BMC Genomics

CrossMark

Comparative transcriptomics of elasmobranchs and teleosts highlight important processes in adaptive immunity and regional endothermy

Nicholas J. Marra[1,2], Vincent P. Richards[3], Angela Early[4], Steve M. Bogdanowicz[4], Paulina D. Pavinski Bitar[1], Michael J. Stanhope[1*] and Mahmood S. Shivji[2*]

Abstract

Background: Comparative genomic and/or transcriptomic analyses involving elasmobranchs remain limited, with genome level comparisons of the elasmobranch immune system to that of higher vertebrates, non-existent. This paper reports a comparative RNA-seq analysis of heart tissue from seven species, including four elasmobranchs and three teleosts, focusing on immunity, but concomitantly seeking to identify genetic similarities shared by the two lamnid sharks and the single billfish in our study, which could be linked to convergent evolution of regional endothermy.

Results: Across seven species, we identified an average of 10,877 Swiss-Prot annotated genes from an average of 32,474 open reading frames within each species' heart transcriptome. About half of these genes were shared between all species while the remainder included functional differences between our groups of interest (elasmobranch vs. teleost and endotherms vs. ectotherms) as revealed by Gene Ontology (GO) and selection analyses. A repeatedly represented functional category, in both the uniquely expressed elasmobranch genes (total of 259) and the elasmobranch GO enrichment results, involved antibody-mediated immunity, either in the recruitment of immune cells (Fc receptors) or in antigen presentation, including such terms as "antigen processing and presentation of exogenous peptide antigen via MHC class II", and such genes as MHC class II, *HLA-DPB1*. Molecular adaptation analyses identified three genes in elasmobranchs with a history of positive selection, including legumain (*LGMN*), a gene with roles in both innate and adaptive immunity including producing antigens for presentation by MHC class II. Comparisons between the endothermic and ectothermic species revealed an enrichment of GO terms associated with cardiac muscle contraction in endotherms, with 19 genes expressed solely in endotherms, several of which have significant roles in lipid and fat metabolism.

(Continued on next page)

* Correspondence: mjs297@cornell.edu; mahmood@nova.edu
Co-senior and co-corresponding authors: Michael J. Stanhope and Mahmood
S. Shivji
[1]Department of Population Medicine and Diagnostic Sciences, College of
Veterinary Medicine, Cornell University, Ithaca, NY 14853, USA
[2]Save Our Seas Shark Research Center and Guy Harvey Research Institute,
Nova Southeastern University, 8000 North Ocean Drive, Dania Beach, FL
33004, USA
Full list of author information is available at the end of the article

Marra *et al. BMC Genomics* (2017) 18:87

Page 2 of 18

(Continued from previous page)

Conclusions: This collective comparative evidence provides the first multi-taxa transcriptomic-based perspective on differences between elasmobranchs and teleosts, and suggests various unique features associated with the adaptive immune system of elasmobranchs, pointing in particular to the potential importance of MHC Class II. This in turn suggests that expanded comparative work involving additional tissues, as well as genome sequencing of multiple elasmobranch species would be productive in elucidating the regulatory and genome architectural hallmarks of elasmobranchs.

Keywords: Regional endothermy, Adaptive immunity, Gene ontology, Positive selection, RNA-seq, Elasmobranchs

Background

The class Chondrichthyes includes all of the cartilaginous fishes: the chimaeras, sharks, skates, and rays. The extant members of the class comprise at least 1,207 species [1], divided into two major groups: Subclasses Holocephali (chimaeras) and Elasmobranchii (sharks, skates, and rays). Recent molecular dating analyses suggest a split between holocephalans and elasmobranchs at about 420 Mya [2, 3]. Chondrichthyans as a whole, are thought to have diverged from bony vertebrates (Osteichthyes: ray-finned fishes, coelacanths, lungfishes, and tetrapods) approximately 450–475 Mya [2–4]; see however Giles et al. 2015 [5], for evidence of a chondrichthyan/osteichthyan ancestry of 415 Mya. Because of their phylogenetic position in vertebrate evolution, chondrichthyans provide an important reference for our understanding of vertebrate genome evolution. In addition to their ancient history and fundamental position in vertebrate systematics, chondrichthyans possess a wide variety of biological characteristics of note. Such traits include, but are not limited to, the presence of a primitive adaptive immune system, efficient wound healing, and the evolution of regional endothermy in several species.

One of the most rapidly expanding areas of research in elasmobranch biology is in understanding the function of the immune system [6]. Cartilaginous fishes are the most ancient group of vertebrates that possesses an adaptive immune system based on the same B and T cell receptor genes that form the foundation of adaptive immunity in higher vertebrates [7]. However, adaptive immunity in chondrichthyans differs from higher vertebrates (including teleost fishes) in the lack of bone marrow (where B cells typically develop), in the types of immunoglobulins (Ig hereinafter), and in the genomic organization of the underlying genes [8–11]. Additionally, elasmobranchs contain a novel immunoglobulin, referred to as new antigen receptor (or IgNAR), which differs from traditional immunoglobulins in that it lacks light chain molecules and is comprised entirely of heavy chain domains [12, 13]. IgNARs have received considerable interest recently, in regard to their unique structure and the possibility of adapting these molecules for future

diagnostic work or drug delivery systems [14–16]. Despite this interest, transcriptomic analyses of the similarities and differences between the elasmobranch immunome and that of higher vertebrates are not currently available.

Regional (or partial) endothermy arose independently in each of the billfishes and tunas (both highly migratory, large bodied teleosts), and the highly migratory, large-bodied lamnid sharks. All three groups have convergently evolved adaptations for increased aerobic capacity, continuous swimming, elevated cruising speed, and heat production and/or retention [17–21]. Although the heart is at ambient temperature in regionally endothermic species, its function is critical to endothermic physiology because of its role in modulating blood flow and oxygen delivery to heat producing tissues (i.e. red muscle) and through the vasculature responsible for the counter-current heat exchange [18]. However, to date the genetic loci that might be associated with this remarkable example of convergent evolution in fishes remain obscure and there are few studies that specifically attempt to investigate this. A recent study of the cytochrome oxidase C subunit (COX) genes found no evidence of molecular convergence across endothermic fishes in these mitochondrial loci involved in aerobic metabolism [22]. Another recent study has shown differences in expression of genes involved in calcium storage and cycling (*serca2* and *ryr2*) in heart tissue of tuna species with different temperature tolerance, with the greatest expression in Pacific bluefin tuna (*Thunnus orientalis*), the species with the greatest cold tolerance of the three tested [23]. In a comparative study of gene expression in heart tissue of Pacific bluefin tuna that were acclimated to cold and warm temperatures, Jayasundara and colleagues found upregulation of genes associated with protein turnover, lipid and carbohydrate metabolism, heat shock proteins, and genes involved in protection against oxidative stress in cold acclimated individuals [24]. This study also detected elevated levels of the *SERCA* enzyme in cold acclimated individuals [24]. Collectively, this suggests the importance of regulating genes involved in metabolism, control of heart contraction and function, and cellular protection against

Marra *et al. BMC Genomics* (2017) 18:87

Page 3 of 18

oxidative stress in heart tissue of an organism with an endothermic physiology. Our goal here was to use the heart transcriptome to examine a large repertoire of genes for possible evidence of convergent evolution in regional endothermy, in terms of either genes expressed, or shared genes with a history of molecular adaptation.

Comparative genomics of chondrichthyans remains limited, with a single genome sequence available for the holocephalan, *Callorhinchus milii* [25, 26], and a few additional genome projects in progress (reviewed in [27], including the whale shark *Rhincodon typus* (http://whaleshark.georgiaaquarium.org), white shark, *Carcharodon carcharias* (our laboratory), catshark, *Scyliorhinus canicula* (Genoscope: http://www.genoscope.cns.fr), and the batoid, the little skate, *Leucoraja erinacea* [28]. There are a larger number of transcriptomic and RNA-seq studies, however, these genetic resources are still limited compared to those of other vertebrate taxa [27]. Transcriptome sequence examples include a heart transcriptome of the white shark [29]; brain, liver, pancreas, and embryo from the small-spotted catshark, *S. canicula*, [30, 31]; an embryo of cloudy catshark, *Scyliorhinus torazame* [32]; whole embryo from the little skate [28]; and spleen and thymus from nurse shark, *Ginglymostoma cirratum* [26] and spleen, thymus, testis, ovary, liver, muscle, kidney, intestine, heart, gills, and brain from elephant shark (a holocephalan), *C. milii* [26]. In addition, EST (expressed sequence tag) sequences exist for cell lines derived from *L. erinacea* and the spiny dogfish, *Squalus acanthias* [33].

Interspecific transcriptomic comparisons of many taxonomic groups, and in particular groups with limited genetic resources such as elasmobranchs, are confounded by both the haphazard sampling of different tissues associated with different studies as well as the different technologies used to obtain the sequence data. At present limited comparative data sets of the same tissue type and technology are available across many taxa, however, this is beginning to change and there exist a few important exceptions; see for example, [34–36].

To examine transcriptomic differences between elasmobranchs vs. teleosts and endothermic vs. ectothermic (i.e. non-endothermic) species, we sampled heart tissue since it is a metabolically active tissue, and expression of major components in innate and adaptive immunity have been demonstrated in heart and associated blood tissues [37, 38]. Compared to ectothermic fishes, regionally endothermic fishes such as tunas tend to have an elevated heart rate and this in part supports the maintenance of elevated temperature in some tissues [18, 39]. We hypothesize, therefore, that there are differences in expressed gene content of heart tissue of endothermic species relative to ectothermic species, to compensate for this increased heart rate. Our study included the following seven species: elasmobranchs - white shark (*Carcharodon carcharias*), shortfin mako (*Isurus oxyrinchus*; hereinafter termed mako), great hammerhead (*Sphyrna mokarran*; hereinafter termed hammerhead), and yellow stingray (*Urobatis jamaicensis*); teleosts - swordfish (*Xiphias gladius*), hogfish (*Lachnolaimus maximus*), and ocean surgeonfish *Acanthurus bahianus*; hereinafter termed surgeonfish). Both white shark and mako (Lamnidae, Lamniformes), like other lamnids, display elevated internal temperatures indicative of regional endothermy [40]; the great hammerhead (Sphyrnidae, Carcharhiniformes) and the yellow stingray (hereinafter referred to as ray) (Urotrygonidae, Myliobatiformes) represent the two ectothermic elasmobranchs included in our study. Molecular phylogenetic studies support rays and skates as the sister group to sharks and suggest that this split was approximately 300 Mya [2, 3, 41]. The swordfish (Xiphiidae, Perciformes) is a representative of a regionally endothermic teleost, while both hogfish (Labridae, Perciformes) and surgeonfish (Acanthuridae, Perciformes) are ectotherms.

In comparing these seven heart transcriptomes we had several specific aims. First, we sought to identify differences in expressed gene content that are a reflection of evolutionary taxonomy (e.g. elasmobranchs vs. teleosts). Secondly, we aimed to identify whether there were significant differences involving the comparative groups– i.e., elasmobranchs vs. teleosts and endotherms vs. ectotherms—in the types of genes (as identified by differences in Gene Ontology, or GO) that are expressed, especially in regards to particular phenomena of interest (e.g. adaptive immunity and wound healing in elasmobranchs, metabolic function in endotherms). Finally, we sought to identify genes with a history of molecular adaptation in elasmobranchs and the endothermic taxa in our data set, through the identification of genes under positive selection in the respective lineages.

Methods

Tissue and RNA

The shark and swordfish hearts were opportunistically obtained from freshly deceased animals captured by recreational or commercial fishermen independent of our study. Dissection was followed by immediate cold storage in order to limit RNA degradation (see below). The ray heart was opportunistically obtained from independent researchers conducting a study on its reproductive organs. Hearts from the hogfish and surgeonfish were similarly obtained from independent researchers conducting a study on age and growth, and stored in RNA-*later*® (ThermoFisher). Heart tissue from all other species was stored at −80 °C. No ethical approval or permit for animal experimentation was required, as the individuals were not sacrificed specifically for this study.

