Historical contingency shapes adaptive radiation in Antarctic fishes

Jacob M. Daane[1,*], **Alex Dornburg**[2], **Patrick Smits**[3], **Daniel J. MacGuigan**[4], **M. Brent Hawkins**[3,5,6], **Thomas J. Near**[4,7], **H. William Detrich III**[1,†,*], **Matthew P. Harris**[3,6,†,*]

[1]Department of Marine and Environmental Sciences, Northeastern University Marine Science Center, Nahant, MA 01908, USA

[2]North Carolina Museum of Natural Sciences, Raleigh, NC 27601, USA

[3]Orthopaedic Research Laboratories, Department of Orthopaedic Surgery, Boston Children's Hospital, Boston, MA 02115, USA

[4]Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT 06520, USA

[5]Museum of Comparative Zoology, Harvard University, Cambridge, MA 02138, USA

[6]Department of Genetics, Harvard Medical School, Boston, MA 02115, USA

[7]Peabody Museum of Natural History, Yale University, New Haven, CT 06520, USA

Abstract

Adaptive radiation illustrates links between ecological opportunity, natural selection, and the generation of biodiversity. Central to adaptive radiation is the association between a diversifying lineage and the evolution of phenotypic variation that facilitates the utilization of novel environments or resources. However, is not clear whether adaptive evolution or historical contingency is more important for the origin of key phenotypic traits in adaptive radiation. Here we use targeted sequencing of >250,000 loci across 46 species to examine hypotheses concerning the origin and diversification of key traits in the adaptive radiation of Antarctic notothenioid fishes. Contrary to expectations of adaptive evolution, we show that notothenioids experienced a punctuated burst of genomic diversification and evolved key skeletal modifications before the onset of polar conditions in the Southern Ocean. We show that diversifying selection in pathways associated with human skeletal dysplasias facilitates ecologically important variation in buoyancy among Antarctic notothenioid species, and demonstrate the sufficiency of altered *trip11, col1a2* and *col1a1* function in zebrafish (*Danio rerio*) to phenocopy skeletal reduction in Antarctic

notothenioids. Rather than adaptation being driven by the cooling of the Antarctic, our results highlight the role of historical contingency in shaping the adaptive radiation of notothenioids. Understanding the historical and environmental context for the origin of key traits in adaptive radiations extends beyond reconstructing events that result in evolutionary innovation as it also provides a context in forecasting the effects of climate change on the stability and evolvability of natural populations.

During adaptive radiation, ecological opportunity and key phenotypic traits interact to facilitate the expansion of populations into new niches[1]. Deciphering the genotype-phenotype relationships of these key traits provides important insight into the historical circumstances that result in phenotypic diversification[7,8]. Advances in sequencing capabilities now allow for the efficient collection of genomic sequence data for scores of species. This potential for increased taxonomic, genome-wide sampling provides opportunities to investigate the macroevolutionary mechanisms driving evolution across clades. For example, a major question in the study of adaptive radiation is the relative role of trait novelty versus the modification of existing phenotypes in facilitating the diversification of species. Ecological opportunities, such as those brought about by environmental changes, can promote the origin of novel phenotypes that accelerate lineage diversification[2,3]. However, trait novelty may arise at any time, and adaptive change can result from a shift in selection regimes that enable lineages to opportunistically explore new ecospace using a trait that had previously evolved under different circumstances[9]. These two hypotheses have long been central in the debate concerning the role of punctuated versus gradualistic evolution in the generation of biodiversity[9,10], and posit opposing views on how the genomic substrate of subsequent phenotypic diversification evolved. Here, we provide a genomic perspective of ecologically important trait variation and phylogenetic origin in a species-rich adaptive radiation. Through consideration of closely related lineages that are not a part of the adaptive radiation, we can specifically investigate the relative importance of contingency in adaptive radiation and test whether the genomic basis of ecological trait variation coincides with or precedes the onset of radiation.

Results and Discussion

To test hypotheses concerning the origin of traits that facilitate adaptive radiation, we focused on Antarctic notothenioids (Cryonotothenioidea), an iconic example of adaptive radiation in marine vertebrates[6,11]. The diversification of cryonotothenioids followed the progressive cooling of Antarctica that initiated ~33 million years ago[6], and was coincident with the extinction of a phylogenetically diverse and cosmopolitan fish fauna[12]. This combination of climate change and vacated niches presented notothenioids with the opportunity to diversify into a large range of benthic and water column habitats. All notothenioids lack a swim bladder, the primary buoyancy organ in most teleost fishes; however, there are substantial differences in buoyancy among notothenioid species that are correlated with habitat and resource utilization[6,12]. These differences in buoyancy are achieved through the reduction of skeletal density coupled with the accumulation of corporeal lipids that provide static lift[12–14]. In the narrative of the Antarctic cryonotothenioid adaptive radiation, these ecologically important traits are hypothesized to

have arisen during the onset of polar conditions; however, the evolutionary origin of reduced skeletal density in the context of early-diverging non-Antarctic notothenioid lineages challenges this scenario[14].

To develop a comprehensive phylogenomic perspective on the Antarctic notothenioid adaptive radiation, we employed a cross-species targeted sequence enrichment approach to specifically sequence conserved and functionally-annotated genetic loci[15]. We combined sequence information from the genome of *Notothenia coriiceps* and select other percomorph teleost species to design a comparative DNA probe set for systematic targeted enrichment of ~250,000 coding and conserved non-coding (CNE) elements comprising over 40 Mb of genomic sequence (Supplementary Figure 1). CNEs were defined as miRNA hairpins, orthologs to human ultraconservative elements, and constrained genetic regions (GERP) identified in the Ensembl compara 11-way teleost genome alignment that did not overlap with protein-coding regions. Using this probe set, we captured coding and non-coding sequences from a phylogenetically-rich sampling of 46 notothenioid species that includes all three early-diverging non-Antarctic lineages and two species from the closely related Percidae (Supplementary Figure 1). Importantly, we used pooled population samples of individual species to permit identification of fixed and variable SNPs for each species.

We achieved 89–95% coverage of targeted regions for each sampled species (Supplementary Figure 1; Supplementary Tables 1,2). This permitted identification of an average of 95,000 fixed species-specific SNPs and 60,000 heterozygous SNPs for each lineage (Supplementary Table 3). As proof of the sensitivity of the dataset, we confirmed known cases of adult hemoglobin gene loss within the icefishes (Supplementary Figure 2)[16]. The power of this dataset was further illustrated by identification of a previously uncharacterized deletion of two putative embryonic hemoglobin genes (Supplementary Figure 2). Thus, this approach provides a robust and efficient method by which to characterize variation across taxonomically rich clades, even when separated by significant evolutionary distances.

Given expectations from the adaptive radiation of East African cichlids[17], we tested whether shifts in the rate of nucleotide evolution correspond to the onset of rapid lineage diversification in the cryonotothenioid radiation. Using a time-calibrated phylogeny of most living notothenioids and a Bayesian framework to assess shifts in speciation rates[11,18], we confirmed a shift in lineage diversification at the origin of the antifreeze-bearing Antarctic cryonotothenioids and an additional acceleration in diversification within the Plunderfishes (Artedidraconidae)[6] (Fig. 1a). Intriguingly, these shifts do not correspond with accelerated rates of nucleotide evolution. Instead, we found that the majority of extant cryonotothenioid sequence diversity is derived from a period of significantly high rates of genomic evolution that occurred in the ancestral lineage that includes the most recent common ancestor of the non-Antarctic distributed *Eleginops maclovinus* and the Antarctic cryonotothenioids (Eleginopsioidea, Fig. 1b, Supplementary Table 4). This demonstrates that a major change in molecular rate variation preceded the onset of global cooling and adaptive radiation of cryonotothenioids by well over 10 million years (Fig. 1c). An assessment of the accumulation of genomic divergence through relaxed molecular clock models (Fig. 1d), along with an analysis of synonymous (dS) and nonsynonymous substitution (dN) rates (Fig. 1e), further substantiates a high rate of nucleotide evolution prior to the origin and

diversification of cryonotothenioids. This acceleration of overall genomic change may have been critical for accumulating the genetic diversity that provided the substrate for phenotypic diversification during the early radiation of cryonotothenioids.

