Genomics 114 (2022) 305–315

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

Genome sequencing and transcriptomic analysis of the Andean killifish *Orestias ascotanensis* reveals adaptation to high-altitude aquatic life

Alex Di Genova [a,b,1], Gino Nardocci [a,c,1], Rodrigo Maldonado-Agurto [a,c,1], Christian Hodar [a,d,1], Camilo Valdivieso [a,e], Pamela Morales [a,e], Felipe Gajardo [a,e], Raquel Marina [a,c], Rodrigo A. Gutiérrez [a,f], Ariel Orellana [a,g], Veronica Cambiazo [a,d], Mauricio González [a,d], Alvaro Glavic [a,e], Marco A. Mendez [a,e,h,i], Alejandro Maass [a,b], Miguel L. Allende [a,e,**], Martin A. Montecino [a,c,*]

[a] FONDAP Center for Genome Regulation, Santiago, Chile
[b] Center for Mathematical Modeling, Department of Mathematical Engineering, Faculty of Physical and Mathematical Sciences, Universidad de Chile and IRL CNRS, 2807 Santiago, Chile
[c] Institute of Biomedical Sciences, Faculty of Medicine and Faculty of Life Sciences, Universidad Andres Bello, Santiago, Chile
[d] Institute of Nutrition and Food Technology, Universidad de Chile, Santiago, Chile
[e] Faculty of Sciences, Universidad de Chile, Santiago, Chile
[f] Department of Molecular Genetics and Microbiology, Pontificia Universidad Catolica de Chile, Santiago, Chile
[g] Center of Plant Biotechnology, Universidad Andres Bello, Santiago, Chile
[h] Center of Applied Ecology and Sustainability (CAPES), Facultad de Ciencias Biológicas, Pontificia Universidad Católica de Chile, Chile
[i] Institute of Ecology and Biodiversity, Chile

ARTICLE INFO

Keywords:
Orestias ascotanensis
DNA repair
High-altitude
Desert pupfish
Altiplano
microRNAs

ABSTRACT

Orestias ascotanensis (Cyprinodontidae) is a teleost pupfish endemic to springs feeding into the Ascotan saltpan in the Chilean Altiplano (3,700 m.a.s.l.) and represents an opportunity to study adaptations to high-altitude aquatic environments. We have *de novo* assembled the genome of *O. ascotanensis* at high coverage. Comparative analysis of the *O. ascotanensis* genome showed an overall process of contraction, including loss of genes related to G-protein signaling, chemotaxis and signal transduction, while there was expansion of gene families associated with microtubule-based movement and protein ubiquitination. We identified 818 genes under positive selection, many of which are involved in DNA repair. Additionally, we identified novel and conserved microRNAs expressed in *O. ascotanensis* and its closely-related species, *Orestias gloriae*. Our analysis suggests that positive selection and expansion of genes that preserve genome stability are a potential adaptive mechanism to cope with the increased solar UV radiation to which high-altitude animals are exposed to.

1. Introduction

The South American Altiplano is the westernmost part of a large volcanic-sedimentary plateau in the central Andes, with an average altitude of 4,000 meters above sea level (m.a.s.l). As a high-altitude environment, the altiplano exhibits harsh conditions for life, including wide day–night temperature oscillations, high UV radiation, low humidity and reduced partial pressure of oxygen. Despite these harsh conditions, several organisms are known to be well adapted to living and thriving in these high-altitude environments including snails of the genus *Biomphalaria* [1] and *Heleobia* [2], frogs from the genus *Telmatobius* [3], fishes from the genus *Orestias* [4], Andean flamingos [5] and South American camelids including llamas and alpacas [6], among others. In recent years, genomic and transcriptomic data have been used to obtain insights into the molecular adaptations to high-altitude environments in different species. Nevertheless, most of these efforts have been focused on organisms inhabiting the high-altitude environment of the Tibetan Plateau, including mammals [7,8], birds [9,10], reptiles

* Correspondence to: M. Montecino, Instituto de Ciencias Biomedicas, Facultad de Medicina, Universidad Andrés Bello, Echaurren 183, Santiago, Chile.
** Correspondence to: M. Allende, Departamento de Biologia, Facultad de Ciencias, Universidad de Chile, Las Palmeras 3425, Ñuñoa, Santiago, Chile.
E-mail addresses: mallende@uchile.cl (M.L. Allende), mmontecino@unab.cl (M.A. Montecino).
[1] Authors that equally contributed to this work.

https://doi.org/10.1016/j.ygeno.2021.12.018
Received 25 August 2020; Received in revised form 26 November 2021; Accepted 17 December 2021
Available online 23 December 2021

[11,12] and fish [13–15]. In contrast, little is known about the mechanisms that animals of the South American Altiplano have developed to cope with this high-altitude environment. Furthermore, the Altiplano harbors unique environments consisting of a high-altitude desert aquatic ecosystem. In one of these, the springs of the Ascotan saltpan at 3,700 m. a.s.l. (Fig. 1A), lives the teleost pupfish *O. ascotanensis* (Cyprinodontidae, Fig. 1B). The genus *Orestias* (Order Cyprinodontiformes; Family Cyprinodontidae) is an endemic group of species inhabiting continental aquatic systems (freshwater and salt lakes) in the inter-Andean basins of Peru, Bolivia and Chile. This genus is remarkable in the number of species described (46, so far), being the most speciose in the family Cyprinodontidae and the third in the order Cyprinodontiformes. This genus has been characterized by an extensive radiation that occurred within Lake Titicaca [15], as well as by several allopatric speciation events that occurred in the southern part of its distributional range [16]. The genus *Orestias* has evolved and adapted to the diverse aquatic systems found in this region which, in addition to the high-altitude environment characteristics such as high UV radiation and low oxygen concentration, also exhibit elevated concentrations of heavy metals in the water as well as variable salinity due to the fluctuation in water levels before and after the South American Summer Monsoon (SASM) [16]. In particular, it has been described that *O. ascotanensis* can tolerate high salinity (3000–4000 μS•cm −1), low oxygen concentration (< 7 mg•L −1), extreme daily temperature changes and ultraviolet radiation above 2000 kWh [17]. Furthermore, a recent characterization of *O. ascotanensis* embryogenesis revealed an early development of pigmented cells and a circulatory system including a high number of red blood cells, which were proposed to represent adaptations to the increased ultraviolet radiation and low oxygen pressure of the Andean Altiplano [18]. These unique features introduce this group of fishes as an attractive biological model to address the mechanisms of adaptations of fish to high-altitude habitats/environments.

