The genomics of exceptional longevity in rockfishes refine genetic foundations of human lifespan variation

Stephen Treaster (Stephen_Treaster@hms.harvard.edu)

Department of Orthopaedic Surgery, Boston Children's Hospital, Boston, MA, 02124, USA; Department of Genetics, Harvard Medical School, Boston, MA, 02124, USA https://orcid.org/0000-0002-1651-4570

Joris Deelen

Max Planck Institute for Biology of Ageing, Joseph-Stelzmann-Str. 9b, D-50931 Köln, Germany; Molecular Epidemiology, Department of Biomedical Data Sciences, Leiden University Medical Center, Leiden, The Netherlands; Cologne Excellence Cluster on Cellular Stress Responses in Aging-Associated Diseases (CECAD), University of Cologne, Cologne, Germany https://orcid.org/0000-0003-4483-3701

Jacob Daane

Department of Orthopaedic Surgery, Boston Children's Hospital, Boston, MA, 02124, USA; Department of Genetics, Harvard Medical School, Boston, MA, 02124, USA; Department of Marine and Environmental Sciences, Northeastern University Marine Science Center, Nahant, MA, 01908, USA

Joanne Murabito

Section of General Internal Medicine, Department of Medicine, Boston University School of Medicine, Boston, MA, USA; Framingham Heart Study, Framingham, MA 01701

David Karasik

Azrieli Faculty of Medicine, Bar-Ilan University, Safed, Israel; Marcus Institute for Aging Research, Hebrew Senior Life, Boston, MA, USA

Matthew Harris

Department of Orthopaedic Surgery, Boston Children's Hospital, Boston, MA, 02124, USA; Department of Genetics, Harvard Medical School, Boston, MA, 02124, USA

Article

Keywords: comparative genomics, convergence, longevity, human gwas, rockfish

Posted Date: December 13th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-1126287/v1

Genomics of exceptional longevity in rockfishes refine genetic foundations of human lifespan variation

Stephen Treaster[1,2]*, Joris Deelen[3,4,5], Jacob M. Daane[1,2,6], Joanne Murabito[7,8], David Karasik[9,10], Matthew P. Harris[1,2]*

[1] Department of Orthopaedic Surgery, Boston Children's Hospital, Boston, MA, 02124, USA
[2] Department of Genetics, Harvard Medical School, Boston, MA, 02124, USA
[3] Max Planck Institute for Biology of Ageing, Joseph-Stelzmann-Str. 9b, D-50931 Köln, Germany
[4] Molecular Epidemiology, Department of Biomedical Data Sciences, Leiden University Medical Center, Leiden, The Netherlands
[5] Cologne Excellence Cluster on Cellular Stress Responses in Aging-Associated Diseases (CECAD), University of Cologne, Cologne, Germany
[6] Department of Marine and Environmental Sciences, Northeastern University Marine Science Center, Nahant, MA, 01908, USA
[7] Section of General Internal Medicine, Department of Medicine, Boston University School of Medicine, Boston, MA, USA
[8] Framingham Heart Study, Framingham, MA 01701
[9] Azrieli Faculty of Medicine, Bar-Ilan University, Safed, Israel
[10] Marcus Institute for Aging Research, Hebrew Senior Life, Boston, MA, USA

*communicating authors, stephen_treaster@hms.harvard.edu, harris@genetics.med.harvard.edu

Key words: comparative genomics, convergence, longevity, human gwas, rockfish

Abstract

Longevity is a defining, heritable trait that varies dramatically between species. To resolve the genetic regulation of this trait, we have mined genomic variation in rockfishes, ranging in longevity from 11 to over 205 years. Shifts in rockfish longevity occurred multiple times independently, and in a short evolutionary time frame, thus empowering convergence analyses. Our analyses reveal a common network of genes under convergent restricted evolution in long-lived lineages, encompassing established aging regulators such as insulin-signaling, yet also identify flavonoid (aryl-hydrocarbon) metabolism as a novel pathway modulating longevity. Further, these genes were used to refine human longevity GWAS, identifying the aryl-hydrocarbon metabolism pathway to be significantly associated with the 99th percentile of human longevity, independently validating its importance and conservation. This evolutionary intersection highlights a novel, conserved genetic architecture that associates with the evolution of longevity across vertebrates and provides actionable targets for research into lifespan and healthspan modulation.

Introduction

Aging pathologies may be delayed, ameliorated, or prevented in aggregate by targeting the foundational declines in homeostasis and function arising over time. Knowledge of foundational targets suitable for this intervention remains limited, yet evolution has already leveraged such targets in the production of a vast diversity of longevities, including those well beyond a human lifespan. Various species undergo aging, and the associated functional declines, at wildly different rates and timings. As these traits are heritable and defining for many species, the underlying genetic mechanisms can be tracked through comparative genomic approaches.

Research into the genetic regulation of longevity has primarily focused on model organisms with lifespans on the order of weeks, months, or a few years, enabling broad screens and timely study of experimental interventions. However, short-lived models represent a fundamentally different evolutionary strategy than humans, and the idiosyncrasies influencing their longevity may not apply to our end of the aging spectrum. Despite great strides in unraveling mechanisms of aging in practicable rodent models, they are still are limited to 20-50% life extension when targeting canonical aging pathways (Liang et al., 2003). Results to date may have been insulated from the broader regulation of longevity itself by the utility of, and focus on, short-lived models.

The best dietary, pharmaceutical, and/or genetic manipulations in model organisms pales in comparison to the lifespan variation between species. There are many examples of long-lived animals that have evolved to endure the rigors of time - including those well past a human lifespan - such as the 507-year-old bivalve *Arctica islandica* (Ridgway et al., 2010), nearly 400 years for the Greenland shark, and over 200 years for bowhead whales (Keane et al., 2015). The Rougheye Rockfish, *Sebastes aleutianus,* is one such vertebrate species, with a maximum lifespan over 205 years (Munk, Kristen M., 2001) as determined from growth ring annuli in otoliths (Craig R. Kastelle et al., 2000). Regardless of the aging theory – oxidative damage, proteostasis collapse, DNA damage, telomere/genomic maintenance, epigenetic drift, etc. – *S.aleutianus* resists the deleterious effects of age for over two centuries, enduring the variety of internal and external stressors assured with time (Kikis et al., 2010; Koga et al., 2011).

S. aleutianus is not the only rockfish lineage with this exceptional capability. The clade encompasses at least 107 extant species, ranging in maximum longevity from 12 to 205 years (**Figure 1**) (Munk, Kristen M., 2001). Propitiously, multiple, independent lineages exhibit impressive lifespans **(Figure 2B**), imparting power into comparative approaches (Heras and Aguilar, 2019; Kolora et al., 2021). Such approaches are further augmented by the relative youth of the *Sebastes* clade, emerging about eight million years ago, as calibrated by molecular clocks and supported with well-characterized paleo-geographic landmarks (Hyde and Vetter, 2007). This recent emergence helps to minimize genetic noise between species; the comparisons are as similar as possible and genetic differences between the short- and long-lived lineages are more likely to be facilitating that phenotypic change.

Critically, the variation in rockfish longevity is only weakly correlated with size (**Figure 1**), or ecological determinates such as temperature and depth (Hua et al., 2015). The independence from common longevity correlates will minimize the contamination of putative longevity

mechanisms with those facilitating the other traits (Austad and Fischer, 1991). This independence is especially evident with ecological comparisons; comparator fish lineages rarely make it past two decades, even those that reach substantially larger sizes and live in overlapping environments (**Figure 1**). The largest fishes living in these waters, like the 2.53m, 363kg pacific halibut, only reaches a quarter of the lifespan of *S. aleutianus*. Rockfish longevity is truly exceptional, and their diversity, availability, and recent emergence combine to provide a unique opportunity for comparative genomics to elucidate the genetic mechanisms of phenotypic change (Seeb, L.W., 1986), particularly the regulation of longevity.

The conservation of genetic mechanisms between vertebrates, particularly of aging, suggests that mechanisms underlying exceptional longevity may apply across species. This concept empowers potent meta-analyses, where broad evolutionary comparisons of exceptional longevity, within and between species, can refine our knowledge of fundamental mechanisms. Here, we leverage the rich diversity of rockfish longevities along with the massive sampling and characterization available for human population genetics of longevity (Deelen et al., 2019). The intersection of these approaches reveals actionable targets to modulate vertebrate lifespan.

Results

Targeted Capture, Sequencing, Assembly, and Rockfish Phylogeny

Modern methods for convergence analyses rely on high quality genomes and annotations at high taxonomic densities. This requirement often precludes analysis of novel, poorly studied groups (Gayral et al., 2013). Such investigations would necessarily lack genomic resources of sufficient quality and taxonomic breadth to resolve trait-associated variation from other shared or species-specific variation. This is particularly problematic, yet prevalent, in the investigation of exceptional longevity; most notable occurrences are singular events, either in one isolated lineage, or shared by the entire group (Kim et al., 2011; Wirthlin et al., 2018). The consequence is a shortage of meaningful comparisons to subtract variation unrelated to the trait of interest. The *Sebastes* genus, with their multiple independent losses of exceptional longevity, can provide these comparisons.

To fully capitalize on the *Sebastes* natural experiment in modulating longevity, we expanded upon our recently developed phylogenomic mapping strategy (Daane et al., 2016), permitting broad sampling and sequencing of novel lineages with minimal established genomic resources. Twenty-three rockfish species, representing the shortest- (average 22 years) and longest-lived extremes (average 108 years) were sequenced using a pan-perciforme oligoarray containing 285,872 coding elements, 118,406 conserved noncoding elements, 2,508 ultraconserved noncoding elements, and 298 miRNAs. These targets were derived from broad synthesis of available genomic data from close relatives (**Supplemental F1**). As an outgroup, the northern sea robin *Prionotus carolinus* was included to root the analyses. Using the well-annotated Stickleback genome as the core reference, we recovered and reconstructed an average of 93% of conserved elements (coding and non-coding) at an average depth of 24x, spanning an average of 43,120,434 bases. It should be noted that less than 100% coverage is expected, as many elements will be unique to the reference genomes and/or lost in rockfish evolution. As pooled populations

were used for each lineage, both fixed and variable alleles were identified for each lineage (**Figure 2A, Supplemental T2**).

To generate a species tree for comparative analyses, a set of universal, high-coverage exons were concatenated, spanning 6,530,274 bases, and used to generate a high-support phylogeny with IQTree using model finder (Hoang et al., 2018; Kalyaanamoorthy et al., 2017) All nodes have greater than 99% bootstrap support, providing a valuable foundation for downstream analyses and resolving rockfish evolutionary relationships (**Figure 2B**).