As buoyancy adaptations are key traits that facilitated notothenioid diversification into the water column during the adaptive radiation, we assessed patterns of skeletal density throughout the phylogeny using computerized tomography (Fig. 2a, Supplementary Figure 3). We find that, in parallel with the observed increase in nucleotide substitution rate, a broad reduction in bone density occurred prior to the recent common ancestor of *Eleginops maclovinus*[19] and the Antarctic cryonotothenioids (Eleginopsioidea, Fig. 2a) and was retained throughout the clade. Consistent with these decreases in bone density, we found biased diversifying selection in genes associated with human skeletal dysplasias and mineralization defects in Eleginopsioidea relative to the well-ossified sister lineage, *Pseudaphritis urvillii* (Fig. 2b,c). This evolutionary reconstruction is contrary to the expectation of the emergence of novel traits temporally coinciding with the onset of adaptive radiation and changing ecological opportunities, but rather indicates that prior to the shift to polar conditions, notothenioid lineages possessed key traits and may have begun to utilize new habitats in the water column through a modification of buoyancy enabled through a derived reduction in bone density.

To further explore the genetic basis of the evolved reduction in bone density, we assessed patterns of diversifying selection within the phylogeny. Among *Eleginops maclovinus* and all cryonotothenioids, we identified shared selective signatures in clinically relevant skeletal genes, such as *collagen1a1* and *collagen1a2* (Fig. 3a, Supplementary Figures 4–6, Supplementary Table 5). The function of these genes is conserved among disparate lineages of vertebrates[20] and nonsynonymous mutations in these collagens can lead to severe osteogenesis imperfecta in humans[21]. Notably, previous studies show that *collagen1* expression is reduced in the developing skeleton of notothenioid embryos, providing further evidence that broad changes at collagen loci are associated with skeletal variation[22]. In addition to selection on collagens, we identified diversifying selection in an unlikely gene candidate for skeletal variation, *trip11*(*gmap-210*) (Fig 3a, Supplementary Figures 4–6, Supplementary Table 5). This gene is conserved across eukaryotes and functions in vesicle tethering in the cis-Golgi membrane[23]. In humans, mutations in *TRIP11* lead to severe skeletal deficiencies that lead to peri-natal lethality[24]. Intriguingly, both *collagen1a1a* and *trip11* are also under diversifying selection and/or accelerated sequence evolution within the further diversification of the Channichthyidae, which have evolved an additional and significant reduction in bone density (Fig. 2a; Fig 3a; Supplementary Figures 4–6, Supplementary Tables 5–8)[14]. This parallelism suggests constraint in the types of mutations in evolution that can drive changes in skeletal density while maintaining viability.

To experimentally test the potential for alterations in *trip11, col1a1* and *col1a2* function to impart non-lethal skeletal phenotypes that are consistent with those observed in notothenioids, we analyzed mutants stemming from genetic screens and CRISPR-Cas9 gene editing in zebrafish (*Danio rerio*). Contrary to expectations from humans, zebrafish homozygous for loss-of-function alleles of *trip11* were viable with no obvious external morphological abnormalities (Fig. 3b,c). However, whereas heterozygous siblings had

density patterns comparable to wild-type zebrafish, similarly size and age-matched homozygous *trip11* mutants had significantly reduced skeletal density (Fig. 3b,c). In affected individuals, all bones investigated, including those of the skull roof, vertebrae and operculum were reduced in density (Fig. 3b,c) and generally phenocopied the pattern observed in adult *Eleginops maclovinus* and cryonotothenioids (Fig. 2a). Notably, expressivity of the *trip11* skeletal phenotype was variable, which suggests the presence of background genetic modifiers. Similarly, we show mutant models of *col1a1*[dmh14/+] (G1144E) and *col1a2* [dmh15/+] (G882N) in zebrafish also cause a reduction in bone density (Fig. 3c)[20]. These experiments confirm that changes to the loci under selection yield phenotypes consistent with those observed in *E. maclovinus* and species of the cryonotothenioid radiation.

Our results reveal that historical contingency was a major factor in shaping the adaptive radiation of notothenioids. As the onset of polar conditions ~33 mya decimated the teleost fauna of the Southern Ocean, notothenioids were the only surviving lineage poised to occupy newly open niche space in a range of benthic and water column habitats[25]. Rather than being driven by directional selection to evolve extreme phenotypes in response to the onset of polar conditions[26,27], the genomic substrate for reduced ossifications and buoyancy modifications had long been established and was selected upon to facilitate the ecological diversification that characterizes the notothenioid adaptive radiation. These results provide an alternative view on the impact of climate change in driving extreme adaptations in the Southern Ocean[28,29]. As we march further into the Anthropocene, our results caution that as both ecosystems and climate continue to change worldwide, the expectation that rapid adaptation will be dominant factor in predicting the response of biodiversity requires careful consideration[30].

Methods

Custom targeted sequence enrichment design

We based the design of the DNA enrichment baits primarily on the *Notothenia coriiceps* genome, the most closely related species to those targeted for sequencing with a published reference assembly[31]. To account for regions that are unannotated, under drift, or not easily identified within the *N. coriiceps* assembly, we included targeted regions from several outgroup genomes as detailed below. By having the same element potentially represented by more than one genome, this strategy allowed us to mitigate against genome assembly and annotation artifacts while facilitating hybridization of diverse species to the capture probes. Elements were identified in the *N. coriiceps* genome using BLASTN (ncbi-blast-2.2.30+; parameters '*-max_target_seqs 1 -outfmt 6*'). If the BLASTN hit had a E-value < 0.0001 and covered >80% of the query sequence, we included this region from the *N. coriiceps* genome. If the region was not identified, or had <85% identity in *N. coriiceps*, we retained the version from the genome of origin.

Coding exons were identified from annotations of the *Notothenia coriiceps*[31], stickleback (*Gasterosteus aculeatus*; BROADS1), and European sea bass (*Dicentrachus labrax*) genomes[32]. Conserved non-coding elements (CNEs) were defined from the constrained elements identified in the stickleback and tilapia (*Oreochromis niloticus*; Orenil1.0)

genomes from the Ensembl compara 11-way teleost whole genome alignment[33]. We also included predicted miRNA hairpins from miRbase[34] and ultraconservative non-coding (UCNE) elements from UCNEbase[35]. miRNA hairpins were padded to be >100bp. CNEs, miRNAs, and UCNEs that overlapped coding exons were removed using Bedtools (v2.26.0) intersectBed[36], and CNEs <100bp were additionally excluded to facilitate space in the sequence capture design. Where the constrained regions that defined the CNEs overlapped with annotations pertaining to specific miRNAs and UCNEs, the latter annotations were prioritized.

Targeted elements were submitted to Nimblegen for final probe design and the manufacturing of a Nimblegen SeqCap EZ Developer Library (Roche cat 06471684001) had 63,838,670 bp of capture space targeting 318,929 elements from four reference genomes (88.9% *Notothenia coriiceps*, 7.7% *Gasterosteus aculeatus*, 3.0% *Dicentrarchus labrax*, 0.4% *Oreochromis nicoleatus*). Accounting for redundancy of orthologous target regions between the genomes, the final design targeted 258,176 unique elements, of which 206,503 were predicted protein coding exons and 51,673 constrained non-coding regions with 85.0% coverage of targets not found in the *N. coriiceps* reference genome (47,097 elements; Supplementary Table 2).

