Nova Southeastern University
NSUWorks

Biology Faculty Articles | Department of Biological Sciences

3-5-2019

White Shark Genome Reveals Ancient Elasmobranch Adaptations Associated with Wound Healing and the Maintenance of Genome Stability

Nicholas J. Marra
Cornell University; Nova Southeastern University

Michael J. Stanhope
Cornell University

Nathaniel K. Jue
California State University

Minghui Wang
Cornell University

Qi Sun
Cornell University

See next page for additional authors

Follow this and additional works at: https://nsuworks.nova.edu/cnso_bio_facarticles

Part of the Biology Commons, Genetics and Genomics Commons, and the Marine Biology Commons

NSUWorks Citation

Authors

Nicholas J. Marra, Michael J. Stanhope, Nathaniel K. Jue, Minghui Wang, Qi Sun, Paulina D. Pavinski Bitar, Vincent P. Richards, Aleksey S. Komissarov, Mike Rayko, Sergey Kliver, Bryce J. Stanhope, Chuck Winkler, Stephen James O'Brien, Agostinho Antunes, Salvador Jorgensen, and Mahmood S. Shivji

White shark genome reveals ancient elasmobranch adaptations associated with wound healing and the maintenance of genome stability

Nicholas J. Marra[a,b,c,1,2], Michael J. Stanhope[c,1,3,4], Nathaniel K. Jue[d], Minghui Wang[e], Qi Sun[e], Paulina Pavinski Bitar[c], Vincent P. Richards[f], Aleksey Komissarov[g], Mike Rayko[h], Sergey Kliver[g], Bryce J. Stanhope[c], Chuck Winkler[i], Stephen J. O'Brien[g,j,4], Agostinho Antunes[k,l], Salvador Jorgensen[m], and Mahmood S. Shivji[a,b,3,4]

[a]Save Our Seas Shark Research Center, Nova Southeastern University, Dania Beach, FL 33004; [b]Guy Harvey Research Institute, Nova Southeastern University, Dania Beach, FL 33004; [c]Population Medicine and Diagnostic Sciences, Cornell College of Veterinary Medicine, Cornell University, Ithaca, NY 14853; [d]School of Natural Sciences, California State University, Seaside, CA 93955; [e]Institute of Biotechnology, Bioinformatics Facility, Cornell University, Ithaca, NY 14853; [f]Department of Biological Sciences, Clemson University, Clemson, SC 29634; [g]Theodosius Dobzhansky Center for Genome Bioinformatics, St. Petersburg State University, 199004 St. Petersburg, Russia; [h]Center for Algorithmic Biotechnology, Institute for Translational Biomedicine, St. Petersburg State University, 199004 St. Petersburg, Russia; [i]Aquatic Research Consultants, San Pedro, CA 90732; [j]Guy Harvey Oceanographic Center, Halmos College of Natural Sciences and Oceanography, Nova Southeastern University, Ft. Lauderdale, FL 33004; [k]Interdisciplinary Centre of Marine and Environmental Research, University of Porto, 4450–208 Porto, Portugal; [l]Department of Biology, Faculty of Sciences, University of Porto, 4169-007 Porto, Portugal; and [m]Conservation Research Department, Monterey Bay Aquarium, Monterey, CA 93940

The white shark (*Carcharodon carcharias*; Chondrichthyes, Elasmobranchii) is one of the most publicly recognized marine animals. Here we report the genome sequence of the white shark and comparative evolutionary genomic analyses to the chondrichthyans, whale shark (Elasmobranchii) and elephant shark (Holocephali), as well as various vertebrates. The 4.63-Gbp white shark genome contains 24,520 predicted genes, and has a repeat content of 58.5%. We provide evidence for a history of positive selection and gene-content enrichments regarding important genome stability-related genes and functional categories, particularly so for the two elasmobranchs. We hypothesize that the molecular adaptive emphasis on genome stability in white and whale sharks may reflect the combined selective pressure of large genome sizes, high repeat content, high long-interspersed element retrotransposon representation, large body size, and long lifespans, represented across these two species. Molecular adaptation for wound healing was also evident, with positive selection in key genes involved in the wound-healing process, as well as Gene Ontology enrichments in fundamental wound-healing pathways. Sharks, particularly apex predators such as the white shark, are believed to have an acute sense of smell. However, we found very few olfactory receptor genes, very few trace amine-associated receptors, and extremely low numbers of G protein-coupled receptors. We did however, identify 13 copies of vomeronasal type 2 (V2R) genes in white shark and 10 in whale shark; this, combined with the over 30 V2Rs reported previously for elephant shark, suggests this gene family may underlie the keen odorant reception of chondrichthyans.

comparative genomics | genome stability | elasmobranch evolution

Chondrichthyan fishes (Elasmobranchii: sharks, rays, skates; Holocephali: chimaeras) represent one of the oldest vertebrate lineages, arising over 400 million y ago (Fig. 1). Sharks, specifically, comprise ~45% of the known Elasmobranchii species (1) and include many of the meso- and apex-level oceanic predators. Perhaps the most recognized of all these predators is the white shark (*Carcharodon carcharias*), capturing extraordinary attention from the public and media. The white shark has a largely cosmopolitan distribution, but the species is thought to be of low abundance throughout this extensive range, and is classified as globally "vulnerable" (IUCN Red List category) (2).

Genomic resources for chondrichthyans are limited, with published genome sequences confined until recently, to a single shark, the whale shark (*Rhincodon typus*; genome size, 3.44 Gbp) (3), and a holocephalan, the elephant shark (*Callorhinchus milii*; genome size, 975 Mb) (4); there is also an ongoing little skate (*Leucoraja erinacea*; Superorder Batoidea, Order Rajiformes) genome project (5). Most

Significance

Sharks represent an ancient vertebrate lineage whose genomes have been only minimally investigated. We here characterize the genome of the white shark, an apex marine predator. Its genome is 4.63 Gbp, over half of which is represented by repeat sequences, including a large proportion of transposable elements. Comparative analysis of white shark, whale shark, chimaera, and several nonchondrichthyan vertebrate genomes reveals positive selection and enrichment of gene functional categories and pathways involved in wound healing, and in the maintenance of genome stability in sharks. Sharks show a limited repertoire of olfactory genes but an expanded vomeronasal (VR2) gene family, suggesting an alternative mechanism underlying their vaunted sense of smell.
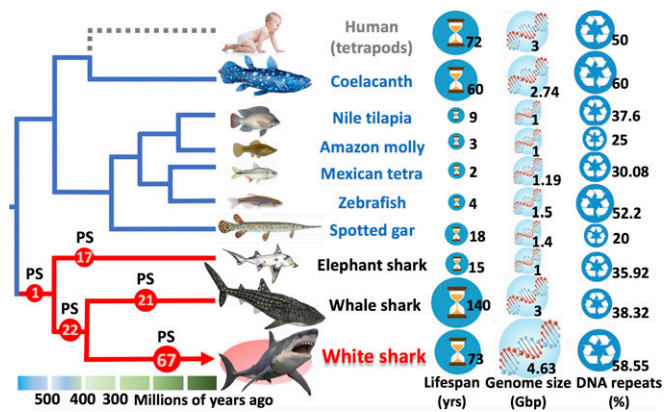
Fig. 1. Schematic, annotated, phylogenetic tree including the fish species used in our positive-selection analysis. Human is included only for comparative and evolutionary reference. Lifespan, genome size, and percent repeat content included for each taxon. Red circles refer to number of positively selected genes (zebrafish, *Danio rerio*; Amazon molly, *Poecilia formosa*; blind cave fish, *Astyanax mexicanus*; Nile tilapia, *Oreochromis niloticus*; coelacanth, *Latimeria chalumnae*; spotted gar, *Lepisosteus oculatus*).

recently, two additional elasmobranch genomes were published—the brownbanded bamboo shark (*Chiloscyllium punctatum*), and the cloudy catshark (*Scyliorhinus torazame*) (6)—these latter two were published only a few weeks before the submission of this report and therefore could not be included in the comparative genomic analyses herein presented. Nonetheless, the availability of the whale shark and elephant shark genomes affords a range of important comparative possibilities to white shark.

The white shark possesses notable physical and physiological characteristics that make it an interesting biological study, including an estimated genome size (C value = 6.45 pg) (7) nearly twice that of humans, large adult sizes reaching up to ∼6 m in length and 3,232 kg in weight, a thermal regulatory capability uncommon in fishes, a slow reproductive cycle with oophagous embryos, a life span of ∼73 y, rapid swimming speeds, extensive migratory capabilities, and an ability to utilize a wide thermal niche, including diving to near 1,000-m depths (8, 9). Sharks have long been noted for their use of smell to locate prey (10), and they locate odorant sources by tracking changes in odorant concentrations (11). Recent genome studies of largely bottom-dwelling chondrichthyans (elephant shark, brownbanded bamboo shark, and cloudy catshark) have reported a distinct paucity of odorant receptor genes (4, 6), raising the question of whether this would also be the case in a pelagic apex predator, which migrates great distances, such as the white shark. There are anecdotal reports that elasmobranchs have a low incidence of documented cancers (12), but these observations remain unconfirmed given the absence of systematic surveys to investigate the question (13). There are also empirical studies reporting on a medium generated from the culture of cells from the epigonal organ of bonnethead sharks (*Sphyrna tiburo*), exhibiting cytotoxic activity against human tumor cell lines, via the induction of apoptosis in the target cells (12, 14).