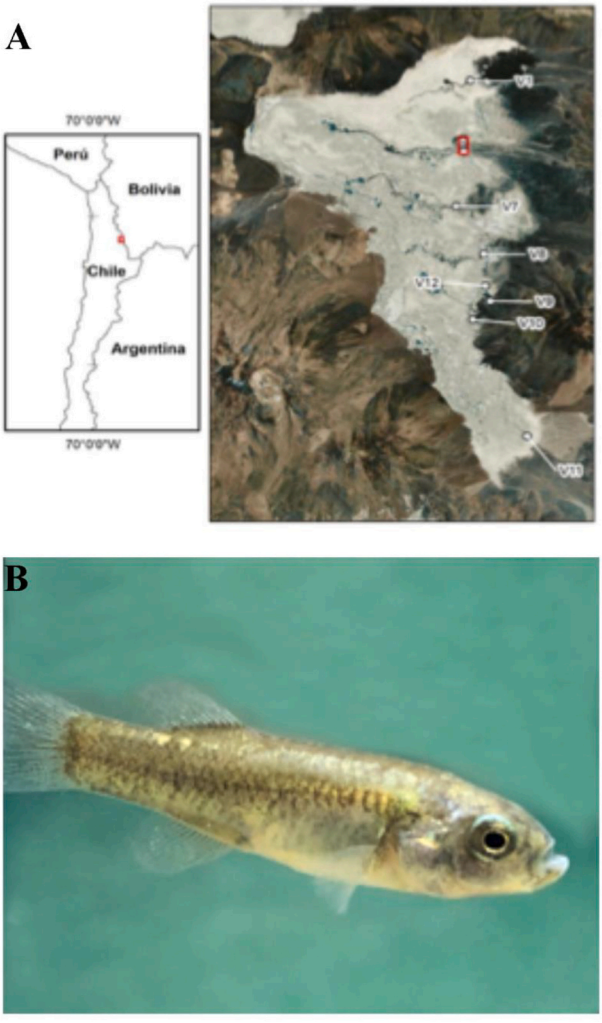

Fig. 1. Geographical location and distribution of *Orestias ascotanensis* in the Ascotan Saltpan. (A) The Ascotan Saltpan (UTM coordinates: 19S 576,991 7,622,725) is located at 3700 m.a.s.l. and is part of the Chilean Altiplano (right). Distribution of the twelve isolated springs (left) located in the eastern margin of the Ascotan saltpan (designated V1-V12) where *O. ascotanensis* inhabits. Indicated in red is spring number 5 (V5), where specimens were collected for this work. (B) Adult specimen of *O. ascotanensis*. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Thus far, systematic studies of *Orestias* species inhabiting the Andean Altiplano have been focused on morphological and embryo-developmental characterizations [4,18–20], phylogenetic and phylogeographic analyses [21–25] and cytogenetic comparisons [26]. These studies indicate that morphology and cariological differences only partially resolve species characterization of the *Orestias* genus. Hence, there is a necessity to add more attributes to study this species group, including genomic studies.

Here, we have fully sequenced, assembled and annotated the genome *O. ascotanensis* to search for genomic signatures that could shed light on the adaptive strategies used by this species to cope with the environmental conditions imposed by the high-altitude desert environment. Despite that changes in the protein-coding sequences can produce major effects on phenotypes, gene regulatory mechanisms such as microRNAs (miRNAs) allow fine-tuning of gene expression and can serve as drivers of speciation [27–29]. For this reason, we have also sequenced the microRNAome of *O. ascotanensis* and a closely-related species from the same genus, *O. gloriae*. Genomic and transcriptomic comparisons with other teleosts, including other Cyprinidontiformes, revealed several candidate loci that may support the ability of *O. ascotanensis* to live in this unique environment. This is the first fully sequenced genome of a high-altitude desert fish, that can be used for future studies on the speciation of the genus *Orestias* during periods of uplift of the Andean Plateau as well as the characterization of the molecular adaptations to high-altitude in aquatic systems.

2. Results

2.1. Sequencing and assembly of the O. ascotanensis genome

A total of 1.45 billion reads (~156 Gb) were obtained using an Illumina HiSeq2500 platform, representing a coverage of 292× of the estimated 750 Mb genome. The resulting assembly is 696.3 Mb (92.84% of the estimated full genome) with a scaffold N50 value of 2.67 Mb. Approximately 90% of the assembly was contained in 364 scaffolds (>383 Kb), with the largest scaffold spanning 14 Mb (Table 1). Core Eukaryotic Genes Mapping Approach (CEGMA) [30] results showed that our assembly captures 98% of 248 highly conserved eukaryotes coding genes, thereby showing that the *O. ascotanensis* sequence reported here represents a very high-quality genome when compared to other non-model teleost fishes (Table S1). Approximately 21% (~ 142 Mb) of the genome corresponded to repetitive sequences, including 113 LINEs, 24 LTRs, 12 SINES and 259 sequences defined as unknown elements (data not shown). The draft assembly of *O. ascotanensis* (DDBJ/ENA/GenBank under the accession: SRP062595/PRJNA293101) was annotated using a combination of homology-based prediction, *ab initio* prediction and transcriptome-based prediction methods. Our final consensus gene set contained a total number of 21,024 protein-coding genes and 33,485 transcripts (Table 2). Subsequently, gene functions were annotated using SwissProt, NR database (from NCBI), UniProt and Kyoto Encyclopedia of Genes and Genomes (KEGG). These analyses revealed that 31,542 transcripts correspond to known curated protein

Table 1
Orestias ascotanensis genome assembly statistics.