Sample preparation and sequencing

Frozen tissue samples were acquired from the HWD and TJN labs, and from the Yale Peabody Museum frozen tissue collection (Supplementary Table 9). DNA from each species was isolated using Qiagen DNeasy Blood and Tissue kits, sequencing multiple individuals per species to account for population variation. For each species, equal amounts of DNA from each individual were pooled prior to shearing and library preparation. The pooled-population DNA was then sheared to an average size of 200bp using the Covaris E220 ultrasonicator in 130ul Covaris microTUBEs (Duty Cycle: 10%, Intensity: 5, Cycles/Burst: 200, Time: 300s, Temp: 4°C). Shearing was performed in shearing buffer: 10mM Tris, 0.1mM EDTA, pH8.3.

Targeted sequence enrichment and next generation sequencing

Sequencing libraries were constructed from 1μg of DNA using KAPA Library Prep kit (Roche cat# 07137923001), following standard protocol with barcoding and dual-SPRI size selection to generate libraries between 200–450bp. Sequencing libraries were hybridized, recovered, and amplified according to the standard protocol (Nimblegen SeqCap EZ Library SR User's Guide v4.3) with the following changes: hybridization was performed at 45°C instead of 47°C to allow for more mismatches between sequencing libraries and probes, and we used SeqCap Developer Reagent (Roche cat# 06684335001) instead of Human CotI DNA to block non-specific hybridization as recommended by the protocol.

Since the majority of the capture probes were designed based on the sequence of a *Nothothenia coriiceps*, we hybridized species in groups to limit potential competition between sequencing libraries of varied relatedness to the capture baits. Captured libraries were pooled for 100bp single-end sequencing using Illumina HiSeq 2500. We targeted multiplexing of 8–9 species per HiSeq 2500 flow cell, totaling six flow cells.

Reference contig assembly

The contig assembly approach is modified from the previously defined Phylomapping pipeline for cross-species targeted sequence enrichment datasets[15]. Briefly, sequencing reads are grouped into bins by homology to a targeted element (exon, CNE, etc.) and then assembled into contigs *de novo* (Supplementary Figure 7)

Processing of sequencing reads

Prior to contig assembly, low quality bases within sequencing reads were masked using the FASTX-Toolkit (fastq_masker; parameters '*-Q 33*') (http://hannonlab.cshl.edu/fastx_toolkit). Illumina adapter sequences were then trimmed using Trimmomatic v0.36[37]. Identical sequencing reads were then collapsed using the fastx toolkit v0.0.13 (fastx_collapser; parameters '*-Q 33*').

Read binning into orthology groups by BLAST

Reads were grouped by homology before contig assembly, using both blastn and dc-megablast (v2.6.0+, parameters '*-max_target_seqs 2 -outfmt 6*'). This dual-BLAST approach accounts for variation between sequencing reads and the reference genome from which the sequencing baits were defined[38]. As short target exons and CNEs can produce disproportionately small blastn E-values, we used an adaptive E-value cutoff based on the size of the target region. For target regions >25 bp, the cutoff was E-value $\leq$e-05. For targets that were $\leq$25 bp the E-value cutoff was $\leq$e-04 and for targets that were $\leq$20bp the E-value cutoff was $\leq$e-03. Reads were further excluded if a substantial portion of the read (>10bp) overlapped the target interval without being included in a blastn hit. The best resulting E-value from either blastn or dc-megablast was selected, with blastn selected in the event of a tie. As dc-megablast utilizes a mismatch-tolerant seed template, inclusion of dc-megablast resulted in the additional recovery of 30,000–50,000 sequencing reads per species (out of an average of 20,000,000 with total blastn hits) and the assembly of 50–150 more target regions than would be assembled by blastn alone.

de novo contig assembly

CAP3 was used to assemble contigs *de novo* from the bins of reads that have high homology to specific target regions (i.e. the same exon, CNE, etc.) that were identified by BLAST[39]. Reads were reverse complemented if necessary in order to put everything into the same complement strand as the target region from the reference genome. To accelerate CAP3 assembly, overlapping reads were first merged into smaller contigs using Usearch and then mixed with original reads as input for CAP3 (parameters '*-id 0.97 -fastq_maxdiffs 3 -fastq_minovlen 5*'). For CAP3 assembly, we required a minimum read overlap of 16bp and 96% identity between reads during contig assembly (parameters '*-o 16 -p 96*'). This cutoff has an effect of separating the reads stemming from duplication events into separate contigs as long as there is >4–6% variance between the paralogous regions.

We simulated the ability of this pipeline to distinguish copy number variants (Supplementary Figure 8a), generating a 300bp random DNA sequence *in silico* and making a second copy of this sequence at with specific levels of variation from the original. Sequencing reads (100bp) were then generated *in silico* at a depth of one read every five base pairs. Reads

were run through the assembly pipeline to assess whether the original DNA sequences were reconstructed from the read data, or if the reads formed a chimeric sequence. This simulation was repeated 250 times. The current assembly approach reliably reproduced single copy exons at all levels of variation and was able to re-assemble the individual paralogs where there was >6% divergence between original paralogous sequences, with inconsistent results at <5% divergence (Supplementary Figure 8b).

Contig merging and filtering

Sequencing reads were aligned to the assembled contigs using NextGenMap (v0.5.5; parameters '*-R 40*')[40]. This alignment step allows for the recruitment of new reads to the contig that may have not been previously identified by blastn due to high degrees of variance relative to the reference blast database, large indels, or low amounts of overlapping sequence with the target region. This allows the contig to be elongated to include more of the flanking regions surrounding each target element. Reads were removed from the alignment to the contig if there were >3 mismatches with the exception of indels. To refine and extend the boundaries of the original contig, a second *de novo* assembly by CAP3 (parameters '*-o 20 -p 85*') was performed using the aligned reads.

Multiple contigs were present in around 65% of target regions after CAP3 assembly. To remove misidentified contigs, we used blastn to compare each contig to the reference genome, removing contigs whose top hit did not match the original bin from which the reads were assembled. To correct for potential assembly artifacts, the multiple contigs that represent each target were compared to each other and the reference sequence in a multiple sequence alignment using Mafft v7.313[41] (parameters '*--maxiterate 1000 --localpair*'), adding contigs as fragments (parameters '*--addfragments*'). Using the read support at each base in the alignment, we generated a consensus contig sequence for each target region. A mismatch between contigs was considered if the read support for the most common base was <80% of all bases present. Contigs were only merged if there were <3 mismatches, if the sequence identity between the contigs was >95%, or if the contigs did not overlap within the target region. After this refinement step, <2% of target regions were represented by multiple contigs.

Comparison of assembled contigs to reference genome

In comparing our reference exome sequence assembly to the published *N. coriiceps* reference genome, we found 99.8% average percent identity of exome sequence to the reference target (Supplementary Figure 9), that included 98.5% of exons in the *N. coriiceps* reference genome (Supplementary Figure 9). The small differences in sequence identity and putative CNVs between this exome and the genome assembly may reflect meaningful biological variation in our independently sampled *N. coriiceps* populations. Both the sequence composition of the assembled exome and the predicted copy number data closely match the whole genome sequence data, providing confidence in our assembly.

Estimation of read coverage and depth of targeted regions

Coverage was estimated using BEDtools (2.23.0)[36]. Reads were first aligned to the assembled contigs with NextGenMap[40]. The coordinates of the read alignments were then

lifted the corresponding position on the reference genome using information from a pairwise sequence alignment between the contig and the orthologous region on the reference genome. Pairwise alignments were performed using Biopython v1.70 (parameters '*pairwise2; match = 5, mismatch = −4, gap_open = −15, gap_extend = −1*'). Alignments were converted to BAM files, sorted, and indexed using SAMtools v1.9[42]. Reads alignments were manually inspected in the Integrative Genome Viewer (IGV) to verify accurate read alignment[43]. Coverage is defined as the percentage of targeted bases in the primary reference genome having at least one read. Depth was estimated using coverageBed (-d).

