.
[Publisher Full Text](#)
- García VB, Lucifora LO, Myers RA: **The importance of habitat and life history to extinction risk in sharks, skates, rays and chimaeras.** *Proc Biol Sci.* 2008; 275(1630): 83–9.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Field IC, Meekan MG, Buckworth RC, *et al.*: **Chapter 4. Susceptibility of sharks, rays and chimaeras to global extinction.** *Adv Mar Biol.* 2009; 56: 275–363.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Springer S: **Oviphagous Embryos of the Sand Shark, *Carcharias taurus*.** *Copeia.* 1948; 1948(3): 153.
[Publisher Full Text](#)
- Joung SJ, Chen CT, Clark E, *et al.*: **The whale shark, *Rhincodon typus*, is a livebearer: 300 embryos found in one 'megamamma' supreme.** *Environ Biol Fishes.* 1996; 46(3): 219–23.
[Publisher Full Text](#)

8. Stevens J, Bonfil R, Dulvy NK, *et al.*: **The effects of fishing on sharks, rays, and chimaeras (chondrichthyans), and the implications for marine ecosystems.** *ICES J Mar Sci.* 2000; **57**(3): 476–94.
[PubMed Abstract](#) | [Publisher Full Text](#)
9. Liu SY, Chan CL, Lin O, *et al.*: **DNA barcoding of shark meats identify species composition and CITES-listed species from the markets in Taiwan.** *PLoS One.* 2013; **8**(11): e79373.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
10. Dudgeon CL, Blower DC, Broderick D, *et al.*: **A review of the application of molecular genetics for fisheries management and conservation of sharks and rays.** *J Fish Biol.* 2012; **80**(5): 1789–843.
[PubMed Abstract](#) | [Publisher Full Text](#)
11. Pinhal D, Shivji MS, Nachtigall PG, *et al.*: **A streamlined DNA tool for global identification of heavily exploited coastal shark species (genus *Rhizoprionodon*).** *PLoS One.* 2012; **7**(4): e34797.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
12. Barbuto M, Galimberti A, Ferri E, *et al.*: **DNA barcoding reveals fraudulent substitutions in shark seafood products: The Italian case of “palombo” (*Mustelus spp.*).** *Food Res Int.* 2010; **43**(1): 376–81.
[PubMed Abstract](#) | [Publisher Full Text](#)
13. Feldheim KA, Gruber SH, Dibattista JD, *et al.*: **Two decades of genetic profiling yields first evidence of natal philopatry and long-term fidelity to parturition sites in sharks.** *Mol Ecol.* 2014; **23**(1): 110–7.
[PubMed Abstract](#) | [Publisher Full Text](#)
14. Schneider I, Aneas I, Gehrke AR, *et al.*: **Appendage expression driven by the Hoxd Global Control Region is an ancient gnathostome feature.** *Proc Natl Acad Sci U S A.* 2011; **108**(31): 12782–6.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
15. Gillis JA, Dahn RD, Shubin NH: **Shared developmental mechanisms pattern the vertebrate gill arch and paired fin skeletons.** *Proc Natl Acad Sci U S A.* 2009; **106**(14): 5720–4.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
16. Gillis JA, Modrell MS, Baker CVH: **Developmental evidence for serial homology of the vertebrate jaw and gill arch skeleton.** *Nat Commun.* 2013; **4**: 1436.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
17. Hentschel DM, Bonventre JV: **Novel non-rodent models of kidney disease.** *Curr Mol Med.* 2005; **5**(5): 537–46.
[PubMed Abstract](#) | [Publisher Full Text](#)
18. Steele SL, Yancey PH, Wright PA: **Dogmas and controversies in the handling of nitrogenous wastes: osmoregulation during early embryonic development in the marine little skate *Raja erinacea*; response to changes in external salinity.** *J Exp Biol.* 2004; **207**(Pt 12): 2021–31.
[PubMed Abstract](#) | [Publisher Full Text](#)
19. Stolte H, Galaske RG, Eisenbach GM, *et al.*: **Renal tubule ion transport and collecting duct function in the elasmobranch little skate, *Raja erinacea*.** *J Exp Zool.* 1977; **199**(3): 403–10.