Genome assembly		N50 (size/number)	N90 (size/number)	Total length
	Contigs	43.8 Kb/4,459	11.1 Kb/15,885	670 Mb
	Scaffolds	2.67 Mb/78	383.2 Kb/364	696.3 Mb
Noncoding RNAs		Copies	Length	
	tRNAs	265	19.6 Kb	
	miRNA	166	12.1 Kb	
Transposable Elements		Total length	Percent of genome	
	Total	142.6 Mb	20.48	
	DNA transposons	39.5 Mb	5.69	
	Retroelemens	49.9 Mb	7.19	
Protein-coding genes	Total Number	Annotated	Unannotated	
	21,024	19,552	1,072	

Table 2
Gene prediction summary for the *Orestias ascotanensis* genome. Consensus gene models were predicted using a combination of evidences from RNA-Seq and putative protein alignments.

Object	Attribute	Value
Genes	Number of genes	21,024
	Number of mRNAs	33,485
	Number of exons	250,022
	Number of CDSs	236,684
	Number of introns	224,318
	Exons per gene	11.9
	CDSs per gene	11.3
	Introns per gene	10.7
Isoforms	Genes alternatively spliced	6,508
	Percentage of genes alternatively spliced	30.9
	Transcript per gene	1.59
	Transcript per alternatively spliced gene	2.91
	Genes alternatively spliced with splicing in CDS	5,423
	CDS-structure per spliced gene	2.76
Span	Average gene length (bp)	17,291.8
	Average exon length (bp)	278.4
	Average intron length (bp)	1,611.3
	Average base pairs between genes (bp)	15,207.9
UTRs	Genes with UTR (5′ or 3′ UTR)	14,525
	mRNA with 5′ and 3′ UTR	24,487
	mRNA with 5′ UTR	557
	mRNA with 3′ UTR	1,917
	mRNA without UTR	6,524
	mRNA with UTR	26,961

and that 27,610 transcripts covered at least 50% of the known proteins (17,457 transcripts covered at least 90%).

2.2. Phylogenetic tree construction

We used the genome sequence of *O. ascotanensis* and of 12 different species of fish to generate a phylogenetic tree based on a common set of single-copy orthologs. For this purpose, we also retrieved the protein-coding genes for 12 selected and recently sequenced fish species. All-against-all BLAST and OrthoMCL (using default options) were used to identify a set of 20,411 orthologous groups shared by all 12 species from which 1,970 contained putative single-copy gene families. We next generated a maximum likelihood phylogenetic tree with a trimmed and concatenated protein sequence alignment from 1,970 single-copy genes in 12 species (Fig. 2). The resulting phylogeny identified *O. ascotanensis* within a sister taxa of the subfamily *Poeciliinae*, a cyprinodontiform group widely distributed throughout the Americas that includes well-known genera of aquarium fishes such as *Xiphophorus* and *Poecilia* [31].

2.3. Gene family expansion and contraction

Gene gain and loss during genome evolution are a major source of genomic variation, and are important drivers of adaptive phenotypic diversification [32–34]. Comparison of the genomes of *O. ascotanensis* and four close relatives (*Xiphophorus maculatus*, *Poecilia formosa*, *Oreochromis niloticus* and *Oryzias latipes*) identified 7,676 gene families across all five species, from which a total of 79 gene families are significantly expanded (p-value <0.05) in *O. ascotanensis* and 198 gene families that are significantly contracted. Supplementary Table S2 summarizes the gene families that show changes for all five species. On average, the genome of *O. ascotanensis* has an expansion/contraction value of −0.12 (Table S2), which would indicate a contraction in most gene families. Based on Gene Ontology (GO) annotations, expanded gene families were highly enriched in microtubule-based movement (p-value = 9.9E-9), protein deubiquitination (p-value = 7.8E-8) and proteolysis involved in cellular protein catabolic process (p-value = 4.5E-5) (Table 3). On the other hand, the GO annotations for the most significantly contracted gene families in the *O. ascotanensis* genome were found to be functionally related to the G-protein coupled (GPCR) signaling pathways (p-value = 7.4E-8), chemotaxis (p-value = 7.9E-6) and signal transduction (p-value = 1.2E-5).

2.4. Positive selection of genes

To gain insight into unique features of the *O. ascotanensis* genome, we identified protein coding genes that are under positive selection in *O. ascotanensis* compared to 15 other fish species. We used a phylogenetic analysis by maximum likelihood (PAML) to determine positive selection at the codon level [35] and identified 818 genes with at least one residue under positive selection in the *O. ascotanensis* genome (**Table S3**). A gene ontology enrichment analysis revealed that the genes under positive selection in *O.ascotanensis* are enriched for components of signal transduction pathways and DNA replication and repair (Table 4).

Additionally, we performed a gene enrichment analysis and protein-protein interaction (PPI) enrichment analysis of positively selected genes (PSGs) using Metascape [36]. Metascape integrates over 40 unique data sources to perform up-to-date gene enrichment analysis, automatically clustering enriched terms into non-redundant groups. In Fig. 3A, the Metascape's heatmap shows enrichment of PSGs in several categories involving processes related to cellular signaling, mitochondrial translation and nucleic acids related processes.