A characteristic of most cancers is genomic instability (15), which is the accumulation of a high frequency of genomic mutations. Throughout an organism's lifespan their genome is under threat from exogenous, endogenous, and cellular processes that can inflict DNA damage and compromise genome integrity. The result of this common set of continual selective pressures has been the evolution of defense mechanisms to counteract the detrimental effects of these events and safeguard the genetic information, many of which mechanisms tend to be conserved across ancient and diverse spans of the tree of life. However, the adaptive molecular specifics associated with fine-tuning these common defense mechanisms will not be the

same for different lineages because of a history of specific and unique selective pressures for each lineage. Defects in the mechanisms regulating genome integrity result in genomic instability, which predisposes cells to malignant transformation (15), neurological disease (16), and premature aging (17). Little or nothing is known about the adaptive evolution of genome stability maintenance in sharks. Sharks are anecdotally reported to have superior wound-healing capabilities and there is recent empirical evidence from black tip sharks to support this (18). Wound healing, like genome stability, is a highly complex process involving several phases or interrelated steps. Although sharks appear to have exceptional wound-healing capabilities, little is known about why that might be the case.

Comparative sequence data on a genomic scale provides the opportunity to explain major biological differences between organisms at the molecular level. Three fundamental molecular characteristics responsible for biological differences between organisms are: (*i*) presence and absence of particular loci, (*ii*) positive selection, and (*iii*) gene regulation. Our interest in the molecular basis of adaptive features in sharks focuses on the first two of these. Positive selection is the fixation of advantageous mutations driven by natural selection, and is a fundamental process behind adaptive changes in genes and genomes, leading to evolutionary innovations and species differences. Conducting such analyses at the level of complete genomes is of high interest because it has the prospect of telling us much about the organisms in question.

This paper reports the genome sequence of the white shark, and through comparative analyses to other chondrichthyans—as well as other vertebrates—we present evidence of molecular adaptation and evolution in gene content behind several notable biological features of this marine apex predator, and in particular identify some of the loci, functional groups, and pathways that may have been of adaptive significance in the evolution of chondrichthyan genome-stability maintenance, wound healing, and smell.

Results

White Shark Genome Sequence. The white shark genome possesses a chromosome number of 2N = 82 and flow cytometry puts the genome size at 6.3 Gbp (7), suggesting a genome size twice that of human. Analysis of the read data through k-mer plots (*SI Appendix*, Fig. S1) indicated a genome size of 4.63 Gbp, smaller than that estimated by flow cytometry, putting our initial raw sequence coverage close to 117×, with a scaffold assembly size of 4.076 Gbp and a gap percentage of 7.24%. To improve the assembly, we enlisted Dovetail Genomics to sequence five Chicago libraries yielding a final, more continuous assembly of 4.079 Gbp and an N50 of 2.77 Mbp. In total, our sequencing had a coverage of 198× over the 4.63-Gbp estimated genome size, with scaffolds comprising 88.1% of the genome.

The final annotated genome assembly comprised 9,222 scaffolds (greater than 10 kbp in length) with a total length of 3.92 Gbp and contained 24,520 predicted genes, a number similar to that reported for other vertebrates. A comparison of the three chondrichthyan genome assemblies using both the metazoan and vertebrate BUSCO (benchmarking universal single-copy orthologs) core gene sets (*SI Appendix*, Fig. S2) indicated that the white shark assembly had a smaller number of missing metazoan genes compared with whale shark, a similar number compared with elephant shark, and almost three times the duplicated genes compared with the other two chondrichthyans. This pattern was different for the vertebrate gene set, with both the elasmobranchs being similar in terms of duplicated and single-copy genes.

The repeat content of the white shark genome is large relative to most other fish species at 58.55% (Fig. 1) and has a GC content of 43.95%. The largest repeat class was long-interspersed elements (LINEs), which comprised 29.84% of the genome, and in particular LINE-3/CR1 elements, which constituted 18.75% of the genome (see *SI Appendix*, Fig. S2 for further percentage breakdown of genome components). Mapping 150- and 250-bp

paired-end reads to the Dovetail assembly yielded about 3,515,000 SNPs present at ≥30% minor allele frequency in the scaffolds greater than 10 kbp, yielding a heterozygosity estimate of about 0.09%.

Positive Selection on Genes Involved in the Maintenance of Genome Integrity. The largest number of genes judged to be under positive selection (see *Methods* for details regarding the carefully curated positive selection analyses including manual check of alignments) on any single branch (67 genes) was on the white shark lineage (Fig. 1; Dataset S1 includes all relevant statistics associated with the positive selection analysis). Of these, nearly one-third (20 of 67) have Gene Ontology (GO) terms and supporting literature that indicate they play a role in genome stability (Table 1; detailed specifics of these many genes, their relevant GO terms and their roles in genome stability are presented in an annotated *SI Appendix*, Table S1) (our examination of the starting ortholog pool of 1,541 and the white shark genome indicates that each includes a roughly comparable proportion of genome stability related genes), with the majority of these genes implicated in DNA damage response, DNA repair, or translesion synthesis (Table 1 and *SI Appendix*, Table S1) and including some of the most fundamental genes involved in these processes. The next most common set of

GO terms deal with ubiquitination, with five corresponding genes, although several additional genes are associated with both ubiquitination and genome stability. Protein ubiquitination is involved in a wide range of processes; however, there is also ample evidence of the importance of ubiquitination and de-ubiquitination in the realm of genome stability (19), and this includes several of the positively selected genes in our analysis (e.g., USP13 and UFD1) (Table 1 and *SI Appendix*, Table S1). Fewer genes were judged to be under positive selection on the whale shark, elephant shark, and elasmobranch branches (Fig. 1 and Dataset S1); however, genes critical to the maintenance of genome integrity were still prominently represented and, in the case of whale shark, like that for white shark, comprised about one-third of the total set of positively selected genes (Table 1 and Dataset S1).

GO Patterns of Positively Selected Genes. As an alternative to tallying GO terms, we examined the nature of GO patterns in the complete set of positively selected genes for each chondrichthyan branch. Employing the GO clustering tool REVIGO (20) on the terms associated with the 67 positively selected genes from white shark, revealed that terms related to genome stability were among those with the greatest average similarity (i.e., least unique) (Fig. 2A), compared with the set of terms as a whole. Signaling

Table 1. List of positively selected genes with roles in genome stability, for all branches of the three species chondrichthyan phylogeny

Gene	Protein name	Role in genome stability
CHEK2*	Serine/threonine-protein kinase Chk2	DNA repair; apoptosis; tumor suppressor
RFC5*	Replication factor C subunit 5	DNA repair; translesion synthesis; nucleotide excision repair
FBXO45*,†	F-box/SPRY domain-containing protein 1	Regulates/degrades tumor suppressor TP73
DICER1*	Endoribonuclease Dicer	siRNA and miRNA biogenesis; DNA repair; apoptosis
INO80B*	INO80 complex subunit B	Chromatin remodeling; DNA repair
DTL*	Denticleless protein	DNA damage response; translesion DNA synthesis
POLD3*	DNA polymerase delta subunit 3	DNA repair
FEM1B*	Protein fem-1 homolog B	Apoptosis; DNA repair
SIRT7*	NAD-dependent protein deacetylase sirtuin-7	DNA repair; chromatin remodeling; apoptosis; regulates p53
PLK2*	Serine/threonine-protein kinase PLK2	Cell cycle control; regulates tumor growth and apoptosis
CENPS*	Centromere protein S	DNA repair
CASS4*	Cas scaffolding protein family member 4	Cell adhesion and cell spreading; apoptosis
UFD1*	Ubiquitin recognition factor in endoplasmic reticulum-associated degradation protein 1	Protein deubiquitination; core component of p97-UFD1-NPL4 complex involved in protein extraction from chromatin
AGT*	Angiotensinogen	Apoptosis; cell proliferation
RPS6*	40S ribosomal protein S6	Apoptosis
MYOG*	Myogenin	Regulation of cell proliferation and cell cycle arrest
USP13*	Ubiquitin carboxyl-terminal hydrolase 13	Controls autophagy and p53 levels; cell proliferation
PRIM1*	DNA primase small subunit	Okazaki fragment synthesis
ALKBH7*	Alpha-ketoglutarate dependent dioxygenase alkB homolog 7	Necrosis
BUD23*	18S rRNA (guanine-N(7))-methyltransferase	Chromatin remodeling
FIGL1‡	Fidgetin–like-1	Double strand break repair; regulation of meiotic recombination
CTNNBL1‡	β-Catenin–like protein 1	Transcription coupled repair; apoptosis
CMTM7‡	CKLF-like MARVEL transmembrane domain-containing protein 7	Tumor suppressor
MDM4‡	Protein MDM4	p53 regulator
ARL6IP5‡	PRA1 family protein 3	Apoptosis
KIAA1324‡	UPF0577 protein KIAA1324	Autophagy; tumor suppressor
SALL4‡	Sal-like protein 4	DNA damage response; stem cell maintenance
PDCD2†,§	Programmed cell death protein 2	Apoptosis; regulation of stem cell proliferation
PDCD4†	Programmed cell death protein 4	Apoptosis; tumor suppressor
NHP2†	H/ACA ribonucleoprotein complex subunit 2	Telomere maintenance
RRS1†	Ribosome biogenesis regulatory protein homolog	p53 regulator