Marra *et al. BMC Genomics* (2017) 18:87

Page 4 of 18

At Cornell University, total RNA was extracted from the frozen heart tissue for each species using the Agencourt® RNAdvance™ Tissue Kit. Extractions were conducted according to manufacturer instructions. Briefly, as part of the extraction protocol tissue was homogenized and digested in lysis buffer containing proteinase K. RNA from this digested tissue was bound to paramagnetic beads to remove contaminants prior to treatment with DNase I and subsequent elution of the extracted RNA in nuclease free water. Due to the collection of some samples from fishermen and uncertainty regarding the time since death, we checked for RNA degradation using an Agilent 2100 BioAnalyzer or AATI Fragment Analyzer™ and quantified extractions using a Qubit™ spectrofluorometer. Prior to further processing these extractions were shown to pass internal quality standards for the Agilent BioAnalyzer and AATI Fragment Analyzer and had limited evidence of degradation. The total RNA extracted from each species was then used to prepare Illumina TruSeq RNA sequencing libraries according to manufacturer's instructions at the genomics facility in the Cornell Biotechnology Resource Center.

Sequencing and assembly

Two lanes of 2x100 bp paired-end sequencing were conducted on an Illumina Hi-Seq 2500 by the Genomics facility in the Biotechnology Resource Center at Cornell University. Four species were pooled per lane (including an eighth species whose library yielded poor sequencing data and was excluded from further analysis). Following sequencing, reads were separated by species and the program FastqMcf within ea-utils [42] was used to remove sequencing adaptors, trim poor quality bases, and remove poor quality reads using a minimum Phred quality score of 30, minimum trimmed length of 50 bp, and removing duplicate reads with 35 or more identical bases. Each species read pool was then used to assemble a species-specific heart transcriptome using Trinity (default parameters, version r2013-02-25 [43]). Following transcriptome assembly, the program TransDecoder (within the Trinity package) was used to extract the longest likely open reading frame (ORF) for each Trinity transcript using default parameters.

Transcriptome assessment and annotation

To get an estimate of the completeness of each transcriptome we analyzed each assembly with the tool CEGMA (under default parameters) to assess the presence of 248 Core Eukaryotic Genes (CEGs) [44, 45]. Subsequently, initial annotation for each transcriptome was done with a BLASTP [46] search (e-value ≤1e-06, minimum match length ≥ 33 amino acids) against the Swiss-Prot database. Blast hits were imported into Blast2GO version 3.1 [47, 48], which was used to assign Gene Ontology (GO) terms to transcripts for each species. Following this BLASTP search, we removed all duplicate sequences within each species that shared the same BLASTP hit in the Swiss-Prot database, retaining the longest sequence with the greatest sequence similarity to the reference sequence. This was done to remove sequences that arose from possible assembly errors and to restrict our analyses to gene level comparisons, rather than also include comparisons across putative isoforms. We refer to this as our most conservative data set and these annotations were used for all analyses unless otherwise indicated. For most of the species concerned here it was not possible to collect RNA-seq data from multiple individuals, which precluded the confirmation of true isoform sequences from assembly errors, by looking for cases of shared intraspecific isoform expression.

Comparison of expressed gene content

To assess expressed gene content shared among species we conducted a clustering analysis to identify sequence clusters between all seven species and an additional chondrichthyan, the elephant shark, *Callorhinchus milii*. This species is a member of the Holocephali, a separate suborder of the Chondrichthyes; we obtained heart RNA-seq data for this species from a recently updated genome assembly of this organism [26]. We subjected the *C. milii* read data to the same FastqMcf trimming and Trinity assembly methods as the RNA-seq data generated in this study. For the clustering analysis we conducted an all against all BLASTP (e-value ≤ 1e-05) search of all protein sequences from the eight species. The results from these BLASTP searches were used in MCLBlastLINE [49], as implemented in [29, 50], to identify homologous sequences between species using an MCL algorithm to cluster protein sequences with an inflation parameter of 1.8 and all other parameters set to defaults. It is possible, using this approach, for paralogues and orthologues to be grouped together in the same sequence cluster; species were considered to share sequence clusters if one or more transcripts from each species were grouped together in the same cluster (hereinafter referred to as an MCL cluster). We tallied all pairwise MCL clusters between all species and those shared between groups of interest (teleost vs. elasmobranch and regional endotherms vs. ectotherms).

In addition to this clustering approach, we sought to identify the enrichment of particular GO terms in elasmobranchs vs. teleosts or involving the regional endothermic species vs. ectotherms. Towards this end, we conducted Fisher's exact tests through FatiGO (filtering mode set to FDR adjusted *p*-value) [51] within BLAST2GO version 3.1 [47, 48] to test whether particular GO terms were overrepresented in comparisons of the four elasmobranch to the three teleost species or in

Marra et al. BMC Genomics (2017) 18:87

Page 5 of 18

comparisons of the three regional endotherms to the four ectotherm species. Each test was two-tailed and allowed us to assess which terms were either overrepresented or underrepresented in our focal group, and filtering at $p < .05$ after FDR correction. Using BLAST2GO, these results were also filtered to obtain a list of the most specific GO terms that were significantly enriched. In this filtering step, if there are multiple GO terms in the same GO hierarchy and they are significantly enriched, then only the most specific term will be retained. For example, the term 'ion channel complex' would be a more specific term than 'transmembrane transporter complex' and 'transmembrane transporter complex' would be a more specific term than 'cellular component'. If 'ion channel complex' and 'transmembrane transporter complex' were both enriched, then the filtering for the most specific terms would remove the term 'transmembrane transporter complex' from the list and only show the term 'ion channel complex' as enriched.

Identification of candidate immunity genes

To classify genes involved in immunity, we assembled a master list of candidate genes involved in both adaptive and innate immune function. This gene list was derived from numerous large-scale mammalian studies that are curated on two databases: InnateDB [52] and the Immunome knowledge base [53]. Using the Swiss-Prot IDs for these candidate genes, we queried our teleost and elasmobranch BLAST data for their presence in the annotations of our most conservative gene set. To ensure that we captured the genes relevant to chondrichthyans, we cross-referenced the adaptive immunity list against the genes identified in the elephant shark genome [26]. Any immunity genes identified in the elephant shark genome that were not in our list were subsequently added to the adaptive immunity list before comparison with our BLAST data.

Positive selection

Positive selection is the fixation of advantageous mutations driven by natural selection, and is one of the fundamental processes behind adaptive changes in genes and genomes, leading to evolutionary innovations and species differences. We sought to identify cases of positive selection in elasmobranchs and in regional endotherms by conducting the branch sites tests for positive selection on putative orthologues shared between all eight species using the codeml package within PAML version 4.8 [54, 55]. For this analysis, orthologues were defined using the clustering analysis described above, with the additional restriction of only considering clusters with a single sequence from each of the eight species (the seven sequenced here, plus elephant shark). For each cluster we aligned the corresponding coding sequence (cds) using the program Probalign v1.1 [56]

with default settings but removing sites where the posterior probability was < 0.6 and retaining only alignments with continuous blocks of aligned sites that covered >50% of the reference protein sequence in the Swiss-Prot database.

Each alignment was then analyzed with two separate models: a null model and an alternative model. For the alternative model, the dn/ds ratio was allowed to vary across the gene and a proportion of sites were allowed to have dn/ds > 1 (model = 2, NSsites = 2, fix omega = 0, initial omega = 1). In the null model, dn/ds was allowed to vary across the gene but fixed at one for the proportion of sites that are allowed to be >1 in the alternative model (model = 2, NSsites = 2, fix omega = 1). We identified selection when the alternative model identified a proportion of sites with a dn/ds >1 and was significantly more likely as determined by a Likelihood Ratio Test (LRT) using the Chi-squared distribution and one degree of freedom. Sites under selection were identified with the Bayes empirical Bayes method (BEB) [57]. Separate runs of each model were conducted, testing for the incidence of positive selection on the branches leading to endothermic taxa (on the branch leading to swordfish and to the lamniformes) and for selection on the branch leading to elasmobranchs. Branch lengths were estimated by PAML for each gene. The program BUSTED [58] was used to confirm the incidence of positive selection via an online server (www.datamonkey.org/busted) using default settings. A multiple sequence alignment is loaded into the BUSTED server, it generates a tree from the alignment, and the user selects foreground branches for assessing positive selection; in each case we selected the branch leading to the elasmobranch ancestor to test for positive selection.

Results

RNA quality was similar across both wild caught and laboratory collected specimens with little RNA degradation as indicated by high RIN and RQN scores (Average of 7.0 across all libraries and above a minimum of 6.0 for RIN and 5.1 for RQN) on an Agilent 2100 BioAnalyzer or AATI Fragment Analyzer™. The basic sequencing and assembly statistics for the seven heart transcriptomes are summarized in Table 1 and the reads are deposited within the bioproject PRJNA313962. The values reported here follow stringent filtering, with an average of 14,737,476 reads retained per species, and which were subsequently assembled into an average of 121,517 Trinity transcripts per species. To remove non-coding RNA and bioinformatic artifacts from the Trinity assembler we assessed the longest open reading frame (ORF) for each transcript and obtained BLASTP annotations for ORFs from each species, ranging in numbers from 22,491-50,494, with an average of 32,474. The

Marra *et al. BMC Genomics* (2017) 18:87

Page 6 of 18

Table 1 Descriptive statistics of quality filtered reads and the subsequent assembly/annotation of 7 heart transcriptomes

Species	Filtered sequence reads	Trinity transcripts	n50	ORFs	Annotated ORFs	Swiss-Prot Proteins	MCL Clusters	Complete coverage of CEGs	Total Coverage of CEGs
White shark	23,288,212	174,288	2,294	46,825	34,986	11,705	7,021	95.2%	98.0%
Great hammerhead	22,977,042	179,367	2,649	49,992	37,700	10,991	6,535	96.4%	98.4%
Mako	10,802,234	101,534	779	30,354	22,491	11,135	6,980	80.2%	90.7%
Ray	12,921,908	104,957	2,004	31,672	23,265	10,651	6,437	98.0%	99.2%
Swordfish	16,524,630	144,990	2,778	63,771	50,494	11,717	6,788	98.0%	100%
Surgeonfish	8,983,274	81,222	2,864	42,302	34,302	9,530	5,999	91.1%	94.4%
Hogfish	7,665,032	64,264	2,168	28,403	24,077	10,412	6,184	87.9%	94.0%