Longer-lived Rockfish lineages have fewer changes from the ancestral state

Given the phylogeny of the rockfish clade and the prevalence of such rare longevities within the genus and greater *Sebastidae* family, the evolution of longevity extending mechanisms likely arose early and are shared among long-lived rockfish, yet were independently lost in the many short-lived lineages. The pattern of branch lengths on the phylogeny hint that long-lived species may have a reduced mutational load as compared to the short-lived, or equivalently, the short-lived species have an enhanced mutational load. To quantify this pattern, ancestral sequences were reconstructed with IQTree, and the percent identity of each terminal branch was calculated as compared to each node ancestral to that branch. At every node, the average percent difference of the long-lived descendants was lower than the short-lived, represented as a consistent ratio below 1.0 (**Figure 2B**). At the root, where all species are included, there is a significant linear relationship between longevity and percent identity ($p = 1.78e\text{-}06$). This relationship is also significant when bundling the short- and long-lived species as a binary trait ($p = 0.0047$)(**Figure 2C**). These data support the notion that the evolution of extended longevity occurred early in the stem rockfish and was subsequently lost in independent lineages.

Specific genes and gene sets are under selective constraint in long-lived Rockfish

The percent identity results indicate that genome-wide evolutionary rate shifts have been altered alongside longevity. We next asked if these shifts were enriched at specific, functional loci, identifying them as longevity mechanisms. The captured and sequenced coding regions were concatenated into genes and gene trees were generated with IQTree. These gene trees were analyzed with TRACCER to quantify convergent rate shifts associated with longevity (Treaster et al., 2021a). As compared to a control set of species, the result was enriched for significant genes, indicated as the bars above the baseline expected at low p-values. The control has a slight enrichment as well, likely due to the difficulties in selecting an unbiased control group when all the lineages were explicitly chosen for sequencing based on their extreme trait status (**Figure 3A**). Critically, this effect is muted when the results are expanded upon with GO Enrichment. We first used direct annotations only, to help minimize redundant terms and increase specificity. We again find a dramatic enrichment at low p-values for longevity, while the control group matches the distribution expected by chance (**Figure 3B**). These results are recapitulated when using indirect annotations as well (**Supplemental T5**). This indicates that not only are specific genes being maintained with longevity, but these patterns underly specific functions.

The top genes include *pla2g10, tcb1, cpsf7, nhlrc2, morn5, gpc6, rbbp6, ccdc25, otud4, hyou1, cdadc1* and *scaf8*. While individually they are only marginally significant, the collection

is enriched and in total have a false discovery rate (FDR) of only 0.27, representing the major FDR inflection point (**Figure 3C, Supplemental T3**). Interestingly, these genes are not driving the signal in the significantly enriched gene sets, which include "positive regulation of glycogen biosynthesis" and "flavonoid metabolism", with low FDRs of 0.019 and 0.009 respectively (**Figure 3D, Supplemental T4**). Both terms are undergoing evolutionary constraint in long-lived lineages, suggesting that the ancestral rockfish was also long-lived.

As a proof of concept of the approach and dataset, the glycogen biosynthesis term has substantial overlap with insulin signaling pathways, and indeed the most significant gene in the set is insulin receptor (*insr*) ($p = 0.003$). Insulin signaling has been implicated in the aging process and longevity in a wide variety of models (Kenyon, 2010; Singh et al., 2019). *Insr* was one of the life-history genes identified as undergoing relaxed selection in the evolution of shortened longevity in killifish radiations (Cui et al., 2019); *ie*. the same gene is undergoing the same evolutionary pressures in independent examples of shortened longevity. Beyond insulin signaling, several genes previously associated with aging were also identified in rockfish, including *sirt2* ($p = 0.001$) (Wang et al., 2019), and *app* ($p = 0.003$), the titular source of amyloid fibrils in Alzheimer's disease, but also involved in diverse signaling pathways (Müller et al., 2017). Notably, the 6th ranked gene, *gpc6* ($p = 1.12E-04$), has also been associated with increased lifespan in human populations (Sebastiani et al., 2012, 2013). The highest ranked gene for convergent evolution with rockfish longevity is *pla2g10*, encoding a secreted phospholipase (sPLA) which has not been previously associated with aging or longevity **(Figure 3E**). Intriguingly, sPLA2s bind to PLA2R1, which has been described as the "master regulator of cellular senescence", acting through reactive oxygen species, DNA damage, and p53 (Augert et al., 2009; Bernard and Vindrieux, 2014; Griveau et al., 2018; Vindrieux et al., 2013).

Rockfish longevity results provides a lens to refine human longevity GWAS to significance. Previous efforts to associate loci with the regulation of longevity in humans have been hindered by the complexity of the trait and the statistical burden of multiple hypothesis testing across the genome. Given the conservation of core mechanisms among vertebrates, we used the resulting genes and gene sets from the rockfish analysis as enriched windows to refine the largest human GWAS of longevity to date (Deelen et al., 2019), dramatically reducing number of hypotheses tested.

Because mapping the functionality of orthologs may not perfectly match across vast evolutionary time, we also used the rockfish enriched GO Term sets directly as a Rosetta stone to translate functionality between species. The directly annotated terms include flavonoid metabolism and glycogen biosynthesis (**Figure 4A**). In addition, we made a custom gene set defined using the top TRACCER genes with a combined FDR of 0.27, along with only the most significant genes ($p < 0.1$) driving the signal of the enriched indirect GO Terms (FDR < 0.1) (**Supplemental Figure 2**, **Supplemental T5**). Network analysis on this Rockfish Longevity Network with GeneMANIA (Mostafavi et al., 2008) demonstrate substantial interactions across functional groups (**Figure 4B**). For instance, APP, one of the genes driving the significance of the carbohydrate metabolism gene set, physically interacts with members of each of the

functional groups. These interactions suggest the convergence rate analysis has resolved an extended network.

These gene sets were intersected with results from the worldwide GWAS for human longevity by Deelen and colleagues (Deelen et al., 2019). While the glycogen term was not significant, SNPs within the flavonoid metabolism gene set were significantly associated with survival to the 99th percentile (FDR-adjusted p = 0.01) in association with human longevity at the 99th percentile. Further, we identified the custom Rockfish Longevity Network with survival to the 90th percentile (FDR-adjusted p = 0.051), while APP (p = 0.004), PTGER (p = 0.018), LHCGR (p = 0.022), SULT1B1 (p = 0.025) and MORN5 (p = 0.044) within that set each passed significance threshold at the 90th or 99th percentile. (**Figure 4C, Supplemental T6, T7**)

Discussion

There is conserved genetic architecture shared by all vertebrates (Lindblad-Toh et al., 2011), including that which outlines longevity (Kowalczyk et al., 2020; Mayne et al., 2019). Nutrient signaling and growth pathways have been implicated as lifespan factors across Metazoa, but the specific changes that set the arc of dwarf goby's life to only sixty days (Depczynski and Bellwood, 2005), a mouse to four years, and humans to roughly eighty, have remained undefined. Evolution modulates and selects for these life-history traits through shaping genomic controls, but disentangling the effective variation from other, species-specific changes is improbable in isolated species genomes. However, with independent evolutionary occurrences, and with elastic measures capable of bridging diverse genetic contexts, shared signatures can be identified across species (Treaster et al., 2021b). Here, we leverage the exceptional longevity and diversity of rockfishes along with human population diversity, and use functional gene sets to translate between genetic contexts.

Fish longevities past 100 years are exceedingly rare, yet it is common in the rockfish clade and in the greater *Sebastidae* family. Given this, at least eight million years ago rockfish likely evolved exceptional longevity and then independent lineages have since lost this ability. The genetic systems that facilitate this remarkable ability can be expected to lose selective pressure in conjunction with the loss of the trait. Our phylogeny and reconstructed ancestral sequences revealed genome-wide evolutionary rate shifts alongside changes in longevity; long-lived lineages have fewer changes overall and are more similar to reconstructed ancestral sequences than the short-lived lineages. As expected, this trend in percent identity suggests a relaxation in selection and/or an increase in mutational burden in the short-lived lineages. This pattern has been demonstrated in other radiations with variable longevities, such as killifish (Cui et al., 2019), and is likely due to reduced investments in germline maintenance (Maklakov and Immler, 2016) increasing the genetic drift between generations.

The ancestral state of exceptional longevity in rockfishes is reinforced by our convergence analysis; specific loci and functional gene sets are convergently undergoing restricted evolution in long-lived lineages to maintain that ancestral state. The TRACCER analysis method has demonstrated success with this strategy in the evolution of mammalian

longevity and marine transitions, with a key feature of the method remaining agnostic to ancestral state (Treaster et al., 2021a). Recent analysis of rockfish genomes (Kolora et al., 2021) and transcriptomes (Heras and Aguilar, 2019) have assumed exceptional longevity is derived and focus on branch-site positive selection that may be mistargeted, or on relative rate analyses that hinge upon problematic ancestral state inferences (Kowalczyk et al., 2018; Treaster et al., 2021a).

Here, our convergence analyses resolved two significant gene sets that are specifically maintained in long-lived rockfish lineages: glycogen biosynthesis and flavonoid metabolism. Glycogen biosynthesis incorporates the iconic insulin-like signaling (ILS) pathways and can be modulated with caloric restriction. These are conserved routes for lifespan modulation across models. Discovered in forward genetic screens in nematode worms (Dorman et al., 1995; Kenyon et al., 1993), the role of glucose and energy metabolism in aging continues to be refined (Gusarov et al., 2017) and retains relevance across vertebrates, mammals (Holzenberger et al., 2003), and humans (Deelen et al., 2013; Flachsbart et al., 2017; Grossi et al., 2018; Soerensen et al., 2012; Suh et al., 2008; Tazearslan et al., 2011; Willcox et al., 2008). Recent analysis of rare coding variants in centenarians converge on ILS pathways (Lin et al., 2021). The most significant gene in the glycogen biosynthesis gene set, and driving its enrichment in long-lived rockfishes, is *insulin receptor* itself. This is an auspicious proof-of-concept of our rockfish comparative analysis and lends credence to the novel "flavonoid metabolism" gene set that has hitherto been unappreciated in the regulations and evolution of longevity.