See *SI Appendix*, Table S1 for a fully referenced and annotated version of this table.
*White shark.
†Elasmobranchs.
‡Whale shark.
§Elephant shark.

pathways appeared prominently in this set of "least unique" terms (Fig. 2A), including those that have important roles in genome stability, such as the ERK1/ERK2 cascade (21), target of rapamycin (TOR) signaling pathway (22), and the smoothened (sonic hedgehog) signaling pathway (23). The latter is a major regulator of cell differentiation, cell proliferation, and tissue polarity, and plays an important role in tumorigenesis, tumor progression, and the therapeutic response to a wide range of cancers (23). Clustering the terms associated with the positively selected genes for each of the other branches (Fig. 2 B–D) suggested an emphasis on genome stability roles in the set of positively selected loci for each elasmobranch species, and their ancestral branch, but most particularly white shark, with less obvious emphasis in elephant shark. Cell redox homeostasis was one such term on the elasmobranch branch; redox regulation has important implications in genome stability through its effects on

DNA damage and repair (24). Of particular note was the recurrence across each elasmobranch branch of a term referring to signal transduction by p53 class mediator. In the case of the white shark the specific term also involves the regulation of p21 transcription (Fig. 2A). p21 is a cyclin-dependent kinase inhibitor and functions as a key regulator of cell cycle progression (25). It is also an important target of p53 activity, linking DNA damage to cell cycle arrest (26).

Positive Selection Throughout the MDM4 Gene, a Key Regulator of p53. Indeed, a common feature of the adaptive emphasis reported throughout this paper involves regulation or interaction with p53, sometimes referred to as the "guardian of the genome." Two of the most important regulators of p53 are MDM2 and MDM4 (27). The primary role for MDM4 is in regulating p53 abundance by modulating both the activity and levels of MDM2. MDM4 was
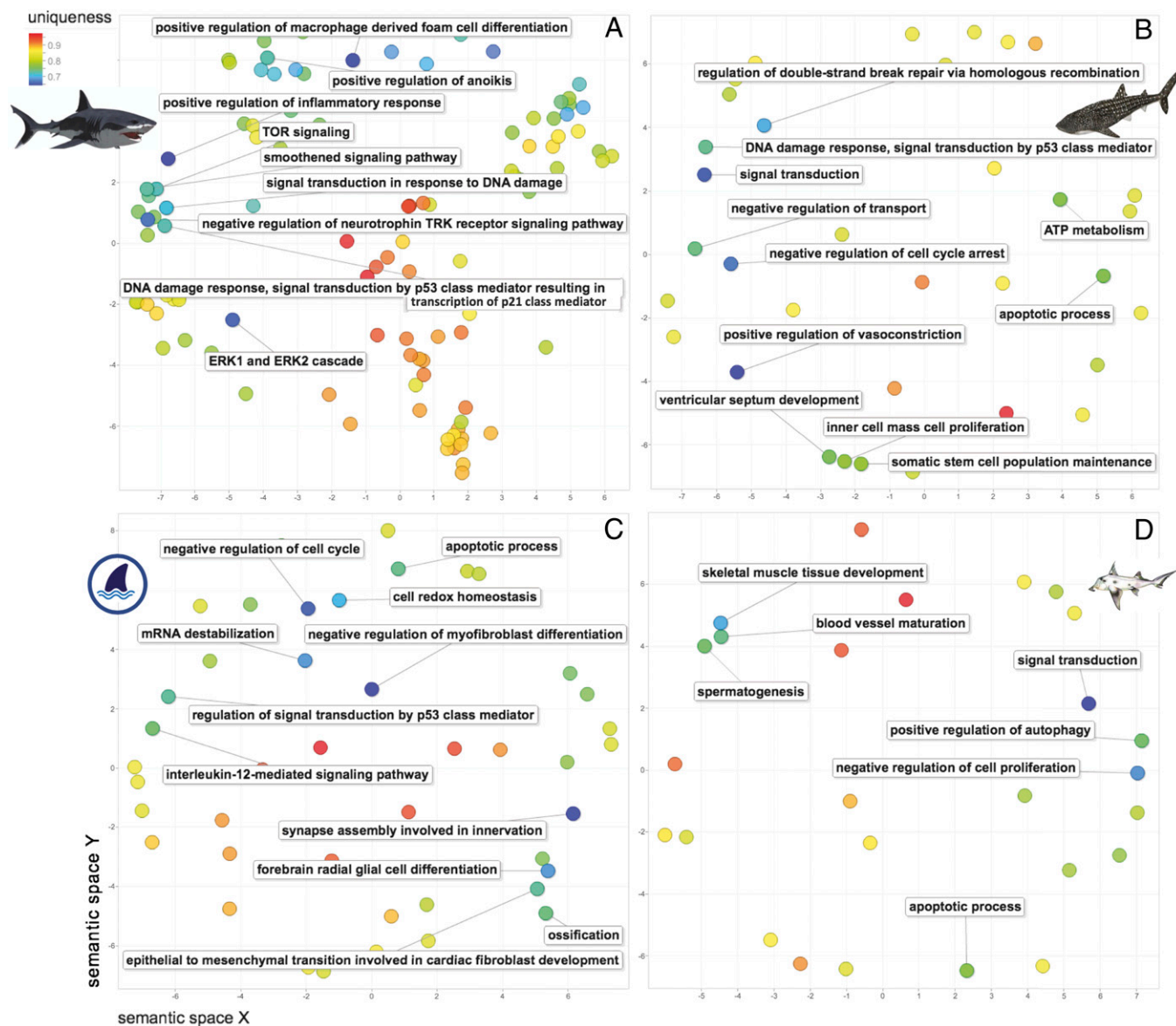
Fig. 2. GO terms of positively selected genes summarized and visualized as a REVIGO scatter plot: (*A*) white shark, (*B*) whale shark, (*C*) elasmobranch, and (*D*) elephant shark. Each circle represents a cluster of related GO terms, with a single term chosen by REVIGO as the cluster representative. Clusters are plotted according to semantic similarities to other GO terms (adjoining circles are most closely related). "Uniqueness" (the negative of average similarity of a term to all other terms) measures the degree to which the term is an outlier when compared semantically to the whole list. Genome stability terms tend to be among the least unique (blue and turquoise dots) and therefore of greatest average similarity to the set as a whole.

GENETICS

under positive selection on the whale shark branch (Table 1). Most of the positively selected sites were in the 3′ end of the gene. White shark was missing the 5′ end (about one-third of the sequence) and therefore that region of the gene was not analyzed for any species in the alignment, because gapped regions were ignored in all analyses (*Methods*). To investigate a more complete picture of positive selection on this gene for whale shark, we removed white shark and reanalyzed. The result identified positive selection throughout the gene. Mapping these sites to a protein model of MDM4 from whale shark revealed that the positively selected sites were in the RING pocket, Zn finger pocket, the p53 binding pocket, and in both the N- and C-terminus disordered regions (Fig. 3). The interactions involving these various domains are many and complex. In brief, the RING domain of MDM4 interacts with MDM2. Casein kinase 1α (CK1α) interacts with the Zn finger domain of MDM4 to aid in the inhibition of p53 activity. The MDM4 p53 binding pocket binds to, and masks the transcriptional activation domain of p53. The N- and C-terminus disordered regions interact with the structured domains, forming intramolecular interactions that enhance or block their activity (28). MDM2 and p53 could not be accurately aligned among the species included in our analysis.

Statistical View of Functional Emphasis in Positive Selection. As an additional gauge of the overall functional emphasis of the positively selected genes, we looked for statistical GO enrichments of these genes in comparisons against GO and pathway databases, as well as against the genome as a whole. Comparing the list of white shark positively selected genes to GO Biological Process (human), yields one enriched genome stability-related term (translesion synthesis) and one other term (cellular macromolecule catabolic process). A comparison against the Reactome database (29) resulted in only the following most specific terms (see *Methods* discussion of Panther and GO comparisons, for explanation of the most specific) being significantly enriched for the white shark positively selected gene list: (*i*) recognition of DNA damage by proliferating cell nuclear antigen (PCNA)-containing replication complex; (*ii*) polymerase switching on the C strand of the telomere; and (*iii*) polymerase switching. PCNA is required for DNA excision repair and is also involved in the DNA damage-tolerance (bypass) pathway known as post-

replication repair. In postreplication repair, there are two subpathways: (*i*) the translesion pathway, which involves switching replicative DNA polymerases for specialized translesion polymerases that mediate damage tolerance by replicating past certain DNA lesions; and (*ii*) the "template switch" pathway, which enables a stalled replication fork to use the daughter strand or a nearby fork with homology as template to bypass the damage. PCNA is pivotal to the cellular choice and activation of both these pathways.