Trinity transcripts refer to the initial number of transcripts in the Trinity assembly, which were then filtered for those containing the longest open reading frames. The translation of these transcripts was then annotated with BLASTP against the Swiss-Prot database and the number of hits to unique Swiss-Prot entries was recorded; if more than one transcript matched the same Swiss-Prot entry then the longest and most significant match was retained. A CEGMA analysis was conducted to evaluate the coverage of Core Eukaryotic Genes with complete coverage representing the proportion of CEGs with "complete" matches and the total coverage representing the percentage of CEGs that had complete or partial matches in the transcriptome
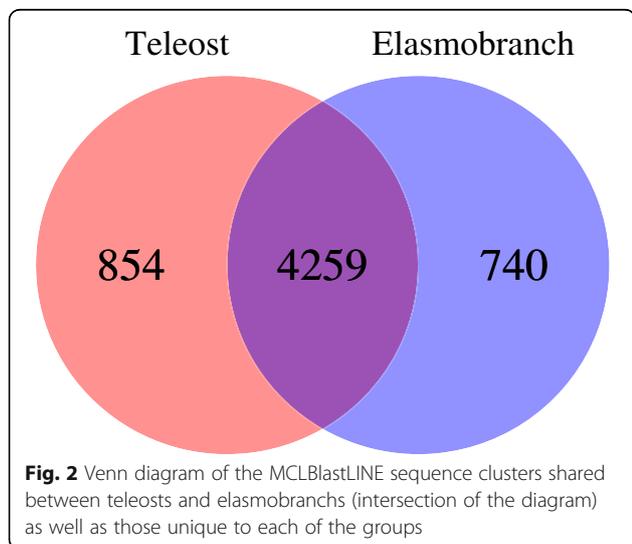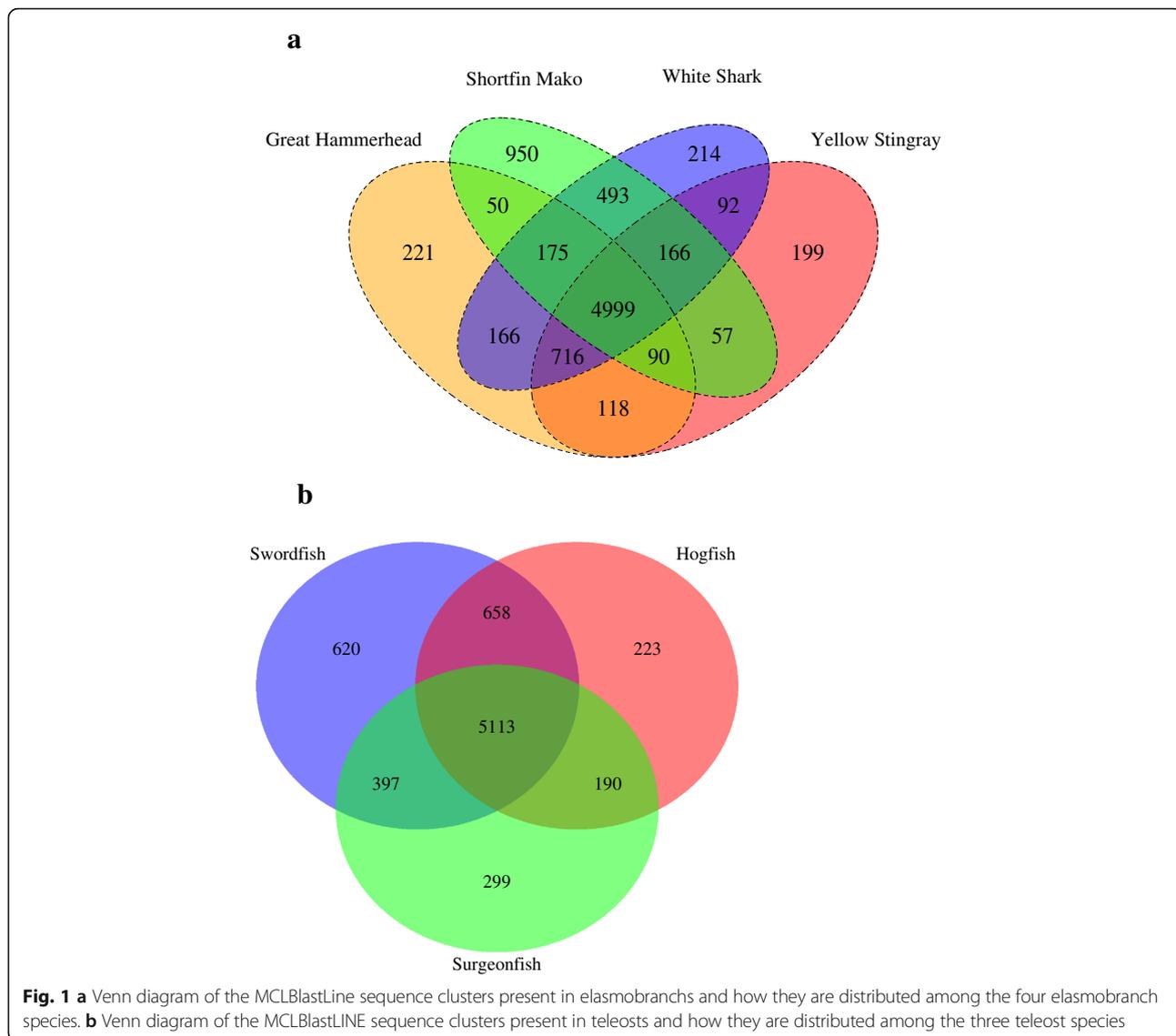
CEGMA analysis (Table 1) indicated that all of our species transcriptomes contained the vast majority of CEGs and these were nearly all "complete" with the possible exception of mako, which still had >90% total coverage but with only 80% of the matches judged as "complete". A "complete" match represents cases where a transcript has an alignment length ≥ 70% of the CEG protein length.

The MCLBlastLINE analysis resulted in sets of homologous clusters of proteins, which were compared across taxa. Considering first elasmobranchs, more clusters were uniquely shared between mako and white shark relative to white shark and hammerhead (493 and 166 respectively; Fig. 1a). Mako and hammerhead shared many fewer homologues; this combination of similarities and differences of clusters across the shark species included in our dataset is both a reflection of the closer evolutionary relationship of mako and white shark, compared to hammerhead, and of the lower output associated with the mako sequencing run. Mako had the largest number of unique clusters (950) among elasmobranchs (possibly a reflection of the somewhat poorer quality of the data for this species; see for example mako n50, Table 1). Despite rays being separated from sharks by about 300 million years of evolution [2, 3, 41], the yellow stingray had a similar number of unique clusters (199) as white shark (214) and great hammerhead (221). The shared set of elasmobranch heart transcriptome MCL clusters was 4,999 (Fig. 1a), and a similar sized core set was apparent in the comparison involving the three teleosts (5,113, Fig. 1b). A large number of these MCL clusters were shared between all seven species (4,259, Fig. 2) with a similar number unique to teleosts, as well as to elasmobranchs (Fig. 2). The number of clusters restricted to only elasmobranchs and only teleosts, represented 14.8% and 16.7% of the complete elasmobranch and teleost clusters, respectively. When looking at the clusters shared among endothermic species we found 5,192 core clusters shared between the three species (white shark, mako, and swordfish). There were 4,259
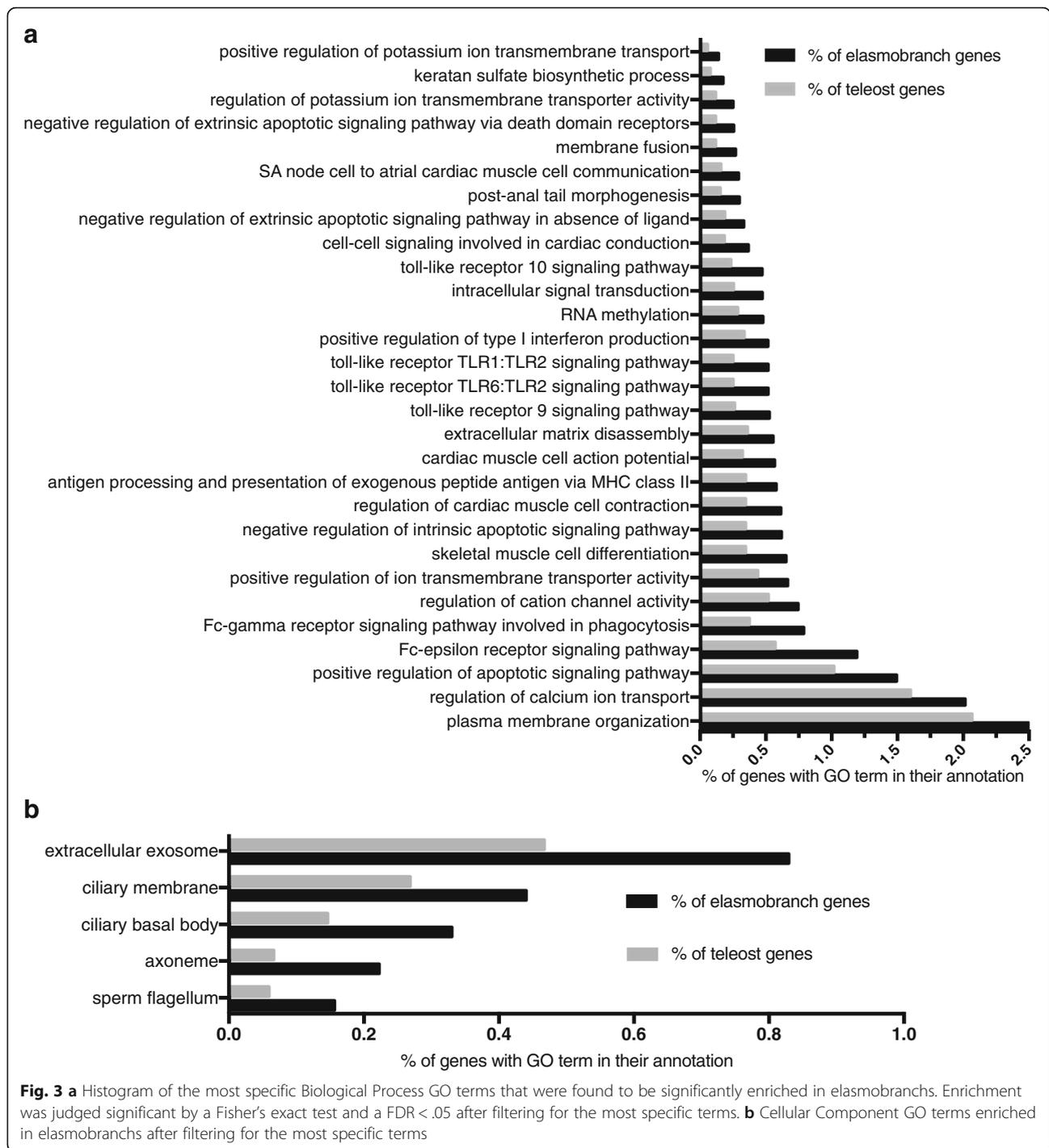
clusters shared between endotherms and ectotherms (i.e. all seven taxa) yielding 933 (18% of the total) clusters unique to endotherms and 473 (10% of the total) unique to ectotherms (Additional file 1: Figure S1 and Additional file 2: Figure S2).

GO content and enrichment tests

Additional file 3: Figure S3, Additional file 4: Figure S4, and Additional file 5: Figure S5 show the top 10 most prevalent GO terms for the three main GO categories: Biological Process (BP), Molecular Function (MF), and Cellular Component (CC) in elasmobranchs and teleosts (a. and b. in each fig., respectively). On the whole, teleosts and elasmobranchs share many of the same GO categories and the same proportions of their transcriptomes are annotated with the most prevalent GO terms. However, there does appear to be limited variation within groups (e.g. within elasmobranchs) for the highest frequency GO terms. Enrichment tests did detect statistical differences (i.e. between both elasmobranchs vs. teleosts and endotherms vs. ectotherms) in the representation of GO terms present in lower frequencies within the transcriptomes. A Fisher's exact test revealed that a total of 93 GO terms were enriched in elasmobranchs (Additional file 6: Table S1) and 97 were enriched in teleosts (Additional file 6: Table S2) after an FDR correction (<.05 post FDR). When filtering these for the most specific GO terms, there were 34 that were enriched in elasmobranchs (four additional terms were removed that were linked to possible symbionts or contaminants). A total of 29 of these terms belonged to the BP category, five were in the CC category; the proportion of genes from elasmobranchs and teleosts that were annotated with these terms is displayed in Figs. 3a and b, respectively. There were 30 GO terms that were enriched in teleosts when filtering for the most specific GO terms (two additional terms were linked to possible symbionts or contaminants), of these 14 were BP terms, seven were CC terms, and nine were MF terms (Figs. 4, a, b, c).

a

b

Fig. 1 a Venn diagram of the MCLBlastLine sequence clusters present in elasmobranchs and how they are distributed among the four elasmobranch species. **b** Venn diagram of the MCLBlastLINE sequence clusters present in teleosts and how they are distributed among the three teleost species

Fig. 2 Venn diagram of the MCLBlastLINE sequence clusters shared between teleosts and elasmobranchs (intersection of the diagram) as well as those unique to each of the groups