Distribution of read coverage across the dataset

Most target regions had either 0% or 100% coverage (Supplementary Figure 10). Though the average depth is similar in the notothenioids compared to the outgroups, there is a wider distribution of depths in outgroup species (Supplementary Figure 10). Though global coverage is >85% in all species, gene classes associated with the immune system, cell adhesion proteins and extracellular matrix were enriched among regions with relatively poor coverage (<25% coverage in >75% of exons; Supplementary Table 10). This suggests that these fast evolving gene classes are less likely to be highly represented in these datasets, and is similar to previous findings with cross-species targeted DNA enrichment[67].

Recovery of population variation

To determine the ability of this approach to recover population variation, we looked for heterozygous SNPs within the targeted regions of the dataset. Sequencing reads were aligned to the reference contigs for each species using NextGenMap[40]. SAM files were converted to BAM files using SAMtools v1.9[42]. Variants were called using SAMtools mpileup and BCFtools v1.9 (parameters '*call -mv*'). We considered sites heterozygous in our small population samples if there is ≥2 reads showing the variant in at least 25% allele frequency within the sequencing reads.

Treatment of exons with a predicted history of duplication

We assembled a single exon/CNE copy for the majority of targets that were directly compared between species. However, less than 3% of targeted regions on average had assembled >1 contig per target region after assembly. For duplicated regions, both exon versions were ignored for that particular species in downstream analyses involving multiple groups, unless those analyses ask specific questions involving copy number.

Identification of orthologous sequences

For ortholog pairing, the contigs generated from the same reference target region were aligned using Mafft v7.313(parameters '*--op 10 --ep 10*')[41], with a maximum likelihood tree topology estimated IQTree[44]. Gene trees were reconciled with the species tree using Notung-2.9[45] (parameters '*--reconcile --rearrange --silent --threshold 90% --treeoutput nhx*') to infer patterns of duplication and loss. The total number of duplication and loss events inferred by Notung were then summed and compared to a null scenario where all copies are local duplicates. If Notung inferred fewer gain/loss events, the duplicate exons were paired based on the reconciled gene tree.

Simulation of ortholog identification approach

To estimate the ability of this approach to parse copy number variation into orthologous groups, we performed a series of simulations (Supplementary Figure 11a) using a random DNA sequence generated *in silico*. This ancestral DNA sequence was duplicated, and mutations were added at defined levels to distinguish each paralog. Both paralogs were then evolved according to a specified phylogeny, with variation added to each paralog in increments at each branch point. We varied copy number, length of contig sequence and also simulated local losses of an individual paralog within downstream lineages. Each simulation was repeated 250 times. These results suggest that as long as there is ≈4–6% variation between paralogous sequences, the approach can properly pair orthologous sequences (Supplementary Figure 11b–d). This coincides with the thresholds at which our pipeline can distinguish copy number variants during contig assembly (see above), meaning there will not be paralog sequences with <4% divergence in the dataset.

Multiple sequence alignment

All paired orthologous sequences were aligned using Mafft v7.313 (parameters '*--maxiterate 1000 --localpair --op 10 --ep 10*'). For coding regions that had out-of-frame or frameshift-causing indels, these alignments were then refined into codon alignments using the frameshift-aware multiple sequence aligner MACSE v2.03 (parameters '*-prog alignSequences -seq -seq_lr -fs_lr 10 -stop_lr 15*')[46]. The multiple sequence alignment was pruned using GUIDANCE v2.02 to mask residues with score <0.6 (parameters '*--bootstrap 25 --mafft --maxiterate 100, --localpair --op 10 --ep 10*')[47].

Reconstruction of gene sequences from exon data

Single-copy coding exons with orthology to *Gasterosteus aculeatus* were concatenated into gene sequences using the annotations of the *G. aculeatus* genome. The exons for each gene were spliced together in the same order in which they appear in the genome on the strand containing the gene. Transcript isoforms were merged into a non-redundant gene sequence containing all possible exons. A total of 18,600 gene sequences with orthology to *G. aculeatus* were reconstructed for each species.

Notothenioid phylogeny

We used two approaches to infer a phylogeny for notothenioids. For both analyses, only single copy genes with >85% coverage in all species were included, resulting in a dataset of 11,627 genes. First, we individually partitioned each gene by codon position and used ModelFinder as implemented in IQTree v.1.6.3[48] to estimate the optimal partitioning scheme and molecular evolution model. Genes were concatenated all a maximul likelihood tree was inferred using IQTree v.1.6.3[44]. To assess support for the phylogenetic relationships, we performed 1,000 ultra-fast bootstrap replicates[49].

In order to account for the effects of incomplete lineage sorting and known issues with concatenation for phylogenetic inference[50,51], we also inferred a species tree. Full species tree inference is not computationally feasible with large genomic datasets, so we relied on the summary species tree approach in ASTRAL v.5.6.2[52]. We first used IQTree v.1.6.3 to infer the maximum likelihood tree for each gene, apply partitioning schemes and molecular

evolution models as described above. We then used ASTRAL to summarize the distribution of gene trees and estimate the species tree. To assess support for the species tree topology, we estimated local posterior probabilities for each quadpartition in the tree[53].

Quantification of lineage diversification dynamics

To test for changes in lineage diversification rates across the temporal history notothenioids, we used a Bayesian analysis of macroevolutionary mixtures implemented in BAMM v2.5 with a previously published time-tree that sampled all major lineages and 87 of ~120 species of notothenioids[11]. Priors were were defined using the function setBAMMpriors contained in the R package *BAMMtools* v2.1.0[54], and missing species were accounted for based on currently described species, with outgroup sampling capturing all major lineages of non-polar notothenioids. BAMM was run for 100 million generations and assessed for target sampling of the posterior distribution (Effective sample size >200). Results were visualized using functions available from the R-package *BAMMtools*[54]. To assess the impact of phylogenetic uncertainty in topology and branch lengths on our estimates of speciation rates through time, we further replicated the above analysis across 500 topologies randomly selected from the posterior distribution of trees from the same prior work[11] (Supplemental Figure 12).

dS and dN estimates of substitution rate

dN, dS and dN/dS were calculated pairwise between each species and the outgroup *Percina caprodes*. This was performed for each reconstructed gene of at least 2,000 bp in both species using cal_dn_ds in the Biopython v1.70 codonseq module (parameters '*method="NG86"*'). The values of dN, dS and dN/dS were then averaged across all genes for each species.

Molecular clock models of substitution rate

The substitution rate was estimated on the reconstructed gene sequences using the random local clock model as implemented in BEAST v2.4.8 (codon partitioned, bModelTest, parameters '*chain length = 30M*')[55–57]. To simplify comparisons between gene trees, and as support for the relationships between the included species is high, we fixed the starting tree topology for each gene tree to match our ASTRAL-inferred species tree (Supplementary Figure 13). Only gene trees with ESS values ≥200 for all parameters were selected for downstream analysis (1,062 total). MRCA age priors were calibrated based on previous age estimates[11,58]: *Psuedaphritis* + Eleginopsioidea 63.0my (52.6–73.4), Bovichtidae + all notothenioids 85.7my (69.5–102.6), *Harpagifer-Pogonophryne* 10.2my (7.7–13.0), *Bathydraco-Chaenocephalus* 11.1my (9.4–13.3), *Notothenia* 17.7my (15.2–20.5), Cryonotothenioidea 21.6my (18.6–23.9), Eleginopsiodea 45.9my (37.2–53.2). The maximum clade credibility was constructed for each gene tree using TreeAnnotator.