Flavonoids are a diverse and abundant chemical group that share two phenyl rings and a heterocyclic ring. They are plant metabolites, although they have a plethora of cellular consequences in animals, including anti-oxidative, anti-inflammatory, anti-mutagenic and anti-carcinogenic properties, and can modulate key cellular signaling and enzymatic pathways (Panche et al., 2016). With such a diversity of potencies, it is unsurprising that flavonoids can influence aging and longevity, and many have been tested to that effect (Pallauf et al., 2017). However, the "flavonoid" nomenclature undermines the broader role of this gene set in detoxification pathways, processing xenobiotic chemicals, and endogenous signal processing. The aryl hydrocarbons in flavonoids are detected and induce machinery to neutralize and remove these active metabolites, with a variety of downstream signaling consequences with both pro- and anti-aging effects (Brinkmann et al., 2020; McElwee et al., 2004). The Aryl Hydrocarbon Receptor, AhR, is the highly conserved master transcription factor for these responses, triggering xenobiotic responsive genes like Cyp1a and sulfur transferases (Ashida et al., 2008; Yueh et al., 2003) - key members of the flavonoid metabolism gene set. Although the precise ligand interactions have proven complex and elusive (Xue et al., 2017), various flavonoids have demonstrated influence on these AhR-pathways, which is how the flavonoid gene set derives its name. The downstream effects of these responsive elements include not only detoxification, but also regulation of hormones and neurotransmitters (Huang et al., 2009). Such xenobiotic metabolism has been linked to longevity (Steinbaugh et al., 2012). The pleiotropic integration of aryl-hydrocarbon metabolism and hormonal signaling will have dramatic consequences for

growth, energy metabolism, developmental timing, and ultimately aging (Li et al., 2017; Mueller et al., 2015; Visser, 1994).

PLA2G10, the most significant gene in our convergent rate analysis, has the highest ability among secreted phospholipases to hydrolyze phosphatidylcholine, which promotes inflammation by driving arachidonic acid (AA) production (Samuelsson, 1991) and elevated eicosanoid levels (Wang et al., 2021). Linking back to aryl-hydrocarbon metabolism, AA can also stimulate AhR (Seidel et al., 2001). AA metabolites such as lipoic acid (LX; LXA4) are tied to inflammation and can ameliorate the aging process (Kniazeva and Han, 2013; O'Rourke et al., 2013). Both regulation of polyunsaturated fatty acid synthesis and lipoic acid are down regulated in human elderly, and the resulting imbalance in AA metabolism removes the brakes from inflammatory reactions (Gangemi et al., 2005). These fatty acid pathways, and their interaction with aryl-hydrocarbon metabolism, define a potentially common axis by which longevity may be modulated. The potential of sPLAs as therapeutic targets in cancer, metabolism, and inflammation have already been noted (Yarla et al., 2016; Zhang et al., 2020). Intriguingly, sPLA2s are entangled with flavonoid metabolism as well. sPLA2 decreases in aged skin, but can be rescued with hesperidin, a flavonone (Man et al., 2015). Red ginseng, rich in the flavonoid catechin (Kim, 2016), extends lifespan in flies and upregulates sPLA2 (Hou et al., 2020).

A utility of such evolutionary refined gene sets is to inform analyses in other models under the principle of conservation. To validate the role of these targets in longevity across vertebrates, we used them as a lens to filter and refine the largest GWAS for longevity in humans (Deelen et al., 2019). Human GWAS is empowered by massive sampling and accurate characterization, but its ability to trace complex, multifactorial traits has been disappointing (Baghdadi, Karasik et al. 2019). Signals for longevity routinely fall short of significance due to multiple hypothesis testing across the genome and the complexity of the trait. However, by focusing on the genes and gene sets implicated in exceptional rockfish longevity, multiple hypothesis testing is reduced. Notably, the flavonoid metabolism gene set is again revealed as significant, with an FDR adjusted p-value of 0.01 for association with human longevity to the 99^{th} percentile. We also identify the custom Rockfish Longevity Network as associating with human longevity to the 90^{th} percentile (FDR adjusted $p = 0.051$). Within these sets, we also identify APP, IGF1, LHCGR, C1QTNF2, SULT1B1, and MORN5, though the strongest signal is at the set level, suggesting the longevity architecture is modulated as a group. These findings independently validate the rockfish sets through conservation and provide new, entrenched mechanisms associating with increased lifespan in diverse vertebrate lineages.

We have capitalized on variation in clades exhibiting exceptional longevity through convergence analyses and translated those results with functional gene sets to identify variation associating with longevity in human populations. Age is the greatest risk factor for the diseases that plague modern society – cancer, Alzheimer's, heart disease, etc. – and medical research has traditionally focused on ameliorating these pathologies. As our understanding of longevity increases, it becomes attractive to instead target the aging process directly. We now have critical information from unconventional species that resist these deleterious effects for centuries, and

functional loci validated with human population variants. These exceptional fish lineages have already arrived at solutions to the aging problem, while others even have age-related improvements in health (Sauer et al., 2021). The specific loci identified here, and their unique variations in both humans and rockfish, are auspicious targets for future interventions to delay, ameliorate, or even prevent aging and age-related diseases in humans.

Acknowledgements

The authors would like to thank Dr. Peter Konstantanidis, Dr. Brittany Schwartzkopf, and staff of Oregon State University for assistance with rockfish sampling, along with Luke Torneborne and the Burke Museum in sharing archived tissue samples of a large number of rockfish taxa to enable this work. Additionally, the authors would like to thank the CHARGE Aging and Longevity Working Group. This work was supported by Guggenheim fellowship to MPH, BSF-NSF grant to MPH and DK; AFAR Postdoctoral Fellowship, and the National Academy of Medicine Healthy Longevity Catalyst Grant to ST.

Methods

Tissue Sampling

Rockfish tissue samples were obtained from collections obtained by Oregon State University and Burke Museum (curt. Luke Torneborne). Sea Robin tissues graciously provided by Department of Marine Resources, Marine Biology Laboratories, Woods Hole. Sample identity and source can be found in Supplemental Table S1. Identities were confirmed by PCR amplification and Sanger sequencing of Cytochrome B.

Capture Design, Sequencing, and Conservome Assembly

Targeted Sequence Capture Design

The sequence capture design targets protein-coding exons as well as a set of conserved non-protein coding elements (CNEs), miRNA hairpins, and ultraconservative non-coding elements (UCNEs). As many of the available perciform genomes were fragmented and poorly annotated at the time of sequence capture design, we generated a list of conserved protein coding and CNE regions from the genomes of the three-spined stickleback *Gasterosteus aculeatus* (BROAD S1), the Japanese medaka (*Oryzias latipes*, MEDAKA1) and green spotted puffer *Tetraodon nigroviridis* (TETRAODON 8.0). Protein coding exons were extracted from Ensembl BioMart (Kinsella et al., 2011). CNEs were defined from the constrained regions in the Ensembl compara 11-way teleost alignment (Ensembl release-91)(Herrero et al., 2016). We removed CNEs < 50bp in length to facilitate space in the capture design. miRNA hairpins were extracted from miRbase and ultraconservative elements (UCNEs) from UCNEbase (Dimitrieva and Bucher, 2013; Kozomara and Griffiths-Jones, 2011). These elements were identified within each reference genome using BLASTN (ncbi-blast-2.2.30+; parameters '-max_target_seqs 1 -outfmt 6') or through direct annotations where available. miRNA hairpins were padded to be at least 100 bp. To verify these sites as non-coding, we eliminated CNEs, miRNAs and UCNEs that overlapped coding exons using Bedtools (v2.26.0)

intersectBed (Quinlan and Hall, 2010). In the event CNEs were also defined as miRNAs or UCNEs, we prioritized miRNA and UCNE annotations.

We aimed to design a series of oligonucleotide capture baits that could efficiently enrich DNA sequencing libraries across Perciformes, a large order of more than 2,200 species (Nelson et al., 2016). We used BLASTN (ncbi-blast-2.2.30+; parameters '-max_target_seqs 1 -outfmt 6') to identify each targeted element within multiple perciform genome assemblies. The majority of capture baits were designed against the genome of the Chabot de Rhénanie *Cottus rhenanus* (ASM145555v1). To account for the possibility that specific genetic regions may be absent or highly divergent in this sculpin genome but conserved in the broader Perciformes order, we iteratively designed capture baits from the genomes of the shorthorn sculpin *Myoxocephalus scorpius* (ASM90031295v1)(Malmstrøm et al., 2016), the sablefish *Anoplopoma fimbria* (AnoFim1.0)(Rondeau et al., 2013), the golden redfish *Sebastes norvegicus* (ASM90030265v1)(Malmstrøm et al., 2016), the flag rockfish *Sebastes rubrivinctus* (SRub1.0), the rougheye rockfish *Sebastes aleutianus* (ASM191080v2), the European perch *Perca fluviatilis* (ASM90030264v1)(Malmstrøm et al., 2016), and the three-spined stickleback *Gasterosteus aculeatus* (BROAD S1). As an example, we included capture baits from the *M. scorpius* genome if the targeted elements were either not identified (coverage <70% or a BLASTN E-value >0.001), and/or had < 85% identity to the orthologous element within the *C. rhenanus* genome. This process is then repeated such that elements from *A. fimbria* are compared to both the *M. scorpius* and *C. rhenanus* targets. As a result of this process, there will be oligonucleotide capture baits of at least 85% identity to each targeted sequence for every perciform genome used in the capture design. This multi-species 'phylochip' design enables usage of these baits to sequence large numbers of distantly-related perciform fish species.

SeqCap EZ Developer (cat #06471684001) capture oligos were designed in collaboration with the Nimblegen design team to standardize oligo annealing temperature, reduce probe redundancy, and to remove low complexity DNA regions. The capture design contained sequence from 492,506 regions (81,493,221 total bp) across all 8 perciform reference genomes. Accounting for probe redundancy between the perciform reference genomes, the final capture design comprised 407,084 distinct elements, including 285,872 protein coding exons, 118,406 conserved non-coding elements, 298 miRNAs and 2,508 UCNEs (**Figure S1**).

Sequencing library preparation

DNA was extracted from multiple individuals of each species using the Qiagen DNeasy Blood & Tissue Kit (cat# 69504). For each species, equal quantities of DNA were pooled from each individual prior to library preparation and sequencing. The DNA pools were diluted in a shearing buffer (10 mM Tris, 0.1 mM EDTA, pH 8.3) and were mechanically sheared to an average size of 200bp in a Covaris E220 ultrasonicator (duty cycle, 10%; intensity, 5; cycles/burst, 200; time, 400 seconds; temperature, 8°C). Barcoded sequencing libraries were generated using a KAPA HyperPrep Kit (Roche, No. 07137923001) following the Nimblegen SeqCap EZ Library protocol (Version 4.3) and using dual-SPRI (solid phase reversible immobilization) size selection

to generate libraries of 200-450 bp. The libraries were hybridized to the capture baits according to the SeqCap EZ Library protocol. To allow for more mistmatches between libraries and baits, the libraries were hybridized to the capture baits and washed at a reduced stringency of 45°C instead of the manufacturer recommended temperature of 47°C. The SeqCap EZ Developer Reagent was used (cat #06684335001) in place of Human CotI DNA during hybridization. Multiple barcoded libraries were then pooled for 100bp single end sequencing with an Illumina HiSeq 2500.