Among the most specific GO terms enriched in teleosts, only two are related to innate immunity ("Toll signaling" and "Toll-like receptor 1 signaling"), and an additional GO is present that may be associated with pathogen removal ("phagocytosis, engulfment"). In contrast, six different GO terms were enriched in elasmobranchs that are involved in innate immunity (five of which are various Toll-like receptor signaling pathways, and the sixth is "positive regulation of type I interferon production"). Additionally, three adaptive immunity GO terms were enriched in elasmobranchs ("Fc-epsilon receptor signaling pathway", "Fc-gamma receptor signaling pathway involved in phagocytosis", and "antigen processing and presentation of exogenous peptide via MHC class II"). These terms are all involved in antibody-mediated immunity; either in recruitment of immune cells (Fc receptors) or in antigen presentation.
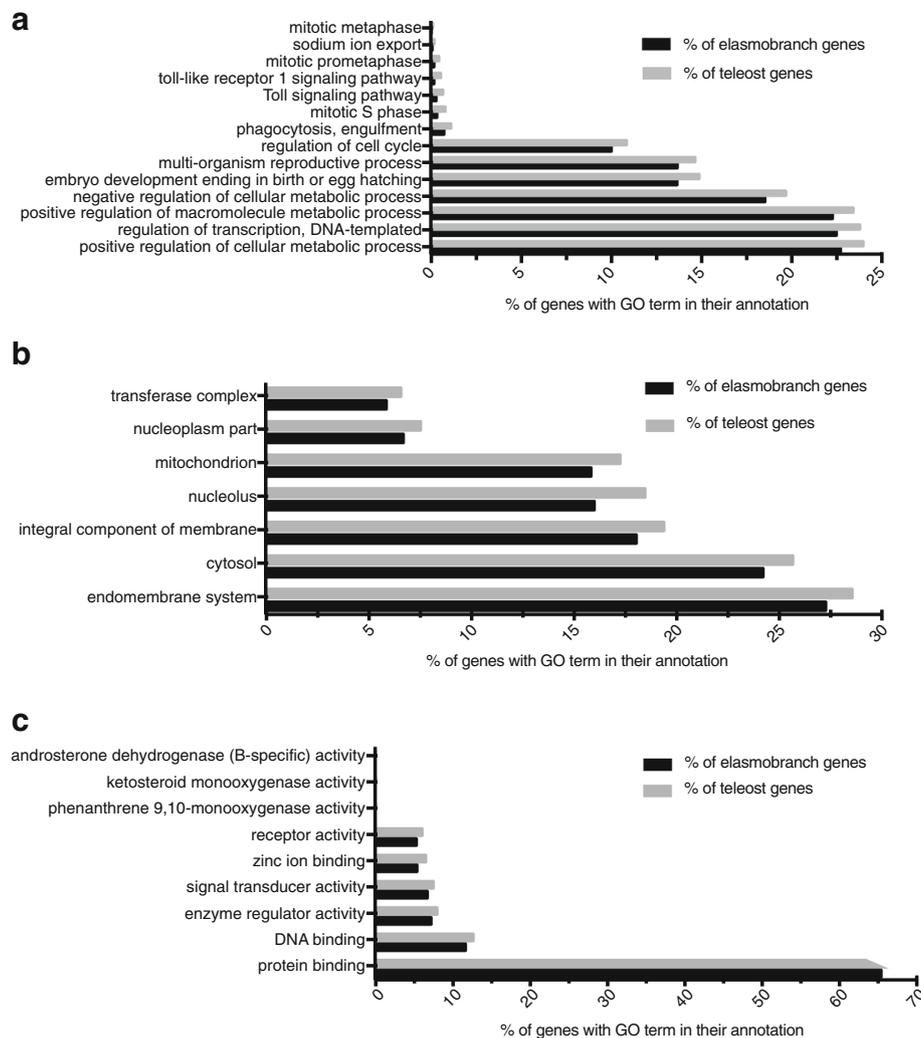
Marra *et al. BMC Genomics* (2017) 18:87

Page 8 of 18

Fig. 3 a Histogram of the most specific Biological Process GO terms that were found to be significantly enriched in elasmobranchs. Enrichment was judged significant by a Fisher's exact test and a FDR < .05 after filtering for the most specific terms. **b** Cellular Component GO terms enriched in elasmobranchs after filtering for the most specific terms

The Fisher's exact test involving the endotherm vs. ectotherm comparison yielded 15 GO terms that were enriched in endotherms (five of which were driven by possible xenobiotics, e.g. bacterial contaminants, pathogens, or commensal organisms, and removed; Fig. 5) and when considering the most specific terms, seven were enriched in endotherms (one CC, six BP); see Additional file 6: Table S3). Although relatively few

GO terms were enriched in endotherms, several are of considerable interest including terms describing genes involved in regulation of cardiac muscle cell contraction.

Unique gene content

In addition to characterizing gene content by clustering analyses and looking at gene ontology information
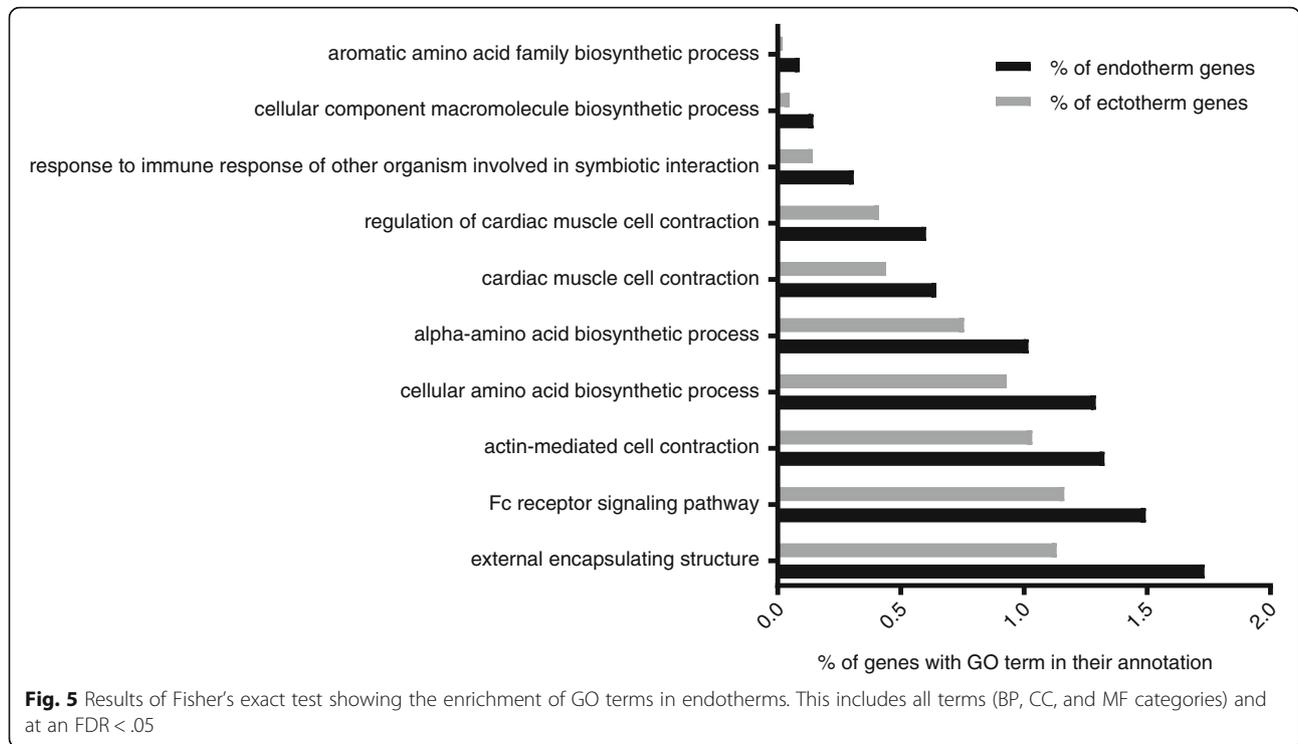
Marra *et al. BMC Genomics* (2017) 18:87

Page 9 of 18

Fig. 4 a Histogram of the most specific Biological Process GO terms that were found to be significantly enriched in teleosts by a Fisher's exact test at an FDR < .05 and after filtering for the most specific terms. **b** Cellular Component GO terms enriched in teleosts after filtering for the most specific terms. **c** Molecular Function GO terms enriched in teleosts after filtering for the most specific terms

to determine possible functional differences between the transcriptomes, we looked at genes whose expression was restricted to elasmobranchs or to endotherms. We identified 262 Swiss-Prot annotated genes that were restricted to elasmobranchs, three of which were from possible xenobiotics (i.e. sequences that resulted from possible microbial contaminants, pathogens, or symbionts present in the tissue sample). The 259 remaining Swiss-Prot annotated genes that were present in all elasmobranch transcriptomes are listed in Additional file 6: Table S4. These genes span a variety of functions, as indicated by their GO annotations, which included metabolic, gene regulatory, and immunity related terms, among others. There were a few genes (19) that were uniquely expressed in endotherms (listed in Additional file 6: Table S5) with several playing possible roles in energy metabolism.

Immune genes

We also searched for the presence of candidate genes involved in innate and/or adaptive immunity in elasmobranchs and teleosts. In particular, we searched for the presence of 911 innate immunity genes and 862 adaptive immunity genes, with 272 of these being present in both categories. When we cross-referenced our list of candidate genes with the annotations of our most conservative gene lists, we identified 736 innate immunity genes and 599 adaptive immunity genes present in at least one of our seven heart transcriptomes (Fig. 6 shows the distribution of these numbers across all seven species). This included 404 innate immunity genes and 217 adaptive immunity genes present in all four elasmobranch species. Within these there were 17 innate and seven adaptive immunity genes whose expression were absent

Marra *et al. BMC Genomics* (2017) 18:87

Page 10 of 18

Fig. 5 Results of Fisher's exact test showing the enrichment of GO terms in endotherms. This includes all terms (BP, CC, and MF categories) and at an FDR < .05
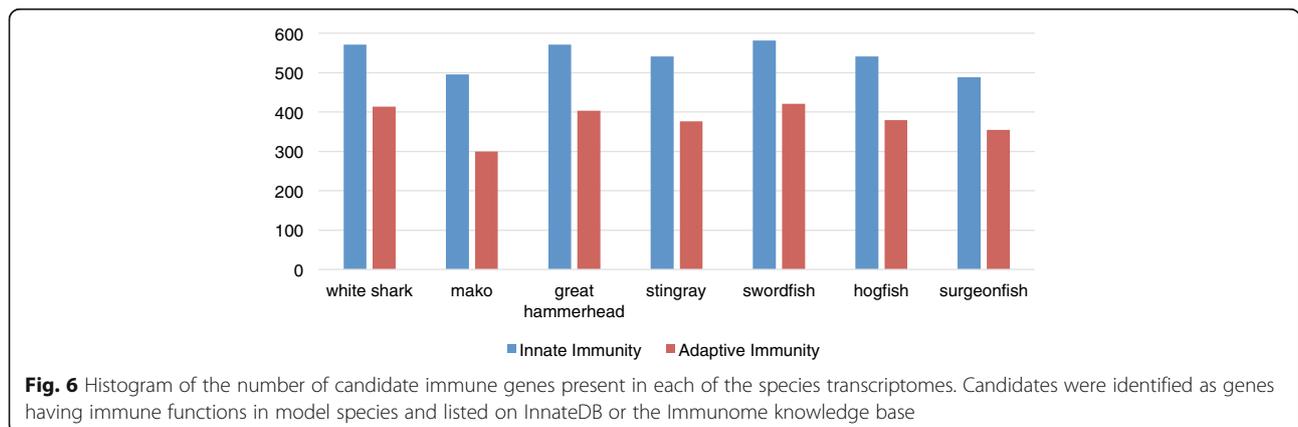
in heart tissue of teleosts, pointing to a substantial proportion of the heart expressed gene content that is unique to elasmobranchs, being due to expression of immunity related genes (24 of 259 unique elasmobranch genes). In addition to the immunity genes that were expressed in all four elasmobranchs and none of the teleosts; there were 48 additional innate immunity and 37 additional adaptive immunity genes whose expression were absent in teleosts but present in two or more of our elasmobranch species.

Positive selection
After requiring a single sequence to be present for each species prior to multiple sequence alignment, we were left with 1,332 MCL clusters as possible input for testing

positive selection. A further 400 MCL clusters were unable to produce alignments that contained continuous sequence for all eight species, leaving 932 genes that were used as input for Probalign to build consensus quality alignments. Probalign does an additional filtering by removing poor quality alignments, and we further required all alignments to cover > 50% of the sites in the Swiss-Prot reference sequence. After all of these careful filtering steps we were left with 472 high quality alignments to test for the presence of positive selection. All alignments that yielded evidence of positive selection were individually inspected for possible alignment errors and those with obvious misalignments (e.g. inappropriate insertions or deletions) were removed from the dataset.

Fig. 6 Histogram of the number of candidate immune genes present in each of the species transcriptomes. Candidates were identified as genes having immune functions in model species and listed on InnateDB or the Immunome knowledge base

Marra *et al. BMC Genomics* (2017) 18:87

Page 11 of 18

Analysis of the elasmobranch branch yielded a number of genes that had significant evidence of positive selection (34 genes) with three genes remaining significant after FDR of 0.05, and six at an FDR of 0.10 (Additional file 7: Table S6). This included (at FDR *p*-value < 0.05) a gene of importance to immune function, legumain (*LGMN*, image of alignment available in Additional file 8: Figure S6*)*, which had several sites under selection as identified by BEB. This gene is of particular interest since it was one of our candidate genes in both the innate and adaptive immunity gene lists, and it plays a role in some of the antigen processing steps that were enriched in the elasmobranch GO annotations. The other two genes that displayed strong evidence of positive selection were *Tim22* (mitochondrial import inner membrane translocase subunit TIM22, image of alignment available in Additional file 9: Figure S7) and *Bag1* (BAG family molecular chaperone regulator 1, image of alignment available in Additional file 10: Figure S8), which are involved in transporting proteins into the mitochondria and in regulating various cellular signaling pathways such as apoptosis, respectively.