[PubMed Abstract](#) | [Publisher Full Text](#)
20. Evans DH: **A brief history of the study of fish osmoregulation: the central role of the Mt. Desert Island Biological Laboratory.** *Front Physiol.* 2010; **1**: 13.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
21. Wyffels JT, Luer CA, Walsh CJ, *et al.*: **In vivo exposure of clearnose skates, *Raja eglanteria*, to ionizing X-radiation: acute effects on the peripheral blood, spleen, and epigonal and Leydig organs.** *Fish Shellfish Immunol.* 2007; **23**(2): 401–18.
[PubMed Abstract](#) | [Publisher Full Text](#)
22. Wyffels JT, Walsh CJ, Luer CA, *et al.*: **In vivo exposure of clearnose skates, *Raja eglanteria*, to ionizing X-radiation: acute effects on the thymus.** *Dev Comp Immunol.* 2005; **29**(4): 315–31.
[PubMed Abstract](#) | [Publisher Full Text](#)
23. Lutton BV, Callard IP: **Morphological relationships and leukocyte influence on steroid production in the epigonal organ-ovary complex of the skate, *Leucoraja erinacea*.** *J Morphol.* 2008; **269**(5): 620–9.
[PubMed Abstract](#) | [Publisher Full Text](#)
24. Walsh CJ, Luer CA, Bodine AB, *et al.*: **Elasmobranch immune cells as a source of novel tumor cell inhibitors: Implications for public health.** *Integr Comp Biol.* 2006; **46**(6): 1072–81.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
25. Eason DD, Litman RT, Luer CA, *et al.*: **Expression of individual immunoglobulin genes occurs in an unusual system consisting of multiple independent loci.** *Eur J Immunol.* 2004; **34**(9): 2551–8.
[PubMed Abstract](#) | [Publisher Full Text](#)
26. Anderson MK, Pant R, Miracle AL, *et al.*: **Evolutionary origins of lymphocytes: ensembles of T cell and B cell transcriptional regulators in a cartilaginous fish.** *J Immunol.* 2004; **172**(10): 5851–60.
[PubMed Abstract](#) | [Publisher Full Text](#)
27. Cai SY, Soroka CJ, Ballatori N, *et al.*: **Molecular characterization of a multidrug resistance-associated protein, Mrp2, from the little skate.** *Am J Physiol Regul Integr Comp Physiol.* 2003; **284**(1): R125–30.
[PubMed Abstract](#)
28. Clusin W, Spray DC, Bennett MVL: **Activation of a voltage-insensitive conductance by inward calcium current.** *Nature.* 1975; **256**(5516): 425–7.
[PubMed Abstract](#) | [Publisher Full Text](#)
29. Elger M, Hentschel H, Litteral J, *et al.*: **Nephrogenesis is induced by partial nephrectomy in the elasmobranch *Leucoraja erinacea*.** *J Am Soc Nephrol.* 2003; **14**(6): 1506–18.
[PubMed Abstract](#) | [Publisher Full Text](#)
30. Zhang C, Du Pasquier L, Hsu E: **Shark IgW C region diversification through RNA processing and isotype switching.** *J Immunol.* 2013; **191**(6): 3410–8.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
31. Beard J: **The origin and histogenesis of the thymus in *Raja batis*.** *Zool Jahrb Anat Ontog Tiere.* 1902; **17**: 403–80.
[Reference Source](#)
32. Luer CA, Walsh CJ, Bodine AB, *et al.*: **The elasmobranch thymus: Anatomical, histological, and preliminary functional characterization.** *J Exp Zool.* 1995; **273**(4): 342–54.
[PubMed Abstract](#) | [Publisher Full Text](#)
33. Wyffels JT, Walsh CJ, Luer CA, *et al.*: **In vivo exposure of clearnose skates, *Raja eglanteria*, to ionizing X-radiation: acute effects on the thymus.** *Dev Comp Immunol.* 2005; **29**(4): 315–31.
[PubMed Abstract](#) | [Publisher Full Text](#)
34. Feldheim KA, Chapman DD, Sweet D, *et al.*: **Shark virgin birth produces multiple, viable offspring.** *J Hered.* 2010; **101**(3): 374–7.
[PubMed Abstract](#) | [Publisher Full Text](#)
35. Chapman DD, Shivji MS, Louis E, *et al.*: **Virgin birth in a hammerhead shark.** *Biol Lett.* 2007; **3**(4): 425–7.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