Analysis of protein interactions can help to allocate PSGs to specific biochemical pathways or signal transduction components. Our analysis shows that the top 3 enriched PPI categories are associated to G-protein coupled receptors signaling, DNA repair and extracellular organization (Fig. 3B). Interestingly, several of these positively selected genes enriched in PPI may be relevant to coping with the high-altitude environment. Positively selected genes enriched in PPI include several genes associated with DNA damage response (Fig. 3B) including PARP1 [37], RPA3 [38], FEN1 [39], POLD1, POLD3 [40] and ERCC5 [41], which are part of both enriched categories of base excision repair and DNA replication.

Interestingly, other positively selected genes (Table S3) associated with DNA damage response include POLI [42], UVSSA [43], FANCF [43], RPA3 [44] and UBE2T [45]. The latter three genes are also part of a DNA repair pathway called the Fanconi Anemia (FA) pathway, which is activated when nucleotides on opposite strands of double stranded DNA undergo covalent linkage, a process also known as inter-strand DNA cross-links (ICLs) [46].

Altogether, these results indicate that, although the coding genome of *O. ascotanensis* has suffered a contraction, there has been a positive selection of genes associated with different DNA repair mechanisms, which may indicate a functional adaptation to maintain the integrity of the genome under an increased UV radiation/exposure in high-altitude environments [47].

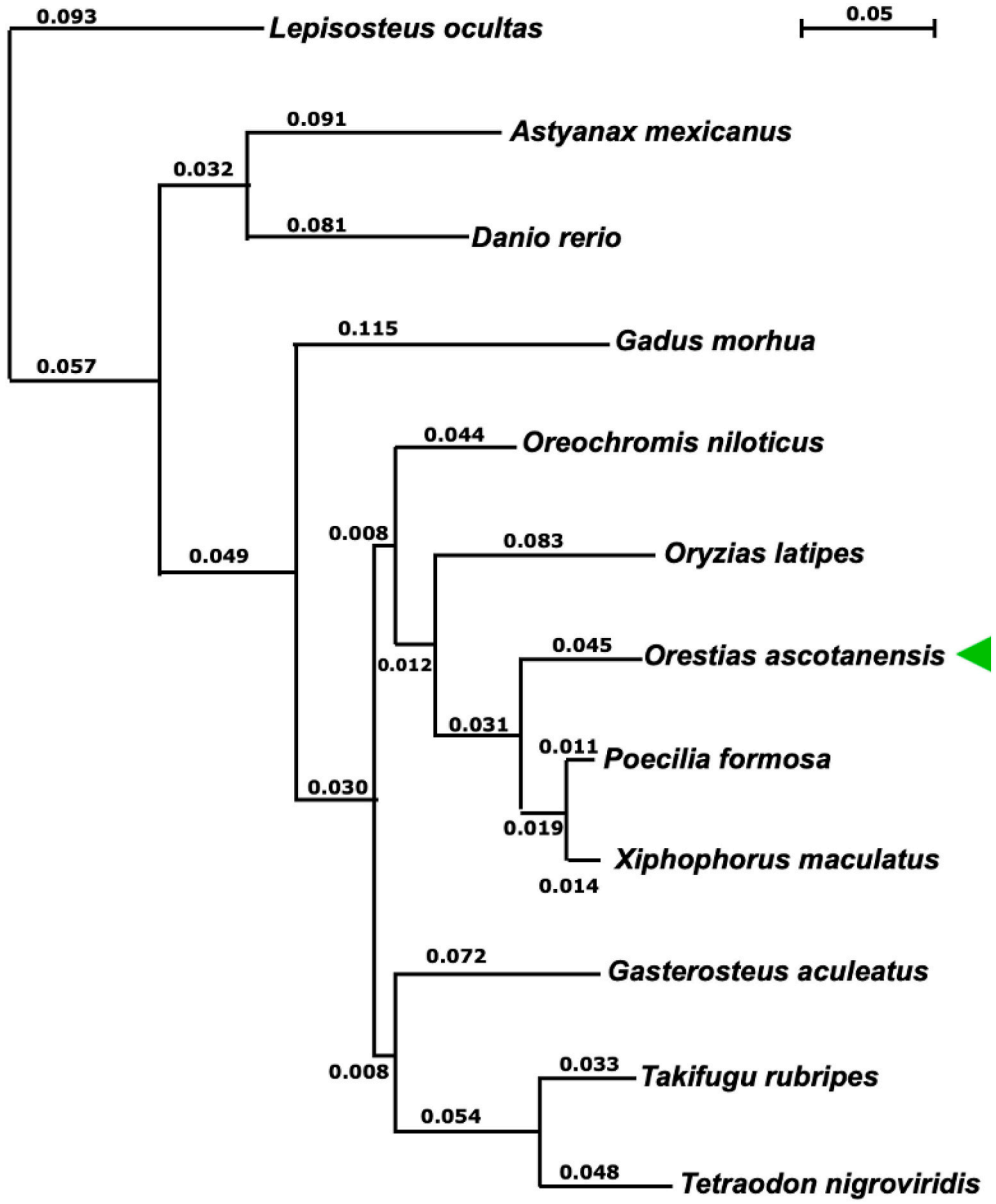

Fig. 2. Evolutionary analysis of *Orestias ascotanensis*. Phylogenetic tree reconstructed from analysis of OrthoMCL using a core of 1,970 orthologous core genes, shows the phylogenetic position of *O. ascotanensis* (indicated by the green triangle) relative to other teleost fish (values indicate branch length). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 3
Enriched biological processes derived from the GO analyses of the most significantly expanded and contracted gene families in *Orestias ascotanensis*.