To provide further confirmatory evidence of positive selection on legumain in elasmobranchs, we employed the program BUSTED [58], looking for episodic positive selection along the branch leading to the elasmobranch ancestor. This test confirmed the branch-site test results, finding significant evidence of diversifying selection (*p* = 0.008) with 11% of sites having a dn/ds of 2.65, indicating an elevated rate of non-synonymous substitutions for our selected branch. The Bayes Empirical Bayes (BEB) analysis conducted during the codeml analysis of legumain indicated that two sites had > 95% probability of being under positive selection in elasmobranchs. These two sites are residues 105 and 261 in the human legumain protein sequence. Neither amino acid is part of the active site (which is composed of a cysteine at residue 189, a histidine at 148, and an asparagine at 42 [59]). However, both sites under selection in elasmobranchs reside at the start or end of a possible helix motif in the secondary structure for the human legumain protein as described on its Swiss-Prot entry. The change at residue 261 is located within the fifth alpha helix of the protein and is also two amino acids before a glycosylation site in the protein [59].

The positive selection analyses on endotherms yielded nine genes where the alternative model was significantly better than the null model and that had a proportion of sites with a dn/ds >1 (Additional file 7: Table S7). One of these genes, catalase (*CAT*), remained marginally significant at a false discovery rate cutoff of 0.10, and is involved in removing oxidative species from the cell, an important complementary process to elevated metabolic function and respiratory rate in an endothermic species [60–63].

Discussion

Our goal here was to characterize transcriptomic differences between elasmobranchs and teleosts, and more specifically, to identify expressed gene differences related to immune system function between these two taxonomic groups. In addition, by including representative species of both ectotherms and endotherms from each of the two lineages, we aimed to identify expressed loci that might have evolved convergent roles linked to endothermy. In so doing, we conducted transcriptome sequencing and assembly, which provides significant genetic resources for several taxa, and particularly the elasmobranchs, where transcriptomic/genomic resources were previously lacking.

From our assembly statistics it was apparent that the species with the greatest number of sequencing reads yielded many more assembled transcripts, ORFs, and annotated ORFs. However, if we consider the n50 statistics, most species were quite similar with n50 values between two and three kb (with the exception of mako shark at 779 bp), which are on par with similar recent RNA-seq experiments in other non-model species [64] and larger than single tissue RNA-seq assemblies in spotted cat-shark [31]. The number of unique Swiss-Prot annotations was similar across species (ranging from 9,530 to 11,717, and a mean = 10,877 genes, Table 1) and agrees with similar estimates in heart transcriptome studies in non-model species [65]. We identified 5,358 Swiss-Prot gene annotations that were shared between all seven of our species, indicating that between 46-56% of gene annotations in each species were shared among the rest. When looking at our MCL clusters, the number of clusters that included all species represented from 61-71% of the transcriptome for each species.

In terms of shared orthologues, comparisons to other studies can be difficult to make because large interspecific comparisons are still somewhat nascent, and there are often differences in sequencing technology, output, or ortholog retrieval. Furthermore, the physiological and developmental state from which the animals were sampled will have a significant bearing on the expressed profile of each species [66]. Nonetheless, several studies are worth comparing to our results. Zhang et al. [67] recently compared interspecific spleen transcriptomes of house finch and zebra finch (same order, different families) and found that less than half of all genes expressed in either species were expressed in both species. Similarly, Elmer et al. [68] reported approximately 50% orthologs in transcriptomes of two congener species of cichlid fishes. Thus, the proportion of expressed heart loci that were shared amongst our species is not atypical compared to other studies, especially considering the number of species sampled herein.

We observed that although most species shared a core set of genes (about 50% of gene annotations in each

Marra *et al. BMC Genomics* (2017) 18:87

Page 12 of 18

species are shared across the seven species) expressed in heart tissues, there was also a set of genes whose expression was limited to each species and/or group of species. It is possible that differences between species and/or groups could be driven by either (or both) the presence/ absence of these hundreds of genes or by regulatory mechanisms driving differential expression [69, 70]. Evaluating whether expression differences between taxa was more of a determining factor than gene presence/ absence would require sampling of multiple individuals of each taxon and confirmation of expression values with additional methods such as qPCR. However, and particularly for this sort of tissue, such sampling is not feasible given the difficulties of accessing these large-bodied species in the wild and the conservation status of the sharks, thus we restricted our analysis to the presence/absence rather than levels of expression. Nonetheless, our characterization of the differences in GO term representation, genes whose expression was limited to particular groups, and analysis of positive selection identified genes and pathways that may be of importance to biological differences between these taxa.

Genetic differences between elasmobranchs and teleosts

The differences in the genes that are expressed solely in elasmobranchs, the enriched GO terms, and the genes under positive selection provide biological insights regarding possible distinguishing factors between elasmobranch and teleost lineages. In particular, a number of immunity related genes differed in their presence and sequence evolution in elasmobranchs. The number of GO terms enriched in elasmobranchs and in teleosts are relatively similar, however, between each of these two sets of genes there are differences that have a direct link to adaptive and innate immunity. It is known that elasmobranchs have a different genomic organization of various adaptive immunity genes compared to other vertebrates [8, 10, 11]. Our GO enrichment results indicate that these genetic differences in adaptive immunity of elasmobranchs exist not only at the genomic but also at the transcriptional level. It is important to identify these differences since it can lead to a better understanding of both the ancestral mechanisms underlying adaptive immunity and those characteristics which have been selectively maintained over an evolutionary timescale that far outstrips that of mammals [6].

The differences in the immunity gene repertoire between teleosts and elasmobranchs was not only apparent in GO category differences but was also evident in the presence of immunity related genes whose expression was restricted to several and/or all elasmobranchs. Similar numbers of candidate adaptive and innate immune genes are present across elasmobranchs and teleosts (Fig. 6). However, when looking at the genes that were

expressed only in elasmobranchs, several interesting immunity genes are apparent. This included genes such as "MHC class II antigen, DP beta 1 chain", 'Ig gamma 2 chain c region', and 'Interferon gamma receptor 1' which echo the enrichment of GO terms associated with antigen presentation and Fc-receptor signaling in elasmobranchs. This also included genes that were not present in the elephant shark genome. Specifically, five of the seven adaptive immune genes restricted to the four elasmobranchs and 30 of the 37 adaptive immunity genes expressed in two or more of the elasmobranchs (but absent from teleosts), were absent in the reported elephant shark genome. This points to possible differences in immune gene content between elasmobranchs and holocephalans that requires confirmation in a comparison between whole genome sequences of representative species from these two suborders of the Chondrichthyes.

Immunoglobulins in elasmobranchs have received attention in the literature due to differences with higher vertebrates in genomic arrangement and the presence of a novel class of immunoglobulins (new antigen receptor, IgNAR, [7, 8, 10, 11]). We did not detect the presence of IgNAR in our Swiss-Prot annotations; however, this does not preclude the presence of the underlying genes in these species. We did detect three different IgW genes whose expression was limited to one or more elasmobranchs. IgW is similar in form to IgD in mammals but is largely restricted to cartilaginous fish [71–73]. This included a secretory form of the IgW protein that was present in all elasmobranchs, except ray, suggesting its consistent expression. Additionally, there are multiple IgG c-domain genes that have significant matches to sequences in the heart transcriptomes of two or more of our elasmobranch species. Although IgG genes are thought to have arisen in amphibians [74, 75], a form of this molecule has been observed in camels [76] that behaves similarly to IgNAR. Instead of the classical two heavy chain, two light chain, formation of most antibodies, both the camelid IgG and the shark IgNAR are composed of only heavy chains. Given this information and the paucity of elasmobranch IgNAR sequences in Swiss-Prot (only one sequence from nurse shark, *Ginglymostoma cirratum*), we suspect that some of these IgG sequences probably represent Ig molecules typical of cartilaginous fishes, with IgNAR or IgW domains that are not well represented in the curated Swiss-Prot database. Indeed, several of these sequences have better matches to IgNAR or IgW proteins in the NR database. This highlights the need for further identification of all of the requisite domains of IgNAR and the need for additional exploration of these genes in elasmobranchs from a genomic perspective.

Our analyses identified significant evidence of positive selection on several elasmobranch genes including

Marra *et al. BMC Genomics* (2017) 18:87

Page 13 of 18

legumain (*LGMN*), which plays an important role in various aspects of immunity, including processing antigens for presentation by MHC class II [59, 77, 78]. Not only is legumain of interest because of its immune functions but it has also been reported to be overexpressed in several cancers [79] and used as a target to eliminate tumor associated macrophages [80]. Elasmobranchs are claimed to have the lowest incidence of malignant neoplasia (tumors) of any vertebrate group [81], however, this remains highly controversial [82], and insufficient information exists from which to base any firm conclusions. At the same time, it is of some note, that there are several publications reporting various shark extracts as having anti-tumorigenic properties [83–85]. Overexpression of legumain occurs in tumor-associated macrophages and is thought to contribute to the tumor promoting inflammation associated with most malignancies [80, 86]. In addition to over or under expression of a gene, altering the structure of the protein, through key substitutions, driven by the action of natural selection, could change its overall function or more subtly alter its role. It is quite possible that the evidence for molecular adaptation that we detect in legumain reflects a modified role in elasmobranchs relative to other vertebrates. We feel it will be important to evaluate the purpose and effect of adaptive changes in this protein in elasmobranchs, providing insight on the impact of these changes on legumain's role in immunity and its interplay with macrophages. Another gene with a potential role in cancer, which was also under positive selection, and remained significant after FDR correction, was *Bag1* (Bcl2 associated athanogene). *Bcl2* is an oncogene that encodes a membrane protein that blocks a step in a pathway leading to apoptosis [87]. *Bag1* encodes a protein that binds to *Bcl2* and enhances the anti-apoptotic effects of *Bcl2* [88]. Apoptosis, or programmed cell death, is a process that eliminates dysfunctional cells and one of the hallmarks of cancer is the ability of malignant cells to evade apoptosis. One of the proposed modes of action of epigonal (shark lymphomyeloid tissue) cell medium as an anti-tumorigenic substance is in inducing apoptosis [85]. This suggests the possible conjectural hypothesis that positive selection on *Bag1* in elasmobranchs may have altered its tendency to enhance the anti-apoptotic effects of *Bcl2*.

In addition to immunity related functions, annotations of elasmobranch specific genes included terms related to various other roles such as cellular responses (e.g. "signal transduction"), transport (e.g. "intracellular protein transport"), and control of gene expression (e.g. "positive regulation of transcription from RNA polymerase II promoter"). This also included genes not only restricted to elasmobranchs, but genes under positive selection on the elasmobranch branch. Of the three genes that had

evidence of positive selection, two of these also lacked an obvious immune function. One of these we already discussed—*Bag1*—the other is *Tim22*, which encodes part of the identically named TIM22 complex. This complex is necessary to transport nuclear encoded proteins that are then embedded in the inner mitochondrial membrane [89]. Differences in the function of this gene could have impacts on basic cellular functions such as cellular respiration and energy production.

It is also possible that some of the differences identified in the elasmobranch transcriptomes, compared to teleosts, are related to their role in wound healing. Sharks, rays, and skates are thought to have the ability to heal wounds relatively quickly, based on various anecdotal reports and observational studies [90–94]. However, the molecular mechanisms of wound healing in elasmobranchs are unclear. In mammals wound repair consists of several distinct stages including an inflammatory response during which various innate and adaptive immune cells such as neutrophils, macrophages, and lymphocytes are recruited to the wound site [95]. The proper timing of these stages is necessary for effective healing, and an important inhibitory factor in the overall process is the development of infection, which can lead to further damage or prolonged inflammation [96]. Thus genes and pathways associated with promoting and controlling inflammation as well as those involved in clearing cellular debris and pathogens are critical to the wound healing process. Some of the processes that we have highlighted as enriched GO terms in elasmobranchs may play an integral complementary role in preventing infection and perhaps even in direct regulation of wound repair. Previously, we discussed the enrichment of several GO terms such as those associated with Fc receptor signaling, and in particular the "Fc-gamma receptor signaling pathway involved in phagocytosis". Fc-gamma receptors are expressed on a variety of immune cells including neutrophils, mast cells, macrophages, and dendritic cells. When these receptors bind to an antibody/antigen complex they can initiate a variety of responses from stimulating inflammation to initiating phagocytosis of foreign cells (reviewed in [97]), that are important factors in preventing infection and ensuring proper progression of the wound healing process. Thus the genes associated with this process may have roles in wound healing that are complimentary to and/or work in concert with their function in elasmobranch immunity.