Detection of diversifying selection

Positive selection was calculated using the adaptive branch-site random effects likelihood (aBSREL) implemented in HyPhy v2.3.9[59,60]. Single copy exon alignments were concatenated into genes as input based on the gene order in the stickleback genome. The

species tree was used for all comparisons (Supplementary Figure 13). Accelerated sequence evolution was assessed using phyloP[61] as implemented in PHAST v1.4[62] (parameters '--*method LRT --no-prune --features --mode ACC*'). The tree model for phyloP was derived separately for CNE and coding gene comparisons using phyloFit and the species (Supplementary Figure 13). Tree models for protein coding regions were based on 3,381 exons ≥1,000bp with ≥85% coverage in all species. CNEs tree models based on 2,912 elements ≥250bp with ≥85% coverage in all species.

Ontology enrichment

As there are no gene ontology (GO) terms relating to *Notothenia coriiceps*, we generated a custom gene ontology database based on the combined GO data from multiple species. GO data was mined from chicken, mouse, rat, human, stickleback, medaka and zebrafish in Ensembl BioMart (downloaded December 2016)[63]. All Ensembl gene IDs were converted to their Ensembl stickleback ortholog and a final merged GO list was created by combining stickleback orthologs for each species. As many of the evolved phenotypes in notothenioids are comparable to human pathologies, we further utilized Human Phenotype Ontology databases to characterize genetic trends within the fish dataset [64]. This HPO database (downloaded April 2018) was then converted from human to stickleback orthologs using Ensembl BioMart.

Ontology enrichment was performed using Fisher's Exact Test (SciPy v0.18.1; fisher_exact). We also assessed patterns of cumulative polygenic enrichment within ontologies using the SUMSTAT approach as implemented in Daub et al[65]. This approach normalizes the distribution of log-likelihood ratio test values (ΔlnL) output from phyloP and HyPhy by taking the fourth root (ΔlnL4). The ΔlnL4 score is then summed for all genes within an ontology and an enrichment p-value is estimated from the empirical sum(ΔlnL4) score through bootstrap resampling (1,500 replicates). For all enrichment analyses, p-values were corrected using FDR (python module statsmodels v0.6.1; fdrcorrection0).

Zebrafish husbandry and genetic lines

Zebrafish wild-type and mutant lines used were housed and maintained as previously described and in accordance with Boston Children's Hospital IACUC regulations[66]. The *col1a1a* (dmh14) and *col1a2* (dmh15) mutants were derived from a forward genetic screen[67].

Zebrafish genome editing

The gRNA site GGTCAGAGTTTGGGTCAGGT<u>CGG</u> in exon 1 of the zebrafish *trip11* gene was targeted. This site is 33 bp downstream of the ATG start codon of *trip11*. gRNA sequences were cloned in the BsaI site of the DR274 plasmid and *in vitro* transcribed with T7 RNAmax kit (ThermoFisher) Cas9 mRNA was obtained from SystemBio (CAS500A-1). Injections of fertilized zebrafish eggs were done with 50 ng/ul gRNA and 150 ng/ul Cas9 mRNA. Genotyping was performed using 5'-CCCTGGTCGGTGATTTAGGGTTAG-3' as forward primer and 5' CACCTCCCACTTCCTCGGCGCTTTCCAGCAGAATATCTTTGGTAAAATTAGAG-3' as reverse primer. PCRs using this primer pair yield a 178 bp wildtype band. Fish were

identified with a 47 bp deletion spanning the gRNA target site and yielding a 131 bp genotyping band. The deleted sequence is 5' - CTGGGTCAGAGTTTGGGTCAGGTCGGGGGAAGCTTGTCTTCATTTAC-3' and generates a frameshift at the 12th amino acid residue of *trip11*.

Analysis of skeletal density through computed tomography (CT)

Adult notothenioid specimens were loaned from the Yale Peabody Museum and Harvard Museum of Comparative Zoology (Supplementary Table 11) and scanned using the Siemens Biograph at Boston Children's Hospital Department of Nuclear Medicine and Molecular Imaging. Scan data was processed in Siemens PETsyngo VG60A software and analyzed in Amira (v 6.0.0; FEI Inc.).

Zebrafish were euthanized in 22% MS-222, fixed in 3.7% formaldehyde overnight, and rinsed in PBS. The fishes were embedded in 1% agarose to reduce movement during imaging. The fish skulls were scanned as in ref. 66, using a Skyscan 1173 (Bruker), 240-degree scan with 0.2 rotational step. X-ray source voltage set to 70 kV and current set to 80 μA. Exposure time was 1500 ms. Resolution of scan was 7.14 microns per pixel. Volume renderings were reconstructed as maximum intensity projections in Amira software. Skeletal density was estimated based on the average pixel intensity measured from the maximum intensity projection of the skull, vertebrae and operculum in ImageJ v1.51s (https://imagej.nih.gov/ij/). A total of n=6 *trip11*$^{+/-}$ and n=10 *trip11*$^{-/-}$ fish were scanned and quantified. The bones of the skull roof (parietal/frontal), operculum and vertebrae were measured from each individual.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

References

1. Schluter D The Ecology of Adaptive Radiation. (Oxford University Press, 2000).

2. Rabosky DL Phylogenetic tests for evolutionary innovation: The problematic link between key innovations and exceptional diversification. Philos. Trans. R. Soc. B Biol. Sci 372, (2017).

3. Stroud JT & Losos JB Ecological Opportunity and Adaptive Radiation. Annu. Rev. Ecol. Evol. Syst 47, 507–532 (2016).

4. Losos JB Lizards in an evolutionary tree: ecology and adaptive radiation of anoles. (University of California Press, 2009).

5. Gould SJ Wonderful Life: the Burgess Shale and the Nature of History Lethaia 17, (W. W. Norton, 1989).

6. Near TJ et al. Ancient climate change, antifreeze, and the evolutionary diversification of Antarctic fishes. Proc. Natl. Acad. Sci 109, 3434–3439 (2012). [PubMed: 22331888]

7. Chan YF et al. Adaptive Evolution of Pelvic Reduction of a Pitx1 Enhancer. Science (80-.). 327, 302–306 (2010).

8. Santos ME et al. The evolution of cichlid fish egg-spots is linked with a cis-regulatory change. Nat. Commun 5, (2014).

9. Gould SJ The Structure of Evolutionary Theory. (Harvard University Press, 2002).

10. Jablonski D Approaches to Macroevolution: 1. General Concepts and Origin of Variation. Evol. Biol 44, 427–450 (2017). [PubMed: 29142333]

11. Dornburg A, Federman S, Lamb AD, Jones CD & Near TJ Cradles and museums of Antarctic teleost biodiversity. Nat. Ecol. Evol (2017).

12. Eastman JT Antarctic Fish Biology: Evolution in a Unique Environment. (Academic Press, Inc, 1993).

13. DeVries AL & Eastman JT Lipid sacs as a buoyancy adaptation in an Antarctic fish. Nature 271, 352–353 (1978).

14. Eastman JT, Witmer LM, Ridgely RC & Kuhn KL Divergence in skeletal mass and bone morphology in antarctic notothenioid fishes. J. Morphol 275, 841–61 (2014). [PubMed: 24590921]

15. Daane JM, Rohner N, Konstantinidis P, Djuranovic S & Harris MP Parallelism and Epistasis in Skeletal Evolution Identified through Use of Phylogenomic Mapping Strategies. Mol. Biol. Evol 33, 162–173 (2015). [PubMed: 26452532]

16. Near TJ, Parker SK & Detrich HW A genomic fossil reveals key steps in hemoglobin loss by the Antarctic icefishes. Mol. Biol. Evol 23, 2008–2016 (2006). [PubMed: 16870682]

17. Brawand D et al. The genomic substrate for adaptive radiation in African cichlid fish. Nature (2014).

18. Rabosky DL Automatic detection of key innovations, rate shifts, and diversity-dependence on phylogenetic trees. PLoS One 9, (2014).