Reference contig assembly

As described in Daane *et al*.(Daane et al., 2019), we used the Phylomapping *de novo* assembly pipeline to generate reference contigs for each species. Briefly, we removed duplicate reads and masked low-quality bases with the FASTX toolkit (http://hannonlab.cshl.edu/fastx_toolkit). Library adaptor sequences were removed using Trimmomatic v.0.36 (Bolger et al., 2014). Sequencing reads were binned by orthology to target regions in the *G. aculeatus*, *T. nigroviridis* and *O. latipes* reference genomes using BLASTN and dc-megaBLAST. The binned sequencing reads were assembled into contigs *de novo* using CAP3 and UCLUST (Edgar, 2010; Huang, 1999). To recruit previously unmapped reads to the analysis, all sequencing reads were aligned back to the assembled contigs using NextGenMap (Sedlazeck et al., 2013). All reads were re-assembled using CAP3 in a second round of *de novo* assembly. If multiple contigs are assembled for any given element (for example, exon, CNE), the multiple contigs were then merged if they overlapped and had >95% identity. This pipeline results in the creation of consensus contig sequence(s) for each target region. Contigs were automatically annotated according to the orthologous element within the reference genome that was identified using BLAST and subsequently used to scaffold and refine contig assembly. See Daane *et al*. for details(Daane et al., 2019).

Identification of orthologs

Contigs from each species are considered orthologs if they are assembled based on the same targeted element. However, if multiple contigs are assembled for a given targeted element, we assigned orthology using a gene tree-species tree reconciliation approach. Contigs for all species were aligned using MAFFT v7.313 (parameters '-op 10 -ep 10')(Katoh and Standley, 2013) and a maximum likelihood tree generated using IQTree v1.7beta2 (parameters '-alrt 1000')(Nguyen et al., 2015). We then used Notung-2.9 (parameters '–reconcile–rearrange–silent–threshold 90%–treeoutput nhx')(Chen et al., 2000) to perform gene tree reconciliation and to infer patterns of gene loss and duplication. The total number of duplication and loss events across the phylogeny as estimated by Notung was then compared to a scenario where all copy number variation is species-specific. The most parsimonious scenario with the fewest total gain/loss events was selected. Using this approach, orthologs can be accurately paired as long as there ≥4–6% variation between paralogous sequences, which coincides with thresholds of variation necessary to distinguish copy number variation during contig assembly (Daane et al., 2019).

Read coverage and depth of targeted regions

The target regions (exons, CNEs, etc.) were originally defined from the *G. aculeatus*, *T. nigroviridis*, and *O. latipes* reference genomes. To estimate sequencing coverage of these targeted regions in our assembly, we lifted our assembled contig read alignment data to the orthologous region within each reference genome. First, sequencing reads were aligned to the assembled contigs using NextGenMap v0.5.5 (Sedlazeck et al., 2013). The assembled contigs were then pairwise aligned to the target regions within the reference genome using Biopython v.1.70 (parameters 'pairwise2; match = 5, mismatch = −4, gap_open = −15, gap_extend = −1'). Finally, data from this pairwise alignment, including indels, were used to map the aligned sequencing reads from the assembled contigs to the reference genome. Alignments were converted from SAM to BAM format and indexed using SAMtools (v1.9)(Li et al., 2009) and were visually inspected for accuracy in the Integrative Genome Viewer(Robinson et al., 2011). Coverage is defined as the proportion of targeted bases overlapping at least one sequencing read. Coverage and depth was calculated with BEDTools v2.23.0 coverageBed (parameters '-d')(Quinlan and Hall, 2010).

Multiple sequence alignment

Orthologous sequences were aligned using MAFFT v 7.313 (parameters '–maxiterate 1000 –localpair –op 10 –ep')(Katoh and Standley, 2013). If the MAFFT alignment predicted an indel within coding sequences, we re-aligned these exons as codon alignments using the frameshift-aware multiple sequence alignment software MACSE v2.03 (parameters 'prog alignSequences -seq -seq_lr -fs_lr 10 -stop_lr 15')(Ranwez et al., 2011).

Reconstruction of gene sequences

Individual exons were reconstructed into full gene sequences. Single copy exons were concatenated in gene order as found in the *G. aculeatus* (three-spined stickleback;BROAD S1), *T. nigroviridis* (green spotted pufferfish; TETRAODON 8.0), or *O. latipes* (Japenese medaka, MEDAKA1) reference genomes. For each rockfish we reconstructed a total of 27,645 gene sequences.

Data availability

Raw short sequencing read data will be deposited on the NCBI Sequence Read Archive (SRA). Assembled contigs and gene tree data will be deposited in the Zenodo data repository.

Tree Construction

Coding sequences with greater than 90% coverage in every sampled species, and longer than 2000 bases, were aligned and concatenated. The resulting 6,530,274 bases were analyzed with IQTree's substitution model finder, identifying GY+F+R3, and generating the rockfish phylogeny with greater than 99% support at every node. Individual gene trees for rate analysis used the same model and were fixed to the species tree topology.

Relative Rate Analysis

Convergent rate analysis was performed with TRACCER (Treaster et al., 2021a) on the 23 rockfish species and sea robin outgroup. *Sebastes borealis, aleutianus, crameri, alutus, ruberrimus, melanostomus, diploproa, pinniger, maliger, melanops, rufus, babcocki, helvomaculatus* were flagged as long-lived, while *Sebastes emphaeus ,glaucus, seminictus, rastrelliger, auriculatus,jordani, hopkinsi, rubrivinctus, umbrosus, chlorostictus* and *Prionotus carolinus* were flagged as short-lived. Gene trees were discarded if they had fewer than 12 lineages, 6 long-lived lineages, or 5 short-lived lineages, as the rates of the remaining representatives were unlikely to be informative for the trait. The ranking transformation was used for both scoring and phylogenetic distance scaling. FDRs are calculated as the expected number of hits at that significance divided by the number of actual hits at that significance or better (Benjamini and Hochberg, 1995).

Gene Set Enrichment

GO Terms were harvested from Ensembl by combining the annotations across orthologs of well-characterized genomes, including human, mice, zebrafish, stickleback, and medaka. Terms with less than three members, or more than one hundred, were discarded as being uninformative or misleading in the context of set enrichment. Gene set enrichment was calculated with the SUMSTAT approach, which has demonstrated power, flexibility, and simplicity over competing gene set enrichment analyses (Ackermann and Strimmer, 2009; Chen et al., 2010; Tintle et al., 2009). SUMSTAT was performed with log transformed p-values from each analysis, with an additional square root transformation to undermine outliers. In short, the square-rooted, log transformed p-values for each gene in a set were summed and compared to a distribution of randomly sampled scores of the same set size to determine an enrichment p-value. FDRs are calculated as above. These methods were applied with both direct and indirect annotations. Direct annotations were used when trying to minimize redundancy between terms and ensure specificity. Indirect annotations were used when feeding into network analyses, ensuring the broader scopes are included, while using only the most significant ($p < 0.1$) genes in each significant term.

Human Genetics of Longevity

Orthologs from fish to humans were taken directly from Ensembl when available (ensembl.org). Those that did not have an annotated ortholog were mapped using pBLAST against the human proteome and confirmed if they found a hit with an E-value less than 1e-30 and percent identity greater than 40. The remaining fish genes were dropped. The orthologous human genes were used with data from GWAS for human longevity (Deelen et al., 2019). MAGMA (Leeuw et al., 2015) was used to assign P-values to genes based on the lowest p-value for a genetic variant within or around a gene coordinates, including 50 kb up- and downstream. The association data of the genetic variants used to assign the p-values was extracted from the 90^{th} and 99^{th} survival percentile analyses summary results from Deelen and colleagues (Deelen et al., 2019). Gene-set

enrichment analysis was performed using the GO-terms coming from the Rockfish with the human orthologous genes using the standard settings in MAGMA.

References

Ackermann, M., and Strimmer, K. (2009). A general modular framework for gene set enrichment analysis. BMC Bioinformatics *10*, 47.

Ashida, H., Nishiumi, S., and Fukuda, I. (2008). An update on the dietary ligands of the AhR. Expert Opin. Drug Metab. Toxicol. *4*, 1429–1447.

Augert, A., Payré, C., de Launoit, Y., Gil, J., Lambeau, G., and Bernard, D. (2009). The M-type receptor PLA2R regulates senescence through the p53 pathway. EMBO Rep. *10*, 271–277.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J. R. Stat. Soc. Ser. B Methodol. *57*, 289–300.

Bernard, D., and Vindrieux, D. (2014). PLA2R1: Expression and function in cancer. Biochim. Biophys. Acta BBA - Rev. Cancer *1846*, 40–44.

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

Brinkmann, V., Ale-Agha, N., Haendeler, J., and Ventura, N. (2020). The Aryl Hydrocarbon Receptor (AhR) in the Aging Process: Another Puzzling Role for This Highly Conserved Transcription Factor. Front. Physiol. *10*.

Chen, K., Durand, D., and Farach-Colton, M. (2000). NOTUNG: A program for dating gene duplications and optimizing gene family trees. J. Comput. Biol. *7*, 429–447.

Chen, X., Wang, L., Hu, B., Guo, M., Barnard, J., and Zhu, X. (2010). PATHWAY-BASED ANALYSIS FOR GENOME-WIDE ASSOCIATION STUDIES USING SUPERVISED PRINCIPAL COMPONENTS. Genet. Epidemiol. *34*, 716–724.

Cui, R., Medeiros, T., Willemsen, D., Iasi, L.N.M., Collier, G.E., Graef, M., Reichard, M., and Valenzano, D.R. (2019). Relaxed Selection Limits Lifespan by Increasing Mutation Load. Cell *178*, 385-399.e20.

Daane, J.M., Dornburg, A., Smits, P., MacGuigan, D.J., Brent Hawkins, M., Near, T.J., William Detrich, H., and Harris, M.P. (2019). Historical contingency shapes adaptive radiation in Antarctic fishes. Nat. Ecol. Evol. *3*, 1102–1109.

