Science
AAAS

Supplementary Materials for

Origins and evolution of extreme lifespan in Pacific Ocean rockfishes

Sree Rohit Raj Kolora[1]†, Gregory L. Owens[1,3]†, Juan Manuel Vazquez[1], Alexander Stubbs[1], Kamalakar Chatla[1], Conner Jainese[4], Katelin Seeto[4], Merit McCrea[4], Michael W Sandel[5], Juliana A. Vianna[6], Katherine Maslenikov[7], Doris Bachtrog[1], James W. Orr[7], Milton Love[4], Peter H. Sudmant[1,2]*
Correspondence to: psudmant@berkeley.edu

This PDF file includes:

Materials and Methods

1. Sequencing and library preparation

Whole genomes from 102 samples from 88 different *Sebastidae* species (**Table S10**) were sequenced using short read Illumina paired-end sequencing on a NovaSeq 6000. These 88 species span across five different genera including 79 *Sebastes*, three *Sebastolobus*, two *Sebastictus*, two *Hozukius,* two *Helicolenus* and one *Adelosebastes*. Whole genome sequencing (WGS) throughput ranged from 8 Gbp to over 250 Gbp (**Fig. S1A**) corresponding to genomic coverage of ~8X to more than 250X with an estimated genome size of 900Mbp. All reference genomes (see below) were sequenced to a minimum of 50X Illumina coverage. We sequenced representative individuals from the majority of *Sebastes* species for which lifespan information has been detailed (**Fig. S1B**). Lifespan information was collected from multiple sources (*6, 41–45*). High quality genomic DNA was extracted to perform long read sequencing for six *Sebastes* species on a PacBio sequel sequencer and one *Sebastolobus* on a PacBio sequel-II with average genomic coverage of above 40X **(Fig. S2A**). To generate chromosome level assemblies, long-range contact information was generated from high quality DNA using DNaseI based HiC (**Fig. S2B**) for five species (*S. aleutianus*, *S. entomelas*, *S. miniatus*, *S. rosaceus* and *S. umbrosus*). DNaseI based Hi-C was performed as described in Bracewell *et al.* (*46*) with a few modifications. Briefly, tissue sliced into small pieces using a scalpel blade while the sample is in dry ice. Then transferred the 10 mg of tissue into a fresh 1.5ml microfuge tube and ground with a micropestle in microtube with 250 µl PBS buffer and we rinsed the pestle with 750 µl of PBS buffer. The resulting 1 ml homogenate was mixed gently and passed through 100 µm Nylon mesh (sterile cell strainer) to remove debris. The homogenate was centrifuged at 2000xg for 5 min at 4°C and the cell pellet fixed with 1 ml of 4% formaldehyde at RT for 10 min, gently inverting tubes every 2 min. The resulting cells were treated as per Ramani *et al.* (*47*) from steps 4 to 74. DNA Libraries were prepared with an Illumina TruSeq Nano library Prep Kit and sequenced on a Hiseq 4000 with 100 bp PE reads. For one species, *Sebastes umbrosus* we performed optical mapping using the Irys technique from Bionano genomics.

2. Mitogenome assembly and annotation

Circular mitogenomes were assembled from the Illumina paired-end libraries of 102 samples using the MitoZ toolkit (*48*). 7Gb of paired end data and a k-mer size of 99 was used to assemble the mitogenomes which were annotated in two iterations using the stand-alone version of the MITOS pipeline (https://gitlab.com/Bernt/MITOS). In the first iteration, the positions of the tRNA were determined using the tRNAscan-SE program with the vertebrate covariance model database from MITOS. The mitogenomes were identically ordered starting with trnaF (Phe) on the sense strand. The second iteration of MITOS on the ordered mitogenome was used to identify the protein coding sequences (13 genes), noncoding RNA (2 rRNA and 22 tRNA) and the origins of replication (heavy and light strand). The assembled mitogenome sizes ranged from 16.5kb to 17.3kb with *Sebastolobus* species containing the smallest mitogenomes.

3. Genome assembly

PacBio reads were assembled in a phase-aware manner using the FALCON assembler (*49*) followed by haplotype separation (FALCON-Unzip). An additional set of genome assemblies were produced using CANU for error correcting PacBio data (*50*) followed by assembly with WTDBG2 (*51*). CANU parameters were adjusted for the PacBio error model (15-18% error) and a genome size of 950Mb. The WTDBG2 assembler parameters were changed to suit our error corrected data and a higher sampling rate of 2. This second set of assemblies were polished with Racon to remove residual errors (*52*). The N50 of contig-level assemblies of *Sebastes* species was at least 8 Mbp, but in the *Sebastolobus alascanus* assembly the N50 was less than 2 Mbp. This lower contiguity in the latter assembly could be due to higher heterozygosity in this species compared to other rockfish (more than 2X of *Sebastes*).

For both sets of long read *de novo* genome assemblies (FALCON-Unzip and CANU+WTDBG2) duplicate sequences were purged and the haplotypes were separated using the purgedups tool (*53*). Further polishing was performed using the Arrow polishing tool and performing two iterations of variant calling with Freebayes (https://github.com/freebayes/freebayes) followed by consensus calling as recommended by the Vertebrate Genome Project (VGP) assembly pipeline (V1.6). Consensus calling after calling variants helped in removing SNP errors which could have been introduced in the PacBio sequencing. Scaffolding of the *S. umbrosus* genome using optical mapping data was performed using Bionano Solve (v3.1). To generate chromosome level scaffolds, further genome scaffolding of *S. umbrosus* was performed with long read interaction data of HiC using SALSA2 (*54*). For the remaining four *Sebastes* genomes (with the exception of *S. pinniger*) without optical mapping data, genome scaffolding was performed using juicer and 3d-dna assembler (*55*) followed by manual corrections using Juicebox (*56*). The reference-quality genomes of *S. umbrosus, S. aleutianus, S. rosaceus, S. pinniger* and *S. miniatus* from FALCON-unzip assembly while those of *S. entomelas* and *Sebastolobus alascanus* from CANU+WTDBG2 assembly were used for downstream analysis due to better quality assemblies in these samples with the CANU+WTDBG2 assembler.

The quality and completeness metrics of the reference-quality genomes (supported with PacBio data) were also assessed based on k-mer abundance (here k=21) using Merqury (*57*) in a reference free manner. Merqury uses the k-mer spectrum from the Illumina short reads (PacBio reads are avoided due to their high error rate) and compares it to the k-mer spectrum of the genome assembly to quantify the Phred quality score (QV) and the error rate in the assembly (errors are based on k-mers specific to assembly but not to Illumina read set). The completeness is based on the reliable k-mers present in both the assembly and the read set. We note however that some parts of the genome (such as AT-rich regions) are often missed by the Illumina based sequencers, hence these quality measures are indications of both assembly quality and intrinsic genome factors. The FALCON based genome assemblies had high QVs (~40) and lower error rates compared to the WTDBG2 (**Table S1**). Except for *Sebastolobus alascanus*, all the reference-quality genomes were >90% k-mer complete and the unphased *S. entomelas* had the best score (96%). The genome assembly quality was also measured with BUSCO (V3) (*58*) using the vertebrate gene set (V9), and the BUSCO based completeness was above 97% for all genomes.

We visualized a minimum of 24 distinct blocks in the five HiC scaffolded genomes (**Fig. S3**) suggesting that genomes were effectively scaffolded into chromosome level assemblies (2n =48 in *Sebastes*). The 102 whole genomes were independently assembled using the paired end Illumina data with MaSuRCA (*59*). Similar to the reference quality genomes, the k-mer completeness and BUSCO metrics were calculated for the 102 Illumina-based assemblies. The genome assembly quality was higher than 90% for a majority of the genomes (**Fig. S4**).

Illumina-only genome assemblies were scaffolded with guided information from the chromosomes of reference-quality genomes using the RagTag (V1) (https://github.com/malonge/RagTag). The reference-guided scaffolding was performed only for the most contiguous sample in cases there were multiple samples for a species. The reference used for scaffolding *Sebastes* genomes was selected based on the closest phylogenetic distance between the sample species and the reference species, while all the outgroup samples were scaffolded with *Sebastolobus* reference (**Table S2**). To minimize differences in genome contiguity between species for MSMC analyses, we also scaffolded against the *Sebastes aleutianus* reference genome for all species.

4. Genome annotation

Genomic repeats were identified using repeatmasker (V4). The reference quality rockfish genome assemblies were annotated with a combination of *ab initio* methods and RNAseq data. Extracted mRNA was sequenced for six reference species (excluding *S. aleutianus*) from either brain or eye tissue. The transcriptomic reads were aligned against the respective species references using RNAseq based aligner STAR V2.7.1 (*60*) in two-pass mode. These read alignments were used as evidence to detect intron-exon boundaries in annotation. Additional transcript information was generated from publicly available RNAseq data from *Sebastes* species (*S. caurinus* - SAMN03757544; *S. carnatus* - SAMN05292523; *S. maliger* - SAMN05292524; *S. nebulosus* - SAMN05292522; *S. schlegelii* - SAMN05893354, SAMN05893355, SAMN05893357; *S. serranoides* - SAMN05292521). The publicly available RNAseq samples were *de novo* assembled independently into transcripts with Trinity v2.8.4 (https://github.com/trinityrnaseq/trinityrnaseq). The longest transcripts of each isoform were retained and the coding sequences and protein translations were extracted using Trinotate v3.1.1. This resulted in 1,435,263 coding proteins from 1,822,510 longest transcripts. The longest transcript isoforms were mapped to the reference genomes using Minimap2 with splice-aware parameters (splice:hq) and the proteins coding sequences were mapped with Spaln V2.3.3 (*61*) with species-specific parameters for tetrapods. Species-specific RNAseq read alignments, transcript and protein sequence alignments against assembled genomes were used as evidence for genome annotation.
The reference-quality genomes, soft masked for repeats, were annotated using *ab initio* gene models generated with BRAKER2 V1.9 (*62*) using Augustus (self-trained for each species) and GeneMark-EM tools. These *ab initio* models and transcript evidence from species-specific RNAseq read alignments, transcript and protein sequence alignments were used to annotate the rockfish genomes with EvidenceModeler (EVM) (higher weights to RNAseq read alignments) using the Funannotate V1.7.4 pipeline (https://github.com/nextgenusfs/funannotate). High quality Augustus predictions were extracted from gene models identified based on BUSCO gene search wherein zebrafish was used as the seed species and vertebrate_odb10 was used as the BUSCO database. Further annotation filtering was performed by retaining those with high similarity to

vertebrate peptide sequences obtained from Ensembl99. The number of annotated gene models in the reference genomes range from 24,000-28,000 with the highest in *S. aleutianus* (>28k) and the lowest in *S. umbrosus* (>24k) (**Table S3**).

The noncoding RNA elements were annotated through secondary structure similarity search using covariance models of the Rfam database (V14) through Infernal tool (V1.1) (https://github.com/EddyRivasLab/infernal) and tRNAscan-SE (https://github.com/UCSC-LoweLab/tRNAscan-SE). We identified three types of non-coding RNA: microRNA (miRNA), small nucleolar RNA (snoRNA) and transfer (tRNA). The highest number of miRNA were in *S. aleutianus* (117), snoRNA in *S. entomelas* (212) and tRNA in *S. rosaceus* (2833) (**Fig. S5**).

Annotation of fragmented Illumina-only contig-level genome assemblies leads to incomplete gene models. Hence, genome annotation was carried out for the reference scaffolded Illumina genomes. These scaffolded genomes (81 in total) were annotated using LiftOff V1.2.1 (https://github.com/agshumate/Liftoff) with minimum genic feature coverage fraction of 0.8 and sequence identity of at least 70%. Coding and protein sequences were extracted from the respective genomes using GFF utilities (https://github.com/gpertea/gffread).

5. Ortholog detection

The orthologs in reference genomes were detected using orthology finding program Proteinortho V6 (*63*) based on minimum 50% sequence similarity (E<1e-4) using fast protein similarity search tool Diamond (*64*) and conserved gene neighborhood information (--synteny option). We identified 47,130 orthogroups between the seven reference genomes with 13,098 groups containing at least one ortholog in every species (including paralogs in one or more species) and 12,755 groups with one-to-one orthologs between all species (13,640 single copy in *Sebastes* species). For genes missing in one or more reference species, lifted annotations were obtained based on other reference genomes. Lifted annotations were also used to impute orthologs for each orthogroup in the non-reference quality genomes using LiftOff. Gene names were assigned based on orthology with either human or zebrafish protein sequences from Ensembl99.

We obtained maximum lifespan information for 55 different rockfish species and for this dataset we identified 17,387 orthogroups with orthologs in all selected species (not all species are present in every orthogroup). These orthologs from the 55 rockfish species were aligned with the HyPhy alignment pre- and post-processing workflows by removing sequences with in-frame stop codons and containing ambiguous bases, while retaining sequences with at least 80% sequence similarity (https://github.com/veg/hyphy-analyses/tree/master/codon-msa). Phylogenetic trees were built for each orthogroup using IQ-TREE2 for protein sequences with auto model selection and ultrafast bootstrapping with 1000 replicates. Further filtering of mis-assigned sequences in orthogroups was performed using composition test in IQ-TREE2.

6. Phylogenetic analysis

To conduct a thorough phylogenetic tree, we ran phylogenetic analysis on a full set of 83 species (excluding 5 low quality genome assemblies) and a *Sebastes* set of 76 *Sebastes* species (excluding 3 low quality genome assemblies) using BUSCO genes that had been identified using BUSCO V3 (see section on genome assembly). Nucleotide sequences were aligned using MACSE (v2.03) with the parameters "-max_refine_iter 3 -local_realign_init 0.3 -local_realign_dec 0.2". Alignments were then filtered to remove positions with >=70% missing data using pxclsq in phyx (https://github.com/FePhyFoFum/phyx). After filtering, we removed genes with < 1000 positions. For each gene, we calculated a maximum likelihood tree with IQ-TREE2 (v2.0.6) (*65*) using default parameters. This set includes cases where there are extremely long branches, likely from mistaken orthology. We used TREESHRINK (v1.3.4) (*66*) with default parameters which can identify and remove abnormally long branches. After this filtering, we found that there were still cases of abnormal and extremely high genetic distances between species. This included cases where a single gene alignment had representatives of two different genes. For these, there would not be a single abnormal branch, but the two clades with large genetic distance between them. We used DiStats to measure pairwise sequence divergence between all *Sebastes* samples in each gene (https://github.com/mptrsen/distats). We required that the maximum pairwise sequence divergence was <= 20%. Finally, the gene trees from this set of genes (**Table S4**) were used to calculate a species tree using ASTRAL (v5.7.1) with default parameters. Prior to ASTRAL, we collapsed nodes with <10% bootstrap support using phyx (v1.1). The species tree produced by ASTRAL uses branch lengths in coalescent units, and since we only included a single sample per species, terminal branches have no length. We required a species tree that included terminal branch lengths for downstream analysis, so we took the topology produced by ASTRAL and used it as the fixed topology for a maximum likelihood tree using IQ-TREE2 on the concatenated gene sequences.