Family genes	Term	Count	%	p-value
	Microtubule-based movement	8	11,9	9,90E-09
	Protein deubiquitination	7	10,4	7,80E-08
Expanded	Ubiquitin-dependent protein catabolic process	7	10,4	1,80E-06
	G-protein coupled receptor signaling pathway	24	19	3,50E-10
	Chemotaxis	6	4,8	7,90E-06
Contracted	Signal transduction	21	16,7	1,20E-05

Table 4
KEGG analysis of positively selected genes in *Orestias ascotanensis.*

Term	Count	%	P-value
Cytokine-cytokine receptor interaction	10	1,8	1,0E-02
DNA replication	5	0,9	1,70E-02
Jak-STAT signaling pathway	8	1,4	2,6E-02
Base excision repair	4	0,7	7,3E-02

2.5. microRNA identification

MicroRNAs are key players in posttranscriptional regulation of protein expression and have been associated with stress-responses in animals and plants. Recently, it was suggested that the regulatory changes driven by species-specific miRNAs can contribute towards phenotypic divergence and speciation [48,49]. To gain insight into miRNA mediated regulation of gene expression in *O. ascotanensis*, we sequenced the small RNA component of the transcriptome of this species. Furthermore,

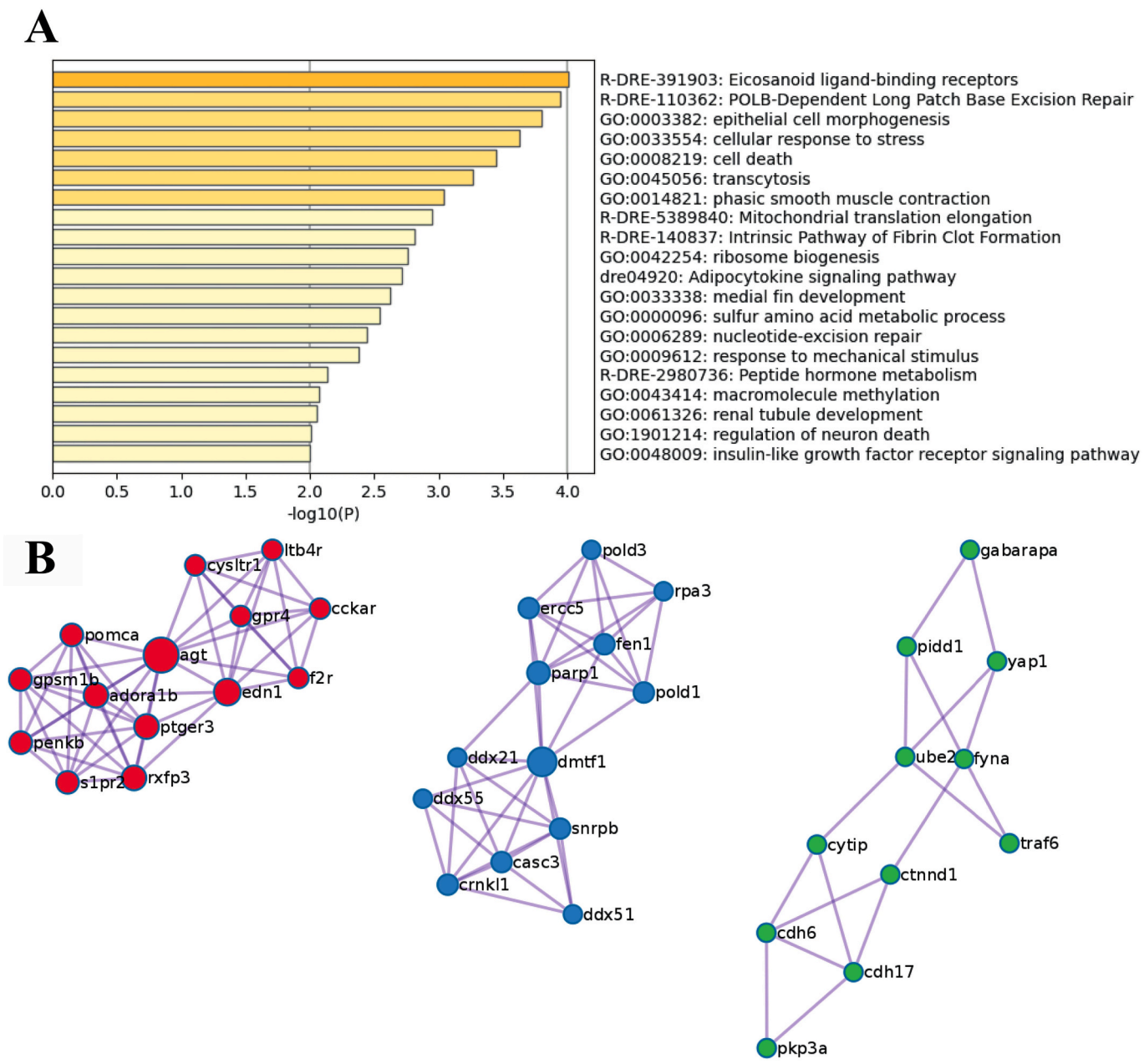

Fig. 3. Gene ontology (GO) enrichment analysis and protein-protein interaction (PPI) enrichment analysis for the genes under positive selection in *Orestias ascotanensis*. (A) Gene ontology enrichment for PSG generated with Metascape. (B) Top 3 PPI networks generated with Metascape. From left to right G-protein coupled receptor signaling (red network), nucleotide excision repair (blue network) and cell-cell communication (green network). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

to further increase the robustness of an identification of potentially novel microRNAs, we also sequenced the small RNAs from a closely-related species: *Orestias gloriae*. *O. gloriae* inhabits the Carcote saltpan (21° 16′ 46”S, 68° 19′ 21”W), located 5 Km north of the northern limit of the Ascotan saltpan.