19. Eastman JT, Witmer LM, Ridgely RC & Kuhn KL Divergence in skeletal mass and bone morphology in antarctic notothenioid fishes. J. Morphol 275, 841–861 (2014). [PubMed: 24590921]

20. Gistelinck C et al. Zebrafish type I collagen mutants faithfully recapitulate human type I collagenopathies Short title: Zebrafish mutants mimic type I collagenopathies. PNAS 1–10 (2018).

21. Van Dijk FS & Sillence DO Osteogenesis imperfecta: Clinical diagnosis, nomenclature and severity assessment. Am. J. Med. Genet. Part A 164, 1470–1481 (2014).

22. Albertson RC et al. Molecular pedomorphism underlies craniofacial skeletal evolution in Antarctic notothenioid fishes. BMC Evol. Biol 10, 4 (2010). [PubMed: 20053275]

23. Witkos TM & Lowe M The Golgin Family of Coiled-Coil Tethering Proteins. Front. Cell Dev. Biol 3, 1–9 (2016).

24. Smits P et al. Lethal skeletal dysplasia in mice and humans lacking the golgin GMAP-210. N. Engl. J. Med 362, 206–216 (2010). [PubMed: 20089971]

25. Eastman JT & McCune AR Fishes on the Antarctic continental shelf: evolution of a marine species flock? J. Fish Biol 57, 84–102 (2000).

26. Chen L, DeVries A & Cheng C Evolution of antifreeze glycoprotein gene from a trypsinogen gene in Antarctic notothenioid fish. Proc. Natl. Acad. Sci 94, 3811–3816 (1997). [PubMed: 9108060]

27. Chen Z et al. Transcriptomic and genomic evolution under constant cold in Antarctic notothenioid fish. Proc. Natl. Acad. Sci. U. S. A 105, 12944–12949 (2008). [PubMed: 18753634]

28. Chown SL et al. The changing form of Antarctic biodiversity. Nature 522, 431–438 (2015). [PubMed: 26108852]

29. Chown SL et al. Antarctica and the strategic plan for biodiversity. PLoS Biol. 15, 1–10 (2017).

30. Bilyk KT, Vargas-Chacoff L & Cheng CHC Evolution in chronic cold: Varied loss of cellular response to heat in Antarctic notothenioid fish. BMC Evol. Biol 18, 1–16 (2018). [PubMed: 29368592]

31. Shin SC et al. The genome sequence of the Antarctic bullhead notothen reveals evolutionary adaptations to a cold environment. Genome Biol. 15, 468 (2014). [PubMed: 25252967]

32. Tine M et al. European sea bass genome and its variation provide insights into adaptation to euryhalinity and speciation. Nat. Commun 5, 5770 (2014). [PubMed: 25534655]

33. Herrero J et al. Ensembl comparative genomics resources. Database 2016, 1–17 (2016).

34. Kozomara A & Griffiths-Jones S MiRBase: Integrating microRNA annotation and deep-sequencing data. Nucleic Acids Res. 39, 1–6 (2011). [PubMed: 20805246]

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

36. Quinlan AR & Hall IM BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841–2 (2010). [PubMed: 20110278]

37. Bolger AM, Lohse M & Usadel B Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120 (2014). [PubMed: 24695404]

38. Altschul S, Gish W & Miller W Basic Local Alignment Search Tool. J Mol Biol. 215, 403–410 (1990). [PubMed: 2231712]

39. Huang X CAP3: A DNA Sequence Assembly Program. Genome Res. 9, 868–877 (1999). [PubMed: 10508846]

40. Sedlazeck FJ, Rescheneder P & Von Haeseler A NextGenMap: Fast and accurate read mapping in highly polymorphic genomes. Bioinformatics 29, 2790–2791 (2013). [PubMed: 23975764]

41. Katoh K, Kuma K, Toh H & Miyata T MAFFT version 5: improvement in accuracy of multiple sequence alignment. Nucleic Acids Res. 33, 511–8 (2005). [PubMed: 15661851]

42. Li H et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–9 (2009). [PubMed: 19505943]

43. Robinson JT et al. Integrative Genome Viewer. Nat. Biotechnol 29, 24–6 (2011). [PubMed: 21221095]

44. Nguyen LT, Schmidt HA, Von Haeseler A & Minh BQ IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol 32, 268–274 (2015). [PubMed: 25371430]

45. Chen K, Durand D & Farach-Colton M NOTUNG: A Program for Dating Gene Duplications and Optimizing Gene Family Trees. J. Comput. Biol 7, 429–447 (2000). [PubMed: 11108472]

46. Ranwez V, Harispe S, Delsuc F & Douzery EJP MACSE: Multiple alignment of coding SEquences accounting for frameshifts and stop codons. PLoS One 6, (2011).

47. Sela I, Ashkenazy H, Katoh K & Pupko T GUIDANCE2: Accurate detection of unreliable alignment regions accounting for the uncertainty of multiple parameters. Nucleic Acids Res. 43, W7–W14 (2015). [PubMed: 25883146]

48. Kalyaanamoorthy S, Minh BQ, Wong TKF, Von Haeseler A & Jermiin LS ModelFinder: Fast model selection for accurate phylogenetic estimates. Nat. Methods 14, 587–589 (2017). [PubMed: 28481363]

49. Hoang DT, Chernomor O, Von Haeseler A, Minh BQ & Vinh LS UFBoot2: Improving the ultrafast bootstrap approximation. Mol. Biol. Evol 35, 518–522 (2018). [PubMed: 29077904]

50. Kubatko LS & Degnan JH Inconsistency of phylogenetic estimates from concatenated data under coalescence. Syst. Biol 56, 17–24 (2007). [PubMed: 17366134]

51. Roch S & Steel M Likelihood-based tree reconstruction on a concatenation of aligned sequence data sets can be statistically inconsistent. Theor. Popul. Biol 100, 56–62 (2015).

52. Zhang C, Rabiee M, Sayyari E & Mirarab S ASTRAL-III: Polynomial time species tree reconstruction from partially resolved gene trees. BMC Bioinformatics 19, 15–30 (2018). [PubMed: 29343218]

53. Sayyari E & Mirarab S Fast Coalescent-Based Computation of Local Branch Support from Quartet Frequencies. Mol. Biol. Evol 33, 1654–1668 (2016). [PubMed: 27189547]

54. Rabosky DL et al. BAMMtools: An R package for the analysis of evolutionary dynamics on phylogenetic trees. Methods Ecol. Evol 5, 701–707 (2014).