This fixed species tree topology was used to estimate the species divergence times under the multispecies coalescent using the A00 model (speciesdelimitation=0 and speciestree=0) of the Bayesian Markov chain Monte Carlo (MCMC) program BPP V4.3.8 (*67*). In total we used 624 loci and the thetaprior was estimated assuming a tauprior (invgamma(3,0.14)) at the base of the tree. 600000 samples (nsample) were run at a sample frequency of 2 (sampfreq) and burnin of 1000. The origin of rockfish ancestor was dated to 8-11.5 million years (My) from fossil evidence from rockfish samples in Japan and California (*10*). Based on these time estimates, the divergence times across the rockfish and outgroup species BPP phylogeny was calibrated to geological time using BPPR package (https://github.com/dosreislab/bppr) (**Fig. S6**). The root of the *Sebatidae* clade was dated to ~27 My, which is close to the median divergence time of 29 My previously estimated (www.timetree.org).

Maximum lifespan information was collected for 55 *Sebastes* species (**Table S11**). To model lifespan evolution throughout the phylogeny we performed continuous trait character mapping using the phytools package (V0.7.70) (*68*) (**Fig. S7**).

7. Selection and evolutionary rates

To investigate the presence of lineage-specific adaptation in either short lived (7 species with max lifespan <=20 years) or long-lived species (7 species with max lifespan >=105 years) we tested for positive selection with either group of species in the foreground using branch-site model in HyPhy with adaptive branch-site random effects likelihood (aBSREL) (*69*). 9,637 and 9,961 orthogroups were tested in total for positive selection in long-lived and short-lived rockfishes respectively, post-filtering for alignment errors. We identified 763 positively selected genes (PSGs) in long-lived, and 873 PSGs in short lived species. Of these, 126 PSGs in long-lived were found in more than one lineage and 118 in more than one short-lived lineage (**Tables S12**, **S13**). Over-representation tests for gene functions and pathways were performed for the positively selected genes independently for short-lived and long-lived categories (with respective backgrounds) using R package clusterprofiler V3.16.1 (*70*). Enrichment tests for gene ontology (GO) revealed significant over-representations for DNA repair related functions in long-lived rockfish ($q<0.05$, **Table S5**), while short-lived rockfish PSGs showed no enrichment. We identified 15 PSGs common between both long-living and short-living species but with no overrepresentations in any functional categories. This indicates that these genes can either be lineage specific adaptations with/without influence on longevity or these gene are influenced by genetic drift rather than positive selection.

The branch-site model of selection helps predict positive selection in foreground species, however genetic drift can be mistaken as positive selection because drift can also promote changes in coding sequences. In order to test the efficiency/strength of selection — i.e., distinguish relaxed selection due to genetic drift versus intensified selection arising from positive selection in genes — we employed RELAX (*71*). We performed this analysis twice, using long-lived taxa as the foreground in the first run and short-lived taxa as foreground for the second. We identified 378 genes in long-lived and 467 genes in short-lived species experiencing intensified selection ($K>1$, $p\leq0.05$) while 420 and 538 genes experienced relaxed selection ($K<1$, $p\leq0.05$) in long- and short-lived species respectively. We did not observe significantly enriched GO terms for relaxed or intensified selection in either long- or short-lived rockfish. Short-lived species possess higher number of genes experiencing relaxed selection compared to long-lived rockfish (**Fig. S8A**). This is similar to short-lived annual killifish accumulating more deleterious gene variants as a consequence of neutral genetic drift which in turn affects their lifespans. Additionally, the proportion of genes under intensified versus relaxed selection is lower in short-lived rockfish compared to long-lived species (**Fig. S8B**). Within genes experiencing intensified selection, 14 genes (*MMP19, MED26, RUBCN, FARSB, MCCC2, NR5A2, FAM183A, SCYL3, FOXN3, GDPD2, POGLUT2, LRRC53, PIGA, NPY2R*) in multiple long- and 15 genes (*VSIG10L, PLA2G4A, PER, GFI1B, IL34, SYNDIG1, FILIP1, GJA1, MRPS16, AN33B, ALG1, IL1B, NMRK2, GALNT15, LRRC3C*) in multiple short-lived species were identified to possess positively selected sites. Only one gene experienced relaxed selection in multiple long- (*SRPX2*) and short-lived (*ADIPOR2*) rockfishes. We did not observe overrepresented biological functions for genes overlapping positively selected genes experiencing relaxed/intensified selective constraints in both lifespan categories.

We identified genes with evolutionary rates significantly correlated with variation in lifespan using RERconverge (*13*). Earlier phylogenetic analysis found prevalent incomplete lineage sorting which means that individual gene trees may not agree with the overall species tree. To account for

this, we ran the analysis with multiple trees and modified this workflow by using Fisher's Z transformation to estimate significance summarizing across tree topologies. 1002 different tree topologies from the species tree were used (including 1000 bootstrap trees, the species tree and the majority rule consensus tree). Maximum lifespan in rockfish is strongly correlated to body size and depth (*72*), hence in addition to maximum lifespan, we compared the evolutionary rates for additional traits which are strongly correlated to lifespan, namely maximum depth, body size at 50% maturity and the residual of body size & maximum depth (**Fig. S9**, see section 10). In this case, the residual represents variation in lifespan that is not associated with body size or depth. At a q-value cutoff of 0.05 we identified 91 genes associated with lifespan, 66 associated with size at maturity (Lmat50), 49 associated with maximum depth, and 56 associated with the linear model residual (**Table S14**). We observed enrichments for glucose metabolism for genes significantly associated with the residual (**Fig. S10**). We tested for functional enrichment of GOs using human orthologs of the residual associated genes and observed significant enrichments ($q<0.05$) for Insulin/IGF pathway (P00033) and Gonadotropin-releasing hormone receptor pathway (P06664). To assign genes a position along these three different *axes of longevity* we computed the proportion of the -log(P) for each axis compared to the sum of the -log(P) values along all axes:

$$S_i = \frac{-\log(P_i)}{\sum_{k=1}^{3} -\log(P_k)} \text{ thus } \sum S_i = 1.$$

In addition to tests for selection and rate of evolution, we identified sites in protein sequences which appear to have convergently evolved in more than one long-lived species ($a \geq 2$) at high frequency ($p \geq 0.7$) compared to the background species ($b \geq 4$), while that particular site occurs at low frequency in background species ($q < 0.3$). This was performed by reading the protein sequence alignments of each gene and testing whether the site occurs at higher frequency in foreground long-lived species compared to the background (rest of the species). We repeated this test in two additional variations where we identified sites in genes which vary in short-lived species compared to the background species or long-lived species compared to the short-lived species. We used a randomization procedure over 1000 iterations where the foreground & background species were randomly sampled to identify sites with significant variation in either long-lived or short-lived species. The number of times each site appears as converged was counted (C) and the *p*-value for each site was defined as the number of times the site doesn't appear as a converged site by chance over all the iterations (n) (1-(C+1/n+1)). Multiple testing was performed using the Benjamini-Hochberg procedure using the *p.adjust* function in R. We identified 31 genes in total (including three uncharacterized) which showed significant ($q \leq 0.05$) converged amino acid sites based on long- and short-lived species (**Table S6**). This test does not control for phylogenetic non-independence which could inflate p-values, therefore we used these results to highlight potentially interesting signals for more robust follow-up analyses.

We observed varying degrees of overlap between positively selected genes in short- and long-lived rockfishes; genes experiencing intensified/relaxed selective constraints and genes with varying evolutionary rates correlated with lifespan (**Figure S11**). The highest overlap (n=162) was between positively selected genes in short-lived species and genes experiencing intensified selection in short-lived rockfish.

We highlight the overlap of genes identified as probably longevity-related from our evolutionary rate and selection analysis and known longevity/disease-causing genes from genomic studies of

other exceptionally long lived vertebrates, namely naked mole rat, giant tortoise and bowhead whale (*73–76*) (**Table S7**).

8. Structural evolution

Adaptive evolution can occur through changes in gene copy number. Copy number changes are challenging to assay genome wide due to the difficulty in assessing orthology in repetitive regions. To get a snapshot of how gene copy number has varied across the genus we used read depth as a proxy for copy number. For each sample, except *S. aleutianus*, we aligned all reads against the *S. aleutianus* reference genome using NextGenMap (v0.5.5) (*77*) to allow for efficient cross-species mapping. We first tested gene regions specifically by extracting total read depth for each *S. aleutianus* gene and normalizing by the median read depth in each sample. For species with multiple samples sequenced, we took the average normalized read depth. Many genes have little variation in read depth, so we first characterized diversity of copy number by classifying each gene in each sample as constant, if read depth was between 0.75X and 1.25X of the median depth, or variable if it had any value outside of that range. We required that at least 5% of species were variable at each gene. This is orthologous to a minor allele frequency filter when conducting genome wide association analysis on SNPs and it retained 11,963 of 28,393 genes. For each remaining gene, we then calculated a Phylogenetic Generalized Least Squares asking how much maximum lifespan explains normalized read depth while controlling for the phylogenetic relationships, and used a q-value correction for multiple testing with a false discovery rate of 5%.

We conducted a qq-plot and found that there was an enrichment of low p-values suggesting more association than expected under the null model and found 47 genes significant after q-value correction (**Fig. S12**, **Table S8**). Two of our top genes were orthologous to butyrophilins which are known immune function genes. For each of these, we see higher read depth in species with longer lifespans and a wide range of depths from zero to four times the median.

We then expanded our analysis to look at the association between read depth and lifespan for the entire genome, rather than just the gene space. We extracted read depths for 100 bp non-overlapping windows across the entire *S. aleutianus* genome using samtools and then repeated the phylogenetic generalized least squares test for each window. If there is a relationship between read depth and lifespan, it can either be positive, where longer lived species have greater read depth and likely higher copy number, or negative where longer lived species have lower read depth and lesser copy number. We find that there is an enrichment for low p-values ($-\log(p)<3$) with a positive relationship ($p = 0.588$; X-squared = 768.92; df = 1; p-value < 2.2e-16). We looked for clusters of windows with low p-values ($-\log p < 3$) using a 10kb non-overlapping sliding window. These should represent tandem repeat regions expanded in our reference genome and also other long-lived species. We find that the region containing our two previously identified butyrophilins (**Fig. S13**), as well as other immune-related genes was our top region by this metric, and also contained the fourth lowest p-value for any individual window ($-\log p = 7.43$). We were then interested in whether the amount of copy number variation in the butyrophilin gene cluster was exceptional compared to the rest of the genome. We calculated the standard deviation of read depth for all windows and compared it to the low p-value windows in the butyrophilin cluster. Across the entire genome, standard deviation of read depth was generally low (mean sd = 0.3) and was

variation was higher for windows in our candidate region with -log p-value > 3 (mean sd = 0.9, sd = 1.05 for lowest p-value individual window). Nevertheless, 2% of all windows had standard deviation > 1, representing 16 Mbp of the genome. This suggests that the level of copy number variation in our candidate region is high, but not exceptional relative to the rest of the genome.

We then repeated the association analysis using the *S. umbrosus* reference genome, to see if reference bias is affecting our signal. We again found a bias towards a positive relationship between read depth and life span, suggesting that this is not being driven by the lifespan of the reference used, since *S. umbrosus* has a below average maximum lifespan (31 years). We found that the orthologous butyrophilin cluster on chromosome 1 also has low p-values (PGLS, $-\log_{10}$ p = 7.58) (**Fig. S14**).

We followed up our analysis of the butyrophilin cluster by asking if read depth in this region was correlated with maximum depth, size at maturity or the residual of maximum lifespan, when controlling for both factors, similar to our RERconverge analysis (See section 7). To do this, we repeated the PGLS across the butyrophilin cluster and surrounding genomic region using these other factors instead of maximum lifespan. We found that while there was enrichment of low p-values when testing maximum depth in the butyrophilin cluster, there was not for size at maturity or the residual of maximum lifespan (**Fig. S15**).

Genes of the butyrophilin family were identified through a HMM based search for protein sequences using domain models from PANTHER (superfamily PTHR24100) using InterProScan5 (*78*). Manual curation of our 7 reference genomes for annotated butyrophilins (Btn) and butyrophilin-like (Btnl) genes revealed the highest estimates in the longest living rockfish (*S. aleutianus*) while the lowest copies were identified in *S. pinniger* and the outgroup *Sebastolobus alascanus* (**Table S9**). However, it is to be noted that both the latter genomes were only contig-level assemblies.

Segmental duplications (SDs) in soft-masked rockfish genomes were detected using SEDEF tool (*79*) which uses minimap2 to self-align the genome sequences and employs a seed finding and chaining approach to detect SDs. The regions on Chr1 and Chr22 which showed butyrophilin clusters in particular were specifically extracted to identify probable SDs. RNAseq data and gene orthology information (at least 4 species in orthogroup) were leveraged to check the levels of expression of butyrophilin genes (all multi-exonic) identified in both these loci. The RNAseq alignment and gene annotation data used for annotation was used to calculate the expression levels as transcripts per million (TPM) using TPMCalculator (https://github.com/ncbi/TPMCalculator). The expression of butyrophilin genes were tissue specific, however some species show a very low (TPM<1) level of expression (**Fig. S16**).