To identify miRNAs in *O. ascotanensis* and *O. gloriae*, four small RNA-seq libraries (two for each species) were constructed from adult females and then sequenced using an Illumina MiSeq sequencer (Table S5). Approximately 7 million raw reads were obtained per library. After removing the adaptors, at least 6 million trimmed reads were determined in each library and the length distribution of the small RNA-seq reads presented a peak at 22 nt as expected for mature miRNAs (Fig. S1). We then used our annotated *O. ascotanensis* genome to identify miRNAs expressed in both species. Using miRDeep2, we identified 453 and 452 putative miRNAs for *O. ascotanensis* and *O. gloriae* respectively. Of these, 186 were expressed in all four libraries which were used for subsequent analysis (Fig. S2 and Table S7).

2.6. microRNA expression and conservation

Based on the number of reads associated with the mature miRNAs, we compared the abundance of miRNAs in *O. ascotanensis* and *O. gloriae* (Fig. 4A and B). We found a conservation of the majority of abundant miRNAs in both species, where the most expressed miRNAs were miR-10a-5p, miR-192, miR-181b-5p, let-7a and miR-26a-5p. Because the changes in gene regulation mediated by species-specific miRNAs can lead to phenotypic variation and speciation, we wanted to identify novel miRNAs that could be common to the genus *Orestias* but not present in other species. We determined whether these common miRNAs identified in *Orestias* were also found among the miRNAs from 16 other teleosts

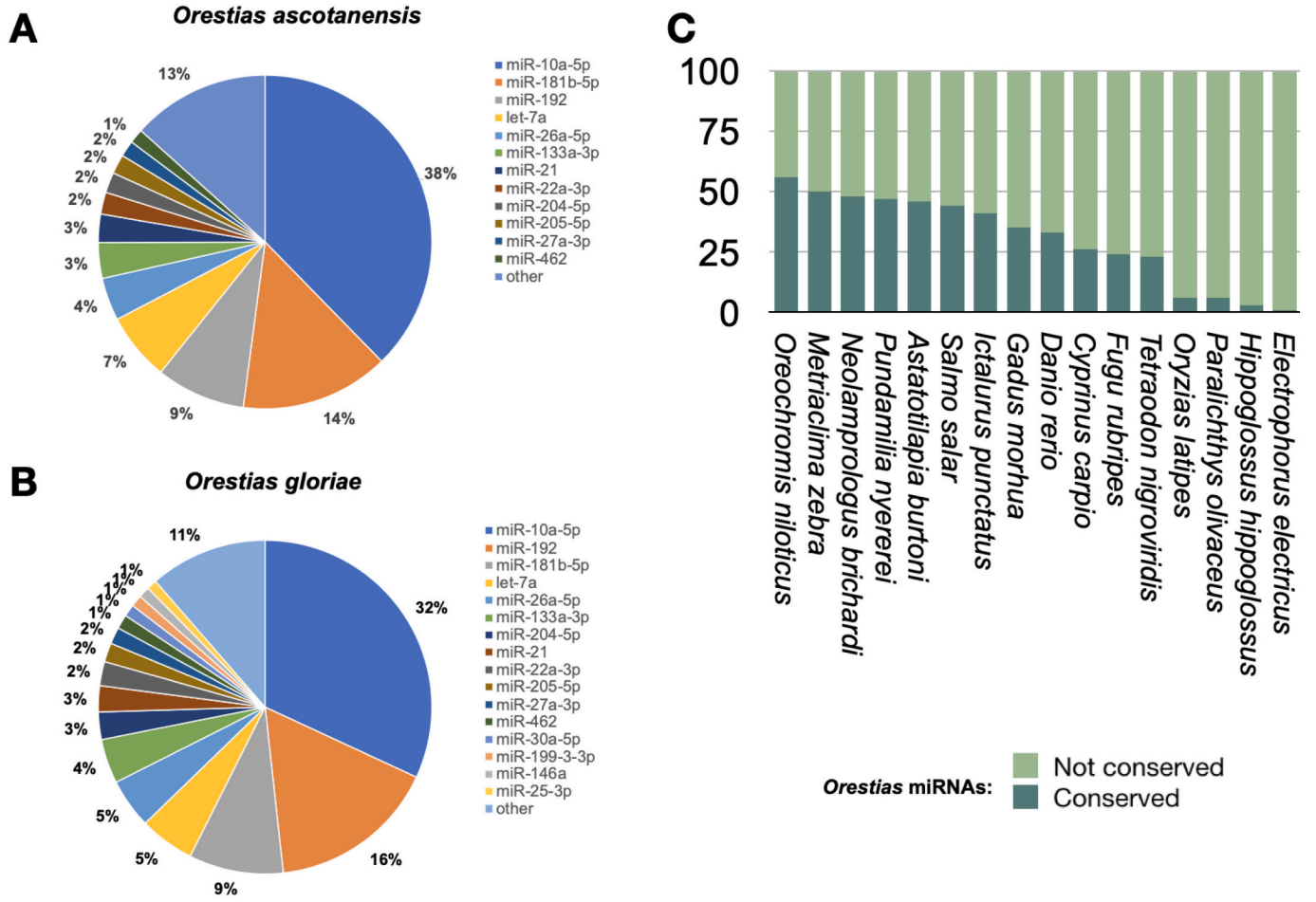

Fig. 4. *Orestias* microRNAs abundance and conservation. Distribution of the miRNA abundance by miRNA family. Top miRNA families with a relative expression above 1% represent the 87% and 89% of the expressed miRNAs respectively in *Orestias ascotanensis* (A) and *Orestias gloriae* (B). (C) Degree of conservation between *Orestias* microRNAs (186 mature miRNA sequences) and those from the other teleost fish species available in miRBase (v22.1). Stack histograms represent the percentage of *Orestias* miRNAs present in each fish species.

contained in miRBase (v22.1) (Fig. 4C). This analysis showed that the species with the most miRNAs in common with *Orestias* is *Oreochromis niloticus*. Thus 56% of the miRNAs identified in *Orestias* had been previously described for *O. niloticus*. This result is in disagreement with our proposed phylogenetic tree, were *Oryzias latipes* appears more closely related to *Orestias* than *O. niloticus*. Moreover, we determined that no more than 6% of the microRNAs expressed in *Orestias* can also be found in the microRNA profile available in miRBase for *O. latipes*. Additionally, we observed that 29% of *Orestias* miRNAs had not been described for other fish species with profiles available in miRbase. Within this group, 21 microRNAs were found expressed in the 4 independent libraries analyzed. As their miRDeep2 score is >1, this group of 21 miRNAs potentially represents a novel set of these regulatory molecules, with reduced similarity to other previously described small non-coding RNAs (Table S6). Detailed information about our mirDeep2 expression analysis in samples from *O. ascotanensis* and *O. gloriae* is shown in Table S7.