Another possible connection between the elasmobranch enrichment results and wound healing is with regard to MHC class II. GO processes associated with antigen presentation by MHC class II were enriched in elasmobranchs, one MHC class II gene had expression restricted to elasmobranch species, and *LGMN* (also

Marra *et al. BMC Genomics* (2017) 18:87

Page 14 of 18

involved in processing antigens for presentation by MHC class II) was under selection in elasmobranchs, serving to highlight the importance of the MHC class II complex in elasmobranchs. Interestingly, there is experimental evidence in mouse models showing that wound healing is impaired in the absence of MHC class II. Schäffer et al. [98] demonstrated the importance of MHC class II in the healing of wounds to the skin; they found that wound collagen deposition and wound breaking strength were impaired in MHC-class-II-knockout mice. A study of myocardial infarction in mice found that the absence of CD4[+] T-cells (which interact with MHC class II) in CD4 knock-outs, resulted in greater mortality and abnormal collagen formation compared to wild-type mice, with even greater mortality in mice that lacked all four MHC class II genes [99]. These studies indicate the importance of MHC class II to not only adaptive immunity but also, more specifically, to tissue repair after injury. To date our understanding of the MHC in elasmobranchs is limited to a few species, and the genomic organization of the region is relatively unknown [100]. Future efforts are needed to understand the organization and function of the MHC in elasmobranchs as part of understanding immune response and the mechanisms by which elasmobranchs are able to regulate efficient healing.

Gene content associated with regional endothermy

Our hypothesis is that there should be patterns, either in terms of gene content/annotation or in terms of positive selection, that are shared by regionally endothermic fishes, reflecting the convergent evolution of mechanisms that raise the temperature of some organs above ambient. Although the heart is often at ambient temperature in endothermic fishes such as tunas [18, 19, 101], it is still required to supply oxygenated blood to the metabolically active tissues, with accompanying demands placed on its function in regionally endothermic fishes, and with associated impacts on gene expression [24]. Thus we expected that the heart transcriptomes would contain gene content and annotation differences in regional endotherms relative to ectotherms. Indeed, we did identify shared molecular characteristics in the regional endotherms, in terms of GO differences compared to ectotherms, and in terms of genes whose expression were restricted to the endothermic species. There were a total of 15 GO terms enriched in the endothermic taxa involved in our comparison, including a number of categories of likely biological importance to an endothermic physiology, such as the terms "cardiac muscle cell contraction" and "regulation of cardiac muscle cell contraction". This coincides with expectations that endothermic species have increased cardiac function to accommodate increased metabolism, heart rate, and overall activity [18, 39, 63, 102–104].

There were a total of 19 genes whose expression was restricted to endotherms. The GO terms that describe these genes included several BP terms linked to metabolism, including for instance, the expression of two genes, *ELOVL6* and *TM6SF2*, restricted to endotherms in our dataset and that are involved in lipid metabolism. *ELOVL6* encodes 'Elongation of very long chain fatty acids protein 6', which catalyzes the elongation of fatty acids [105]. The function of *TM6SF2* is not well characterized, but it has recently been shown that in humans this gene regulates triglyceride secretion and fat metabolism in the liver [106] and certain variants have been associated with lowered risk of myocardial infarction [107, 108]. In addition to these genes, *MGAM2*, a gene annotated with the GO term: "carbohydrate metabolic process", was also expressed only in the endotherms. Though there is little information on this gene, its similarity to its paralogue, *MGAM* and its structure have led to its description on Swiss-Prot as being involved in hydrolysis of 1,4-alpha-D-glucose. Considering the higher energy needs and higher metabolic rates of endothermic fishes relative to ectothermic fish [19], the expression of these metabolic genes in the endothermic species may reflect the need for efficient energy storage and metabolism in these species. Elevated expression of genes dealing with lipid and carbohydrate metabolism were also detected in cold acclimated Pacific bluefin tuna, highlighting the importance of metabolic genes to heart function during maintenance of elevated temperature in other tissues of partially endothermic species [24]. The increase in expression of lipid metabolism genes in cold acclimated fish was suggested to lead to an increase in lipid formation and changes in membrane composition [24]. It is possible that some of the genes identified here (e.g. *ELOVL6*), known for their function in other tissues, are playing a role in these metabolic and structural processes in heart tissue of endothermic fishes.

In addition to this evidence suggesting the importance of cardiac function and lipid metabolism in the regional endotherms, we identified limited evidence of positive selection on the gene that encodes the protein catalase. This enzyme is responsible for breaking down hydrogen peroxide and thus preventing damage to the cell from oxidative species [61]. For a species with increased metabolic rate this would be a key process to prevent damage to the cell from H_2O_2 produced following electron transport at the end of cellular respiration [60, 63]. The evidence for selection on this gene is, however, marginal after correction for FDR and would require further validation with full-length sequence from lamnid sharks (we possessed only a fragment of the sequence for mako), billfishes, and tunas. Nonetheless, the importance of this gene for protection against oxidative damage makes it an excellent candidate for future study. Indeed,

Marra *et al. BMC Genomics* (2017) 18:87

Page 15 of 18

microarray study of the heart transcriptome of the Pacific bluefin tuna at various temperatures have identified changes in the expression of genes dealing with oxidative stress under cold temperatures [24]. This suggests additional study of the evolution of catalase and other genes dealing with reactive oxygen species might well prove a fruitful line of investigation as a secondary adaptation to the metabolic demands in regionally endothermic species.

Conclusions

In this study we utilized RNA-seq of heart transcriptomes in seven species in order to examine differences in gene content and patterns of positive selection in elasmobranchs relative to teleost fishes, as well as in regional endothermic species. We have uncovered several lines of evidence pointing to differences between elasmobranchs and teleosts in genes underlying notable functional properties. This was particularly highlighted by differences in the expression and evolution of immunity genes in the elasmobranchs, representing the earliest vertebrate lineage with adaptive immunity. Our study also identified genes that may have evolved convergent roles in the phenomenon of regional endothermy, characteristic of a select few elasmobranchs and teleosts. In addition to the biological differences discussed herein, these transcriptome data provide significant genetic resources for future studies in a group (elasmobranchs) that is severely lacking in genomic resources.

Additional files

Additional file 1: Figure S1a. MCLBlastLINE sequence clusters shared among ectotherms. **Figure S1b.** MCLBlastLINE sequence clusters shared among endotherms. (PDF 573 kb)

Additional file 2: Figure S2. MCLBlastLINE sequence clusters shared between endotherms (left) and ectotherms (right). (PDF 188 kb)

Additional file 3: Figure S3a. Top 10 Biological Process GO terms in elasmobranchs. Bars represent the proportion of contigs in each transcriptome that are annotated with the respective GO term. **Figure S3b.** Top 10 Biological Process GO terms in teleosts. Bars represent the proportion of contigs in each transcriptome that are annotated with the respective GO term. (PDF 311 kb)

Additional file 4: Figure S4a. Top 10 Molecular Function GO terms in elasmobranchs. Bars represent the proportion of contigs in each transcriptome that are annotated with the respective GO term. **Figure S4b.** Top 10 Molecular Function GO terms in teleosts. Bars represent the proportion of contigs in each transcriptome that are annotated with the respective GO term. (PDF 409 kb)

Additional file 5: Figure S5a. Top 10 most abundant Cellular Component GO terms for elasmobranchs. Histograms represent the proportion of contigs in each transcriptome that are annotated with the respective GO term. **Figure S5b.** Top 10 most abundant Cellular Component GO terms for teleosts. Histograms represent the proportion of contigs in each transcriptome that are annotated with the respective GO term. (PDF 290 kb)

Additional file 6: Table S1. List of GO terms that were significantly enriched in elasmobranchs relative to teleosts. The list includes all terms including multiple GO terms that may be enriched from xenobiotics or annotations of the genes function in other tissues. Within the "Category" column, BP is biological process and CC is cellular component. **Table S2.** List of GO terms that were significantly enriched in teleosts relative to elasmobranchs. The list includes all terms including multiple GO terms that may be enriched from xenobiotics or annotations of the genes function in other tissues. Within the "Category" column, BP is biological process, CC is cellular component, and MF is Molecular Function. **Table S3.** List of GO terms that were significantly enriched in endotherms relative to ectotherms. The list includes all terms including multiple GO terms that may be enriched from xenobiotics or annotations of the genes function in other tissues. Within the "Category" column, BP is biological process, CC is cellular component, and MF is Molecular Function. **Table S4.** List of 259 genes that were present in all four elasmobranchs but not present in any of the three teleost species. The Swiss-Prot ID, suggested protein name, and gene name are listed for each gene. **Table S5.** List of 19 genes that were present in all three endotherms but not present in any of the four teleost species. The Swiss-Prot ID, suggested protein name, and gene name are listed for each gene. (XLSX 71 kb)

Additional file 7: Table S6. Genes with evidence of positive selection in elasmobranchs under the branch sites test after FDR filtering that remain significant at *p* < .10. **Table S7.** Genes with evidence of positive selection in endotherms prior to FDR filtering. (DOCX 19 kb)

Additional file 8: Figure S6. Image of the alignment for legumain (*LGMN*) resulting from Probalign, as displayed in Geneious (v. 7.1.7). Bases that are located in gap regions or areas where the alignment is of low confidence are converted to N's and then removed prior to running the branch sites test. Note that the numbers for residues given in the manuscript correspond to positions in the human reference sequence; the sites under selection according to BEB are flagged with yellow arrows at bases 421 and 889 in the consensus sequence. The alignment file is available upon request. (PDF 4320 kb)

Additional file 9: Figure S7. Image of the alignment for mitochondrial import inner membrane translocase subunit TIM22 (*Tim22*), resulting from Probalign, as displayed in Geneious (v. 7.1.7). Bases that are located in gap regions or areas where the alignment is of low confidence are converted to N's and then removed prior to running the branch sites test. The sites that are under selection according to BEB are flagged with yellow arrows at bases 139, 184, and 190 in the consensus sequence. The alignment file is available upon request. (PDF 1981 kb)

Additional file 10: Figure S8. Image of the alignment for BAG family molecular chaperone regulator 1 (*Bag1*), as displayed in Geneious (v. 7.1.7). Bases that are located in gap regions or areas where the alignment is of low confidence are converted to N's and then removed prior to running the branch sites test. The sites that are under selection according to BEB are flagged with yellow arrows at bases 91 and 133 in the consensus sequence. The alignment file is available upon request. (PDF 2094 kb)

Abbreviations

BAG1: *Bcl2* associated athanogene 1; Bcl2: B-cell lymphoma 2; BEB: Bayes empirical bayes; bp: Base pairs; BP: Biological process; CAT: Catalase; CC: Cellular component; cds: Coding sequence; CEGMA: Core eukaryotic genes mapping approach; ELOVL6: Elongation of very long chain fatty acids protein 6; FDR: False discovery rate; GO: Gene ontology; Ig: Immunoglobulin; igNAR: Immunoglobulin new antigen receptor; LGMN: Legumain; MCL: Markov cluster algorithm; MF: Molecular function; MGAM2: Maltase-Glucoamylase (Alpha-Glucosidase); Mya: Million years ago; ORF: Open reading frame; TIM22: Mitochondrial import inner membrane translocase subunit TIM22; TM6SF2: Transmembrane 6 superfamily member 2

Marra *et al. BMC Genomics* (2017) 18:87

Page 16 of 18

Authors' contributions
NJM conducted GO analyses, sequence alignment, tests of positive selection and drafted the majority of the manuscript, VPR conducted transcriptome assembly and annotation and performed some of the comparative analysis, AE curated candidate gene lists that facilitated comparisons of immune gene content, SMB and PPB conducted RNA extractions and library preparation for sequencing, MJS and MSS conceived, developed, and supervised the project in addition to contributing to drafting the manuscript. All authors have read and approved the final manuscript.