55. Bouckaert R et al. BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. PLoS Comput. Biol 10, 1–6 (2014).

56. Bouckaert RR & Drummond AJ bModelTest: Bayesian phylogenetic site model averaging and model comparison. BMC Evol. Biol 17, 1–11 (2017). [PubMed: 28049419]

57. Drummond AJ & Suchard MA Bayesian random local clocks, or one rate to rule them all. BMC Biol. 8, 114 (2010). [PubMed: 20807414]

58. Near TJ et al. Identification of the notothenioid sister lineage illuminates the biogeographic history of an Antarctic adaptive radiation. BMC Evol. Biol 15, 109 (2015). [PubMed: 26062690]

59. Smith MD et al. Less is more: An adaptive branch-site random effects model for efficient detection of episodic diversifying selection. Mol. Biol. Evol 32, 1342–1353 (2015). [PubMed: 25697341]

60. Kosakovsky Pond SL, Frost SDW & Muse SV HyPhy: Hypothesis testing using phylogenies. Bioinformatics 21, 676–679 (2005). [PubMed: 15509596]

61. Pollard KS, Hubisz MJ, Rosenbloom KR & Siepel A Detection of nonneutral substitution rates on mammalian phylogenies. Genome Res. 20, 110–21 (2010). [PubMed: 19858363]

62. Hubisz MJ, Pollard KS & Siepel A PHAST and RPHAST: phylogenetic analysis with space/time models. Brief. Bioinform 12, 41–51 (2011). [PubMed: 21278375]

63. Kinsella RJ et al. Ensembl BioMarts: A hub for data retrieval across taxonomic space. Database 2011, 1–9 (2011).

64. Köhler S et al. The human phenotype ontology in 2017. Nucleic Acids Res. 45, D865–D876 (2017). [PubMed: 27899602]

65. Daub JT, Moretti S, Davydov II & Excoffier L Detection of Pathways Affected by Positive Selection in Primate Lineages Ancestral to Humans. Mol. Biol. Evol 34, 1391–1402 (2017). [PubMed: 28333345]

66. Nüsslein-Volhard C & Dahm R Zebrafish: a practical approach. (Oxford University Press, 2002).

67. Henke K et al. Genetic Screen for Post-embryonic Development in the Zebrafish (Danio rerio): Dominant Mutations Affecting Adult Form. Genetics 207, genetics.300187.2017 (2017).
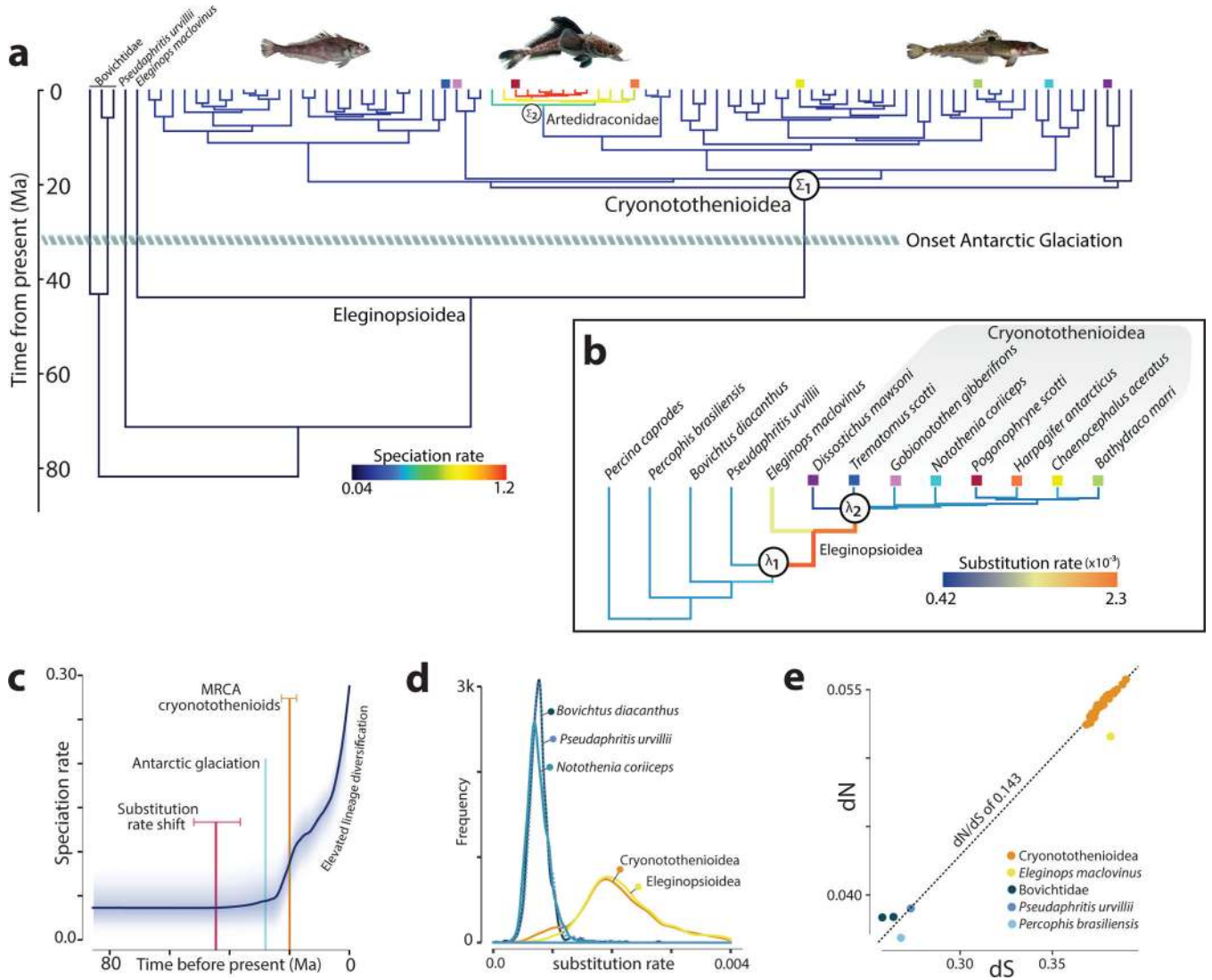
Fig. 1 |. Punctuated elevation in genomic diversification prior to ecological change and adaptive radiation.

(**a**) Bayesian analysis of lineage diversification rate on time-calibrated notothenioid phylogeny[12,18]. S indicates change in speciation rate. Colors on branches correspond to the mean of the marginal posterior density of estimated speciation rates, with a shift from low (0.04 lineages/Ma) to faster rates of speciation (0.046 lineages/Ma) at the onset of the notothenioid adaptive radiation. Colored squares at select tips in phylogeny refer to corresponding species in panel **b**. **b**) Relaxed molecular clock model of nucleotide substitution rate (substitutions/site/Ma) during notothenioid evolution revealing a transient elevation in substitution rate prior to the increase in species diversification. Substitution rates estimated from 1,062 independent gene trees constructed using the random local clock model in BEAST2. λ_1 indicates change in substitution rate, from an ancestral baseline of 7.7×10^{-4} substitutions/site/Ma (95% highest posterior density interval (HPD): $5.0 \times 10^{-4}, 1.1 \times 10^{-3}$) to 2.3×10^{-3} substitutions/site/Ma (95% HPD: $1.0 \times 10^{-3}, 3.9 \times 10^{-3}$) observed on the branch leading to *Eleginops maclovinus* and all cryonotothenioids. λ_2 indicates return to