36. Robinson DP, Baverstock W, Al-Jaru A, *et al.*: **Annually recurring parthenogenesis in a zebra shark *Stegostoma fasciatum*.** *J Fish Biol.* 2011; **79**(5): 1376–82.
[PubMed Abstract](#) | [Publisher Full Text](#)
37. Compagno LJV: **Alternative life-history styles of cartilaginous fishes in time and space.** *Environ Biol Fishes.* 1990; **28**(1–4): 33–75.
[PubMed Abstract](#) | [Publisher Full Text](#)
38. Luer CA, Gilbert PW: **Mating behavior, egg deposition, incubation period, and hatching in the clearnose skate, *Raja eglanteria*.** *Environ Biol Fishes.* 1985; **13**(3): 161–71.
[PubMed Abstract](#) | [Publisher Full Text](#)
39. Luer CA, Walsh CJ, Bodine AB, *et al.*: **Normal embryonic development in the clearnose skate, *Raja eglanteria*, with experimental observations on artificial insemination.** *Environ Biol Fishes.* 2007; **80**(2–3): 239–55.
[PubMed Abstract](#) | [Publisher Full Text](#)
40. Masuda M, Izawa Y, Kametusa S, *et al.*: **Artificial insemination of the cloudy catshark.** *J Japanese Assoc Zool Gard Aquariums.* 2003; **44**(2): 39–43.
[Reference Source](#)
41. Castro JI, Bubucis PM, Overstrom NA, *et al.*: **The Reproductive Biology of the Chain Dogfish, *Scyliorhinus retifer*.** *Copeia.* 1988; **1988**(3): 740.
[PubMed Abstract](#) | [Publisher Full Text](#)
42. Barrett T, Clark K, Gevorgyan R, *et al.*: **BioProject and BioSample databases at NCBI: facilitating capture and organization of metadata.** *Nucleic Acids Res.* 2012; **40**(Database issue): D57–63.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
43. Venkatesh B, Lee AP, Ravi V, *et al.*: **Elephant shark genome provides unique insights into gnathostome evolution.** *Nature.* 2014; **505**(7482): 174–9.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
44. Venkatesh B, Kirkness EF, Loh YH, *et al.*: **Survey sequencing and comparative analysis of the elephant shark (*Callorhynchus milii*) genome.** *PLoS Biol.* 2007; **5**(4): e101.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
45. Luo M, Kim H, Kudrna D, *et al.*: **Construction of a nurse shark (*Ginglymostoma cirratum*) bacterial artificial chromosome (BAC) library and a preliminary genome survey.** *BMC Genomics.* 2006; **7**: 106.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
46. Gwee PC, Tay BH, Brenner S, *et al.*: **Characterization of the neurohypophysial hormone gene loci in elephant shark and the Japanese lamprey: origin of the vertebrate neurohypophysial hormone genes.** *BMC Evol Biol.* 2009; **9**: 47.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
47. Mulley JF, Zhong YF, Holland PW: **Comparative genomics of chondrichthyan Hoxa clusters.** *BMC Evol Biol.* 2009; **9**: 218.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
48. Chiu C, Amemiya C, Dewar K, *et al.*: **Molecular evolution of the HoxA cluster in the three major gnathostome lineages.** *Proc Natl Acad Sci U S A.* 2002; **99**(8): 5492–7.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
49. Mattingly C, Parton A, Dowell L, *et al.*: **Cell and molecular biology of marine elasmobranchs: *Squalus acanthias* and *Raja erinacea*.** *Zebrafish.* 2004; **1**(2): 111–20.
[PubMed Abstract](#) | [Publisher Full Text](#)
50. Merson RR, Mattingly C, Planchart AJ: **Tandem duplication of aryl hydrocarbon receptor (AHR) genes in the genome of the spiny dogfish shark (*Squalus acanthias*).** *Bull MDIBL.* 2009; **48**: 43–4.
[Reference Source](#)
51. Oulion S, Debais-Thibaud M, d'Aubenton-Carafa Y, *et al.*: **Evolution of Hox gene clusters in gnathostomes: insights from a survey of a shark (*Scyliorhinus canicula*) transcriptome.** *Mol Biol Evol.* 2010; **27**(12): 2829–38.
[PubMed Abstract](#) | [Publisher Full Text](#)