Deelen, J., Uh, H.-W., Monajemi, R., van Heemst, D., Thijssen, P.E., Böhringer, S., van den Akker, E.B., de Craen, A.J.M., Rivadeneira, F., Uitterlinden, A.G., et al. (2013). Gene set analysis of GWAS data for human longevity highlights the relevance of the insulin/IGF-1 signaling and telomere maintenance pathways. Age *35*, 235–249.

Deelen, J., Evans, D.S., Arking, D.E., Tesi, N., Nygaard, M., Liu, X., Wojczynski, M.K., Biggs, M.L., van der Spek, A., Atzmon, G., et al. (2019). A meta-analysis of genome-wide association studies identifies multiple longevity genes. Nat. Commun. *10*, 3669.

Depczynski, M., and Bellwood, D.R. (2005). Shortest recorded vertebrate lifespan found in a coral reef fish. Curr. Biol. CB *15*, R288-289.

Dimitrieva, S., and Bucher, P. (2013). UCNEbase - A database of ultraconserved non-coding elements and genomic regulatory blocks. Nucleic Acids Res. *41*, 101–109.

Dorman, J.B., Albinder, B., Shroyer, T., and Kenyon, C. (1995). The age-1 and daf-2 genes function in a common pathway to control the lifespan of Caenorhabditis elegans. Genetics *141*, 1399–1406.

Edgar, R.C. (2010). Search and clustering orders of magnitude faster than BLAST. Bioinformatics *26*, 2460–2461.

Flachsbart, F., Dose, J., Gentschew, L., Geismann, C., Caliebe, A., Knecht, C., Nygaard, M., Badarinarayan, N., ElSharawy, A., May, S., et al. (2017). Identification and characterization of two functional variants in the human longevity gene FOXO3. Nat. Commun. *8*, 2063.

Gangemi, S., Pescara, L., D'Urbano, E., Basile, G., Nicita-Mauro, V., Davì, G., and Romano, M. (2005). Aging is characterized by a profound reduction in anti-inflammatory lipoxin A4 levels. Exp. Gerontol. *40*, 612–614.

Griveau, A., Wiel, C., Calvé, B.L., Ziegler, D.V., Djebali, S., Warnier, M., Martin, N., Marvel, J., Vindrieux, D., Bergo, M.O., et al. (2018). Targeting the phospholipase A2 receptor ameliorates premature aging phenotypes. Aging Cell *17*, e12835.

Grossi, V., Forte, G., Sanese, P., Peserico, A., Tezil, T., Lepore Signorile, M., Fasano, C., Lovaglio, R., Bagnulo, R., Loconte, D.C., et al. (2018). The longevity SNP rs2802292 uncovered: HSF1 activates stress-dependent expression of FOXO3 through an intronic enhancer. Nucleic Acids Res. *46*, 5587–5600.

Gusarov, I., Pani, B., Gautier, L., Smolentseva, O., Eremina, S., Shamovsky, I., Katkova-Zhukotskaya, O., Mironov, A., and Nudler, E. (2017). Glycogen controls Caenorhabditis elegans lifespan and resistance to oxidative stress. Nat. Commun. *8*.

Heras, J., and Aguilar, A. (2019). Comparative Transcriptomics Reveals Patterns of Adaptive Evolution Associated with Depth and Age Within Marine Rockfishes (Sebastes). J. Hered. *110*, 340–350.

Herrero, J., Muffato, M., Beal, K., Fitzgerald, S., Gordon, L., Pignatelli, M., Vilella, A.J., Searle, S.M.J., Amode, R., Brent, S., et al. (2016). Ensembl comparative genomics resources. Database *2016*, bav096.

Hoang, D.T., Chernomor, O., von Haeseler, A., Minh, B.Q., and Vinh, L.S. (2018). UFBoot2: Improving the Ultrafast Bootstrap Approximation. Mol. Biol. Evol. *35*, 518–522.

Holzenberger, M., Dupont, J., Ducos, B., Leneuve, P., Géloën, A., Even, P.C., Cervera, P., and Le Bouc, Y. (2003). IGF-1 receptor regulates lifespan and resistance to oxidative stress in mice. Nature *421*, 182–187.

Hou, W., Pei, J., Wang, Y., Zhang, J., Zheng, H., and Cui, R. (2020). Anti-ageing effects of red ginseng on female Drosophila melanogaster. J. Cell. Mol. Med. *24*, 3751–3755.

Hua, X., Cowman, P., Warren, D., and Bromham, L. (2015). Longevity Is Linked to Mitochondrial Mutation Rates in Rockfish: A Test Using Poisson Regression. Mol. Biol. Evol. *32*, 2633–2645.

Huang, X. (1999). CAP3: A DNA Sequence Assembly Program. Genome Res. *9*, 868–877.

Huang, C., Chen, Y., Zhou, T., and Chen, G. (2009). Sulfation of dietary flavonoids by human sulfotransferases. Xenobiotica Fate Foreign Compd. Biol. Syst. *39*, 312–322.

Hyde, J.R., and Vetter, R.D. (2007). The origin, evolution, and diversification of rockfishes of the genus Sebastes (Cuvier). Mol. Phylogenet. Evol. *44*, 790–811.

Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A., and Jermiin, L.S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. Nat. Methods *14*, 587–589.

Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. *30*, 772–780.

Kenyon, C.J. (2010). The genetics of ageing. Nature *464*, 504–512.

Kenyon, C., Chang, J., Gensch, E., Rudner, A., and Tabtiang, R. (1993). A C. elegans mutant that lives twice as long as wild type. Nature *366*, 461–464.

Kim, J.-S. (2016). Investigation of Phenolic, Flavonoid, and Vitamin Contents in Different Parts of Korean Ginseng (Panax ginseng C.A. Meyer). Prev. Nutr. Food Sci. *21*, 263–270.

Kim, E.B., Fang, X., Fushan, A.A., Huang, Z., Lobanov, A.V., Han, L., Marino, S.M., Sun, X., Turanov, A.A., Yang, P., et al. (2011). Genome sequencing reveals insights into physiology and longevity of the naked mole rat. Nature *479*, 223–227.

Kinsella, R.J., Kähäri, A., Haider, S., Zamora, J., Proctor, G., Spudich, G., Almeida-King, J., Staines, D., Derwent, P., Kerhornou, A., et al. (2011). Ensembl BioMarts: a hub for data retrieval across taxonomic space. Database J. Biol. Databases Curation *2011*, bar030.

Kniazeva, M., and Han, M. (2013). Fat chance for longevity. Genes Dev. *27*, 351–354.

Kolora, S.R.R., Owens, G.L., Vazquez, J.M., Stubbs, A., Chatla, K., Jainese, C., Seeto, K., McCrea, M., Sandel, M.W., Vianna, J.A., et al. (2021). Origins and evolution of extreme life span in Pacific Ocean rockfishes. Science *374*, 842–847.

Kowalczyk, A., Meyer, W.K., Partha, R., Mao, W., Clark, N.L., and Chikina, M. (2018). RERconverge: an R package for associating evolutionary rates with convergent traits. BioRxiv 451138.

Kowalczyk, A., Partha, R., Clark, N.L., and Chikina, M. (2020). Pan-mammalian analysis of molecular constraints underlying extended lifespan. ELife *9*, e51089.

Kozomara, A., and Griffiths-Jones, S. (2011). miRBase: integrating microRNA annotation and deep-sequencing data. Nucleic Acids Res. *39*, D152–D157.

Leeuw, C.A. de, Mooij, J.M., Heskes, T., and Posthuma, D. (2015). MAGMA: Generalized Gene-Set Analysis of GWAS Data. PLOS Comput. Biol. *11*, e1004219.

Li, F., Zhu, W., and Gonzalez, F.J. (2017). Potential role of CYP1B1 in the development and treatment of metabolic diseases. Pharmacol. Ther. *178*, 18–30.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. Bioinforma. Oxf. Engl. *25*, 2078–2079.

Liang, H., Masoro, E.J., Nelson, J.F., Strong, R., McMahan, C.A., and Richardson, A. (2003). Genetic mouse models of extended lifespan. Exp. Gerontol. *38*, 1353–1364.

Lin, J.-R., Sin-Chan, P., Napolioni, V., Torres, G.G., Mitra, J., Zhang, Q., Jabalameli, M.R., Wang, Z., Nguyen, N., Gao, T., et al. (2021). Rare genetic coding variants associated with human longevity and protection against age-related diseases. Nat. Aging *1*, 783–794.

Lindblad-Toh, K., Garber, M., Zuk, O., Lin, M.F., Parker, B.J., Washietl, S., Kheradpour, P., Ernst, J., Jordan, G., Mauceli, E., et al. (2011). A high-resolution map of human evolutionary constraint using 29 mammals. Nature *478*, 476–482.

Maklakov, A.A., and Immler, S. (2016). The Expensive Germline and the Evolution of Ageing. Curr. Biol. CB *26*, R577–R586.

Malmstrøm, M., Matschiner, M., Tørresen, O.K., Star, B., Snipen, L.G., Hansen, T.F., Baalsrud, H.T., Nederbragt, A.J., Hanel, R., Salzburger, W., et al. (2016). Evolution of the immune system influences speciation rates in teleost fishes. Nat. Genet. *48*, 1204–1210.

Man, G., Mauro, T.M., Zhai, Y., Kim, P.L., Cheung, C., Hupe, M., Crumrine, D., Elias, P.M., and Man, M.-Q. (2015). Topical Hesperidin Enhances Epidermal Function in an Aged Murine Model. J. Invest. Dermatol. *135*, 1184–1187.

Mayne, B., Berry, O., Davies, C., Farley, J., and Jarman, S. (2019). A genomic predictor of lifespan in vertebrates. Sci. Rep. *9*, 17866.

McElwee, J.J., Schuster, E., Blanc, E., Thomas, J.H., and Gems, D. (2004). Shared transcriptional signature in Caenorhabditis elegans Dauer larvae and long-lived daf-2 mutants implicates detoxification system in longevity assurance. J. Biol. Chem. *279*, 44533–44543.

Mostafavi, S., Ray, D., Warde-Farley, D., Grouios, C., and Morris, Q. (2008). GeneMANIA: a real-time multiple association network integration algorithm for predicting gene function. Genome Biol. *9 Suppl 1*, S4.

Mueller, J.W., Gilligan, L.C., Idkowiak, J., Arlt, W., and Foster, P.A. (2015). The Regulation of Steroid Action by Sulfation and Desulfation. Endocr. Rev. *36*, 526–563.