3. Discussion

Organisms living in high-altitude environments are subjected to multiple abiotic stresses which may lead to a series of adaptive evolutionary changes in these species. In the present study, we sequenced and annotated the genome of *O. ascotanensis*, a species found exclusively in springs of the Ascotan salt pan. Based on molecular phylogenetic reconstruction, we can place this species within a sister taxa of the subfamily *Poeciliinae*. The habitat where *O. ascotanensis* lives poses unique challenges to life. In addition to the low oxygen and high UV irradiation imposed by high altitude, the springs that these animals inhabit have high concentrations of salts and heavy metals, making Ascotán one of the harshest ecosystems sustaining fish life. To begin to unravel the adaptive mechanisms evolved in *O. ascotanensis* to tolerate these conditions, we set out to sequence the genome of this species. The high quality reference genome reported here can now be used for comparative genomics and as a source of information that will aid in our understanding of the development and physiology of this species.

Interestingly, gene ontology analyses (GO: biological processes) led to the identification of gene families that have suffered significant expansion in this species, particularly genes coding for functions in protein metabolism (protein ubiquitination/deubiquitination) and microtubule-based intracellular movement. Ubiquitination has been demonstrated to be a post-translational modification strongly associated with protein fate and function in cells [50] as well as during the DNA damage response [51]. In addition, it was found that within the expanded microtubule-based movement category there are genes coding for motor proteins like dynein, which plays an essential role during intracellular transport [52]. Positive selection of these motor proteins was also previously observed in the Tibetan Mastiff [53] and it was suggested they play important roles during adaptation to chronic hypoxia.

During their lifetime, organisms' genomes are constantly exposed to various sources of DNA damage that can arise from either endogenous or exogenous sources. One of these exogenous sources of DNA damage is

the solar UV light, which can generate DNA photoproducts and oxidative base modifications [54]. Solar UV damage becomes more relevant in high-altitude environments where UV radiation has been reported to be up to 1.65 times higher than at sea level [55,56]. This is especially relevant in the Chilean Altiplano, where the prevalent cloudless conditions and relatively low water vapor and ozone column makes this one of the places with the highest levels of surface UV irradiance in the world [55]. To maintain the integrity of their genome, cells possess several conserved and effective mechanisms for DNA repair. The Fanconi anemia (FA) pathway, is one of these DNA repair pathways that acts when nucleotides on opposite strands of double stranded DNA undergo covalent linkage [46], also known as inter-strand DNA cross-links (ICLs). These ICLs are a highly toxic form of genomic damage that can inhibit basic cellular processes such as transcription and replication. In this study, we found three members of this pathway (FANCF [43], RPA3 [38] and UBE2T [45]) as positively selected in *O. ascotanensis*.

Additional genes that are also involved in DNA repair pathways were identified as PSG in *Orestias*. This group includes PARP1 [37], POLD1 [40], POLD3 [40] and FEN1 [39] which participate in the base excision repair pathway, and POLI [42], ERCC5 [41] and UVSSA [57] that represent critical components of the DNA damage response to UV in several systems and were here identified during our PPI analysis. Interestingly, FEN1 was previously reported as a PSG in high-altitude snakes inhabiting the Tibetan Plateau [12]. Moreover, functional studies on the FEN1 gene in these snakes showed that this high-altitude allele gives rise to protein that is more stable under UV radiation than from the ancestral allele found in low-elevation species, including lizards and snakes [12]. Genome sequencing of other organisms inhabiting high-altitude environments such as Tibetan hot-spring snake [12], Tibetan Chicken [9], Tibetan antelope [58] and snub-nosed monkeys [8] have also found positive selection pressure on genes associated with the response to DNA damage and repair. Together, these results suggest that the evolution of DNA repair system is an important common mechanism for organisms to adapt to high-altitude environments.

In agreement with these results, a recent characterization of embryonic development of *O. ascotanensis* shows that this fish exhibits early embryonic pigmentation (since day 2 post fertilization), an event that can be associated with prevention to damage caused by the high UV present in the Altiplano [18]. Alternatively, this positive selection of DNA damage-related genes in *O. ascotanensis* might be associated with an enhanced ability to control cell stress responses induced by changes in water salinity which are known to activate DNA repair mechanisms [59]. Together, these data indicate that environmental stressors can trigger cellular stress responses involving pathways that support the preservation of the genome stability [60]. Additionally, these results support the necessity that future studies specifically assess the gene molecular networks that are required to cope with the different stressing conditions at which these fishes live.

Our results also revealed a contraction of gene families related with G protein signaling pathways. This type of contraction has been previously observed in other high-altitude animals. Specifically, loss of members of the olfactory transduction pathway have been reported for hot-spring snake [12], Tibetan chicken [9], wild boars [61], ground tit [62] and musk deer [63]. The loss of the olfactory transduction pathway has been associated with the limited food sources available at high altitude, which in turn results in a limited number of odorant molecules. Accordingly, in *O. ascotanensis* we find a contraction in genes that are associated with both odorant and taste receptors.