The apparent functional emphasis in white shark of translesion synthesis and telomere maintenance are further evidenced by their repetition in GO enrichment comparisons involving the white shark positively selected gene list against the complete white shark genome, with the GO terms (*i*) telomere maintenance via recombination (GO:0000722) and (*ii*) translesion synthesis (GO:0019985) being 2 of the 32 most specific, significantly enriched Biological Process terms (Dataset S2). These particular functional enrichments then, are all interrelated and involve aspects of DNA repair associated with polymerase switching. Performing the same tests for whale shark and elephant shark did not reveal any significant enrichments of positively selected genes against GO or Reactome databases, but did indicate a variety of enriched terms when compared back to each of their respective genomes. Of particular note in whale shark were enrichments regarding TORC1 and -2 signaling, spliceosomal complexes, and the regulation of histone H3 and H4 methylation and acetylation (Dataset S3), all of which can have important implications in the realm of genome stability (30–32). In the case of the elephant shark, most specific enriched terms of note included one involving regulation of histone H3 acetylation, and one regarding regulation of JUN kinase activity (Dataset S4).

GO Enrichments in Genome Content. To examine differences in the functional gene content of white shark in comparison with other vertebrates, we employed the Panther system of GO enrichment (33) using the complete set of white shark Swiss-Prot IDs compared with the genomes of five other species (human, zebrafish, frog, anole, chicken) (*Methods*) for each of the Biological Process, Cellular Component, Molecular Function, and Reactome (in this case just human) databases (Dataset S5). A large number of most specific Biological Process terms were enriched in white shark in these intervertebrate comparisons [ranging from 63 (zebrafish) to 200 (human)], with about 80 terms (of a total of 367, considering all five vertebrate comparisons), judged to be enriched against more than one vertebrate. Terms enriched across multiple vertebrates for all three chondrichthyans (Datasets S5–S7) included a subset of 23 terms, covering a wide-ranging functional landscape, from regulating hydrolase, to very broad terms regarding anatomical structure formation. Within this subset there were, however, several terms related to genome stability, examples of which are presented in Fig. 4*A* and include the same or very similar terms that were prominent in the positive selection list, such as DNA repair, regulation of apoptosis, and negative regulation of cell proliferation.

Additional GO terms related to genome stability and enriched in more than one intervertebrate comparison, included regulation of Wnt signaling pathway, chromatin organization, histone modification, and hemopoiesis. The Wnt signaling pathway increases quantities of β-catenin, which in turn can initiate transcriptional activation of proteins controlling the G1 to S phase cell cycle transition, ultimately leading to cell proliferation (34). Hematopoietic stem cells (HSCs) and their maintenance are critical because the life-time persistence and regenerative potential of HSCs requires tight control of HSC genome stability, and indeed HSCs have highly efficient DNA repair capabilities (35). Several genes important to stem cell maintenance were also under positive selection on different chondrichthyan branches (Table 1). In addition to these enrichments against GO Biological Process, there were also several notable genome-stability term enrichments in comparison of the chondrichthyans to the Reactome database (Fig. 4*B*), including DNA repair, signal transduction, and transcriptional
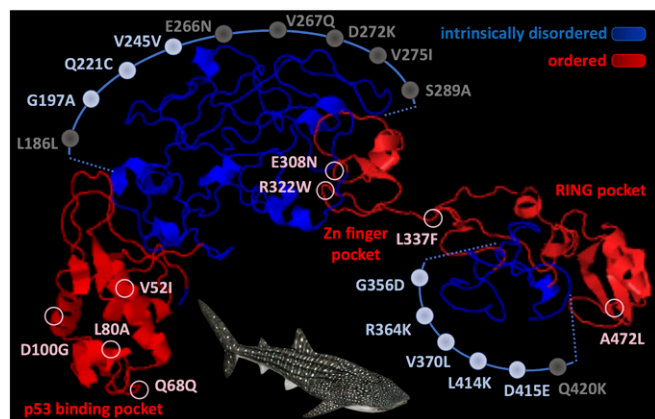
Fig. 3. Protein model of MDM4 from whale shark, showing positively selected sites. The site positions correspond to the human Swiss-Prot reference sequence. The residue to the left of the position number is the human amino acid at that position and to the right the whale shark residue at the corresponding positively selected site. Corresponding sites between whale shark and human were determined from an amino acid alignment. Solid light blue circles in disordered regions represent sites that could be accurately aligned; gray circles could not, however, the position number nonetheless reflects the approximate location, determinable because of alignment anchors closely flanking these regions of variability. All positions in ordered regions of the protein could be accurately aligned between human and whale shark.

Fig. 4. Gene content enrichments of genome stability and wound-healing terms. (A) GO Biological Process and (B) Reactome pathways for all three chondrichthyan genomes compared to model organism vertebrates. (A) Human, *Homo sapiens*; frog, *Xenopus tropicalis*; anole, *Anolis caroliensis*; chicken, *Gallus gallus*; zebrafish, *Danio rerio*.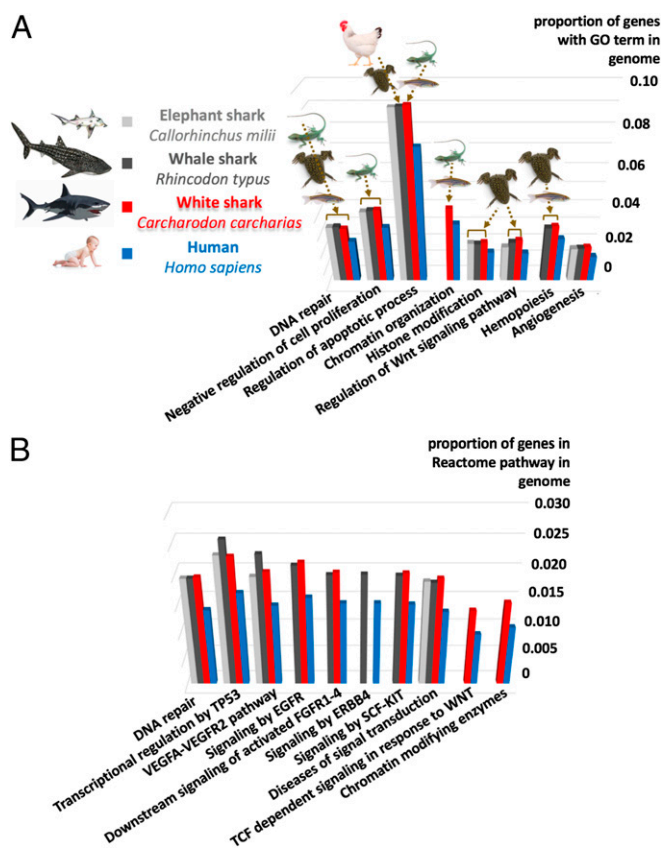 (B) Human only. For comparisons depicted in A, human and each of the chondrichthyan proportions are represented, and an arrow connecting the schematics to histogram bars illustrates the vertebrates to which that term was enriched in comparisons involving that chondrichthyan. With the exception of angiogenesis, only terms that were enriched compared to two or more vertebrates are represented; all presented terms for chondrichthyans were enriched compared with human.

regulation by TP53, terms that are prominent throughout the comparative analyses herein presented. SCF-Kit (stem cell factor-tyrosine kinase receptor) signaling, enriched in the elasmobranchs, plays a role in the regulation of HSCs. Thus, it is not just the positive selection of protein-coding genes with roles pertinent to genome stability, but also accumulated gene content within functional categories related to genome stability, that typifies the history of molecular evolution in chondrichthyans.

Core Histone Gene Expansions and Their Possible Role in Genome Stability. Unlike the majority of the other genome-stability functional enrichments presented in Fig. 4, chromatin organization and histone modification (Biological Process), as well as chromatin-modifying enzymes (Reactome), were only enriched in white shark. This could be a reflection of the much larger genome of white shark and the associated requirements for efficient DNA packaging, and due to the inflated numbers of core histones present in this species (Table 2). In addition to their role in packaging DNA, histones also play key roles in the DNA damage response. Acetylation of histone H3 K56 is thought to be a key factor in affording a favorable chromatin environment for DNA repair (36). Acetylation and methylation of H3-K9 are required for proper chromosome condensation, an integral factor in the maintenance of genomic stability (37). GO terms referring to H3-K9 acetylation and meth-

ylation were among those that were significantly enriched in the set of whale shark positively selected genes when compared back against the whale shark genome. Similarly, terms referring to acetylation of histone H4, and in particular, acetylation of H4-K16, were also among the enriched GO terms in this same comparison. Acetylation of H4-K16 is an important factor in recruiting the mediator of DNA damage checkpoint protein 1 (MDC1) in DNA damage repair (38). A high number of H2A histone genes were noted in both the white and whale shark genomes (Table 2); of particular note is the large number of putative H2AX that are part of this H2A total. The phosphorylation of H2AX is another critical component of the DNA damage response, because it is phosphorylated H2AX, which mediates interaction with MDC1 (39), which in turn binds the DNA repair protein NBS1 (DNA repair and telomere maintenance protein nbs1) (40). Thus, a large number of histone genes, in particular those known to play key roles in DNA damage response, coupled with positive selection related to epigenetic modification of histones related to DNA damage response, suggests an evolutionary history where core histones have played a particularly important role in the maintenance of genome stability.