Competing interests
The authors declare that they have no competing interests.

Consent for publication
Not applicable.

Ethics approval
No ethical approval or permit for animal experimentation was required, as the individuals were not sacrificed specifically for this study.

Author details
[1]Department of Population Medicine and Diagnostic Sciences, College of Veterinary Medicine, Cornell University, Ithaca, NY 14853, USA. [2]Save Our Seas Shark Research Center and Guy Harvey Research Institute, Nova Southeastern University, 8000 North Ocean Drive, Dania Beach, FL 33004, USA. [3]Department of Biological Sciences, Clemson University, Clemson, SC 29634, USA. [4]Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, NY 14853, USA.

References
1. Cotton CF, Grubbs RD. Biology of deep-water chondrichthyans: Introduction. Deep Sea Res Part II Top Stud Oceanogr. 2015;115:1–10.
2. Heinicke MP, Naylor GJP, Hedges SB. Cartilaginous fishes (Chondrichthyes). In: Hedges SB, Kumar S, editors. The Timetree of Life. Volume IX. Oxford: University Press; 2009. p. 320–7.
3. Renz AJ, Meyer A, Kuraku S. Revealing less derived nature of cartilaginous fish genomes with their evolutionary time scale inferred with nuclear genes. PLoS One. 2013;8:e66400.
4. Inoue JG, Miya M, Lam K, Tay B-H, Danks JA, Bell J, Walker TI, Venkatesh B. Evolutionary origin and phylogeny of the modern Holocephalans (Chondrichthyes: Chimaeriformes): a mitogenomic perspective. Mol Biol Evol. 2010;27:2576–86.
5. Giles S, Friedman M, Brazeau MD. Osteichthyan-like cranial conditions in an Early Devonian stem gnathostome. Nature. 2015;520:82–5.
6. Criscitiello MF. What the shark immune system can and cannot provide for the expanding design landscape of immunotherapy. Expert Opin Drug Discov. 2014;9:725–39.
7. Flajnik MF, Kasahara M. Origin and evolution of the adaptive immune system: genetic events and selective pressures. Nat Rev Genet. 2010;11: 47–59.
8. Hinds KR, Litman GW. Major reorganization of immunoglobulin V_H segmental elements during vertebrate evolution. Nature. 1986;320:546–9.
9. Kokubu F, Litman R, Shamblott MJ, Hinds K, Litman GW. Diverse organization of immunoglobulin VH gene loci in a primitive vertebrate. EMBO J. 1988;7:3413–22.
10. Litman GW, Anderson MK, Rast JP. Evolution of antigen binding receptors. Annu Rev Immunol. 1999;17:109–47.
11. Flajnik MF, Rumfelt LL. The immune system of cartilaginous fish. Curr Top Microbiol Immunol. 2000;248:249–70.
12. Greenberg AS, Avila D, Hughes M, Hughes A, McKinney EC, Flajnik MF. A new antigen receptor gene family that undergoes rearrangement and extensive somatic diversification in sharks. Nature. 1995;374:168–73.
13. Roux KH, Greenberg AS, Greene L, Strelets L, Avila D, McKinney EC, Flajnik MF. Structural analysis of the nurse shark (new) antigen receptor (NAR): Molecular convergence of NAR and unusual mammalian immunoglobulins. Proc Natl Acad Sci U S A. 1998;95:11804–9.
14. Stanfield RL, Dooley H, Verdino P, Flajnik MF, Wilson IA. Maturation of shark single-domain (IgNAR) antibodies: evidence for induced-fit binding. J Mol Biol. 2007;367:358–72.
15. Juarez K, Dubberke G, Lugo P, Koch-Nolte F, Buck F, Haag F, Licea A. Monoclonal antibodies for the identification and purification of vNAR domains and IgNAR immunoglobulins from the horn shark *Heterodontus francisci*. Hybridoma. 2011;30:323–9.
16. Griffiths K, Dolezal O, Parisi K, Angerosa J, Dogovski C, Barraclough M, Sanalla A, Casey J, González I, Perugini M, Nuttall S, Foley M. Shark variable new antigen receptor (VNAR) single domain antibody fragments: stability and diagnostic applications. Antibodies. 2013;2:66–81.
17. Block BA, Finnerty JR. Endothermy in fishes: a phylogenetic analysis of constraints, predispositions, and selection pressures. Environ Biol Fishes. 1994;40:283–302.
18. Block BA, Stevens ED. Tuna : Physiology, Ecology, and Evolution. BOOK. San Diego, Calif. London: Academic; 2001.
19. Dickson KA, Graham JB. Evolution and consequences of endothermy in fishes. Physiol Biochem Zool. 2004;77:998–1018.
20. Goldman KJ, Anderson SD, Latour RJ, Musick JA. Homeothermy in adult salmon sharks, *Lamna ditropis*. Environ Biol Fishes. 2004;71:403–11.
21. Watanabe YY, Goldman KJ, Caselle JE, Chapman DD, Papastamatiou YP. Comparative analyses of animal-tracking data reveal ecological significance of endothermy in fishes. Proc Natl Acad Sci. 2015;112:6104–9.
22. Little AG, Lougheed SC, Moyes CD. Evolution of mitochondrial-encoded cytochrome oxidase subunits in endothermic fish: The importance of taxon-sampling in codon-based models. Mol Phylogenet Evol. 2012;63:679–84.
23. Madigan DJ, Carlisle AB, Gardner LD, Jayasundara N, Micheli F, Schaefer KM, Fuller DW, Block BA. Assessing niche width of endothermic fish from genes to ecosystem. Proc Natl Acad Sci. 2015;112:8350–5.
24. Jayasundara N, Gardner LD, Block BA. Effects of temperature acclimation on Pacific bluefin tuna (*Thunnus orientalis*) cardiac transcriptome. Am J Physiol Regul Integr Comp Physiol. 2013;305:R1010–20.
25. Venkatesh B, Kirkness EF, Loh YH, Halpern AL, Lee AP, Johnson J, Dandona N, Viswanathan LD, Tay A, Venter JC, Strausberg RL, Brenner S. Survey sequencing and comparative analysis of the elephant shark (*Callorhinchus milii*) genome. PLoS Biol. 2007;5:932–44.
26. Venkatesh B, Lee AP, Ravi V, Maurya AK, Lian MM, Swann JB, Ohta Y, Flajnik MF, Sutoh Y, Kasahara M, Hoon S, Gangu V, Roy SW, Irimia M, Korzh V, Kondrychyn I, Lim ZW, Tay B-H, Tohari S, Kong KW, Ho S, Lorente-Galdos B, Quilez J, Marques-Bonet T, Raney BJ, Ingham PW, Tay A, Hillier LW, Minx P, Boehm T, et al. Elephant shark genome provides unique insights into gnathostome evolution. Nature. 2014;505:174–9.
27. Wyffels J, King BL, Vincent J, Chen C, Wu CH, Polson SW. SkateBase, an elasmobranch genome project and collection of molecular resources for chondrichthyan fishes. F1000Research. 2014;3:191.
28. Wang Q, Arighi CN, King BL, Polson SW, Vincent J, Chen C, Huang H, Kingham BF, Page ST, Rendino MF, Thomas WK, Udwary DW, Wu CH. Community annotation and bioinformatics workforce development in concert-Little Skate Genome Annotation Workshops and Jamborees. Database (Oxford). 2012; 2012:bar064. doi:10.1093/database/bar064.
29. Richards VP, Suzuki H, Stanhope MJ, Shivji MS. Characterization of the heart transcriptome of the white shark (*Carcharodon carcharias*). BMC Genomics. 2013;14:697.
30. Oulion S, Debiais-Thibaud M, D'Aubenton-Carafa Y, Thermes C, Da Silva C, Bernard-Samain S, Gavory F, Wincker P, Mazan S, Casane D. Evolution of Hox gene clusters in gnathostomes: insights from a survey of a shark (*Scyliorhinus canicula*) transcriptome. Mol Biol Evol. 2010;27:2829–38.
31. Mulley JF, Hargreaves AD, Hegarty MJ, Heller RS, Swain MT. Transcriptomic analysis of the lesser spotted catshark (*Scyliorhinus canicula*) pancreas, liver and brain reveals molecular level conservation of vertebrate pancreas function. BMC Genomics. 2014;15:1074.
32. Takechi M, Takeuchi M, Ota KG, Nishimura O, Mochii M, Itomi K, Adachi N, Takahashi M, Fujimoto S, Tarui H, Okabe M, Aizawa S, Kuratani S. Overview of the transcriptome profiles identified in hagfish, shark, and bichir: current

Marra *et al. BMC Genomics* (2017) 18:87

Page 17 of 18

issues arising from some nonmodel vertebrate taxa. J Exp Zool Part B Mol Dev Evol. 2011;316B:526–46.

33. Parton A, Bayne CJ, Barnes DW. Analysis and functional annotation of expressed sequence tags from in vitro cell lines of elasmobranchs: spiny dogfish shark (*Squalus acanthias*) and little skate (*Leucoraja erinacea*). Comp Biochem Physiol Part D Genomics Proteomics. 2010;5:199–206.

34. Perry GH, Melsted P, Marioni JC, Wang Y, Bainer R, Pickrell JK, Michelini K, Zehr S, Yoder AD, Stephens M, Pritchard JK, Gilad Y. Comparative RNA sequencing reveals substantial genetic variation in endangered primates. Genome Res. 2012;22:602–10.

35. Romiguier J, Gayral P, Ballenghien M, Bernard A, Cahais V, Chenuil A, Chiari Y, Dernat R, Duret L, Faivre N, Loire E, Lourenco JM, Nabholz B, Roux C, Tsagkogeorga G, Weber AA-T, Weinert LA, Belkhir K, Bierne N, Glemin S, Galtier N. Comparative population genomics in animals uncovers the determinants of genetic diversity. Nature. 2014;515:261–3.

36. Künstner A, Wolf JBW, Backström N, Whitney O, Balakrishnan CN, Day L, Edwards SV, Janes DE, Schlinger BA, Wilson RK, Jarvis ED, Warren WC, Ellegren H. Comparative genomics based on massive parallel transcriptome sequencing reveals patterns of substitution and selection across 10 bird species. Mol Ecol. 2010;19 Suppl 1:266–76.

37. Mann DL. The emerging role of innate immunity in the heart and vascular system: for whom the cell tolls. Circ Res. 2011;108:1133–45.

38. Epelman S, Liu PP, Mann DL. Role of innate and adaptive immune mechanisms in cardiac injury and repair. Nat Rev Immunol. 2015;15:117–29.

39. Farrell AP. Features heightening cardiovascular performance in fishes, with special reference to tunas. Comp Biochem Physiol Part A Physiol. 1996;113:61–7.

40. Bernal D, Carlson J, Goldman K, Lowe C. Energetics, metabolism, and endothermy in sharks and rays. In: Biology of Sharks and Their Relatives. Secondth ed. 2012. p. 211–37.

41. Douady CJ, Dosay M, Shivji MS, Stanhope MJ. Molecular phylogenetic evidence refuting the hypothesis of Batoidea (rays and skates) as derived sharks. Mol Phylogenet Evol. 2003;26:215–21.

42. Aronesty E. Comparison of sequencing utility programs. Open Bioinforma J. 2013;7:1–8.

43. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol. 2011;29:644–52.

44. Parra G, Bradnam K, Korf I. CEGMA: A pipeline to accurately annotate core genes in eukaryotic genomes. Bioinformatics. 2007;23:1061–7.

45. Parra G, Bradnam K, Ning Z, Keane T, Korf I. Assessing the gene space in draft genomes. Nucleic Acids Res. 2009;37:289–97.

46. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215:403–10.

47. Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M. Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics. 2005;21:3674–6.

48. Götz S, Garcia-Gomez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, Robles M, Talon M, Dopazo J, Conesa A. High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Res. 2008;36:3420–35.

49. van Dongen S. Graph clustering by flow simulation. Volume PhD thesis. Utrecht: University of Utrecht; 2001.

50. Richards VP, Palmer SR, Bitar PDP, Qin X, Weinstock GM, Highlander SK, Town CD, Burne RA, Stanhope MJ. Phylogenomics and the dynamic genome evolution of the genus *streptococcus*. Genome Biol Evol. 2014;6:741–53.