52. King BL, Gillis JA, Carlisle HR, *et al.*: **A natural deletion of the HoxC cluster in elasmobranch fishes.** *Science*. 2011; **334**(6062): 1517.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
53. Kim CB, Amemiya C, Bailey W, *et al.*: **Hox cluster genomics in the horn shark, *Heterodontus francisci*.** *Proc Natl Acad Sci U S A*. 2000; **97**(4): 1655–60.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
54. Ravi V, Lam K, Tay BH, *et al.*: **Elephant shark (*Callorhynchus milii*) provides insights into the evolution of Hox gene clusters in gnathostomes.** *Proc Natl Acad Sci U S A*. 2009; **106**(38): 16327–32.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
55. Naylor GJP, Caira JN, Jensen K, *et al.*: **Elasmobranch Phylogeny: A mitochondrial estimate based on 595 species.** In: Carrier JC, Musick JA, Heithaus MR, editors. *The Biology of Sharks and Their Relatives*. CRC Press, Taylor & Francis Group. 2012; 31–56.
[Reference Source](#)
56. Naylor GJP, Ryburn JA, Fedrigo O, *et al.*: **Phylogenetic Relationships among the Major Lineages of Modern Elasmobranchs.** 1995.
[Reference Source](#)
57. Naylor GJP, Caira JN, Jensen K, *et al.*: **A DNA Sequence–Based Approach To the Identification of Shark and Ray Species and Its Implications for Global Elasmobranch Diversity and Parasitology.** *Bull Am Museum Nat Hist*. 2012; **367**: 1–262.
[Publisher Full Text](#)
58. Feijão PC, Neiva LS, de Azeredo-Espin AM, *et al.*: **AMIGA: the arthropodan mitochondrial genomes accessible database.** *Bioinformatics*. 2006; **22**(7): 902–3.
[PubMed Abstract](#) | [Publisher Full Text](#)
59. Li C, Hofreiter M, Straube N, *et al.*: **Capturing protein-coding genes across highly divergent species.** *Biotechniques*. 2013; **54**(6): 321–6.
[PubMed Abstract](#)
60. UniProt Consortium. **Activities at the Universal Protein Resource (UniProt).** *Nucleic Acids Res*. 2014; **42**(Database issue): D191–8.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
61. Berman HM, Westbrook J, Feng Z, *et al.*: **The Protein Data Bank.** *Nucleic Acids Res*. 2000; **28**(1): 235–42.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
62. Takeuchi M, Takeuchi M, Ota KG, *et al.*: **Overview of the transcriptome profiles identified in hagfish, shark, and bichir: current issues arising from some nonmodel vertebrate taxa.** *J Exp Zool B Mol Dev Evol*. 2011; **316**(7): 526–46.
[PubMed Abstract](#) | [Publisher Full Text](#)
63. Wang Q, Arighi CN, King BL, *et al.*: **Community annotation and bioinformatics workforce development in concert—Little Skate Genome Annotation Workshops and Jamborees.** *Database (Oxford)*. 2012; **2012**: bar064.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
64. Jennifer TW: **Embryonic development of Chondrichthyan fishes – a review.** In: W KY, Lue CA & Kapoor BG, editors. *Development of Non-Teleost Fishes*. *Enfield: Science Publishers*. 2009; 1–103.
[Publisher Full Text](#)
65. Deng W, Nickle DC, Learn GH, *et al.*: **ViroBLAST: a stand-alone BLAST web server for flexible queries of multiple databases and user's datasets.** *Bioinformatics*. 2007; **23**(17): 2334–6.
[PubMed Abstract](#) | [Publisher Full Text](#)
66. Zadjali F, Pike AC, Vesterlund M, *et al.*: **Structural basis for c-KIT inhibition by the suppressor of cytokine signaling 6 (SOCS6) ubiquitin ligase.** *J Biol Chem*. 2011; **286**(1): 480–90.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
67. Mungall CJ, Batchelor C, Eilbeck K: **Evolution of the Sequence Ontology terms and relationships.** *J Biomed Inform*. 2011; **44**(1): 87–93.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