baseline substitution rate in cryonotothens, ranging from 4.2×10^{-4} substitutions/site/Ma in *Disosstichus mawsoni* (95% HPD: $1.0 \times 10^{-4}, 9.0 \times 10^{-4}$) to 7.7×10^{-4} substitutions/site/Ma in *Chaenocephalus aceratus* (95% HPD: $2.0 \times 10^{-4}, 2.2 \times 10^{-3}$). See Supplementary Table 4 for additional rates and 95% HPD. Colors on branches correspond to mean substitution rate. (**c**) Elevation in speciation rate well after increase in nucleotide substitution rate and climate change events that precipitated adaptive diversification. Shading indicates 10–90% Bayesian credible region across time. Error bars for age of cryonotothenioid most recent common ancestor (MRCA) and time of substitution rate shift at MRCA of eleginops and cryonotothenioids based on 95% Bayesian credible interval. (**d**) Distribution of substitution rates from select species in (**b**). (**e**) Average synonymous (dS) and non-synonymous (dN) substitution rate from over 4,000 pairwise alignments for each species relative to the outgroup, *Percina caprodes*. This shows an increase in overall substitution rate (dN and dS) without a genome-wide change in diversifying selection (dN/dS).
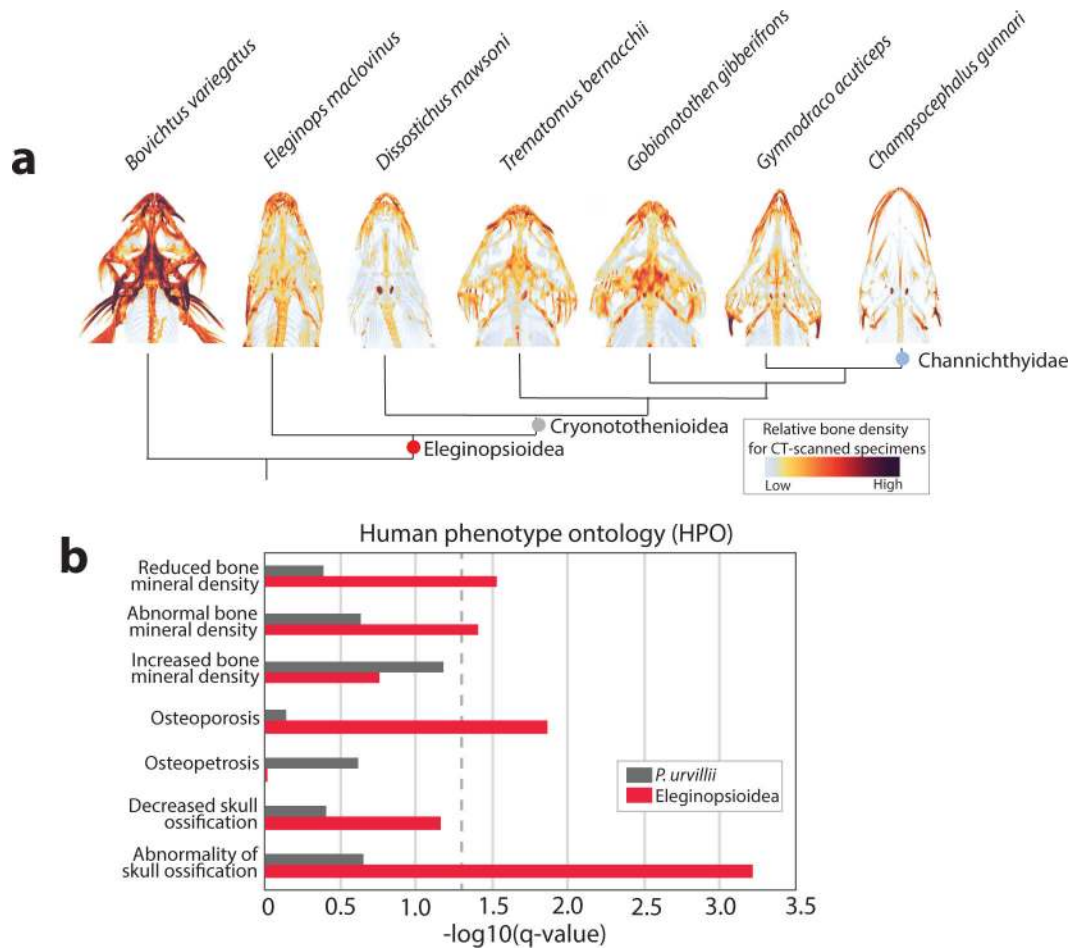
Fig. 2 |. Skeletal reduction occurs prior to the cryonotothenioid radiation.
(**a**) Computed tomography (CT) of notothenioid skulls showing relative skeletal density across the phylogeny and decrease in skeletal density prior to cryonotothenioid radiation. (**b**) Comparison of significance of enrichment for polygenic selection on representative bone-density associated HPO terms between Eleginopsioidea and well-ossified sister group *Pseudaphritis urvillii*. Dashed line indicates FDR q-value < 0.05 for SUMSTAT gene set enrichment test.
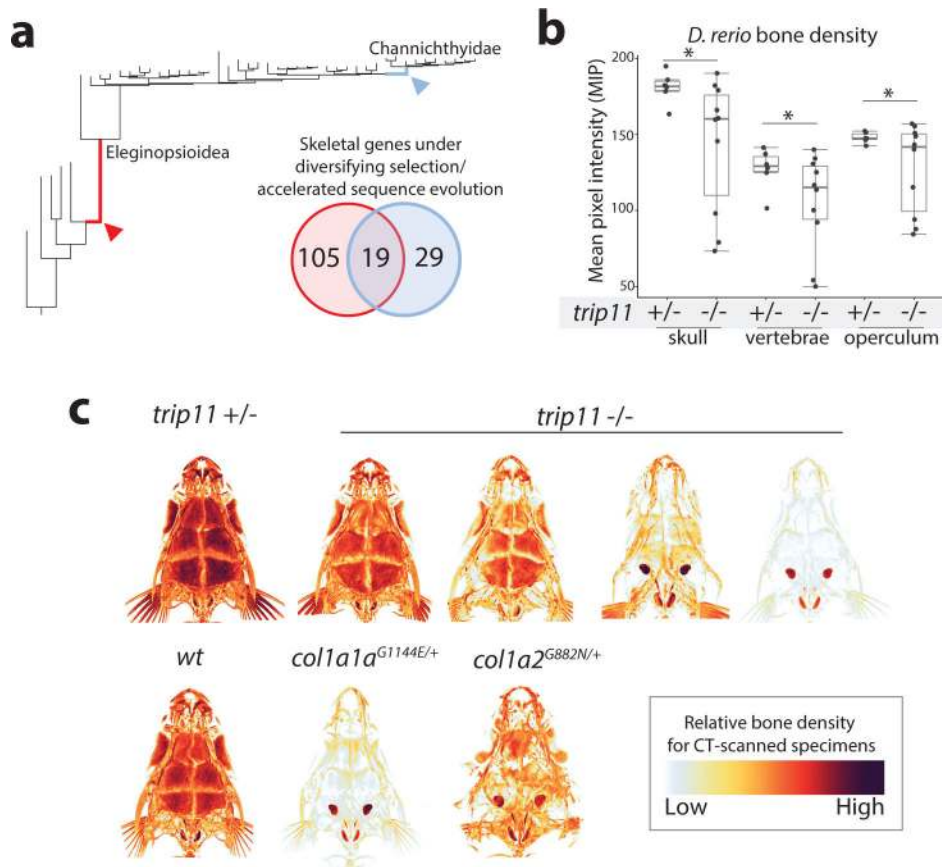
Fig. 3 |. Skeletal genes under diversifying selection uncover genetic mechanisms regulating bone density.

(**a**) Comparison of genes under positive selection (aBSREL, p<0.05) and/or accelerated sequence evolution (phyloP, p<0.05) on the branch leading to *Eleginops maclovinus* and Antarctic cryonotothenioids (Eleginopsioidea) with the branch leading to the icefishes, Channichthyidae. (**b**) Quantitation of skeletal density from μCT analysis of zebrafish mutants in *trip11*. Density is the average pixel intensity. n=6 scanned *trip11⁺/⁻* and n=10 *trip11⁻/⁻* individuals. Center line is mean density, box bounds indicate lower and upper quartiles, whiskers extend to a maximum of 1.5 times the interquartile range. *: one tailed t-test p-value <0.05. (**c**) Micro-computed tomography (μCT) of zebrafish showing reduction in bone density in the cranium of zebrafish mutant models of genes under selection in Eleginopsioidea. Representative μCT images of zebrafish skulls from heterozygous and homozygous *trip11* mutants, showing reduced bone density and variable penetrance. Reduction in density in the cranium caused by mutations in *col1a1a^dmh14/+* (G1144E) and *col1a2^dmh15/+* (G882N).