Rapidly evolving gene families across our reference rockfish and outgroup *Sebastolobus* species were inferred using the CAFE5 (*80*). We limited this analysis to the high-quality genomes alone since the annotations in contig-level genomes were lifted from the high-quality ones. The PANTHER superfamilies for the genome annotations and ultrametric trees for these 7 species were used as input assuming a Poisson distribution for the root frequency distribution and two gamma categories (k=2). This analysis was repeated excluding *S. aleutianus* since the annotations lack transcriptomic support. Gene families undergoing significant (p<=0.01) contractions/expansions

in both (including and excluding *S. aleutianus*) and those with significant likelihood of belonging to our provided gamma category were retained. Compared to the *Sebastolobus* outgroup, species in the *Sebastes* clade seem to have experienced less gains across gene families. We observed lineage-specific variation in gene family content rather than based on maximum lifespan, with the exception of T-lymphocyte activation antigen (PTHR25466:SF1) (**Fig. S17**) which showed a significant positive correlation with gene family size (PGLS $p<0.05$). However, the study of gene family variation from the perspective of longevity within this clade requires more high-quality genomes rockfish genomes.

9. SNV calling and population size analysis

To calculate effective population sizes and quantify genetic diversity within different species, we aligned paired-end Illumina reads for each sample to its own species' scaffolded genome using BWA-MEM2 (v2.0pre2). Reads were sorted and indexed using samtools (v1.8). We then used GATK (v3.8.1) to remove duplicate reads and realign around indels with the tools `RealignerTargetCreator` and `IndelRealigner` because the variant calling tool used did not include a subsequent realignment around indels step. SNPs were called using bcftools (v1.10.2) mpileup with the parameters `-B -q 20 -Q 20 -C 50 -O b -f`, and bcftools call with the parameter `-v indels`. MSMC input files and mask files based on read depth were created using bamCaller.py from msmc-toolkit (https://github.com/stschiff/msmc-tools). We further refined our input files by calculating mappability of the genome with GenMap (v1.2.0) with the parameters `-K 50 -E 2` (*81*). This estimates k-mer uniqueness and identifies regions of the genome where Illumina reads are unable to map uniquely. We masked all regions larger than 50 bp with less than 1— i.e., max— mappability. Final input files were created by combining our refined mask files with the msmc inputs using generate_multihetsep.py from msmc-toolkit. We further filtered this set to retain only chromosome scale contigs larger than 10 Mbp. Finally, MSMC (v1.1.0) was run on each sample (*82*) to estimate effective population size (Ne). We visualized the Ne estimates and filtered the data to remove outliers. For each sample, we removed the most recent timepoint Ne estimate because it often had impossibly high population size estimates. We also removed two samples (*S. rosaceus* and *S. mystinus*) that had population size estimates an order of magnitude higher than all other samples.

Since we were interested in the effects of lifespan differences between species and our phylogenetic analysis found that lifespan evolved rapidly across the genus, we limited our examinations of Ne to only values within the last 1e6 generations. Further back it is more likely that past maximum lifespan has diverged from current maximum lifespan. With these, we calculated the average recent Ne per species over all retained timepoints. When there were multiple samples per species, these values were averaged to retain a single Ne value per species. We accounted for phylogenetic non-independence by using a Phylogenetic Generalized Least Squares using NLME v3.1-151 (https://cran.r-project.org/web/packages/nlme) with the species level phylogeny generated (see section 4). We found a significant negative relationship ($p=0.0048$) suggesting that species with longer max lifespan have a lower effective population size.

10. Life history trait analyses and modelling

Mortality in different species is commonly modeled using the Gompertz-Makeham hazard function $\mathrm{m}(t) = A_0 e^{Gt}$ which states that the probability of mortality ($\mathrm{m}(t)$) can be modeled by an age-dependent component that increases exponentially with age (the Gompertzian component, *G*), and an age-independent component (the Makeham component, A_0). While maximum lifespan is known for many rockfish species (see main figures), the specific mortality functions and their parameters are not. However, using the mathematical approach outlined by Finch *et al.* (*83*), we estimated these coefficients along with the age specific mortality hazard and survival curves for 74 different rockfish species. Briefly, the Gompertzian equation can be expressed as the following survival function:

$$\mathrm{S}(t) = \exp\left(\frac{A_0}{G}(1 - e^{Gt})\right)$$

We can then consider the maximum observed lifespan (t_{max}) as representing the time at which a population of size *N* has diminished to 1/N which can be approximated by *S(t):*

$$\mathrm{S}(\mathrm{t_{max}}) \approx 1/\mathrm{N} = \exp\left(\frac{\mathrm{A_0}}{\mathrm{G}}(1 - \mathrm{e}^{\mathrm{Gt_{max}}})\right)$$

Which we can rearrange as:

$$\mathrm{t_{max}} = \frac{1}{\mathrm{G}}\ln[1 + \mathrm{Gln(N)/A_0}]$$

Additionally, average mortality *m* can be expressed as:

$$\mathrm{m} = 1/\int_0^\infty S(t)dt$$

We estimate average mortality using the approach of Then *et al.* (*84*) which has been shown to provide an accurate estimate of *m*. We can then solve the above system of non-linear equations for various values of N (**Fig. S18**). We implemented the fitting procedure using the *nleqslv* package in R (**Table S15**).

Because rockfish are targeted by commercial fishing throughout the United States, Canada, and Japan, age-specific growth rates and fecundity have been accurately quantified for numerous species (*6, 85*). Growth rates are modelled as a function of age using the von Bertalanffy growth model:

$$\mathrm{L_a} = \mathrm{L_\infty}\left(1 - \mathrm{e}^{-\mathrm{k(a-a_0)}}\right)$$

Where $\mathrm{L_a}$ is the length at age *a,* and L_∞, *k*, and a_0 are species specific parameters from (*6*). Fecundity is modelled as a function of length as $F=aL^b$ where *a* and *b* are again species specific parameters as found in Dick *et al*. (*85*) and the probability of having reached reproductive maturity was modelled using cumulative log normal distribution centered at p=0.5 equivalent to the length at which 50% fish have reached maturity. Using these parameters, we were able to model both mortality and fecundity (**Table S16**) in 36 different rockfish species (**Fig. 5A-D**).

We also modelled generation time (**Table S17**) and reproductive value as a function of age using the method of Felsenstein (*86*) assuming a constant population size that has reached a stable age distribution. In this model l_x is the survival and b_x is the fecundity at age x and

$$\sum l_x b_x = 1$$

The generation time T is thus given by

$$\mathrm{T} = \sum \mathrm{x} l_x b_x$$

And the reproductive value is then calculated as

$$\mathrm{v_x} = (1/l_x) \sum_{\mathrm{i} \geq \mathrm{x}} \mathrm{l_x b_x}$$

We compared the correlations between various life history traits and environmental factors to understand their relative contributions to lifespan variation in rockfish (**Fig. S19**). The two most variables most predictive of lifespan were size at maturity and maximum depth, similar to previous results found in previous studies (*87*). We constructed a predictive linear model of lifespan using these two variables which explained 59% of the variance in lifespan overall. A more complex model which included an interaction term between size at maturity and maximum depth was not significantly better (ANOVA P=0.6).

The correlation of lifespan and body size in rockfish further highlights the increase in fecundity as a function of age (**Fig. 5B**). It is known that most fish exhibit non-linear scaling in reproductive output with size modelled by a power-function with an exponent β. Values of β >1 indicate non-linear reproductive output as a function of mass. Analyzing recently quantified β values from 342 marine fish (*33*) including 36 different rockfish species we find that rockfish exhibit significantly higher β values than most marine fish (**Fig. S20**, Wilcoxon, P=6.3e-6). Thus, older, larger rockfish have significantly more offspring even compared to other fish of similar sizes.

11. Lifespan and substitution rate

The rate of mitochondrial substitution accumulation has previously been shown to be negatively correlated to lifespan in rockfish (*72*). We used the Poisson regression (PR) approach to test for correlation between nuclear substitution rates and lifespan. Additionally, we tested for correlation with a linear regression model (LM) comparing the ratio of log lifespans to the ratio of substitutions between species pairs. Independent sets of species pairs were selected for multiple bootstraps such that no ancestral branches were shared among different pairs. We used the dN and dS trees generated with HyPhy for the protein coding genes (9,772 in total) in the nuclear genome and calculated the number of synonymous, nonsynonymous and total mutations for each species pair. Both PR and LM were tested for 1000 iterations with different species pairs (avoiding shared ancestries). We observed that the nuclear mutation rate was negatively correlated with max lifespan in rockfish i.e. reduced substitution rates in long lived species (**Fig. S21**) with the lowest values in non-synonymous sites.

12. Mutation spectrum analysis

We sought to understand how the mutation spectrum shifts with changes in life span. To do this we took advantage of heterozygous sites identified in each sample. These sites are new mutations that have occurred specifically in each species lineage as well as older mutations segregating in multiple species. To identify specifically new mutations and to determine which of the two alleles is derived we compared assembled genome sequence. Creating a multi-species alignment of all reference genomes is computationally challenging, so we used the *S. aleutianus* genome as a stable reference. For each species, we used the version of the reference genome scaffolded against the chromosome scale *S. aleutianus* genome using RagTag. Raw read data for each sample was aligned to its species reference genome and heterozygous sites were called and filtered as was done for Ne calculations (see above). Each reference genome was aligned to the *S. aleutianus* genome using minimap2 and alignments were filtered for 1-to-1 matches to exclude ambiguously aligned regions using singlecov2 tool from MultiZ pipeline. For each heterozygous position in each sample, we extracted the matching position in the *S. aleutianus* genome. These positions were compiled for all samples, and then matching base was identified for all species reference genomes. This produced a database with the allelic state for all species for all sites heterozygous in at least one sample.

For each sample, we then identified the neighboring bases for each heterozygous position to get the tri-nucleotide context. We filtered out heterozygous sites that were next to another heterozygous site, because in these cases the neighboring bases is uncertain. To ensure that each heterozygous site was a new mutation in the species, we required that the position have a called base in at least 20% of other species and that for the two alleles in the heterozygote, one be found in other species and one be novel and not found in other species. The novel allele was called as derived and the other as ancestral for polarizing the mutation. Together this gives us counts of each mutation type for each sample. The total number of heterozygous sites is partially determined by population size, so we normalized mutation counts into proportions (**Fig. S22**). We then used a phylogenetic generalized least squares regression to identify the effect of maximum lifespan on the proportion of each mutation independently. We found the strongest signal for *CG->*TG mutations which can be caused by CpG->TpG mutations, so we specifically tested the proportion of *CG->*TG mutations.

Our analysis of the mutation spectrum used proportions of each mutation types rather counts because the total number of heterozygous sites is highly dependent on demographic factors like population size. One potential bias is that proportions are all interdependent; an increase in one site proportion requires a decrease in others. We approached this issuing using a sequential test approach by Harris and Pritchard (*39*). Briefly, we took our initial estimates of the effect of lifespan on each mutation type and ordered mutation types by p-value. We then sequential tested each mutation, while removing the counts of each mutation type with a lower p-value from proportion calculations. In simple terms, for example we tested if the proportion of AAT->ACT mutations was still affected by lifespan after excluding counts for CCG->CTG mutations and repeated this for each mutation type sequentially by p-value. With this analysis, the effect on *CG->*TG was still significant, although some mutation types that were previously less significant but passed the $q < 0.05$ threshold in the original analysis no longer passed the threshold (**Fig. S23**).

Supplementary Figures

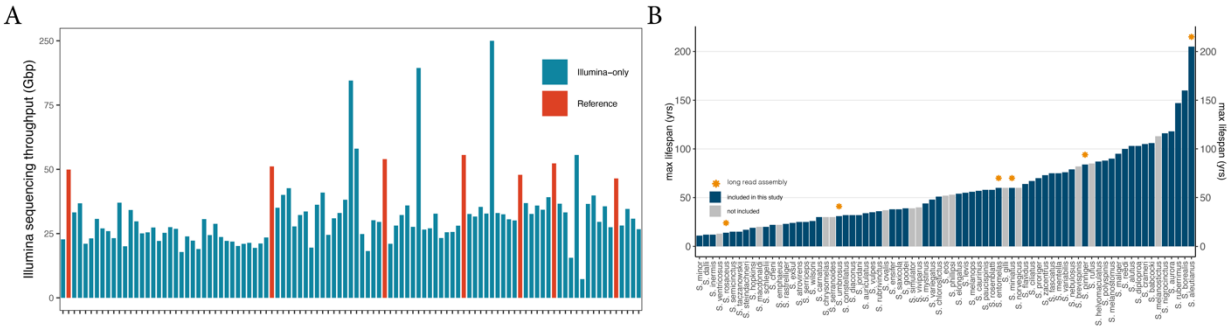

Figure S1: Sequencing depth and maximum lifespan of rockfish. (A) The sequence throughput from Illumina paired-end sequencing for the 102 rockfish samples. The samples whose genomes were sequenced further with Pacific Bioscience long-reads are highlighted in red. Multiple individuals were sequenced for *S. crameri*, *S. entomelas*, *S. mystinus*, *S. oculatus*, *S. rosaceus*, *S. ruberrimus*, *S. rubrivinctus* and *Sebastolobus alascanus.* (B) Lifespans for 75 different rockfish species. Individuals in blue were sequenced and assembled in this study. Stars indicate long-read assemblies.

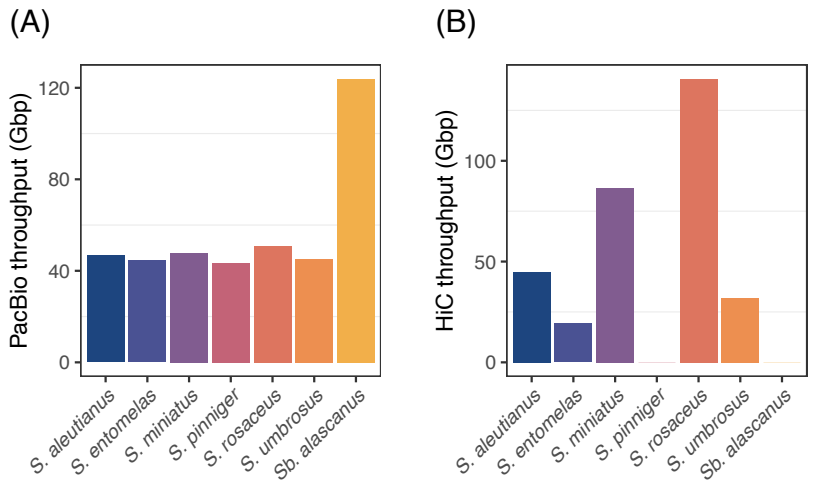

Figure S2: Genome scaffolding sequence effort for reference genomes. Sequencing throughput from (A) PacBio genome sequencing with sequel-II and (B) Illumina paired-end sequencing for DNase-I based HiC libraries. PacBio genome coverages were above 45X for all the seven reference genomes and HiC coverages were above 30X for the five sequenced samples.