Müller, U.C., Deller, T., and Korte, M. (2017). Not just amyloid: physiological functions of the amyloid precursor protein family. Nat. Rev. Neurosci. *18*, 281–298.

Nelson, J.S., Grande, T.C., and Wilson, M.V.H. (2016). Fishes of the World (Hoboken, NJ: John Wiley & Sons, Inc.).

Nguyen, L.T., Schmidt, H.A., Von Haeseler, A., and Minh, B.Q. (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. *32*, 268–274.

O'Rourke, E.J., Kuballa, P., Xavier, R., and Ruvkun, G. (2013). ω-6 Polyunsaturated fatty acids extend life span through the activation of autophagy. Genes Dev. *27*, 429–440.

Pallauf, K., Duckstein, N., and Rimbach, G. (2017). A literature review of flavonoids and lifespan in model organisms. Proc. Nutr. Soc. *76*, 145–162.

Panche, A.N., Diwan, A.D., and Chandra, S.R. (2016). Flavonoids: an overview. J. Nutr. Sci. *5*.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinforma. Oxf. Engl. *26*, 841–842.

Ranwez, V., Harispe, S., Delsuc, F., and Douzery, E.J.P. (2011). MACSE: Multiple Alignment of Coding SEquences Accounting for Frameshifts and Stop Codons. PLoS ONE *6*, e22594.

Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative Genome Viewer. Nat. Biotechnol. *29*, 24–26.

Rondeau, E.B., Messmer, A.M., Sanderson, D.S., Jantzen, S.G., von Schalburg, K.R., Minkley, D.R., Leong, J.S., Macdonald, G.M., Davidsen, A.E., Parker, W.A., et al. (2013). Genomics of sablefish (Anoplopoma fimbria): expressed genes, mitochondrial phylogeny, linkage map and identification of a putative sex gene. BMC Genomics *14*, 452.

Samuelsson, B. (1991). Arachidonic acid metabolism: role in inflammation. Z. Rheumatol. *50 Suppl 1*, 3–6.

Sauer, D.J., Heidinger, B.J., Kittilson, J.D., Lackmann, A.R., and Clark, M.E. (2021). No evidence of physiological declines with age in an extremely long-lived fish. Sci. Rep. *11*, 9065.

Sebastiani, P., Solovieff, N., DeWan, A.T., Walsh, K.M., Puca, A., Hartley, S.W., Melista, E., Andersen, S., Dworkis, D.A., Wilk, J.B., et al. (2012). Genetic Signatures of Exceptional Longevity in Humans. PLOS ONE *7*, e29848.

Sebastiani, P., Bae1, H., Sun, F.X., Andersen, S.L., Daw, E.W., Malovini, A., Kojima, T., Hirose, N., Schupf, N., Puca, A., et al. (2013). Meta-analysis of genetic variants associated with human exceptional longevity. Aging *5*, 653–661.

Sedlazeck, F.J., Rescheneder, P., and Von Haeseler, A. (2013). NextGenMap: Fast and accurate read mapping in highly polymorphic genomes. Bioinformatics *29*, 2790–2791.

Seeb, L.W. (1986). Biochemical systematics and evolution of the scorpaenid genus Sebastes. Unpubl. Ph.D. diss., Univ. of Washington, Seattle. Univ. Of Washington, Seattle.

Seidel, S.D., Winters, G.M., Rogers, W.J., Ziccardi, M.H., Li, V., Keser, B., and Denison, M.S. (2001). Activation of the Ah receptor signaling pathway by prostaglandins. J. Biochem. Mol. Toxicol. *15*, 187–196.

Singh, P.P., Demmitt, B.A., Nath, R.D., and Brunet, A. (2019). The Genetics of Aging: A Vertebrate Perspective. Cell *177*, 200–220.

Soerensen, M., Dato, S., Tan, Q., Thinggaard, M., Kleindorp, R., Beekman, M., Jacobsen, R., Suchiman, H.E.D., de Craen, A.J.M., Westendorp, R.G.J., et al. (2012). Human longevity and variation in GH/IGF-1/insulin signaling, DNA damage signaling and repair and pro/antioxidant pathway genes: cross sectional and longitudinal studies. Exp. Gerontol. *47*, 379–387.

Steinbaugh, M.J., Sun, L.Y., Bartke, A., and Miller, R.A. (2012). Activation of genes involved in xenobiotic metabolism is a shared signature of mouse models with extended lifespan. Am. J. Physiol.-Endocrinol. Metab. *303*, E488–E495.

Suh, Y., Atzmon, G., Cho, M.-O., Hwang, D., Liu, B., Leahy, D.J., Barzilai, N., and Cohen, P. (2008). Functionally significant insulin-like growth factor I receptor mutations in centenarians. Proc. Natl. Acad. Sci. U. S. A. *105*, 3438–3442.

Tazearslan, C., Huang, J., Barzilai, N., and Suh, Y. (2011). Impaired IGF1R signaling in cells expressing longevity-associated human IGF1R alleles. Aging Cell *10*, 551–554.

Tintle, N.L., Borchers, B., Brown, M., and Bekmetjev, A. (2009). Comparing gene set analysis methods on single-nucleotide polymorphism data from Genetic Analysis Workshop 16. BMC Proc. *3 Suppl 7*, S96.

Treaster, S., Daane, J.M., and Harris, M.P. (2021a). Refining Convergent Rate Analysis with Topology in Mammalian Longevity and Marine Transitions. Mol. Biol. Evol. *38*, 5190–5203.

Treaster, S., Karasik, D., and Harris, M.P. (2021b). Footprints in the Sand: Deep Taxonomic Comparisons in Vertebrate Genomics to Unveil the Genetic Programs of Human Longevity. Front. Genet. *12*.

Vindrieux, D., Augert, A., Girard, C.A., Gitenay, D., Lallet-Daher, H., Wiel, C., Le Calvé, B., Gras, B., Ferrand, M., Verbeke, S., et al. (2013). PLA2R1 mediates tumor suppression by activating JAK2. Cancer Res. *73*, 6334–6345.

Visser, T.J. (1994). Role of sulfation in thyroid hormone metabolism. Chem. Biol. Interact. *92*, 293–303.

Wang, B., Wu, L., Chen, J., Dong, L., Chen, C., Wen, Z., Hu, J., Fleming, I., and Wang, D.W. (2021). Metabolism pathways of arachidonic acids: mechanisms and potential therapeutic targets. Signal Transduct. Target. Ther. *6*, 1–30.

Wang, Y., Yang, J., Hong, T., Chen, X., and Cui, L. (2019). SIRT2: Controversy and multiple roles in disease and physiology. Ageing Res. Rev. *55*, 100961.

Willcox, B.J., Donlon, T.A., He, Q., Chen, R., Grove, J.S., Yano, K., Masaki, K.H., Willcox, D.C., Rodriguez, B., and Curb, J.D. (2008). FOXO3A genotype is strongly associated with human longevity. Proc. Natl. Acad. Sci. U. S. A. *105*, 13987–13992.

Wirthlin, M., Lima, N.C.B., Guedes, R.L.M., Soares, A.E.R., Almeida, L.G.P., Cavaleiro, N.P., Loss de Morais, G., Chaves, A.V., Howard, J.T., Teixeira, M. de M., et al. (2018). Parrot Genomes and the Evolution of Heightened Longevity and Cognition. Curr. Biol. *28*, 4001-4008.e7.

Xue, Z., Li, D., Yu, W., Zhang, Q., Hou, X., He, Y., and Kou, X. (2017). Mechanisms and therapeutic prospects of polyphenols as modulators of the aryl hydrocarbon receptor. Food Funct. *8*, 1414–1437.

Yarla, N.S., Bishayee, A., Vadlakonda, L., Chintala, R., Duddukuri, G.R., Reddanna, P., and Dowluru, K.S.V.G.K. (2016). Phospholipase A2 Isoforms as Novel Targets for Prevention and Treatment of Inflammatory and Oncologic Diseases. Curr. Drug Targets *17*, 1940–1962.

Yueh, M.-F., Huang, Y.-H., Hiller, A., Chen, S., Nguyen, N., and Tukey, R.H. (2003). Involvement of the Xenobiotic Response Element (XRE) in Ah Receptor-mediated Induction of Human UDP-glucuronosyltransferase 1A1*. J. Biol. Chem. *278*, 15001–15006.

Zhang, F.-Y., Li, R.-Z., Li, J.-X., Fan, X.-X., Xie, C., Liu, L., Yao, X.-J., and Lai-Han Leung, E. (2020). Phospholipase A2 as a novel therapeutic target in lung cancer. Integr. Cancer Sci. Ther. *7*.

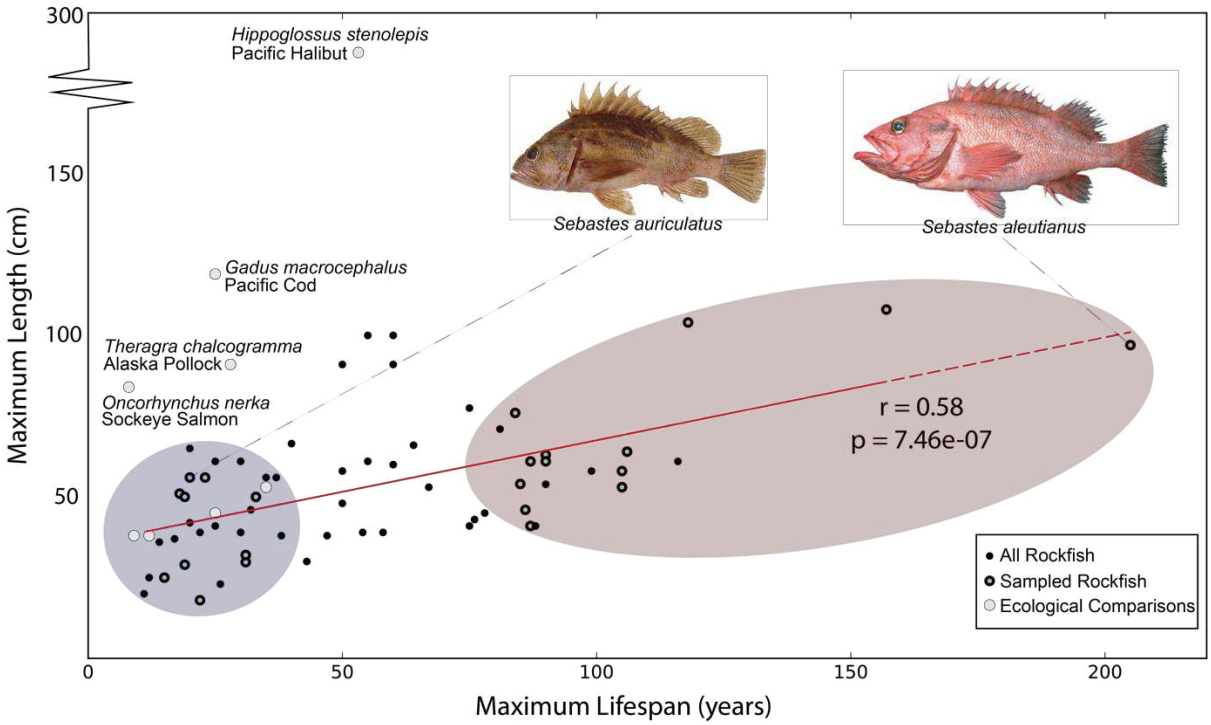


Figure 1. Rockfish exhibit a twenty-fold range in longevities.
Available maximum longevities and lengths and for sixty-five rockfish species (black dots) and ecological comparators (gray dots). While there is a significant relationship between size and longevity, it cannot fully explain the twenty-fold range in longevities, with substantial overlap in size across the spectrum. This capability is clearly exceptional, as *S. aleutianus* can survive for over two hundred years, while the much larger Cod and Pacific Halibut survive for only 25 and 55 years respectively, despite having overlapping ecologies. Trend line is specific for rockfish. Bold dots were sampled and sequenced for analysis, focusing upon the short-lived (blue) and long-lived (red) extremes of the group.