68. Finn RD, Bateman A, Clements J, *et al.*: **Pfam: the protein families database.** *Nucleic Acids Res*. 2014; **42**(Database issue): D222–30.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
69. Krogh A, Larsson B, von Heijne G, *et al.*: **Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes.** *J Mol Biol*. 2001; **305**(3): 567–80.
[PubMed Abstract](#) | [Publisher Full Text](#)
70. Petersen TN, Brunak S, von Heijne G, *et al.*: **SignalP 4.0: discriminating signal peptides from transmembrane regions.** *Nat Methods*. 2011; **8**(10): 785–6.
[PubMed Abstract](#) | [Publisher Full Text](#)
71. Lupas A, Van Dyke M, Stock J: **Predicting coiled coils from protein sequences.** *Science*. 1991; **252**(5009): 1162–4.
[PubMed Abstract](#) | [Publisher Full Text](#)
72. Blom N, Gammeltoft S, Brunak S: **Sequence and structure-based prediction of eukaryotic protein phosphorylation sites.** *J Mol Biol*. 1999; **294**(5): 1351–62.
[PubMed Abstract](#) | [Publisher Full Text](#)
73. Rice P, Longden I, Bleasby A: **EMBOSS: the European Molecular Biology Open Software Suite.** *Trends Genet*. 2000; **16**(6): 276–7.
[PubMed Abstract](#) | [Publisher Full Text](#)
74. Heist E: **Genetics of Sharks, Skates, and Rays.** In: Musick J, Carrier J, Heithaus M, editors. *Biology of Sharks and Their Relatives*. CRC Press. 2004; 471–85.
[Publisher Full Text](#)
75. Chapman D, Pinhal D, Shivji M: **Tracking the fin trade: genetic stock identification in western Atlantic scalloped hammerhead sharks *Sphyrna lewini*.** *Endanger Species Res*. 2009; **9**: 221–8.
[Publisher Full Text](#)
76. Clarke SC, Magnussen JE, Abercrombie DL, *et al.*: **Identification of shark species composition and proportion in the Hong Kong shark fin market based on molecular genetics and trade records.** *Conserv Biol*. 2006; **20**(1): 201–11.
[PubMed Abstract](#) | [Publisher Full Text](#)
77. Liu SY, Chan CL, Lin O, *et al.*: **DNA barcoding of shark meats identify species composition and CITES-listed species from the markets in Taiwan.** *PLoS One*. 2013; **8**(11): e79373.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
78. Shivji MS: **DNA Forensic Applications in Shark Management and Conservation.** In: Carrier JC, Musick JA, Heithaus MR, editors. *Sharks and Their Relatives II Biodiversity, Adaptive Physiology, and Conservation*. CRC Press. 2010; 593–610.
[Publisher Full Text](#)
79. Rodrigues-Filho LF, Pinhal D, Sodré D, *et al.*: **Applications and Impacts on Genetic Diversity.** In: Caliskan M, editor. *Analysis of Genetic Variation in Animals*. InTech. 2012; 269–86.
[Publisher Full Text](#)
80. Chabot CL, Allen LG: **Global population structure of the tope (*Galeorhinus galeus*) inferred by mitochondrial control region sequence data.** *Mol Ecol*. 2009; **18**(3): 545–52.
[PubMed Abstract](#) | [Publisher Full Text](#)
81. Keeney DB, Heist EJ: **Worldwide phylogeography of the blacktip shark (*Carcharhinus limbatus*) inferred from mitochondrial DNA reveals isolation of western Atlantic populations coupled with recent Pacific dispersal.** *Mol Ecol*. 2006; **15**(12): 3669–79.
[PubMed Abstract](#) | [Publisher Full Text](#)
82. Schultz JK, Feldheim KA, Gruber SH, *et al.*: **Global phylogeography and seascape genetics of the lemon sharks (genus *Negaprion*).** *Mol Ecol*. 2008; **17**(24): 5336–48.
[PubMed Abstract](#) | [Publisher Full Text](#)
83. Ward RD, Holmes BH, White WT, *et al.*: **DNA barcoding Australasian chondrichthyans: results and potential uses in conservation.** *Mar Freshw Res*. 2008; **59**(1): 57–71.
[Publisher Full Text](#)