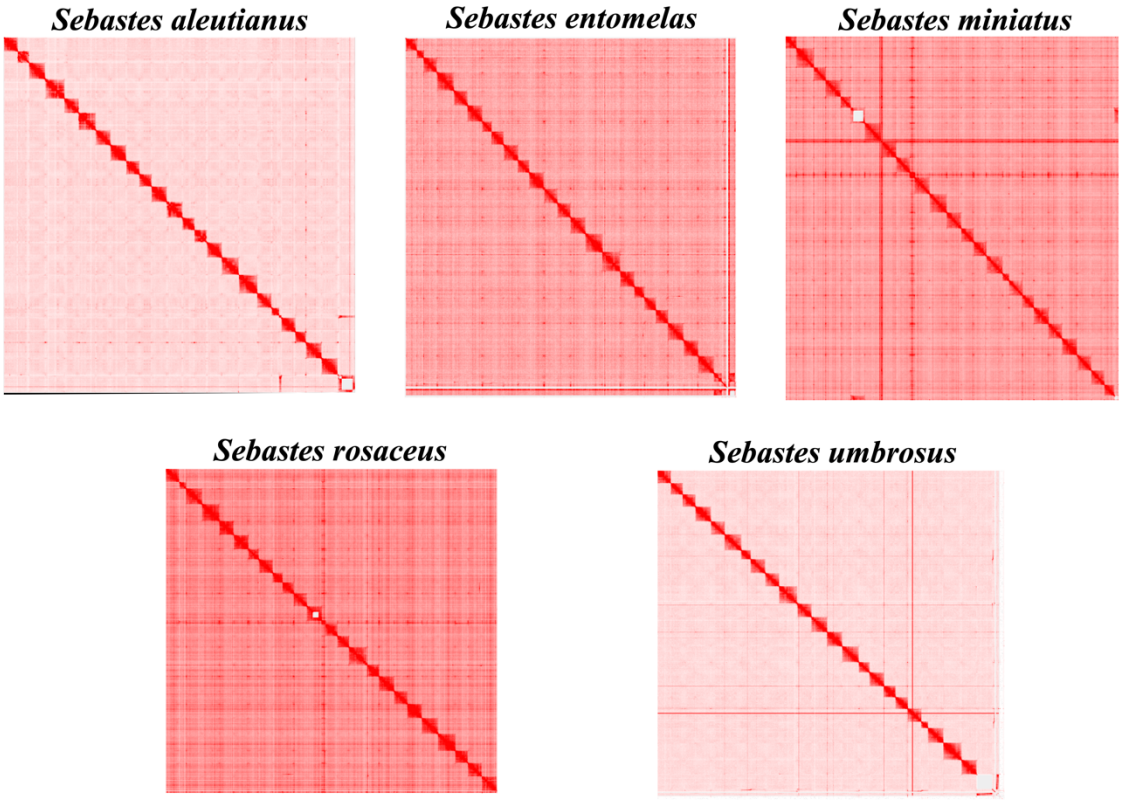

Figure S3: Chromosome scale scaffolds of five *Sebastes* genomes. 24-26 distinct blocks were visualized in HiC interaction matrixes using Juicebox tool. Color depth indicates the strength of interactions found using HiC sequencing.

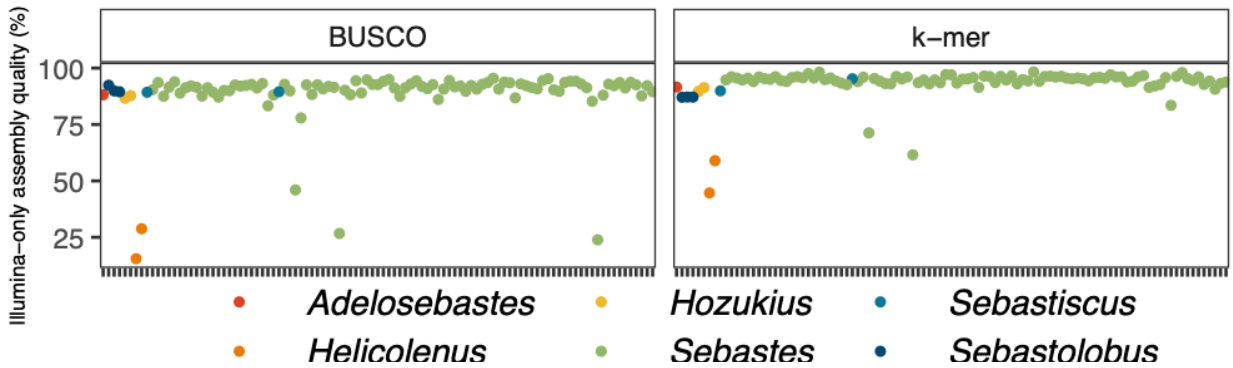

Figure S4: K-mer based and BUSCO genomic completeness metrics for *de novo* assemblies of 102 samples generated from only Illumina paired-end data. The genome assemblies with lower completeness metrics were highly fragmented.

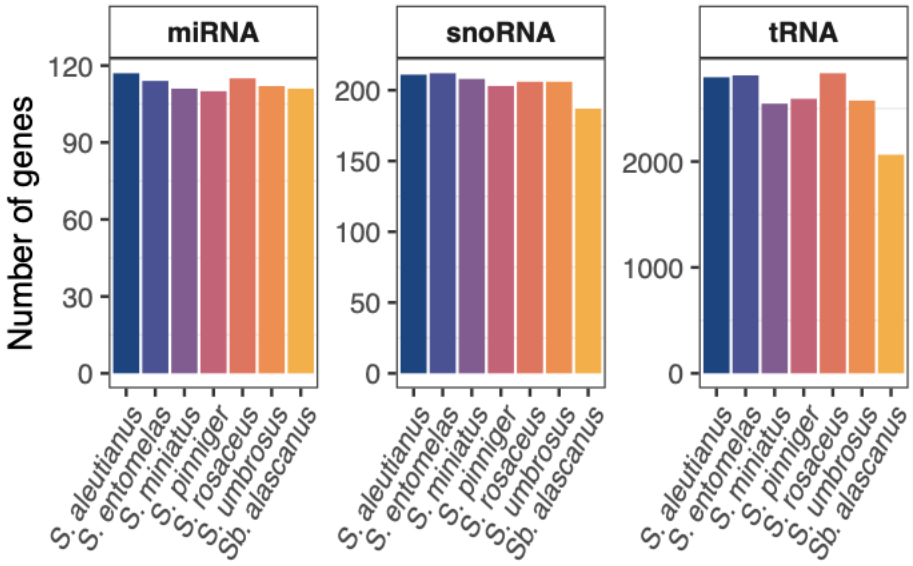

Figure S5: The number of noncoding RNAs (miRNA, snoRNA and tRNA) identified across the seven reference genomes.

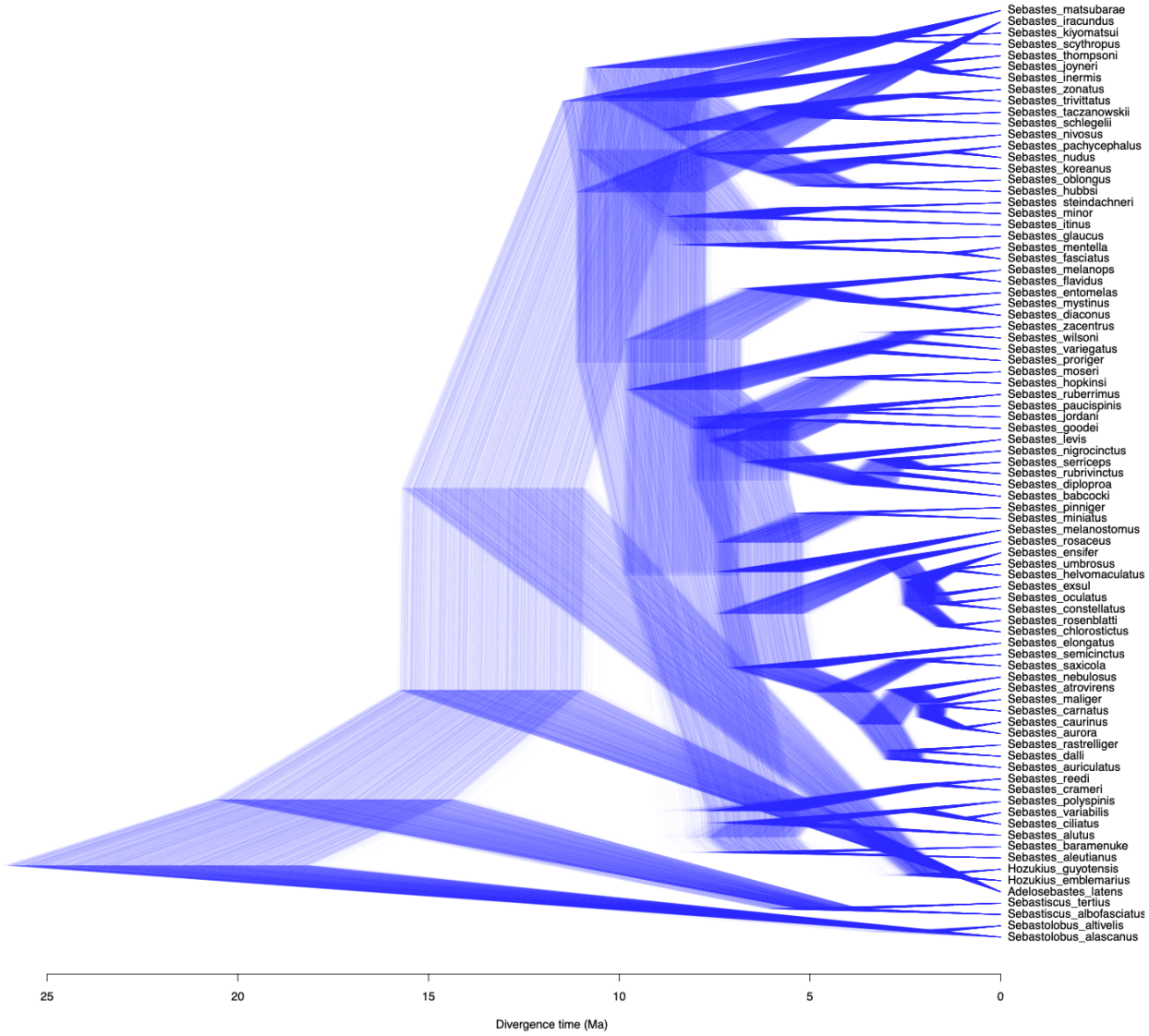

Figure S6: Densitree of species within Sebastidae with calibrated divergence times. Divergence times were taken from the posterior trees of BPPR with a set species topology.

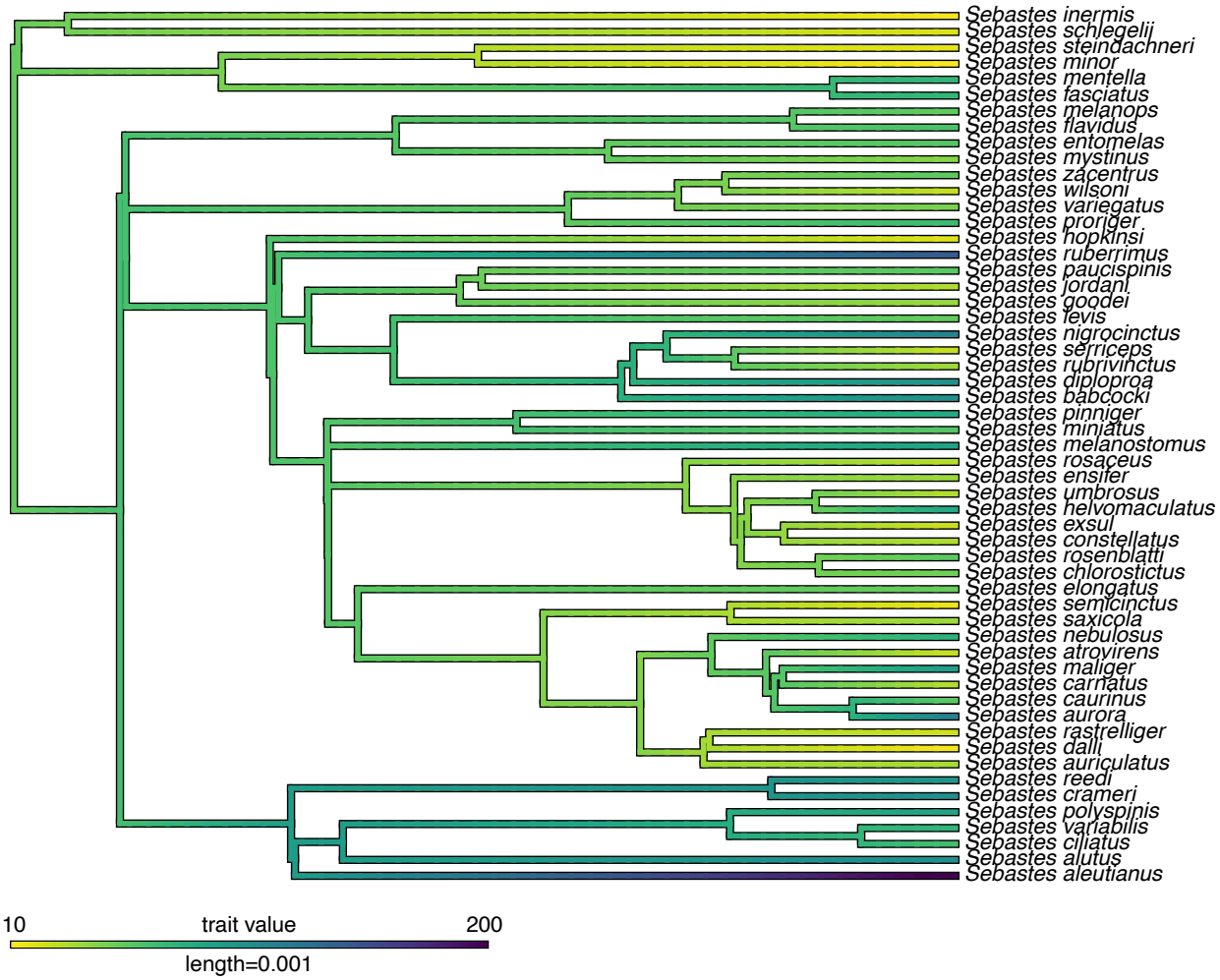

Figure S7: Continuous trait character mapping analysis of lifespan in rockfish shows multiple independent gains of long lifespan.