A

MLSP	Coverage	Depth
157	0.94	32
205	0.94	35
19	0.94	23
100	0.94	23
105	0.93	21
22	0.93	22
31	0.93	22
33	0.93	22
87	0.93	18
19	0.93	19
85	0.93	22
31	0.93	22
106	0.94	33
18	0.94	28
86	0.93	20
118	0.94	23
23	0.93	21
20	0.93	22
90	0.93	23
15	0.94	27
87	0.94	25
84	0.93	20
90	0.93	21
12	0.85	25
22 | 108	0.93	24

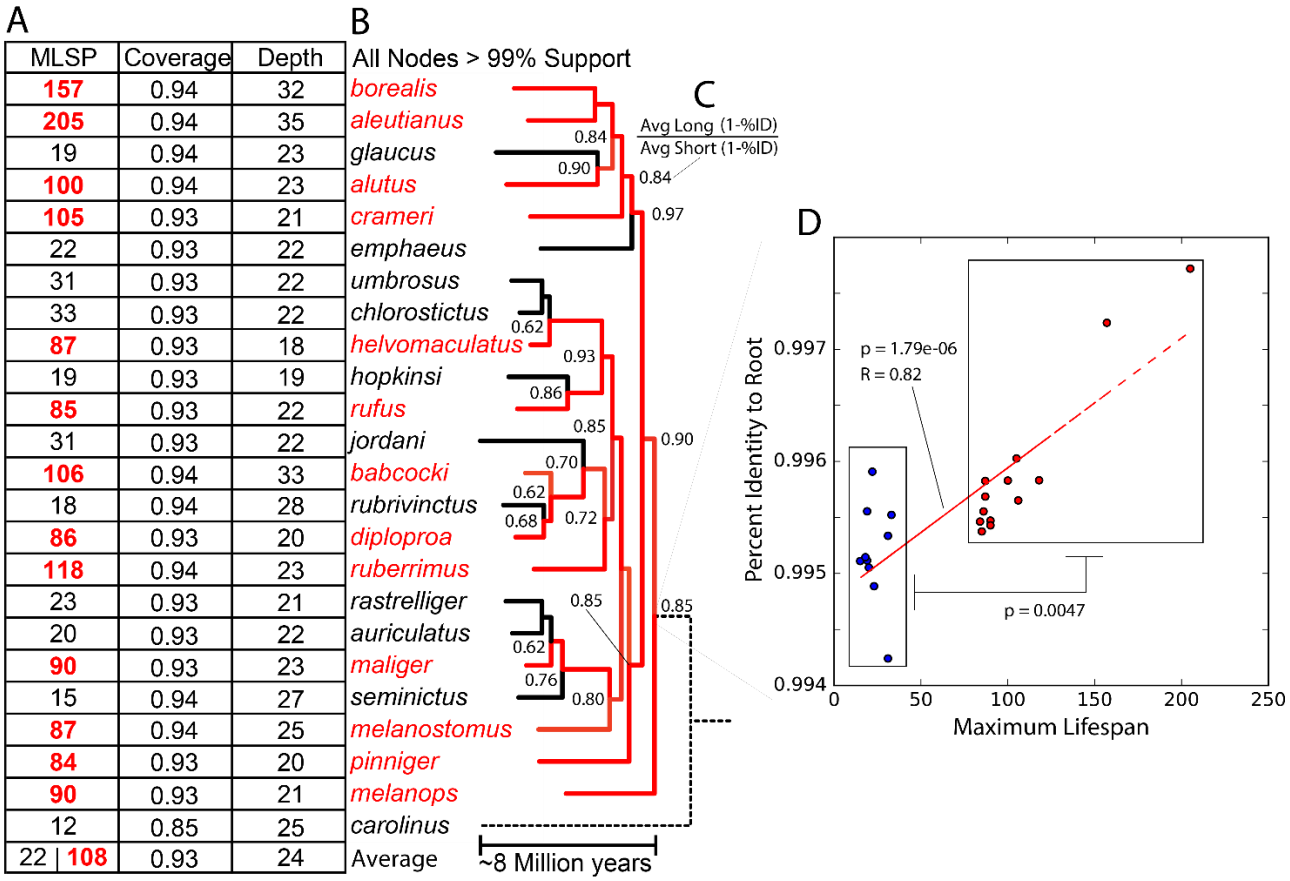
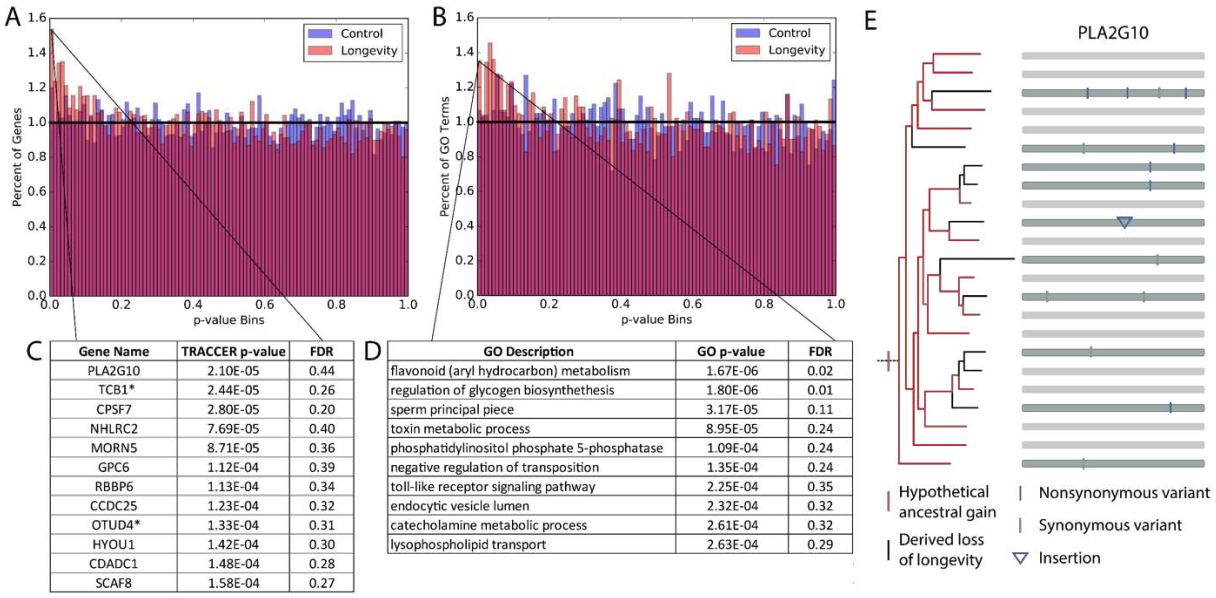


Figure 2. Targeted genomic sequencing of Rockfishes with multiple independent reductions of longevity. A. Targeted capture and sequencing from short- (black) and long-lived (red) rockfish. ~93% of the targeted exons and CNEs were reconstructed at an average depth of 24x, covering an average of 43,120,434 bases. Less than 100% coverage is expected, as many elements will be unique to reference genomes used and/or lost in rockfish evolution. **B.** Rockfish phylogeny generated from a set of high coverage elements with support values greater than 99% at each node. **C**. Node values are the ratio of the average long- to average short-lived percent identity. This value is always below 1.0, suggesting the mutational rate is inversely related to longevity across the phylogeny. **D.** At the rockfish root, where every species can be compared, there is a significant linear relationship (p = 1.79e-06) and difference (p = 0.0047) between longevity and percent identity to the ancestral state.



C

Gene Name	TRACCER p-value	FDR
PLA2G10	2.10E-05	0.44
TCB1*	2.44E-05	0.26
CPSF7	2.80E-05	0.20
NHLRC2	7.69E-05	0.40
MORN5	8.71E-05	0.36
GPC6	1.12E-04	0.39
RBBP6	1.13E-04	0.34
CCDC25	1.23E-04	0.32
OTUD4*	1.33E-04	0.31
HYOU1	1.42E-04	0.30
CDADC1	1.48E-04	0.28
SCAF8	1.58E-04	0.27

D

GO Description	GO p-value	FDR
flavonoid (aryl hydrocarbon) metabolism	1.67E-06	0.02
regulation of glycogen biosynthethesis	1.80E-06	0.01
sperm principal piece	3.17E-05	0.11
toxin metabolic process	8.95E-05	0.24
phosphatidylinositol phosphate 5-phosphatase	1.09E-04	0.24
negative regulation of transposition	1.35E-04	0.24
toll-like receptor signaling pathway	2.25E-04	0.35
endocytic vesicle lumen	2.32E-04	0.32
catecholamine metabolic process	2.61E-04	0.32
lysophospholipid transport	2.63E-04	0.29

Figure 3. Convergent rate analysis reveals specific genes and gene sets under convergent evolution with rockfish longevity. TRACCER identifies convergent shifts in the evolutionary rate of specific genes (**A**) and gene sets (**B**) associating with longevity in rockfish, and both are enriched at low p-values, indicated by the red bars above the baseline expected by chance. **C**. The top twelve genes in total have a false discovery rate of 0.27. **D.** GO enrichment of the TRACCER results reveals significant gene sets evolving with longevity, including "flavonoid (aryl hydrocarbon) metabolism" and "glycogen biosynthesis" with low FDRs of 0.019 and 0.009. **E**. Analysis of variation in *pla2g10* highlights the independent accumulation of mutations in short-lived lineages (black), while long-lived lineages (red) maintain the ancestral state without changes.