Adaptation for Wound Healing. Our comparative genomic analyses identified both positive selection and gene content enrichment results which suggest possible molecular genetic support relevant to wound healing. In terms of positive selection, three loci stood out in particular, FGG (fibrinogen γ-chain), EXTL2 (exostosin-like 2), and KRT18 (keratin, type I cytoskeletal 18). FGG is the γ-component of the blood-borne glycoprotein, fibrinogen, which when cleaved by thrombin to form fibrin, comprises the main component of blood clots (Fig. 5 *A* and *B*) (41). FGG was under positive selection on both the white and whale shark branches (Fig. 5*C*). With the exception of one site in whale shark (in the central nodule), all of the positively selected sites were in the γ-nodule (Fig. 5*C*). Fibrinogen has several binding sites for calcium ions, which are important for its function, including fibrin polymerization; many of these Ca^{2+} binding sites are in the γ-nodule (42). All of the positively selected sites in the γ-nodule were in regions of α-helix or β-strand structure (Fig. 5*D*). EXTL2 is a glycosyltransferase required for the biosynthesis of heparan-sulfate, which is a proteoglycan that binds a variety of ligands and regulates various processes, including angiogenesis and blood coagulation (44). EXTL2 was under positive selection on the elephant shark branch. KRT18, is a type 1 cytokeratin, and a member of the intermediate filament gene family. Keratin proteins provide mechanical support and protection against injury and several of them play an active part in healing wounds (45). KRT18 was under positive selection on the elasmobranch branch.

In addition to these positive selection results, GO enrichments relevant to wound healing were also evident in the Panther intervertebrate genomic comparisons and the Reactome comparisons (Fig. 4), including terms such as angiogenesis and the VEGFA-VEGFR2 signaling network. Angiogenesis, the formation of new blood vessels from preexisting vasculature, is central to a number of

Table 2. Number of predicted proteins in each chondrichthyan species that had significant blast hits to core histone proteins from Swiss-Prot, with greater than 90% similarity and within 20 amino acids in length of the subject protein sequence

Core histone	White shark	Whale shark	Elephant shark
H2A	48*	41†	7
H2B	15	4	2
H3	25	26	2
H4	58	32	6

*Of these sequences, 37 are H2AX.
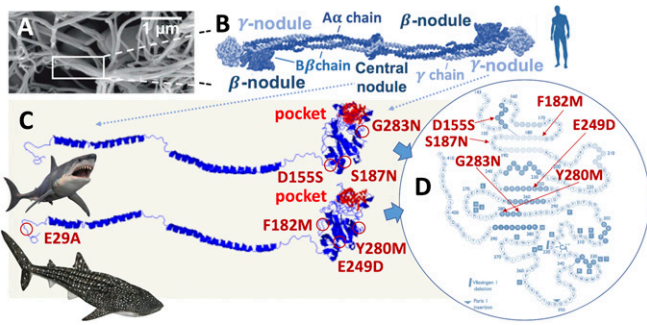†Of these, 14 are H2AX.

GENETICS

Fig. 5. Unfolding of the coiled-coils of fibrin, illustrating FGG and the location of positively selected sites in white and whale sharks. (*A*) A scanning electron micrograph of a fibrin clot, with a box enclosing part of a fiber (zoom-in cartoon in *B*). (*B*) Schematic representation of the human fibrinogen (FG) molecule in the naturally folded state (PDB ID code 3GHG), consisting of pairs of Aα chains (in dark blue), Bβ chains (in medium blue), and γ chains (in light blue: zoom-in shown in *C*), linked by S-S bonds. Structural details include the central nodule, γ-nodules, β-nodules. (*C*) Protein models of the white and whale shark FGG highlighting the residues with evidence of positive selection in the γ-nodule, and for whale shark, also the central nodule. The site positions correspond to the human Swiss-Prot reference sequence. The residue to the left of the position number is the human amino acid at that position and to the right the shark residue at the corresponding positively selected site. Corresponding sites between shark and human were determined from an amino acid alignment. (*D*) Bead model of the globular carboxyl-terminal region (γ-nodule) of the human fibrinogen γ chain, from Val143 to Val411. Darker colored amino acid beads indicate stretches of α-helix or β-strand structure. Sites under positive selection in the white shark (*Left*) and the whale shark (*Right*) are highlighted. (*A* and *B*) Adapted by permission from ref. 42, Springer Nature: Subcellular Biochemistry, copyright 2017. (*D*) Republished with permission of American Society of Hematology, from ref. 43; permission conveyed through Copyright Clearance Center, Inc.

conditions, including wound healing. Vascular endothelial growth factor (VEGF) is the key angiogenic growth factor that modulates angiogenesis via receptor tyrosine kinase VEGF receptors (VEGFRs) (46). Several other enrichments involving signaling pathways fundamental to wound healing were evident in the elasmobranch comparisons against Reactome (Fig. 4*B*), including epidermal growth factor receptor (EGFR), fibroblast growth factor receptor (FGFR), and receptor tyrosine-protein kinase erbB-4 (ERBB4). EGF and EGFR play an essential role in wound healing through the stimulation and proliferation of fibroblast, keratinocyte, and endothelial cells, facilitating epidermal and dermal regeneration (47). FGF are signaling proteins that bind and activate a series of FGF receptors (FGFR); several FGF, for example FGF-1 and -2, play a critical role in the wound-healing process, including fibroblast and keratinocyte proliferation, wound contraction, and angiogenesis (48). ERBB4 is a receptor tyrosine kinase that is a member of the EGFR subfamily and is activated by a number of ligands, including heparin-binding EGF-like growth factor (HB-EGF); ligand binding induces a number responses, including mitogenesis and differentiation, both important aspects of skin wound healing (49). Furthermore, HB-EGF is regarded as the principal growth factor in the epithelialization of skin wound healing (50).

Smell. Many aquatic and terrestrial species with an enhanced sense of smell have a proliferation of loci, referred to as olfactory receptor (OR) genes (51). Sharks locate prey using an apparently keen sense of smell (10, 11, 52). Thus, it might be expected that sharks and other chondrichthyans would have a large OR repertoire; however, the predicted protein sequences of white shark, whale shark, and elephant shark included only two putative OR proteins in white shark, two in whale shark, and two in elephant shark. [Venkatesh et al. (4) reported two matching OR gene families and six additional OR-like genes.] More detailed examination of sequence scaffolds in the white shark revealed one additional OR sequence, as well as an

OR pseudogene. Brownbanded bamboo shark and cloudy catshark also only have three OR genes (6), so it appears likely that chondrichthyans as a whole have a distinct paucity of OR sequences, including predatory epipelagic sharks, such as white shark. This raises the possibility of an alternative gene family responsible for enhanced smell in white sharks or perhaps selection on genes responsible for detecting a particular odorant and picking up the signal from that odorant at great distances, rather than a large gene repertoire to detect a wide range of odorants. Regarding the former explanation, it has been proposed that the vomeronasal system may be used by elasmobranchs in olfaction (6, 53). We found a single copy of the vomeronasal type 1 receptor (V1R) in the white shark proteins, three in whale shark, and two in elephant shark; for vomeronasal type 2 receptors (V2R), we found 13 V2R proteins in white shark, 10 in whale shark, and 5 in elephant shark, although Venkatesh et al. (4) report two clades of 6 and 25 V2R genes for elephant shark, after detailed examination of sequence scaffolds. White shark sequence scaffolds did not reveal additional V2R sequences, although we did note that white shark V2Rs are distributed across only four scaffolds, suggesting a clustered expansion. Based on ultrastructural, immunohistochemical, and the elephant shark genome sequence, Ferrando and Gallus (53) suggested that the primary olfactory system of Chondrichthyes primarily relies on V2Rs. Based on the copy number of V2Rs in our examination of the white and whale shark genomes, V2Rs may indeed represent the most parsimonious hypothesis for olfaction reception in chondrichthyans. Vomeronasal gene copy number was not reported in the brownbanded bamboo and cloudy catshark paper (6).

We also searched for the presence of additional alternative genes that can serve as receptors of olfactory signals, namely trace amine associated receptors (TAARs), which are G protein-coupled receptors that function as vertebrate ORs (54, 55). We found one TAAR in white shark proteins, six whale shark proteins matching TAARs, and five elephant shark proteins matching TAARs. A search of the scaffolds in white shark did not identify additional white shark TAAR sequences. The Panther intervertebrate comparative genomic comparisons corroborated this overall paucity of olfactory gene content, with dramatic underrepresentation of gene content in the GO categories "detection of chemical stimulus involved in sensory perception of smell (GO:0050911)," or simply "sensory perception of smell (GO:0007608)" in all five intervertebrate comparisons made for all three chondrichthyans (Datasets S5–S7). Another important functional category very underrepresented in all shark/intervertebrate comparisons, and of which the odorant receptors are a subtype was: "G-protein coupled receptor signaling pathway (GO:0007186)" (Datasets S5–S7). G protein-coupled receptors are a large family of proteins found throughout eukaryotes; the ligands that bind and activate them include odors, pheromones, light-sensitive compounds, and neurotransmitters. Chondrichthyans have many fewer representatives of this class of protein receptors involved in numerous human diseases, and which are the target of about 35% of current medicinal drugs (56).