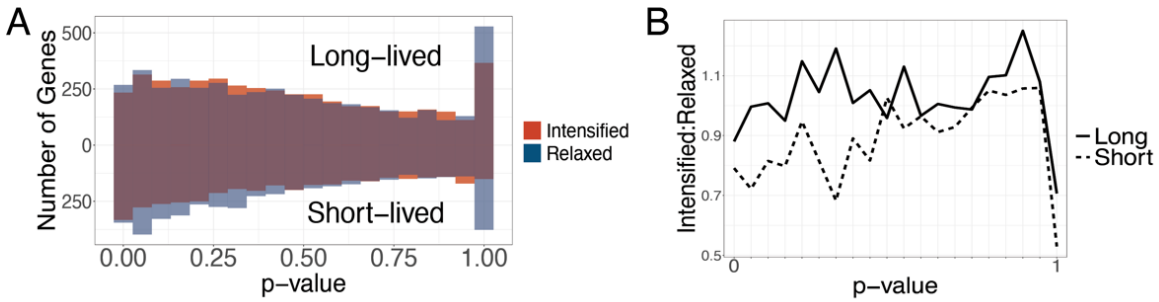

Figure S8: Results of RELAX analysis for short- and long-lived species. (A) The counts of genes identified as having relaxed or intensified selection in either short- or long-lived species sorted by p-value. P-values are derived from the likelihood ratio test in RELAX. (B) The ratio of intensified to relaxed selection of genes, sorted by p-value.

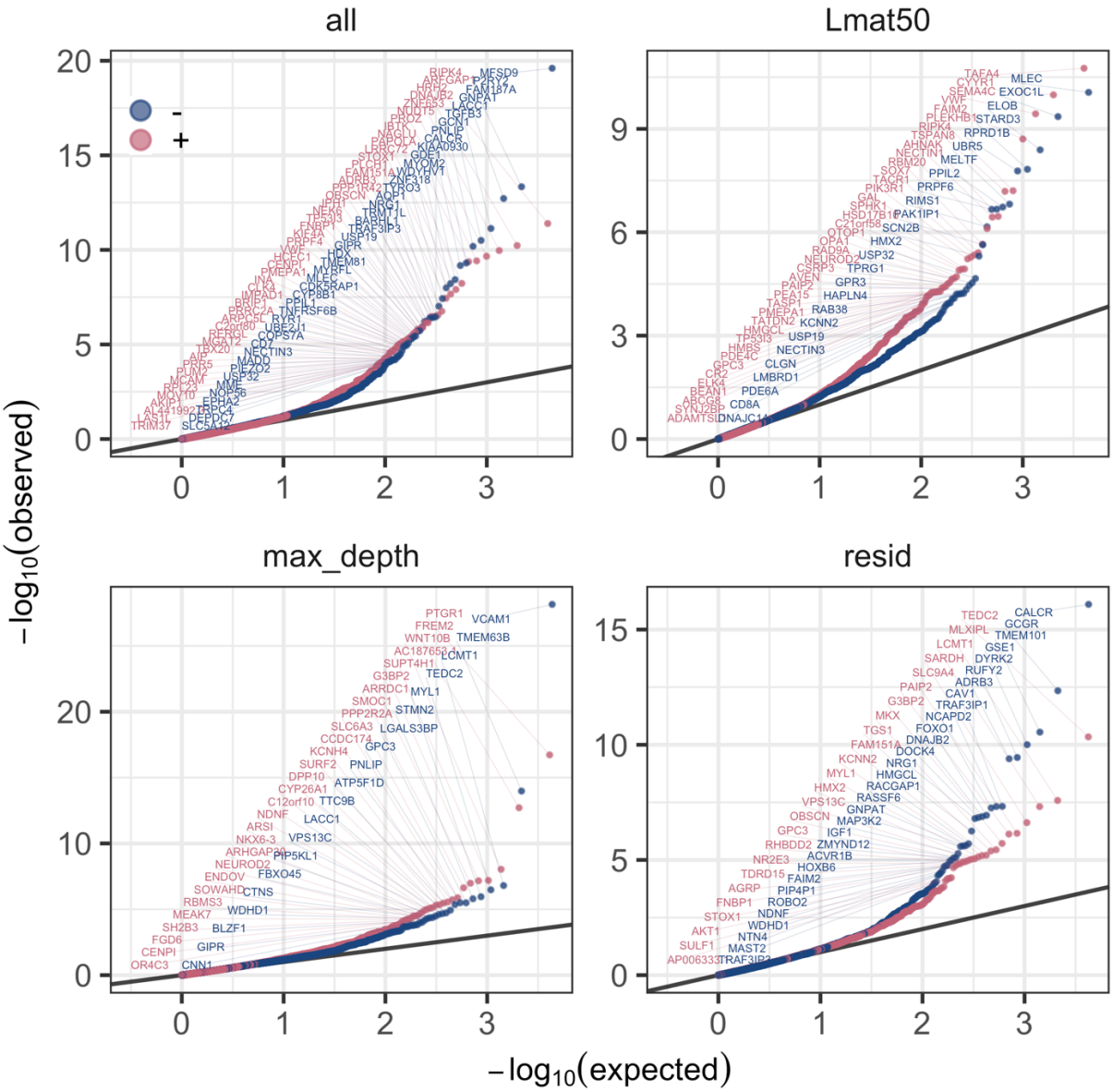

Figure S9: Genes with relative evolutionary rates correlated with lifespan (all), size at maturity (Lmat50), maximum depth, or the residual of the linear model predicting lifespan from maximum depth and size at maturity. The -$\log_{10}$(observed P-value) is plotted against the -$\log_{10}$(expected P-value) for the correlation between the relative evolutionary rate of genes and individual phenotypes. See full list of genes and P-values in Table S14. Blue color indicates genes where lifespan is negatively correlated with evolutionary rate, while red indicates a positive correlation.

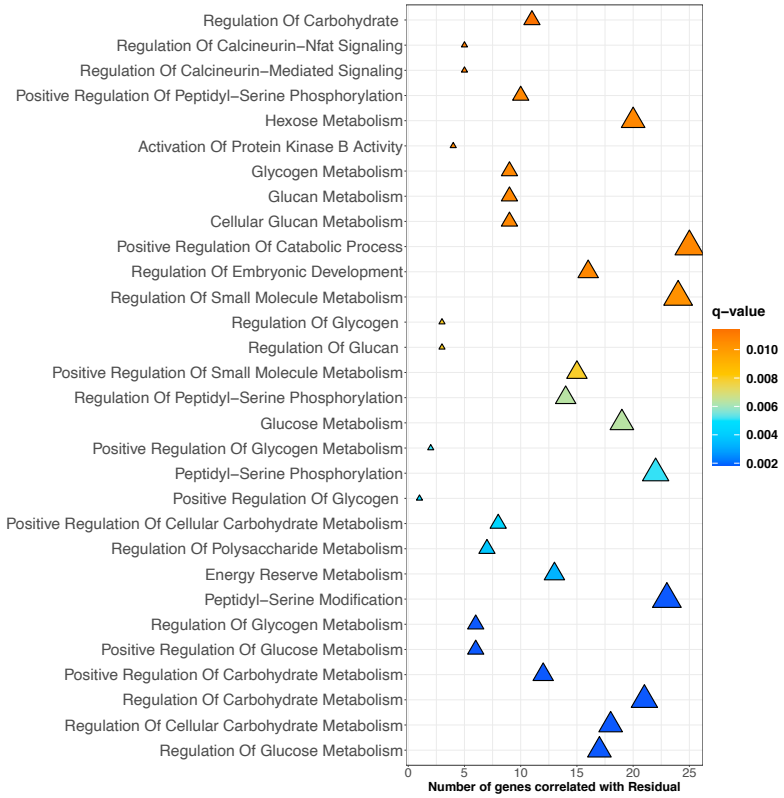

Figure S10: Overrepresented GO terms for genes significantly associated with the residual of the linear model predicting lifespan from maximum depth and size at maturity. The size of the triangle reflects the number of genes correlated.

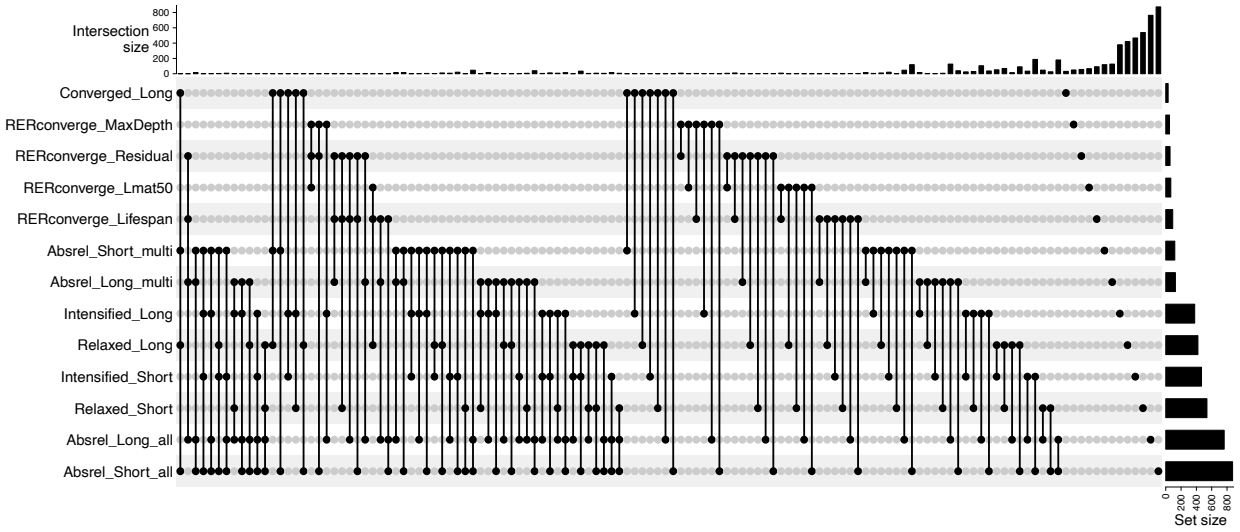

Figure S11: The overlap between different methods of detecting genes associated with lifespan. Converged_Long: Convergent amino acids in long-lived species. RERconverge_*: RERconverge analysis with different traits and the residual of the linear model predicting lifespan from maximum depth and size at maturity. Absrel_*_multi: Positive selection in either long- or short-lived species using a branch-site model in aBSREL, where positive selection occurs in more than one species. Absrel_*_all: Positive selection in either short- or long-lived species using a branch-site model in aBSREL, where positive selection occurs in one or more species. Intensified_*: Intensified selection identified from RELAX in short- or long-lived species. Relaxed_*: Relaxed selection identified from RELAX in short- or long-lived species.

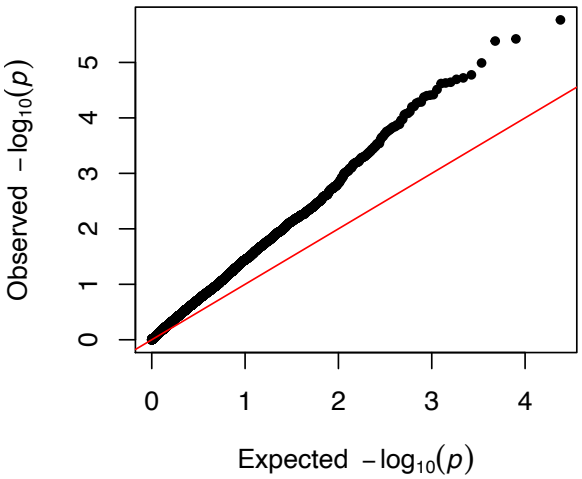

Figure S12: QQ-plot of expected and observed p-values for PGLS comparing read depth to maximum lifespan. Values above the red line suggest an excess of low p-values. Observed p-values are derived from phylogenetic least squares test and expected p-values are derived from a normal distribution.

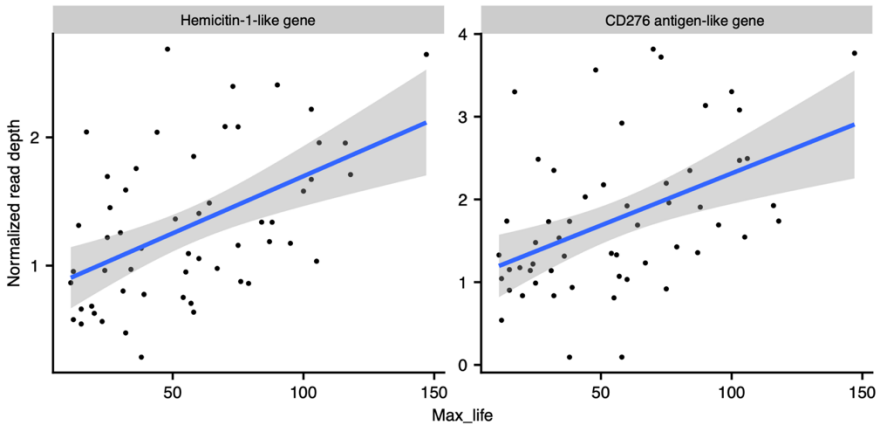

Figure S13: The relationship of normalized read depth and maximum lifespan for two butyrophilin-like genes. Regression line does not account for phylogenetic relatedness and is used for illustrative purposes.