Open Peer Review

Current Referee Status:



Version 1

Referee Report 01 September 2014

doi:10.5256/f1000research.5333.r5798



Ed Heist

Department of Zoology, Southern Illinois University, Carbondale, USA

This manuscript has two components: a description of current genomic, proteomic, and transcriptomic resources for chondrichthyan fishes and a description of the SkateBase project. Of the two, I would rather have seen more emphasis on SkateBase since the rate at which new data are being generated will make the survey data presented in the manuscript increasingly incomplete. Nevertheless I think it is a good paper worthy of publication and SkateBase seems to be a valuable resource.

Specific comments include:

Abstract – capitalize Sarcopterygii, Actinopterygii misspelled. Throughout the manuscript there are numerous inconsistencies in capitalization, e.g. “order Chimaeriformes” and “Fish barcode of life (FISH-BOL) a working group of the International Barcode of Life Project (iBOL)” both on page 8.

Introduction – In the first sentence replace “Chondrichthyan fishes” with “Living chondrichthyan fishes.”

Paragraph 2 – The text implies that the only benefits of chondrichthyans are ecotourism and fisheries. Perhaps a mention of ecological services (e.g. food-web dynamics) or other benefits of chondrichthyan conservation would be appropriate here.

Paragraph 4 – “asexual” parthenogenesis seems redundant.

Paragraph 5 – perhaps some more details on the distribution, life history, reproductive mode, genome size etc. of little skate is in order. Why is the geographic location of the species relevant?

Figure 1 is hard for me to decipher. I can't figure out what the two separate outer rings indicate and why the spacing between them is different for the orders vs. the families, genera, and species. I really think this figure is too complex and tries to cram too much information that is peripheral to the goals of the paper into a single figure. I may be old fashioned, but I think a table that listed the numbers of families, genera, and species would have been sufficient. There also needs to be a citation on the source of the information as the taxonomy and number of recognized species of chondrichthyans is changing.

Figure 4B – A citation is needed for the cladogram.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Competing Interests: No competing interests were disclosed.

Referee Report 26 August 2014

doi:[10.5256/f1000research.5333.r5793](https://doi.org/10.5256/f1000research.5333.r5793)



Kevin Feldheim

Pritzker Laboratory for Molecular Systematics and Evolution, Field Museum of Natural History, Chicago, USA

The authors describe a publicly accessible database SkateBase, a public site for the little skate genome project. This paper is well written, and I have just a few comments and suggestions:

- Why was the little skate picked for genome size? Is it small and manageable? Is it a typical size compared to other chondrichthyans?
- Page 3, 1st column: change “compliance with fishing quotas” to “maintain compliance with fishing quotas”.
- Page 3, 2nd column: Heterodontiformes is misspelled (Heterodontiformes).
- Page 3, 2nd column: Lamniformes is misspelled (Lamniormes).
- Page 3, under “Existing resources” I suggest including Chimaeriformes here as well.
- Page 7: *Heterodontus* is misspelled (*Heterodontus*).
- Page 8: species name for hagfish should be *burgeri* (not *burger*).
- Page 8: The last three sentences under the “Protein databases” section can be incorporated into the legend for Figure 4.
- Page 8: bichir is misspelled (birchir).
- Page 8: change “SkateBase data is linked” to “SkateBase data are linked” and “SkateBase data includes” to “SkateBase data include”.
- Page 19, 1st column: change “this data” to “these data”.

Figures and Tables:

- In the figure 1 legend — most of this information is stated in the text. I suggest keeping the sentence: “The distribution of chondrichthyan species in each of the 14 orders is shown relative to

the total number of species, genera and families for the clade.” I also suggest explaining the figure in a little more detail. What do the numbers mean? Do the colors represent anything?

- Figure 2 legend: “is” should not be italicized.
- Table 1: What do the numbers in the columns mean? Are these numbers of entries?
- Table 2: Contigs for *L. erinacea* is 2,62,265. Also, under N50, change 1466 to 1,466.
- Figure 4: It is difficult to tell apart colors for the following pairs:
 - Squatiniformes and Carcharhiniformes
 - Lamniformes and Myliobatiformes
 - Chimaeriformes and Rajiiformes
- This can be rectified by using black, white and yellow for three of the above orders in the figure.
- Figure 5 legend: I suggest changing “Four important fields in the output should be examined carefully to interpret the alignments and determine which returned alignment best represents the skate ortholog to SOCS6. First, the alignment score, E-value, alignment length and percent identity can be used to interpret the overall alignment significance.” to “Four important fields in the output should be examined carefully to interpret the alignments: the alignment score, E-value, alignment length and percent identity.”

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Competing Interests: No competing interests were disclosed.