51. Al-Shahrour F, Díaz-Uriarte R, Dopazo J. FatiGO: A web tool for finding significant associations of Gene Ontology terms with groups of genes. Bioinformatics. 2004;20:578–80.

52. Breuer K, Foroushani AK, Laird MR, Chen C, Sribnaia A, Lo R, Winsor GL, Hancock REW, Brinkman FSL, Lynn DJ. InnateDB: systems biology of innate immunity and beyond—recent updates and continuing curation. Nucleic Acids Res. 2013;41(Database issue):D1228–33.

53. Ortutay C, Vihinen M. Immunome knowledge base (IKB): an integrated service for immunome research. BMC Immunol. 2009;10:3.

54. Yang Z. PAML: a program package for phylogenetic analysis by maximum likelihood. Comput Appl Biosci. 1997;13:555–6.

55. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 2007;24:1586–91.

56. Roshan U, Livesay DR. Probalign: multiple sequence alignment using partition function posterior probabilities. Bioinformatics. 2006;22:2715–21.

57. Yang Z, Wong WSW, Nielsen R. Bayes empirical Bayes inference of amino acid sites under positive selection. Mol Biol Evol. 2005;22:1107–18.

58. Murrell B, Weaver S, Smith MD, Wertheim JO, Murrell S, Aylward A, Eren K, Pollner T, Martin DP, Smith DM, Scheffler K, Kosakovsky Pond SL. Gene-wide identification of episodic selection. Mol Biol Evol. 2015;32:1365–71.

59. Dall E, Brandstetter H. Mechanistic and structural studies on legumain explain its zymogenicity, distinct activation pathways, and regulation. Proc Natl Acad Sci. 2013;110:10940–5.

60. Boveris A, Oshino N, Chance B. The cellular production of hydrogen peroxide. Biochem J. 1972;128:617–30.

61. Chance B, Sies H, Boveris A. Hydroperoxide metabolism in mammalian organs. Physiol Rev. 1979;59:527–605.

62. Rial E, Zardoya R. Oxidative stress, thermogenesis and evolution of uncoupling proteins. J Biol. 2009;8:1–5.

63. Clarke A, Pörtner HO. Temperature, metabolic power and the evolution of endothermy. Biol Rev. 2010;85:703–27.

64. Kumar V, Kutschera VE, Nilsson MA, Janke A. Genetic signatures of adaptation revealed from transcriptome sequencing of Arctic and red foxes. BMC Genomics. 2015;16:585.

65. Babik W, Stuglik M, Qi W, Kuenzli M, Kuduk K, Koteja P, Radwan J. Heart transcriptome of the bank vole (*Myodes glareolus*): towards understanding the evolutionary variation in metabolic rate. BMC Genomics. 2010;11:390.

66. Pantalacci S, Sémon M. Transcriptomics of developing embryos and organs: A raising tool for evo–devo. J Exp Zool Part B Mol Dev Evol. 2015;324:363–71.

67. Zhang Q, Hill GE, Edwards SV, Backström N. A house finch (*Haemorhous mexicanus*) spleen transcriptome reveals intra- and interspecific patterns of gene expression, alternative splicing and genetic diversity in passerines. BMC Genomics. 2014;15:305.

68. Elmer KR, Fan S, Gunter HM, Jones JC, Boekhoff S, Kuraku S, Meyer A. Rapid evolution and selection inferred from the transcriptomes of sympatric crater lake cichlid fishes. Mol Ecol. 2010;19 Suppl 1:197–211.

69. Carroll SB. Evo-Devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. Cell. 2008;134:25–36.

70. Harrison PW, Wright AE, Mank JE. The evolution of gene expression and the transcriptome–phenotype relationship. Semin Cell Dev Biol. 2012;23:222–9.

71. Berstein RM, Schluter SF, Shen S, Marchalonis JJ. A new high molecular weight immunoglobulin class from the carcharhine shark: implications for the properties of the primordial immunoglobulin. Proc Natl Acad Sci U S A. 1996;93:3289–93.

72. Greenberg AS, Hughes AL, Guo J, Avila D, McKinney EC, Flajnik MF. A novel "chimeric" antibody class in cartilaginous fish: IgM may not be the primordial immunoglobulin. Eur J Immunol. 1996;26:1123–9.

73. Ota T, Rast JP, Litman GW, Amemiya CT. Lineage-restricted retention of a primitive immunoglobulin heavy chain isotype within the Dipnoi reveals an evolutionary paradox. Proc Natl Acad Sci U S A. 2003;100:2501–6.

74. Warr GW, Magor KE, Higgins DA. IgY: clues to the origins of modern antibodies. Immunol Today. 1995;16:392–8.

75. Zhao Y, Pan-Hammarström Q, Yu S, Wertz N, Zhang X, Li N, Butler JE, Hammarström L. Identification of IgF, a hinge-region-containing Ig class, and IgD in *Xenopus tropicalis*. Proc Natl Acad Sci. 2006;103:12087–92.

76. Desmyter A, Transue TR, Ghahroudi MA, Dao Thi M-H, Poortmans F, Hamers R, Muyldermans S, Wyns L. Crystal structure of a camel single-domain VH antibody fragment in complex with lysozyme. Nat Struct Mol Biol. 1996;3:803–11.

77. Manoury B, Hewitt EW, Morrice N, Dando PM, Barrett AJ, Watts C. An asparaginyl endopeptidase processes a microbial antigen for class II MHC presentation. Nature. 1998;396:695–9.

78. Watts C. The endosome-lysosome pathway and information generation in the immune system. Biochim Biophys Acta. 1824;2012:14–21.

79. Murthy RV, Arbman G, Gao J, Roodman GD, Sun X-F. Legumain expression in relation to clinicopathologic and biological variables in colorectal cancer. Clin Cancer Res. 2005;11:2293–9.

80. Luo Y, Zhou H, Krueger J, Kaplan C, Lee S-H, Dolman C, Markowitz D, Wu W, Liu C, Reisfeld RA, Xiang R. Targeting tumor-associated macrophages as a novel strategy against breast cancer. J Clin Invest. 2006;116:2132–41.

Marra et al. BMC Genomics (2017) 18:87

Page 18 of 18

81. Ballantyne JS. Jaws: The inside story. The metabolism of elasmobranch fishes. Comp Biochem Physiol - B Biochem Mol Biol. 1997;118:703–42.

82. Ostrander GK, Cheng KC, Wolf JC, Wolfe MJ. Shark cartilage, cancer and the growing threat of pseudoscience. Cancer Res. 2004;64:8485–91.

83. Rao MN, Shinnar AE, Noecker LA, Chao TL, Feibush B, Snyder B, Sharkansky I, Sarkahian A, Zhang X, Jones SR, Kinney WA, Zasloff M. Aminosterols from the dogfish shark Squalus acanthias. J Nat Prod. 2000;63:631–5.

84. Bhargava P, Marshall JL, Dahut W, Rizvi N, Trocky N, Williams JI, Hait H, Song S, Holroyd KJ, Hawkins MJ. A phase I and pharmacokinetic study of Squalamine, a novel antiangiogenic agent, in patients with advanced cancers. Clin Cancer Res. 2001;7:3912–9.

85. Walsh CJ, Luer CA, Bodine AB, Smith CA, Cox HL, Noyes DR, Gasparetto M. Elasmobranch immune cells as a source of novel tumor cell inhibitors: Implications for public health. Integr Comp Biol. 2006;46:1072–81.

86. Edgington LE, Verdoes M, Ortega A, Withana NP, Lee J, Syed S, Bachmann MH, Blum G, Bogyo M. Functional imaging of legumain in cancer using a new quenched activity-based probe. J Am Chem Soc. 2013;135:174–82.

87. Hata AN, Engelman JA, Faber AC. The BCL2 family: key mediators of the apoptotic response to targeted anticancer therapeutics. Cancer Discov. 2015;5:475–87.

88. Takayama S, Sato T, Krajewski S, Kochel K, Irie S, Milian JA, Reed JC. Cloning and functional analysis of BAG-1: A novel Bcl-2-binding protein with anti-cell death activity. Cell. 1995;80:279–84.

89. Bauer MF, Rothbauer U, Mühlenbein N, Smith RJH, Gerbitz K-D, Neupert W, Brunner M, Hofmann S. The mitochondrial TIM22 preprotein translocase is highly conserved throughout the eukaryotic kingdom. FEBS Lett. 1999;464:41–7.

90. Bird PM. Tissue regeneration in three Carcharhinid sharks encircled by embedded straps. Copeia. 1978;1978:345–9.

91. Reif W-E. Wound healing in sharks. Zoomorphologie. 1978;90:101–11.

92. Domeier M, Nasby-Lucas N. Annual re-sightings of photographically identified white sharks (Carcharodon carcharias) at an eastern Pacific aggregation site (Guadalupe Island, Mexico). Mar Biol. 2007;150:977–84.

93. Marshall AD, Bennett MB. The frequency and effect of shark-inflicted bite injuries to the reef manta ray Manta alfredi. African J Mar Sci. 2010;32:573–80.

94. Towner A, Smale M, Jewell O. Boat strike wound healing in Carcharodon carcharias. In: Domeier ML, editor. Global Perspectives on the Biology and Life History of the White Shark. New York: CRC Press; 2012. pp. 77–83.

95. Gosain A, DiPietro LA. Aging and wound healing. World J Surg. 2004;28:321–6.

96. Guo S, DiPietro LA. Factors affecting wound healing. J Dent Res. 2010;89:219–29.

97. Nimmerjahn F, Ravetch JV. Fc[gamma] receptors as regulators of immune responses. Nat Rev Immunol. 2008;8:34–47.

98. Schäffer M, Bongartz M, Hoffmann W, Viebahn R. MHC-Class-II-Deficiency impairs wound healing. J Surg Res. 2007;138:100–5.

99. Hofmann U, Beyersdorf N, Weirather J, Podolskaya A, Bauersachs J, Ertl G, Kerkau T, Frantz S. Activation of CD4 + T lymphocytes improves wound healing and survival after experimental myocardial infarction in mice. Circulation. 2012;125:1652–63.

100. Bartl S, Nonaka M. HC Molecules of Cartilaginous Fishes. In: Smith SL, Sim RB, Flajnik MF, editors. Immunobiology of the Shark. Boca Raton: CRC Press; 2014. p. 173–98.

101. Shiels HA, Di Maio A, Thompson S, Block BA. Warm fish with cold hearts: thermal plasticity of excitation–contraction coupling in bluefin tuna. Proc R Soc London B Biol Sci. 2010;278:18–27.

102. Bennett AF, Ruben JA. Endothermy and activity in vertebrates. Science. 1979;206:649–54.

103. Nespolo RF, Bacigalupe LD, Figueroa CC, Koteja P, Opazo JC. Using new tools to solve an old problem: The evolution of endothermy in vertebrates. Trends Ecol Evol. 2011;26:414–23.

104. Little AG, Seebacher F. The evolution of endothermy is explained by thyroid hormone-mediated responses to cold in early vertebrates. J Exp Biol. 2014;217(Pt 10):1642–8.

105. Moon YA, Shah NA, Mohapatra S, Warrington JA, Horton JD. Identification of a mammalian long chain fatty acyl elongase regulated by sterol regulatory element-binding proteins. J Biol Chem. 2001;276:45358–66.

106. Mahdessian H, Taxiarchis A, Popov S, Silveira A, Franco-Cereceda A, Hamsten A, Eriksson P, Van't Hooft F. TM6SF2 is a regulator of liver fat metabolism influencing triglyceride secretion and hepatic lipid droplet content. Proc Natl Acad Sci U S A. 2014;111:8913–8.

107. Pirola CJ, Sookoian S. The dual and opposite role of the TM6SF2-rs58542926 variant in protecting against cardiovascular disease and conferring risk for nonalcoholic fatty liver: A meta-analysis. Hepatology. 2015;62:1742–56.

108. Holmen OL, Zhang H, Fan Y, Hovelson DH, Schmidt EM, Zhou W, Guo Y, Zhang J, Langhammer A, Lochen M-L, Ganesh SK, Vatten L, Skorpen F, Dalen H, Zhang J, Pennathur S, Chen J, Platou C, Mathiesen EB, Wilsgaard T, Njolstad I, Boehnke M, Chen YE, Abecasis GR, Hveem K, Willer CJ. Systematic evaluation of coding variation identifies a candidate causal variant in TM6SF2 influencing total cholesterol and myocardial infarction risk. Nat Genet. 2014;46:345–51.