As an alternative to odorant receptor gene content, the overexpression of receptors and molecular adaptation of protein-coding genes involved in detection of olfactory signals, or the transmission of signals, could be involved in any enhanced sense of smell. Of potential significance in this latter regard is evidence of positive selection on a Bardet–Biedl syndrome protein (Bbs5) on the white shark branch. Bbs5 is one of the proteins that comprise the BBSome, a protein complex that functions in primary cilium biogenesis. Bardet–Biedl syndrome is a ciliopathy with a pleiotropic phenotype, including anosmia (57). An additional positively selected locus in white shark relevant to odorant reception is I5T52, which is one of the subcomponents of the intraflagellar transport complex B, which is essential for the formation and maintenance of cilia. Mutations in at least one of these complex B subcomponents has been linked to Bardet–Biedl syndrome (58). It is possible that positive selection in one or more of these BBSome and intraflagellar transport complex

proteins, could affect the transfer of odorant signals and might represent part of a compensatory or alternative adaptive strategy for odor detection.

Discussion

The results herein presented provide evidence of a history of selection pressure underlying the maintenance of elasmobranch genome stability and the development of shark superior wound-healing capabilities. The evidence comes from a consideration of both the function of positively selected genes and gene-content enrichment compared with other vertebrates, with the two types of evidence complementing one another. We regard our estimates of the number of positively selected genes to be conservative assessments. This is not only because we adopted a stringent manual inspection and editing of alignments, but also based on a consideration of the possible influence of synonymous saturation. When conducting positive selection analysis, comparing species that have diverged over a few hundred million years, saturation at synonymous sites can be evident, and indeed even after our conservative treatment of alignments, some of our genes did show evidence for a slight level (small tailing off from linear at extreme end of saturation plots) of third-position saturation (see *Methods* for more detail). Importantly, however, simulation studies have demonstrated that the branch-site test employed here is robust to synonymous saturation and that false negatives are much more likely than false positives (59), thus yielding an overall more conservative assessment of genes under positive selection (complete details of positive selection analysis provided in *Methods*).

Recent studies with plants have shown that there is a correlation between genome size and DNA damage, with larger genomes suffering greater damage, but now also providing evidence that this is likely not the consequence of less-efficient DNA repair in the species with larger genomes (60). This implies that organisms with large genomes may have adapted to the higher probability of DNA damage through the evolution of enhanced DNA damage response and repair. The genome size of white shark, at 4.63 Gbp, is large, but not unusually so for elasmobranchs; the two new genomes for brownbanded bamboo shark and cloudy catshark are 4.7 and 6.7 Gbp, respectively (6). White shark did have a greater proportion of genes under positive selection with functions related to genome stability than the other two chondrichthyans in our analysis. The larger genome size of white shark is at least partly due to a large repeat content and the proliferation of LINEs, particularly LINE-3/ CR1 elements (*SI Appendix*, Fig. S2). LINE-1 elements generate double-strand breaks (DSBs), and notably, the number of LINE-1– created DSBs has been shown to be greater than the number of successful insertions, suggesting a degree of inefficiency in the integration process (61). We are not aware of similar experiments with LINE-3 elements; however, they are regarded as a more ancient family of LINEs, with similar retro-transposition capabilities to LINE-1 (62). A proliferation of LINEs on the white shark lineage could represent a strong selective agent for the evolution of efficient DSB repair. This raises the question of whether this large genome (and indeed large genomes in general) evolved because of superior DNA damage repair, or do large genomes and their accompanying repetitive elements act as selective agents for the evolution of more proficient DNA damage-repair systems.

The maintenance of genome stability is critically important in the aging process. It is widely accepted that the main cause of aging is the gradual, accumulation of molecular and cellular damage. López-Otìn et al. (17) identified three hallmarks of aging, the first of which they called "primary hallmarks," including such processes as DNA damage, telomere loss, and epigenetic alterations that trigger the aging process. The accumulation of DNA damage plays a key role in triggering the aging process, because it can result in aging in a number of ways: (*i*) mutagenic lesions that result in cancer, (*ii*) defects of cellular functions, and (*iii*) cell death and senescence (63). Furthermore, the decrease in the efficiency of DNA repair with age results in a

feedback loop that reinforces aging. Maintenance of genome stability may also play an important role in the development of large body size. Theoretically, the risk of developing cancer should increase with both the number of cells and an organism's lifespan, and there is statistical support for a positive relationship of size and cancer risk within a species (64). However, this does not tend to hold up across species (called Peto's Paradox) (65), and very large animals do not get cancer more often than humans, suggesting that superior cancer fighting abilities have evolved numerous times across the tree of life. Long-lived, large mammal species, such as bowhead whale and elephant, have recently been shown to possess copy number variants and positive selection of important genome stability related genes that could reflect their solution to Peto's Paradox (66–68). We did not find copy number variants of genome stability-related genes [such as that reported for TP53, in the elephant (67)] in the white shark genome, and indeed with the exception of histones, limited evidence of gene family expansions. It is possible that the positive selection and gene-content enrichments we report here may reflect adaptations that act to fine-tune mechanisms related to the maintenance of genome integrity in these sharks and could be at least part of the overall molecular character, facilitating the evolution of their large bodies and long lifespans. Elasmobranchs as a group exhibit a great deal of variation in genome size, body size, lifespan, and quite likely, repeat content. These suggestions related to the evolution of elasmobranch genome integrity are conjectural hypotheses, but nonetheless consistent with our data, and represent hypotheses that could be tested with the proper comparative genomic datasets to provide a more definitive picture.

The evidence we present on the positive selection of genes involved in wound healing and gene content enrichments involve several key loci and some of the most fundamental pathways in wound healing. The remarkable capabilities of elasmobranch wound healing are well known to anyone working in elasmobranch field biology and, as mentioned earlier, empirical evidence now supports this (18). Our ortholog recovery process for positive selection necessitated that all of the same eight species were represented for every gene included in the analysis and therefore resulted in testing only a small set of possible genes. A broader targeted sampling of genes involved in wound healing, for positive selection analysis, concomitant with transcriptomic studies of control and wounded tissue would be invaluable in identifying some of the key molecular loci underlying this unusual ability. Indeed, comparative genomic and transcriptomic work on sharks varying in not only the characteristics discussed in this paper, but many of their other life-history characteristics, could provide much information relevant to not only an understanding of basic biology of these vertebrates of great antiquity, but also provide information of potentially valuable biomedical significance.

Methods

Sample Collection. DNA from two separate *C. carcharias* individuals were used to build a hybrid genome assembly, one of these comprising the primary genome assembly produced at Cornell University (see assembly methods, below). DNA was extracted from heart tissue of this individual, a female caught in the Atlantic Ocean off the coast of Delaware (see previous RNA-seq studies involving this same individual for further details) (69, 70). DNA extractions from this individual yielded a mixture of high molecular weight, as well as more fragmented pieces, and it was decided that a second sample would be necessary to obtain sufficient amounts of high molecular-weight DNA for scaffolding at Dovetail Genomics (see assembly methods, below). A second individual (198-cm male) was captured and released live off the Pacific coast of southern California on November 6, 2014. An extraction of blood was preserved on dry ice, and subsequently frozen at −80 °C; it was this male that was used for Dovetail Genomics scaffolding. Additionally, a biopsy (muscle, subdermis, and epidermis) was extracted from a third, free-swimming individual (300-cm male) off Tomales Point in central California on September 26, 2016. RNA sequencing was conducted on the additional tissues (blood, muscle, subdermis, and epidermis) from both these Pacific individuals to supplement the

heart transcriptome of the Atlantic individual. Samples regarding the Pacific individuals were taken under permit from the California Department of Fish and Wildlife (Monterey Bay Aquarium Entity Permit 1349) and all procedures were reviewed and approved by the Monterey Bay Aquarium Research Oversight Committee. The sample from the Atlantic individual was obtained from the National Oceanic and Atmospheric Administration; details regarding this specific heart sample are outlined in Richards et al. (69).

Sequencing and Genome Assembly. The Atlantic individual was used for production of an initial genome assembly through deep sequencing on the Illumina 2500 sequencing platform. Sequencing libraries included a variety of 150-bp single-end, 2 × 150-bp paired end, 2 × 250-bp paired end, overlapping 2 × 250-bp paired end (producing 450-bp single-end reads), and mate paired sequencing libraries using 3–5 kbp, 8–10 kbp, and 15–20 kbp inserts (see Dataset S8 for statistics on each library type). These reads were assembled in SOAPdenovo2 (71) (this assembler yielded the best assembly of the programs able to handle the entire set of read data) using a mixed k-mer strategy following trimming of adaptors and poor quality sequence using Trimmomatic (72) (see Dataset S8 for settings of bioinformatics programs used in the assembly). This assembly was used as input for scaffolding by Dovetail Genomics with Chicago library sequencing of DNA extracted from the Pacific individual (198-cm male). The final assembly consisted of the original assembly subsequently linked into larger scaffolds by these Chicago libraries. To assess genome quality and completeness, we ran BUSCO (73) on the Dovetail genome assembly, as well as using it to obtain white shark-specific training data for the AUGUSTUS (74) gene-prediction program. Additional methods for transcriptome and genome annotation are in *SI Appendix*.