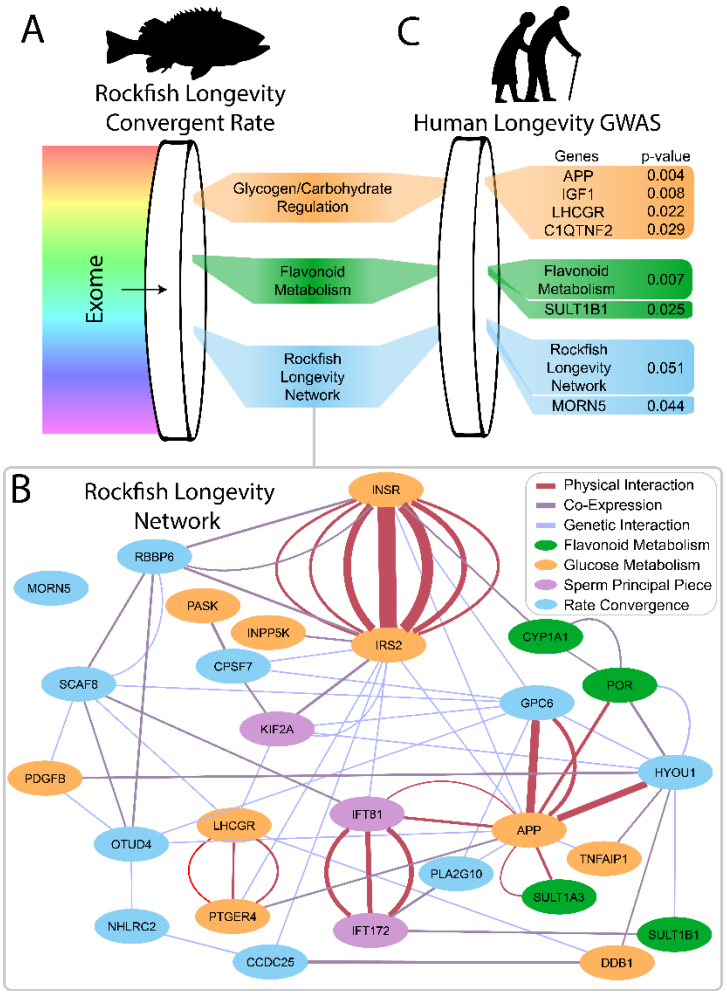
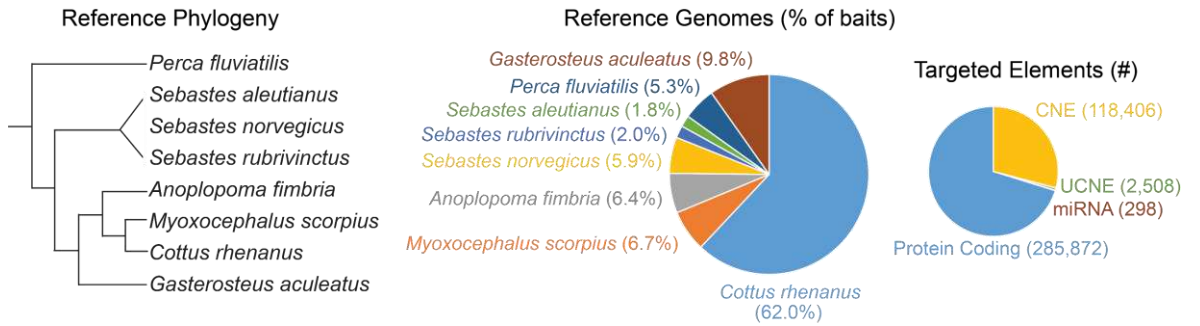


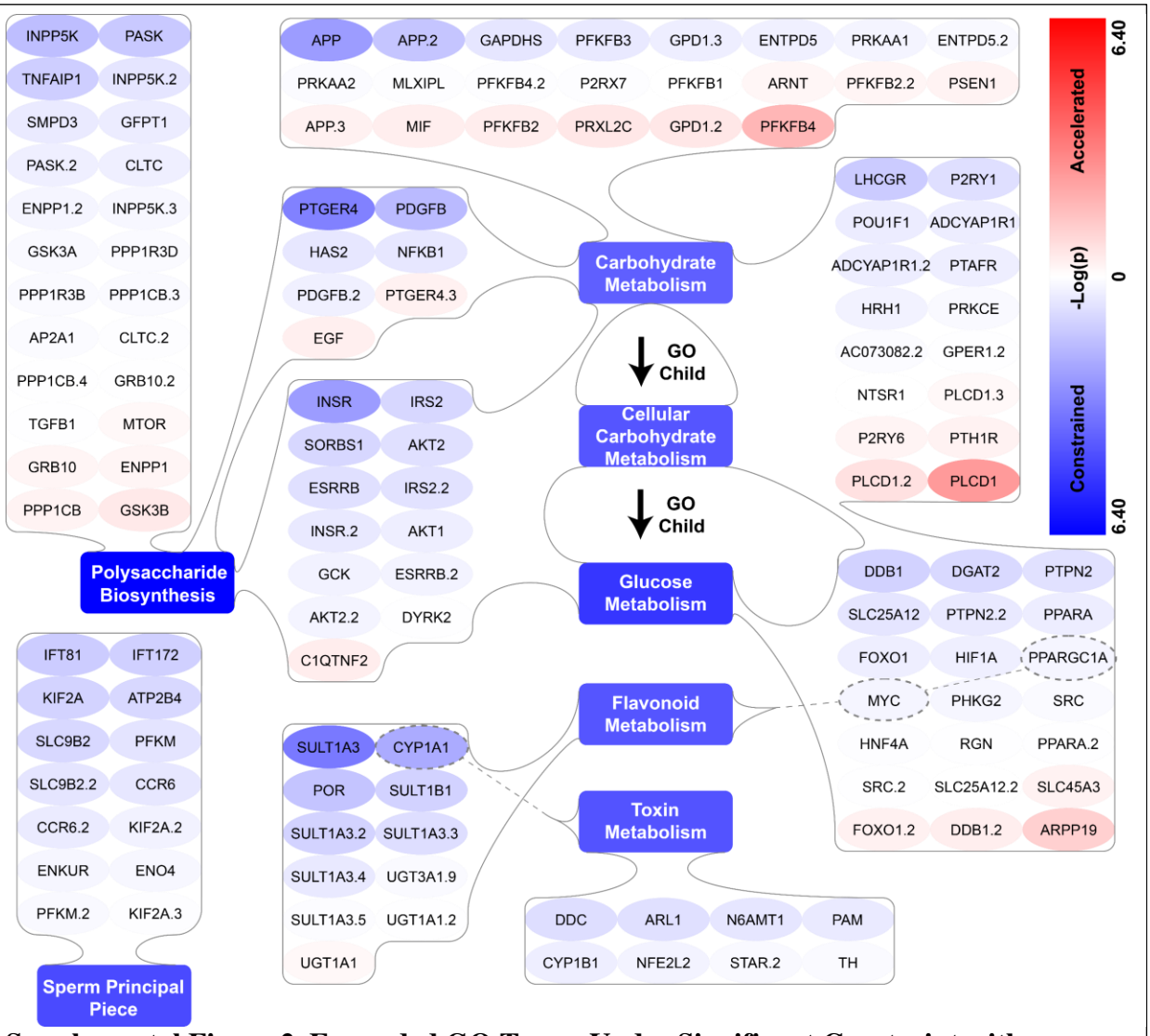
Figure 4. Rockfish Longevity Genes are Enriched for Cross-Function Interactions and Contain SNPs Associating with Human Longevity. A. Convergent rate analysis of Rockfish coding regions reveals enrichment of genes involved in flavonoid metabolism and glycogen metabolism. **B.** Network analysis of the top Rockfish Convergence genes (FDR < 0.27) along with the top genes ($p < 0.1$) in the top indirect GO Terms (FDR < 0.1) demonstrates extensive interactions, including across terms. For instance, APP physically interacts IFT81, POR, and SULT1B1, representing the other enriched indirect GO terms, along with GPC6 and HYOU1, which are not characterized in these terms. Individual lines are individual experiments confirming that interaction. **C.** When human GWAS of longevity is focused upon the top gene sets from the rockfish longevity analysis, the multiple hypothesis testing is dramatically reduced. The flavonoid metabolism gene set is significantly associated with human survival to the 99th percentile (FDR-adjusted $p = 0.01$), while the Rockfish Longevity Network is associated with survival to the 90th percentile (FDR-adjusted $p = 0.05$).

Supplementary Figures



Supplemental Figure 1. Perciforme Capture Design. The sequence capture design targets protein-coding exons as well as a set of conserved non-protein coding elements (CNEs), miRNA hairpins, and ultraconservative non-coding elements (UCNEs). As many of the available perciform genomes were fragmented and poorly annotated at the time of sequence capture design, we generated a list of conserved protein coding and CNE regions from the genomes of the three-spined stickleback (*Gasterosteus aculeatus*), the Japanese medaka (*Oryzias latipes*) and green spotted puffer (*Tetraodon nigroviridis*). A diversity of reference perciformes were then iteratively mined for these regions if their orthologous sequence was less than 85% identical to the baits already in the capture. As a result of this process, there will be oligonucleotide capture baits of at least 85% identity to each targeted sequence for every perciform genome used in the capture design. This multi-species 'phylochip' design enables usage of these baits to sequence large numbers of distantly-related perciform fish species.

1



Supplemental Figure 2. Expanded GO Terms Under Significant Constraint with Rockfish Longevity. GO Enrichment with propagated (indirect) annotations. Gene groupings demonstrate the degrees of interdependence between the various enriched functional groups. Genes in these functional groups are consistently and significantly under constrained (blue) evolutionary rates, while accelerated genes (red) are rare. These trends yield even greater signals in the functional groups than in the individual genes. The genes driving these signals are sorted to the top left of each grouping, and include hallmark aging candidates, such as INSR, APP, and LHCGR. Due to the complexities of mapping fish paralogues, some genes are appended additional identifiers.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Rockfishsupplemental.xlsx