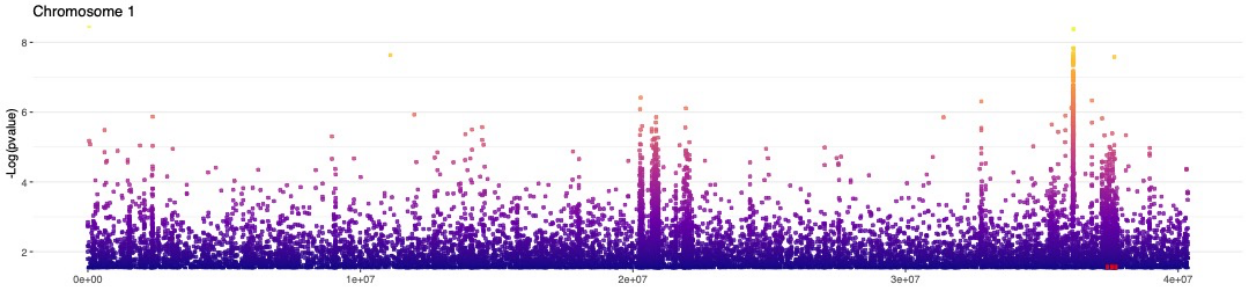

Figure S14: Phylogenetically corrected correlations between read depth and lifespan, using *S. umbrosus* as a reference genome. The red bars indicate butyrophilin genes. Points with a negative correlation between read depth and life span and those with log(p) < 1.5 were not plotted.

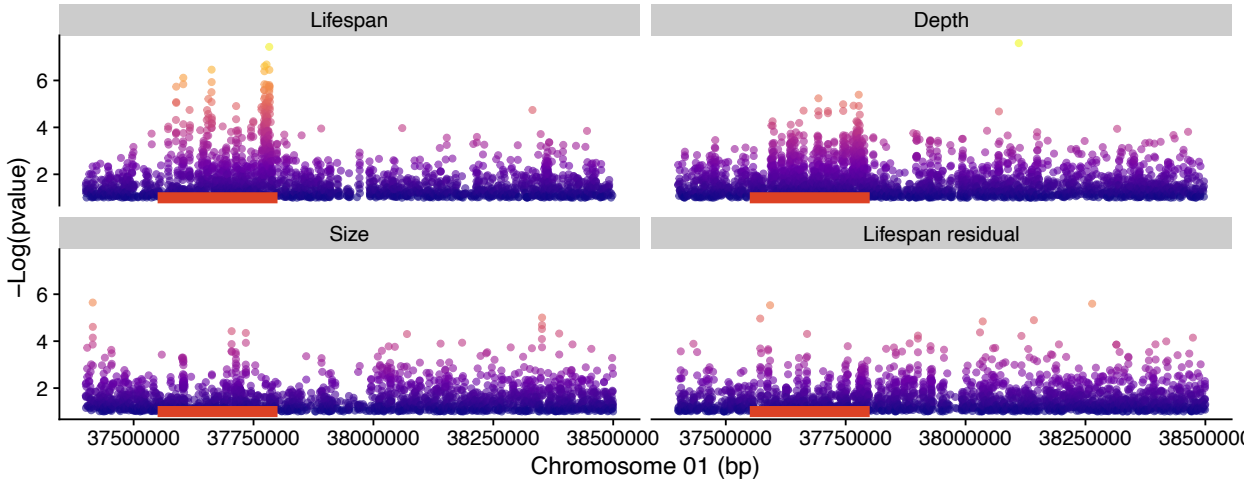

Figure S15: Phylogenetically corrected correlations between read depth and species features associated with lifespan. The orange bar highlights the region of low p-values found when testing maximum lifespan containing butyrophilin genes.

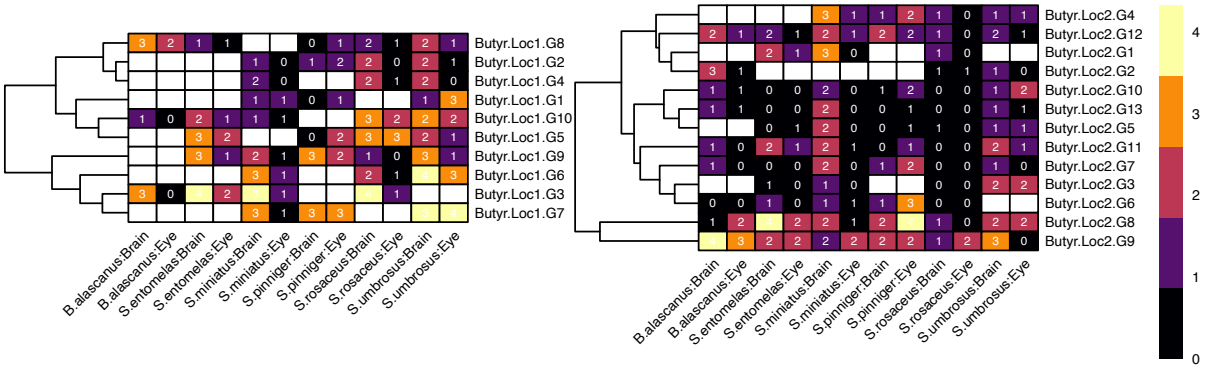

Figure S16: Heatmap of gene expression (log2(TPM+1)) across the two butyrophilin loci (Chr1 and Chr22 respectively). The column names represent the sample species and tissue. The cells are blank for species without a gene ortholog in the respective gene group (represented by rows).

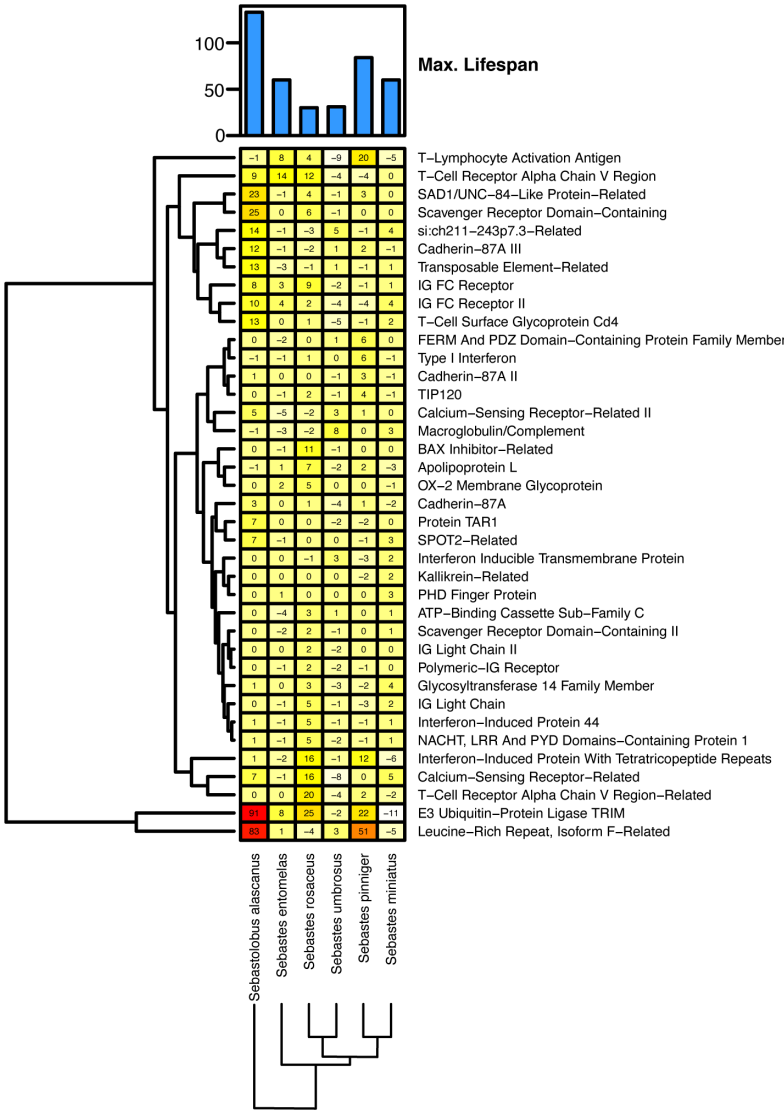

Figure S17: Heatmap of gene family changes across the *Sebastes* clade including the *Sebastolobus* outgroup. The rows are clustered based on Euclidean distance and the columns are ordered by phylogeny. The top bar plot shows the maximum lifespan (in years) in these species.

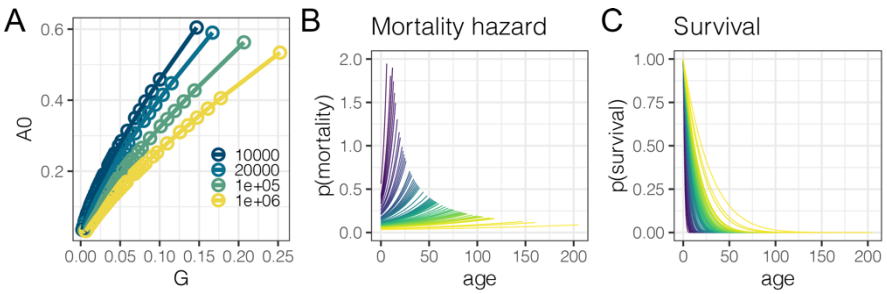

Figure S18: Life history parameter estimates for rockfish. A) Estimated Gompertzian parameters for 74 rockfish species for various values of N and corresponding mortality hazard (B) and survival function (C) for N=1e6.

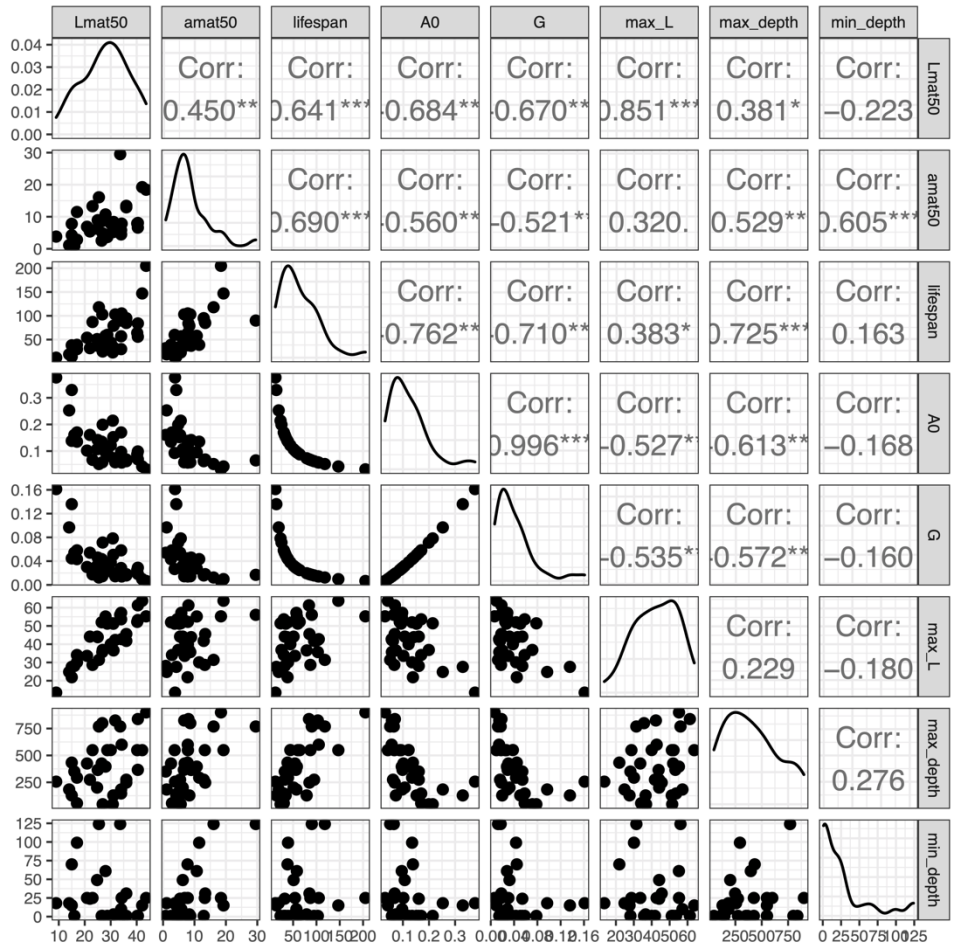

Figure S19: Correlation among different life history and environmental parameters for rockfish. *Lmat50* is the average length at which 50% of individuals have reached maturity, *amat50* is the average age at which 50% of individuals have reached maturity, *A0* and *G* are the Gompertzian parameters estimated for different species, *max_L* is the maximum length, *min_depth* and *max_depth* are the minimum and maximum depths at which species are found respectively, *lifespan* is the maximum lifespan of species.

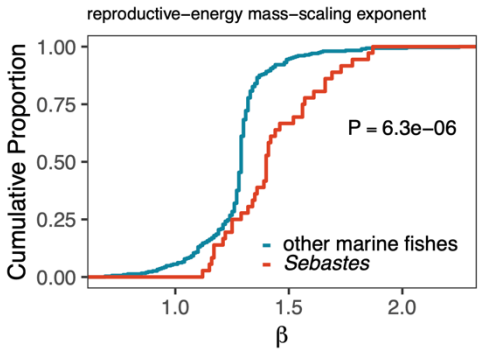

Figure S20: Cumulative distribution of reproductive-energy mass-scaling exponents for marine fishes with rockfishes highlighted in red. Data from Barneche *et al.* 2018. (*33*)

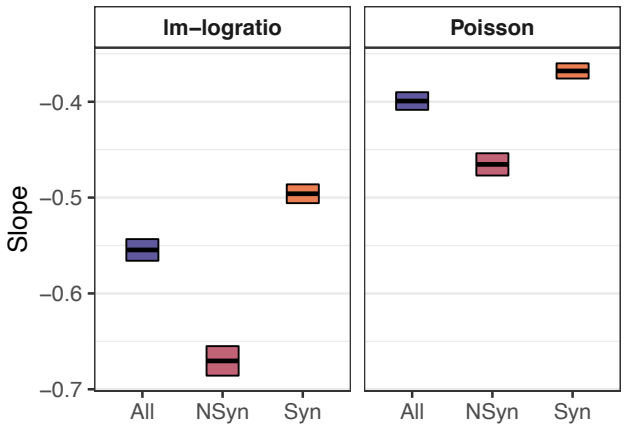

Figure S21: The relationship between evolutionary rate and maximum lifespan. The correlation between lifespan and substitution rate in single copy orthologous nuclear genes compared using both a linear model and Poisson regression. The signal for all substitutions (All) is contrasted with synonymous (Syn) and non-synonymous (NSyn) mutations.