PANTHER and GO Comparisons. To identify differences in the types of genes present in the three chondrichthyan genomes relative to model vertebrates, we conducted a comparison of GO between the predicted proteins of each chondrichthyan genome (white shark, whale shark, and elephant shark) against five different model vertebrates (*Homo sapiens*, *Anolis carolinensis*, *Gallus gallus*, *Xenopus tropicalis*, and *Danio rerio*) in separate pairwise comparisons for each of the GO categories (Biological Process, Molecular Function, and Cellular Component) using the Panther database (33) overrepresentation test. We also ran similar tests against the Reactome (29) database, which is based on human. For each comparison we obtained the Swiss-Prot IDs for the orthologs between the chondrichthyan and reference model species, and then tested this list of chondrichthyan IDs against the complete genome list of proteins from that reference species. This approach of intervertebrate comparison using the Panther system is very similar to that recently performed with a hummingbird transcriptome (75). We ran the Panther statistical overrepresentation test employing the Fishers exact test with false-discovery rate multiple-test correction for each of these 48 (3 chondrichthyans × 5 model species × 3 GO categories + 3 Reactome) pairwise comparisons and identified the most specific GO terms (the most specific term in a hierarchy of terms exhibiting the same statistical pattern; for example, if cell membrane, protein integral to cell membrane, and sodium ion channel were all significant, then sodium ion channel would be the most specific term) that were overrepresented in the tested chondrichthyan.

Positive Selection. To identify cases of molecular adaptation, we conducted tests for positive selection using the branch-sites test employed within the codeml package of PAML (76). As input for this analysis we identified orthologous sequences from each of the three chondrichthyan genomes (elephant shark, whale shark, and our assembled white shark genome) as well as several other fish species covering a wide range of fish groups with existing genomes (zebrafish, *Danio rerio*; Amazon molly, *Poecilia formosa*; blind cave fish, *Astyanax mexicanus*; Nile tilapia, *Oreochromis niloticus*; coelacanth, *Latimeria chalumnae*; spotted gar, *Lepisosteus oculatus*). This choice of species was based on several factors, including quality of genome assembly, the inclusion of adequate number of species to provide the analysis with power, and avoiding the inclusion of species so divergent that the number of recovered orthologs would be diminished, while concomitantly increasing the chances of alignment ambiguity for those few that were recovered. The analysis included the genome coding sequences available on GenBank as of September 1, 2017. Ortholog

recovery for the chondrichthyans involved taking the longest protein corresponding to each gene of white shark, whale shark, and elephant shark and aligning to profile hidden Markov models of orthologous groups of the veNOG subset from the *eggNOG* database (77) using *HMMER*. Top hits from the alignments were extracted and used for assignment of corresponding proteins to orthologous groups. For the other six species, proteins and the corresponding assignments were obtained from the veNOG database. Protein sequences were initially aligned using MAFFT and then the corresponding coding sequences were used to reconstruct a codon based alignment. This alignment was then used to build a maximum-likelihood tree RaxML (78) with a GTR+gamma model of molecular evolution and 1,000 bootstrap replicates. Before testing in PAML, Gblocks was used to remove poor regions of codon alignments. Each alignment was retained only if the gene of the species corresponding to the branch tested, covered greater than 60% of the coding sequence length, as determined by its comparison with the Swissprot reference sequence for that gene. Finally, 1,541 remaining alignments and the maximum-likelihood tree were used to test for positive selection. All alignments were tested for evidence of selection using the branch-sites test in separate runs on the following lineages: separately on each of the three chondrichthyan species (elephant shark, whale shark, white shark), the elasmobranch lineage (branch leading to whale shark and white shark), and the ancestral chondrichthyan branch (branch leading to the Chondrichthyes). All positive-selection tests were run with the "cleandata" option of PAML set to ignore all gapped regions. Adjustments were made to correct for multiple testing before identifying a gene as significantly under positive selection using the qvalue R package by Storey et al. (79) as implemented on qvalue.princeton.edu/. It is well documented that misaligned genes greatly inflate estimates of false positives in this analysis (80, 81). The "gold standard" for this assurance is manual inspection of all the alignments judged to be under positive selection. All alignments that were identified as under positive selection were manually inspected and edited to remove the possibility of false positives from alignment errors. PAML was then rerun on all manually checked alignments and the q-value correction reimplemented to identify the genes confidently judged to be under positive selection. When conducting positive-selection analysis that includes species that have diverged over a few hundred million years, saturation at synonymous sites can be evident. An examination of third position saturation plots of our genes (uncorrected P distance against HKY corrected distances) did indicate for some genes a slight level of saturation (slight tailing away from linear). Importantly, however, simulation studies have demonstrated that the branch-site test is robust to synonymous saturation and that false negatives are much more likely than false positives (18).

Three-dimensional protein modeling of whale shark MDM4 was performed on the PHYRE2 Protein Fold Recognition Server, www.sbg.bio.ic.ac.uk/~phyre2.

1. Naylor G-J-P, et al. (2012) A DNA sequence-based approach to the identification of shark and ray species and its implications for global elasmobranch diversity and parasitology. *Bull Am Mus Nat Hist* 367:1–262.
2. Fergusson IK, Compagno LJV, Marks MA (2009) *Carcharodon carcharias*. IUCN red list of threatened species. Version 2018.2. Available at www.iucnredlist.org. Accessed August 3, 2018.
3. Read TD, et al. (2017) Draft sequencing and assembly of the genome of the world's largest fish, the whale shark: *Rhincodon typus* Smith 1828. *BMC Genomics* 18:532.
4. Venkatesh B, et al. (2014) Elephant shark genome provides unique insights into gnathostome evolution. *Nature* 505:174–179.
5. Wang Q, et al.; North East Bioinformatics Collaborative Curation Team (2012) Community annotation and bioinformatics workforce development in concert—Little skate genome annotation workshops and jamborees. *Database (Oxford)* 2012:bar064.
6. Hara Y, et al. (2018) Shark genomes provide insights into elasmobranch evolution and the origin of vertebrates. *Nat Ecol Evol* 2:1761–1771.

7. Schwartz FJ, Maddock MB (1986) Comparisons of karyotypes and cellular DNA contents within and between major lines of elasmobranch. *Indo-Pacific Fish Biology*, eds Uyeno T, Arai R, Tuniuchi T, Matsuura K (Ichthyological Society Japan, Tokyo), pp 148–157.

8. Castro JI (2011) *The Sharks of North America* (Oxford Univ Press, Oxford).

9. Hamady LL, Natanson LJ, Skomal GB, Thorrold SR (2014) Vertebral bomb radiocarbon suggests extreme longevity in white sharks. *PLoS One* 9:e84006.

10. Sheldon RE (1911) The sense of smell in selachians. *J Exp Zool* 10:51–62.

11. Gardiner JM, Atema J (2010) The function of bilateral odor arrival time differences in olfactory orientation of sharks. *Curr Biol* 20:1187–1191.

12. Walsh CJ, et al. (2013) Epigonal conditioned media from bonnethead shark, *Sphyrna tiburo*, induces apoptosis in a T-cell leukemia cell line, Jurkat E6-1. *Mar Drugs* 11:3224–3257.

13. Ostrander GK, Cheng KC, Wolf JC, Wolfe MJ (2004) Shark cartilage, cancer and the growing threat of pseudoscience. *Cancer Res* 64:8485–8491.

14. Walsh CJ, et al. (2004) Characterization of shark immune cell factor (*Sphyrna tiburo* epigonal factor, STEF) that inhibits tumor cell growth by inhibiting S-phase and inducing apoptosis via the TRAIL pathway. *FASEB J* 18:A60.

15. Negrini S, Gorgoulis VG, Halazonetis TD (2010) Genomic instability—An evolving hallmark of cancer. *Nat Rev Mol Cell Biol* 11:220–228.

16. McKinnon PJ (2017) Genome integrity and disease prevention in the nervous system. *Genes Dev* 31:1180–1194.

17. López-Otín C, Blasco MA, Partridge L, Serrano M, Kroemer G (2013) The hallmarks of aging. *Cell* 153:1194–1217.

18. Chin A, Mourier J, Rummer JL (2015) Blacktip reef sharks (*Carcharhinus melanopterus*) show high capacity for wound healing and recovery following injury. *Conserv Physiol* 3:cov062.

19. Schwertman P, Bekker-Jensen S, Mailand N (2016) Regulation of DNA double-strand break repair by ubiquitin and ubiquitin-like modifiers. *Nat Rev Mol Cell Biol* 17:379–394.

20. Supek F, Bošnjak M, Škunca N, Šmuc T (2011) REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One* 6:e21800.

21. Chen H, et al. (2015) Erk signaling is indispensable for genomic stability and self-renewal of mouse embryonic stem cells. *Proc Natl Acad Sci USA* 112:E5936–E5943.

22. Shimada K, et al. (2013) TORC2 signaling pathway guarantees genome stability in the face of DNA strand breaks. *Mol Cell* 51:829–839.

23. Rimkus TK, Carpenter RL, Qasem S, Chan M, Lo HW (2016) Targeting the sonic hedgehog signaling pathway: Review of smoothened and GLI inhibitors. *Cancers (Basel)* 8:22.

24. Mikhed Y, Görlach A, Knaus UG, Daiber A (2015) Redox regulation of genome stability by effects on gene expression, epigenetic pathways and DNA damage/repair. *Redox Biol* 5:275–289.

25. Gartel AL, Radhakrishnan SK (2005) Lost in transcription: p21 repression, mechanisms, and consequences. *Cancer Res* 65:3980–3985.

26. Bunz F, et al. (1998) Requirement for p53 and p21 to sustain G2 arrest after DNA damage. *Science* 282:1497–1501.

27. Perry ME (2010) The regulation of the p53-mediated stress response by MDM2 and MDM4. *Cold Spring Harbor Perspect Biol* 2:a000968.

28. Chen J (2016) Intra molecular interactions in the regulation of p53 pathway. *Transl Cancer Res* 5:639–649.

29. Fabregat A, et al. (2018) The reactome pathway knowledgebase. *Nucleic Acids Res* 46:D649–D655.

30. Weisman R, Cohen A, Gasser SM (2014) TORC2-a new player in genome stability. *EMBO Mol Med* 6:995–1002.

31. Wickramasinghe VO, Venkitaraman AR (2016) RNA processing and genome stability: Cause and consequence. *Mol Cell* 61:496–505.

32. Peng JC, Karpen GH (2009) Heterochromatic genome stability requires regulators of histone H3 K9 methylation. *PLoS Genet* 5:e1000435.

33. Mi H, et al. (2017) PANTHER version 11: Expanded annotation data from gene ontology and reactome pathways, and data analysis tool enhancements. *Nucleic Acids Res* 45:D183–D189.

34. Polakis P (2000) Wnt signaling and cancer. *Genes Dev* 14:1837–1851.

35. Vitale I, Manic G, De Maria R, Kroemer G, Galluzzi L (2017) DNA damage in stem cells. *Mol Cell* 66:306–319.

36. Masumoto H, Hawke D, Kobayashi R, Verreault A (2005) A role for cell-cycle-regulated histone H3 lysine 56 acetylation in the DNA damage response. *Nature* 436:294–298.

37. Park JA, et al. (2011) Deacetylation and methylation at histone H3 lysine 9 (H3K9) coordinate chromosome condensation during cell cycle progression. *Mol Cells* 31:343–349.

38. Li X, et al. (2010) MOF and H4 K16 acetylation play important roles in DNA damage repair by modulating recruitment of DNA damage repair protein Mdc1. *Mol Cell Biol* 30:5335–5347.

39. Stucki M, et al. (2006) MDC1 directly binds phosphorylated histone H2AX to regulate cellular responses to DNA double-strand breaks. *Cell* 124:1299.

40. Chapman JR, Jackson SP (2008) Phospho-dependent interactions between NBS1 and MDC1 mediate chromatin retention of the MRN complex at sites of DNA damage. *EMBO Rep* 9:795–801.

41. Mosesson MW (2003) Fibrinogen γ chain functions. *J Thromb Haemost* 1:231–238.

42. Weisel JW, Litvinov RI (2017) Fibrin formation, structure and properties. *Subcell Biochem* 82:405–456.

43. Côté HCF, Lord ST, Pratt KP (1998) Gamma-chain dysfibrinogenemias: Molecular structure-function relationships of naturally occurring mutations in the gamma chain of human fibrinogen. *Blood* 92:2195–2212.

44. Kitagawa H, Shimakawa H, Sugahara K (1999) The tumor suppressor EXT-like gene EXTL2 encodes an alpha1, 4-N-acetylhexosaminyltransferase that transfers N-acetylgalactosamine and N-acetylglucosamine to the common glycosaminoglycan-protein linkage region. The key enzyme for the chain initiation of heparan sulfate. *J Biol Chem* 274:13933–13937.

45. Kim S, Wong P, Coulombe PA (2006) A keratin cytoskeletal protein regulates protein synthesis and epithelial cell growth. *Nature* 441:362–365.

46. Abhinand CS, Raju R, Soumya SJ, Arya PS, Sudhakaran PR (2016) VEGF-A/VEGFR2 signaling network in endothelial cells relevant to angiogenesis. *J Cell Commun Signal* 10:347–354.

47. Bodnar RJ (2013) Epidermal growth factor and epidermal growth factor receptor: The Yin and Yang in the treatment of cutaneous wounds and cancer. *Adv Wound Care (New Rochelle)* 2:24–29.

48. Yun YR, et al. (2010) Fibroblast growth factors: Biology, function, and application for tissue regeneration. *J Tissue Eng* 2010:218142.

49. Grazul-Bilska AT, et al. (2003) Wound healing: The role of growth factors. *Drugs Today (Barc)* 39:787–800.

50. Shirakata Y, et al. (2005) Heparin-binding EGF-like growth factor accelerates keratinocyte migration and skin wound healing. *J Cell Sci* 118:2363–2370.

51. Niimura Y (2009) Evolutionary dynamics of olfactory receptor genes in chordates: Interaction between environments and genomic contents. *Hum Genomics* 4:107–118.

52. Parker GH, Sheldon RE (1913) The sense of smell in fishes. *Bull US Bur Fish* 32:33–46.

53. Ferrando S, Gallus L (2013) Is the olfactory system of cartilaginous fishes a vomeronasal system? *Front Neuroanat* 7:37.

54. Liberles SD (2009) Trace amine-associated receptors are olfactory receptors in vertebrates. *Ann N Y Acad Sci* 1170:168–172.

55. Liberles SD (2015) Trace amine-associated receptors: Ligands, neural circuits, and behaviors. *Curr Opin Neurobiol* 34:1–7.

56. Sriram K, Insel PA (2018) G protein-coupled receptors as targets for approved drugs: How many targets and how many drugs? *Mol Pharmacol* 93:251–258.

57. Kulaga HM, et al. (2004) Loss of BBS proteins causes anosmia in humans and defects in olfactory cilia structure and function in the mouse. *Nat Genet* 36:994–998.

58. Aldahmesh MA, et al. (2014) IFT27, encoding a small GTPase component of IFT particles, is mutated in a consanguineous family with Bardet-Biedl syndrome. *Hum Mol Genet* 23:3307–3315.

59. Gharib WH, Robinson-Rechavi M (2013) The branch-site test of positive selection is surprisingly robust but lacks power under synonymous substitution saturation and variation in GC. *Mol Biol Evol* 30:1675–1686.

60. Einset J, Collins AR (2018) Genome size and sensitivity to DNA damage by X-rays-plant comets tell the story. *Mutagenesis* 33:49–51.

61. Gasior SL, Wakeman TP, Xu B, Deininger PL (2006) The human LINE-1 retrotransposon creates DNA double-strand breaks. *J Mol Biol* 357:1383–1393.

62. Burch JB, Davis DL, Haas NB (1993) Chicken repeat 1 elements contain a pol-like open reading frame and belong to the non-long terminal repeat class of retrotransposons. *Proc Natl Acad Sci USA* 90:8199–8203.

63. Nicolai S, et al. (2015) DNA repair and aging: The impact of the p53 family. *Aging (Albany NY)* 7:1050–1065.

64. Albanes D (1998) Height, early energy intake, and cancer. Evidence mounts for the relation of energy intake to adult malignancies. *BMJ* 317:1331–1332.

65. Peto R, Roe FJ, Lee PN, Levy L, Clack J (1975) Cancer and ageing in mice and men. *Br J Cancer* 32:411–426.

66. Keane M, et al. (2015) Insights into the evolution of longevity from the bowhead whale genome. *Cell Rep* 10:112–122.

67. Sulak M, et al. (2016) TP53 copy number expansion is associated with the evolution of increased body size and an enhanced DNA damage response in elephants. *eLife* 5:e11994.

68. Vazquez JM, Sulak M, Chigurupati S, Lynch VJ, Zombie A (2018) LIF Gene in elephants is upregulated by TP53 to induce apoptosis in response to DNA damage. *Cell Rep* 24:1765–1776.

69. Richards VP, Suzuki H, Stanhope MJ, Shivji MS (2013) Characterization of the heart transcriptome of the white shark (*Carcharodon carcharias*). *BMC Genomics* 14:697.

70. Marra NJ, et al. (2017) Comparative transcriptomics of elasmobranchs and teleosts highlight important processes in adaptive immunity and regional endothermy. *BMC Genomics* 18:87.

71. Luo R, et al. (2012) SOAPdenovo2: An empirically improved memory-efficient short-read de novo assembler. *Gigascience* 1:18.

72. Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120.

73. Waterhouse RM, et al. (2017) BUSCO applications from quality assessments to gene prediction and phylogenomics. *Mol Biol Evol* 35:543–548.

74. Stanke M, Diekhans M, Baertsch R, Haussler D (2008) Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics* 24:637–644.

75. Workman RE, et al. (2018) Single-molecule, full-length transcript sequencing provides insight into the extreme metabolism of the ruby-throated hummingbird *Archilochus colubris*. *Gigascience* 7:1–12.

76. Zhang J, Nielsen R, Yang Z (2005) Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Mol Biol Evol* 22:2472–2479.

77. Huerta-Cepas J, et al. (2016) eggNOG 4.5: A hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Res* 44:D286–D293.

78. Stamatakis A (2014) RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30:1312–1313.

79. Storey JD, Taylor JE, Siegmund D (2004) Strong control, conservative point estimation, and simultaneous conservative consistency of false discovery rates: A unified approach. *J R Stat Soc B* 66:187–205.

80. Schneider A, et al. (2009) Estimates of positive Darwinian selection are inflated by errors in sequencing, annotation, and alignment. *Genome Biol Evol* 1:114–118.

81. Markova-Raina P, Petrov D (2011) High sensitivity to aligner and high rate of false positives in the estimates of positive selection in the 12 *Drosophila* genomes. *Genome Res* 21:863–874.

GENETICS