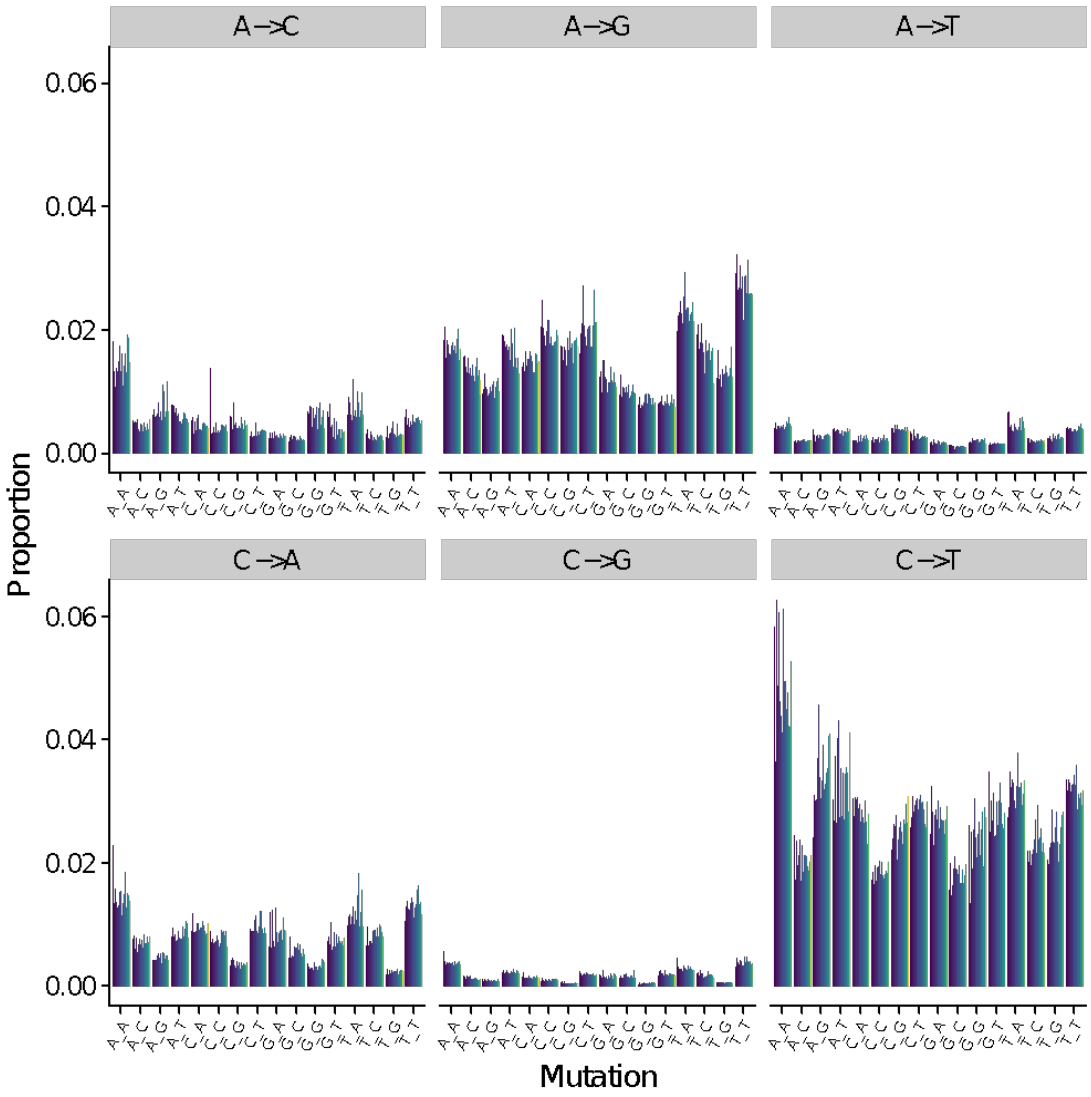

Figure S22: The mutation spectrum for all *Sebastes* samples. The x-axis labels show the 5' and 3' neighboring basepair. Samples are ordered and colored by maximum lifespan.

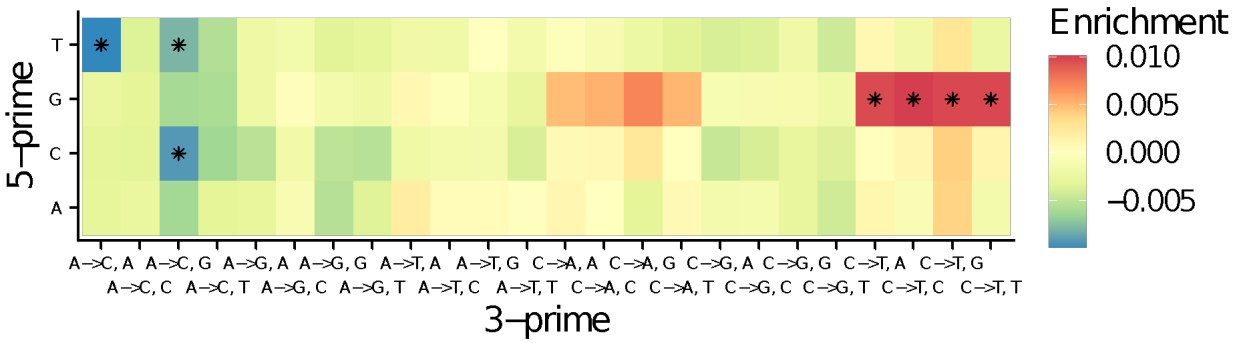

Figure S23: The effect of maximum lifespan on different mutation types. A phylogenetic generalized least squares test with sequential data removal for the relationship between mutation type proportion and maximum lifespan (*39*). Asterisks indicate q-value < 0.05. Color represents effect size estimate.

Supplementary Tables

Table S1: Reference free quality metrics for the high-quality rockfish genome assemblies. The WTDBG2 assemblies have lower base quality values (<30), however all the seven genomes have high BUSCO completeness.

Species	*de novo* Assembler	Quality value (QV)	k-mer error rate	k-mer completeness (%)	BUSCO completeness (%)
S. aleutianus	FALCON	40.84	8e-05	93.60	98.4
S. entomelas	WTDBG2	31.22	8e-04	95.94	97.9
S. miniatus	FALCON	43.35	5e-05	95.56	98.3
S. pinniger	FALCON	44.26	4e-05	95.73	97.7
S. rosaceus	FALCON	40.5	9e-05	90.75	97.9
S. umbrosus	FALCON	39.62	1e-04	92.24	97.1
Sebastolobus alascanus	WTDBG2	33.63	4e-04	85.20	98.2

Table S2: Reference species used to guide the scaffolding of de novo genome assemblies generated from Illumina paired-end sequencing.

Sample species	Reference species	Sample species	Reference species
Adelosebastes latens	*Sebastolobus alascanus*	*Sebastes matsubarae*	*Sebastes aleutianus*
Helicolenus avius	*Sebastes aleutianus*	*Sebastes melanops*	*Sebastes entomelas*
Helicolenus hilgendorfii	*Sebastes aleutianus*	*Sebastes melanostomus*	*Sebastes pinniger*
Hozukius emblemarius	*Sebastes aleutianus*	*Sebastes mentella*	*Sebastes aleutianus*
Hozukius guyotensis	*Sebastes aleutianus*	*Sebastes minor*	*Sebastes aleutianus*
Sebastes alutus	*Sebastes aleutianus*	*Sebastes moseri*	*Sebastes pinniger*
Sebastes atrovirens	*Sebastes pinniger*	*Sebastes mystinus*	*Sebastes entomelas*
Sebastes auriculatus	*Sebastes pinniger*	*Sebastes nebulosus*	*Sebastes pinniger*
Sebastes aurora	*Sebastes pinniger*	*Sebastes nigrocinctus*	*Sebastes rosaceus*
Sebastes babcocki	*Sebastes rosaceus*	*Sebastes nivosus*	*Sebastes aleutianus*
Sebastes baramenuke	*Sebastes aleutianus*	*Sebastes nudus*	*Sebastes aleutianus*
Sebastes borealis	*Sebastes aleutianus*	*Sebastes oblongus*	*Sebastes aleutianus*
Sebastes carnatus	*Sebastes pinniger*	*Sebastes oculatus*	*Sebastes rosaceus*
Sebastes caurinus	*Sebastes pinniger*	*Sebastes pachycephalus*	*Sebastes aleutianus*
Sebastes chlorostictus	*Sebastes umbrosus*	*Sebastes paucispinis*	*Sebastes rosaceus*
Sebastes cheni	*Sebastes aleutianus*	*Sebastes polyspinis*	*Sebastes aleutianus*
Sebastes ciliatus	*Sebastes aleutianus*	*Sebastes proriger*	*Sebastes entomelas*
Sebastes constellatus	*Sebastes rosaceus*	*Sebastes rastrelliger*	*Sebastes pinniger*
Sebastes crameri	*Sebastes entomelas*	*Sebastes reedi*	*Sebastes entomelas*
Sebastes dalli	*Sebastes pinniger*	*Sebastes rosenblatti*	*Sebastes umbrosus*
Sebastes diaconus	*Sebastes entomelas*	*Sebastes ruberrimus*	*Sebastes pinniger*
Sebastes diploproa	*Sebastes rosaceus*	*Sebastes rubrivinctus*	*Sebastes rosaceus*
Sebastes elongatus	*Sebastes pinniger*	*Sebastes saxicola*	*Sebastes pinniger*
Sebastes ensifer	*Sebastes umbrosus*	*Sebastes schlegelii*	*Sebastes aleutianus*
Sebastes exsul	*Sebastes rosaceus*	*Sebastes scythropus*	*Sebastes aleutianus*
Sebastes fasciatus	*Sebastes aleutianus*	*Sebastes semicinctus*	*Sebastes pinniger*
Sebastes flavidus	*Sebastes entomelas*	*Sebastes serriceps*	*Sebastes rosaceus*
Sebastes glaucus	*Sebastes aleutianus*	*Sebastes steindachneri*	*Sebastes entomelas*
Sebastes goodei	*Sebastes rosaceus*	*Sebastes taczanowskii*	*Sebastes aleutianus*
Sebastes helvomaculatus	*Sebastes umbrosus*	*Sebastes thompsoni*	*Sebastes aleutianus*
Sebastes hopkinsi	*Sebastes pinniger*	*Sebastes trivittatus*	*Sebastes aleutianus*
Sebastes hubbsi	*Sebastes aleutianus*	*Sebastes variabilis*	*Sebastes aleutianus*
Sebastes inermis	*Sebastes aleutianus*	*Sebastes variegatus*	*Sebastes entomelas*
Sebastes iracundus	*Sebastes aleutianus*	*Sebastes vulpes*	*Sebastes aleutianus*
Sebastes itinus	*Sebastes aleutianus*	*Sebastes wilsoni*	*Sebastes entomelas*
Sebastes jordani	*Sebastes rosaceus*	*Sebastes zacentrus*	*Sebastes entomelas*

Sebastes joyneri	*Sebastes aleutianus*	*Sebastes zonatus*	*Sebastes aleutianus*
Sebastes kiyomatsui	*Sebastes aleutianus*	*Sebastiscus albofasciatus*	*Sebastes aleutianus*
Sebastes koreanus	*Sebastes aleutianus*	*Sebastiscus tertius*	*Sebastes aleutianus*
Sebastes levis	*Sebastes rosaceus*	*Sebastolobus altivelis*	*Sebastolobus alascanus*
Sebastes maliger	*Sebastes pinniger*		

Table S3: Gene models for protein coding genes generated using the BRAKER2 and Funannotate pipelines. The final gene models were retained based on similarity to known vertebrate peptide sequences.

Type of evidence	*S. aleutianus*	*S. entomelas*	*S. miniatus*	*S. pinniger*	*S. rosaceus*	*S. umbrosus*	*Sebastolobus alascanus*
#Protein alignments	569,960	569,136	1,280,015	565,111	573,760	569,920	452,191
#Transcript alignments	1206,721	1,212,794	1,210,028	1,200,986	1,196,801	1,190,554	1,034,629
Augustus models	191,932	143,906	178,542	145,124	168,690	152,098	153,902
Genemark models	68,728	60,656	63,202	59,606	74,258	63,128	65,206
BUSCO models	3,187	3,176	3,184	3,159	3,196	3,164	3,198
Total EVM models	61,226	37,538	67,037	43,784	72,685	63,340	46,470
Final models	28,393	24,278	27,109	25,018	26,171	24,115	26,729

Table S4: The number of genes retained during filtering for species phylogeny.

Stage	Full set	*Sebastes* set
Initial BUSCO	877	949
Alignment length filter	659	711
Distance filter	626	670

Table S5: Genes with DNA repair related functions with evidence of positive selection in long-lived species.

Gene	Description	Function/Disease	Species
NBN1	Bibrin protein	TP53 activity; Chromosomal instability syndrome	*Sebastes ruberrimus; Sebastes crameri*
SLX4	SLX4 structure-specific endonuclease	DNA lesion repair	*Sebastes nigrocinctus; Sebastes crameri*
RBBP8	RB binding protein 8	Retinoblastoma; Pancreatic cancer	*Sebastes ruberrimus; Sebastes crameri*
FEN1	Flap structure-specific endonuclease 1	DNA synthesis; Cell cycle	*Sebastes alutus; Sebastes crameri*
MCM6	Minichromosome maintenance complex component 6	DNA replication initiation; Craniopharyngioma	*Sebastes nigrocinctus; Sebastes ruberrimus*
RAD51AP1	RAD51 associated protein-1	RNA binding; Anemia	*Sebastes ruberrimus*
WRAP53	WD repeat containing antisense to TP53	p53 antisense transcript; chromosome maintenance	*Sebastes aleutianus*
XRCC1	X-ray repair cross complementing 1	Germ cells recombination; DNA processing; Cardiac cancer	*Sebastes aurora*
BARD1	BRCA1 associated ring domain 1	TP53 pathway; Breast and ovarian cancer	*Sebastes ruberrimus*
SWI5	SWI5 homologous recombination repair protein	Recombination repair; Ceroid lopofusconosis	*Sebastes crameri*
KDM2A	Lysin demethylase 2A	Ubiquitination; Epigenetic silencing; F-box protein	*Sebastes alutus*
FAN1	FANCD2 and FANCI associated nuclease 1	Faconi anemia pathway; Interstitial nephritis	*Sebastes aleutianus*
DCLRE1B	DNA cross-link repair 1B	Interstrand crosslinking; Faconi anemia pathway	*Sebastes alutus*
NIPBL	Nipped B-like protein	Chromosome condensation; Microcephaly	*Sebastes nigrocinctus*
WAS	WASP actin nucleation promoting factor	PAK pathway; Integrin pathway; Thrombocytopenia	*Sebastes alutus*
SPIRE1	Spire type actin nucleation factor 1	TGF-beta pathway; Actin organization; Osteogenesis	*Sebastes alutus*

Table S6: Genes with significant convergent AA (q<0.05) in long-lived species compared to short-lived species across *Sebastes*. P-values were calculated by randomization with 1000 permutations and multiple-testing correction performed using the Benjamini-Hochberg procedure.

Gene	Description	Function/Disease	Comment
FPR1	fMet-Leu-Phe receptor		PSG in short-lived; Relaxed in long-lived
SERPINH1	Serpin H1	Collagen biosynthesis; Rheumatoid arthritis	Probably lineage convergence
PI4K2B	Phosphatidylinositol 4-kinase type 2-beta	T cell activation; Macrocephaly	-
STARD9	Star-related lipid transfer protein 9	ATPase activity	-
ATL2	ADAMTS-like protein 2	GTPase activity; Spondylolysis	Convergence in short lived
GRAM4	GRAM domain-containing protein 4	Apoptosis	-
WDR1	WD-repeat containing protein 1	Platelet activation; Immunodeficiency	Relaxed in short lived; Convergent site in WD40 domain
SALL3	Sal-like protein 3	Oncogenesis; CpG island methylation	-
FHDC1	FH2 domain-containing protein 1	Golgi ribbon formation	-
TM104	Transmembrane protein 104	Transporter	Convergent site in SdaC domain
ANGE1	Protein angel homolog 1	Arrhythmogenic right ventricular dysplasia; Glaucoma	-
NDUFS3	Ubiquinone iron-sulphur protein 3	Mito complex I deficiency; Neurodegeneration	Nuclear encoded mitochondrial gene
EDRF1	Erythroid differentiation related factor 1	DNA binding; Erythroid cell differentiation	-
TEP1	Telomerase protein component 1	RNA binding; Telomerase activity	-
APLF	Aprataxin and PNK-like factor	GPCR pathway; DNA double-stranded break repair	-
RTN4	Reticulon 4	GPCR signaling; Demyelinating disease	-
SHOC2	SHOC2 leucine rich repeat protein	RAS/ERK MAP kinase signaling	Intensified in long lived; Relaxed in short lived

SLC12A7	Solute carrier family 12 member	Phosphorylated in response to DNA damage; Insulin receptor recycling	Probably lineage convergence; Converged site in permease domain
PLCH2	Phosphoinositide phospholipase C-eta-2	TP53 pathway; cardiolipin synthesis	-
TSN6	Tetraspinin-6	Erythroid progenesis; Myeloid malignancies	Converged site in Tetraspannin domain
CISH	Cytokine-inducible SH2-containing protein	IL-2 signaling; Prolactin signaling; Microtubule binding	-
FMN1	Sander lucioperca formin 1	E-cadherin signaling; cell adhesion	-
TXLNB	Beta-taxilin	Syntaxin binding; Blepharoconjunctivitis	Converged site in SMC_prokB domain
XN536	Zinc finger protein 536	DNA binding; Ectoderm differentiation	-
FOXP2	Forkhead box P2	Wnt/Notch signaling; Neural mechanism	Converged site in FOXP-CC domain
CNGB3	Cyclic nucleotide-gated cation channel beta-3	Cone photoreceptors; Achromatopsia 3	Converged site in PLN03192 domain
TRIM27	RING-type E3 ubiquitin transferase	DNA binding; MAPK cascade; Male germ cell differentiation	Converged site in SPRY_PRY_C-I_1 domain
PSBG9	Pregnancy-specific beta-1-glycoprotein 9	Hemostasis; Platelet-fibrinogen interactions	Converged site in IG domain

Table S7: Aging-related genes in Rockfish (RF) linked to longevity in other vertebrates.

Other species	Gene	Activity in rockfish(es)	Comment
Bowhead whale	ERCC1	Intensified selection in long-lived RF	DNA repair gene
	FEN1	PSG in multiple RF (both long- and short-lived); Interacts with PCNA which is a PSG in whales	DNA repair gene
	FGFR1	Relaxed in long-lived RF	Immune response
	FOXO3	PSG in multiple short-lived RF	Primordial follicle development
	ST13	Intensified selection in long-lived RF	Cellular senescence
Naked mole rat	CASQ2	Relaxed in long-lived RF	Cardiac functioning
	DEDD	Relaxed in short-lived RF	Apoptosis
	MTNR1A	Intensified selection in short-lived; Relaxed in long-lived RF	Circadian rhythm
	RPGRIP1L	Converged sites in short-lived RF	Cell signaling
	SCN2B	Correlated to size at maturity; Relaxed in short-lived RF	Cell migration
	TEP1	Converged sites in long-lived RF	Telomere maintenance
	TERF1	Intensified selection in long-lived RF	Telomere maintenance
	TERT	Relaxed selection in long-lived RF	Telomere maintenance
Giant tortoise	CHIA	Intensified selection in short-lived RF	Immune response
	GSK3A	Relaxed selection in both short- and long-lived RF	Cell migration
	PTPN11	Relaxed selection in long-lived RF	Cell signaling
	SMAD4	Relaxed selection in short-lived RF	Chromatin binding

Table S8: Genes with copy numbers associated with lifespan in *Sebastes*.

Gene	*p*	*q*	*ENSEMBL hit*	*ENSEMBL description*
FUN_031460	1.7E−06	0.011	ENSLCRP00005010755.1	nuclear factor 7, ovary-like [Source:NCBI gene;Acc:113745853]
FUN_017515	3.8E−06	0.011	ENSLCAP00010027394.1	-
FUN_055709	4.1E−06	0.011	ENSLBEP00000027370.1	complement C3-like [Source:NCBI gene;Acc:110000750]
FUN_031471	1.0E−05	0.019	ENSONIP00000068480.1	-
FUN_028439	1.7E−05	0.019	ENSSORP00005011768.1	homeobox protein MOX-2-like [Source:NCBI gene;Acc:115427936]
FUN_060007	1.9E−05	0.019	ENSSDUP00000013461.1	cullin-associated NEDD8-dissociated protein 1-like [Source:NCBI gene;Acc:111224398]
FUN_010492	2.0E−05	0.019	ENSCGOP00000055632.1	-
FUN_060004	2.3E−05	0.019	ENSLCAP00010022377.1	cullin-associated and neddylation-dissociated 2 (putative) [Source:ZFIN;Acc:ZDB-GENE-060503-645]
FUN_015045	2.3E−05	0.019	ENSSAUP00010057799.1	tripartite motif-containing protein 16-like [Source:NCBI gene;Acc:115593625]
FUN_015024	2.4E−05	0.019	ENSSAUP00010057807.1	tripartite motif-containing protein 16-like [Source:NCBI gene;Acc:115593616]
FUN_056576	3.1E−05	0.022	ENSATEP00000023305.1	dkey-30j10.5 [Source:ZFIN;Acc:ZDB-GENE-091118-81]
FUN_017562	3.8E−05	0.022	ENSLCAP00010027394.1	-
FUN_060100	3.9E−05	0.022	ENSSDUP00000020085.1	butyrophilin subfamily 3 member A2-like [Source:NCBI gene;Acc:111239169]
FUN_017514	4.0E−05	0.022	ENSLCAP00010027394.1	-
FUN_031462	4.2E−05	0.022	ENSLCRP00005043887.1	-
FUN_016067	5.2E−05	0.024	ENSLACP00000004102.1	-
FUN_008242	5.2E−05	0.024	ENSGACP00000000743.1	-
FUN_031464	5.4E−05	0.024	ENSMMDP00005002174.1	-
FUN_031466	6.2E−05	0.025	ENSLCRP00005010462.1	-
FUN_045937	6.3E−05	0.025	ENSLCAP00010002087.1	-
FUN_060099	7.7E−05	0.028	ENSLCAP00010054796.1	butyrophilin-like protein 1 [Source:NCBI gene;Acc:108873900]
FUN_030618	8.3E−05	0.028	ENSATEP00000030907.1	basic helix-loop-helix family member e41 [Source:NCBI gene;Acc:113165857]
FUN_043893	8.4E−05	0.028	ENSSLDP00000015741.1	nuclear factor 7, brain-like [Source:NCBI gene;Acc:111663055]
FUN_060001	8.7E−05	0.028	ENSCGOP00000002308.1	polymeric immunoglobulin receptor-like [Source:NCBI gene;Acc:115008767]
FUN_060011	1.1E−04	0.033	ENSCGOP00000002308.1	polymeric immunoglobulin receptor-like [Source:NCBI gene;Acc:115008767]
FUN_029278	1.1E−04	0.033	ENSACLP00000010891.1	dkey-30j10.5 [Source:ZFIN;Acc:ZDB-GENE-091118-81]
FUN_059096	1.3E−04	0.036	ENSMMDP00005002174.1	-
FUN_031393	1.3E−04	0.036	ENSLCRP00005043887.1	-
FUN_012252	1.3E−04	0.036	ENSCGOP00000036483.1	-
FUN_037699	1.4E−04	0.036	ENSSAUP00010057745.1	sialic acid-binding Ig-like lectin 5 [Source:NCBI gene;Acc:115566902]

FUN_044604	1.5E−04	0.036	ENSLCAP00010058946.1	zinc metalloproteinase-disintegrin-like brevilysin H2a [Source:NCBI gene;Acc:108878417]
FUN_010689	1.5E−04	0.036	ENSCGOP00000009865.1	up-regulator of cell proliferation-like [Source:NCBI gene;Acc:115023890]
FUN_055710	1.6E−04	0.037	ENSATEP00000022248.1	complement C3-like [Source:NCBI gene;Acc:113162367]
FUN_018544	1.6E−04	0.037	ENSLCRP00005060884.1	uncharacterized LOC104929100 [Source:NCBI gene;Acc:104929100]
FUN_044991	1.7E−04	0.037	ENSSMAP00000032331.1	ARVCF delta catenin family member [Source:HGNC Symbol;Acc:HGNC:728]
FUN_062148	1.8E−04	0.037	ENSTRUP00000058883.1	collagen, type VI, alpha 3 [Source:ZFIN;Acc:ZDB-GENE-070501-8]
FUN_060009	1.8E−04	0.037	ENSLCAP00010022377.1	cullin-associated and neddylation-dissociated 2 (putative) [Source:ZFIN;Acc:ZDB-GENE-060503-645]
FUN_062650	1.9E−04	0.039	ENSGACP00000013864.1	-
FUN_042988	2.0E−04	0.039	ENSGWIP00000026157.1	ephrin type-A receptor 6-like [Source:NCBI gene;Acc:114480686]
FUN_037697	2.1E−04	0.042	ENSCGOP00000003077.1	sialic acid-binding Ig-like lectin 14 [Source:NCBI gene;Acc:115020870]
FUN_040399	2.2E−04	0.042	ENSGACP00000013864.1	-
FUN_051799	2.3E−04	0.042	ENSONIP00000065348.1	complement factor H [Source:NCBI gene;Acc:100710693]
FUN_031948	2.5E−04	0.045	ENSLBEP00000027957.1	-
FUN_033114	2.9E−04	0.049	ENSLBEP00000027957.1	-
FUN_019917	2.9E−04	0.049	ENSGACP00000025781.1	-
FUN_055712	2.9E−04	0.049	ENSLCAP00010030542.1	-
FUN_043047	3.0E−04	0.049	ENSCCRP00010114392.1	PR/SET domain 6 [Source:HGNC Symbol;Acc:HGNC:9350]

Table S9: Number of genes annotated as Btn and Btnl genes in the annotated genomes of 6 *Sebastes* and outgroup *Sebastolobus*.

Species	#Genes
Sebastes aleutianus	36
Sebastes entomelas	21
Sebastes miniatus	24
Sebastes pinniger	18
Sebastes rosaceus	23
Sebastes umbrosus	20
Sebastolobus alascanus	18

Table S10: Species names, location and IDs of the samples sequenced in this study.

Table S11: Species names and maximum lifespan information from various sources.

Table S12: Positively selected genes identified through branch-site model and those experiencing relaxed/intensified selection in long lived species.

Table S13: Positively selected genes identified through branch-site model and those experiencing relaxed/intensified selection in short lived species.

Table S14: Correlation values for evolutionary rates of genes with maximum lifespan, length at maturity, maximum depth and linear model residual of lifespan and depth across species.

Table S15: Gompertz parameters fit for different population sizes using the approach of Finch et al 1990 (83).

Table S16: Growth (Von Bertalanffy, VBF) and fecundity parameters for 36 rockfish species (85) used to model life history patterns in rockfish.

Table S17: Estimated generation times from life history parameters.