Using the assembled genome as a reference, we identified 186 microRNAs expressed in *O. ascotanensis* and in the closely-related species *O. gloriae*. Unexpectedly, in our study we found that *Orestias* shows a higher number of microRNAs that are shared with species like *O. niloticus* than with closer related species in the phylogenetic tree. This result is likely to be due to the limited and unbalanced representation of miRNAs reported in miRBase [48]. Hence, whereas 695 mature miRNAs associated with *O. niloticus* are described in miRBase, only 146 can be found for *O. latipes*. Alternatively, we may speculate that this discrepancy is an indication of evolutionary independent processes in fishes that resulted in a higher variability in the number of microRNA-coding genes with respect to the changes in the protein-coding genome. The fact that several fish genomes thus far studied exhibit large differences in the number and type of microRNAs [48,64,65], argues in favor of this hypothesis. This is an issue that requires further systematic investigation. Interestingly, we identified 21 potentially novel microRNAs that may be specific of the genus *Orestias*, as they are not found for other fish in miRBase. Whether this group of microRNAs is associated with regulatory functions that are relevant for the adaptation of *Orestias* to its challenging environment remains to be further investigated.

In summary, our results shed light on the potential mechanisms operating during adaptation of Andean fish species to their unique environmental conditions. In a broader biological perspective, our results support the concept that the Andean aquatic environments represent new natural laboratories suitable for exploring the evolutionary molecular responses developed by the inhabitant fish species.

4. Materials and methods

4.1. Genome sequencing and assembly

Individuals of *O. ascotanensis* were collected at one of the springs feeding into the Ascotan saltpan (Spring 5; UTM coordinates: 19S 576,991 7,622,725). All the experimental procedures were carried with the official permission from the Chilean Government Animal and Agricultural Agency “Servicio Agrícola y Ganadero” (SAG), in accordance with the guidelines and regulations approved by the Ethical Committees of the Faculty of Sciences of the Universidad de Chile and of the Universidad Andrés Bello, as well as in compliance with the guidelines of the Animal Use Ethics Committees of both institutions. Whole genome sequencing was performed using DNA (gDNA) isolated from a single female individual, a sample that was initially stored in RNAlater (Sigma, St. Louis, USA) and ice until arrival to the laboratory, where it was then stored at −80 °C until required. gDNA was extracted an eviscerated whole fish using DNAeasy Blood and Tissue Kits (Qiagen, Hilden, Germany) and the quality (purity and integrity) was measured in an Agilent 2100 Bioanalyzer (Agilent, Santa Clara, CA, USA). Purified gDNA was sent to Macrogen (Macrogen Inc., Seoul, ROK) for library preparation and subsequent Illumina sequencing. Four genomic libraries were sequenced on a HiSeq2500 with read lengths of 151 base pairs and library insert sizes of 280, 3,000, 5,000 and 10,000 base pairs. The *O. ascotanensis* genome was then assembled by using ALLPATHS-LG [66], as well as the programs RepeatModeler [67] and RepeatMasker [68]. The quality of the assembled genome was subsequently validated using the CEGMA program [30].

4.2. Transcriptome assembly

Total RNA was extracted from five adult females (including the individual used for genome sequencing) collected from the same location described above, by using the TRizol reagent (Invitrogen, Carlsbad, CA, USA). To generate expression libraries that represent the majority of the transcripts produced from the *O. ascotanensis* genome, we collected total RNA from a significant portion of the specimens` bodies and organs. These tissue-samples excluded the entire gastrointestinal system, which was removed immediately after harvesting, to avoid incorporating DNA from organisms ingested by the individuals. The quality of the total RNA from each individual was measured in an Agilent 2100 Bioanalyzer and then sent to Macrogen where individual libraries for each fish were prepared and sequenced.

To reconstruct the transcriptome from RNA Seq data, we employed two strategies: *de novo* assembly and genome-guided assembly. The initial total reads were trimmed based on quality parameters using a minimum of Q20 and a minimum length of 50 base pairs, eliminating

about 2.3% of the reads from all libraries. The filtered reads were then aligned to the assembled genome using Tophat, version v2.0.11 [69]. Transcripts were built using Cufflinks version v2.1.1 [70] on each resulting BAM file. A consensus set was then constructed by merging the predictions with Cuffmerge [70]. A total number of 27,612 putative genes (Table S1) were predicted using the genome-guided strategy. Additionally, filtered reads were used to build the transcriptome *de novo* using the Trinity assembler, version r20131110 [71]. The parameters were set at a minimum contig length of 300 base pairs and a minimum Kmer coverage of 2. A total number of 85,712 putative genes were predicted using this approach. This highlights the importance of having constructed a high-quality reference genome of *O. ascotanensis* as it significantly reduces the fragmentation of transcripts generated during the gene prediction step.

4.3. Identification of O. ascotanensis genes under positive selection

To identify genes in the *O.ascotanensis* lineage that are under positive selection, we used PAML (Phylogenetic Analysis by Maximum Likelihood) [35], which implements a maximum likelihood framework to evaluate adaptive selection based on nonsynonymous by synonymous substitution rate ratio. We used a branch-site model implemented in PAML to identify individual amino-acid sites targeted by positive selection in *O. ascotanensis* branch using 15 other fish species as a background according to PAML author recommendations.

We selected 15 different fish species available on Ensembl based on their closeness to the Cyprinodontiformes order (from which *O. ascotanensis* is a member of) on the fish tree of life [72]. We included in our analysis four species belonging to the Beloniformes order, five species from Cichliformes order, four species from Ovalentaria order (*incertae sedis*) and two species from the Pleuronectiformes. Other orders that may be also close to Cyprinodontiformes such as Atheriniformes, Pholidichthyiformes and Mugiliformes have no genome information available on Ensembl (accessed 26th October 2021).

Orthofinder [73] was used to identify single-copy ortholog (SCO) genes common to the 16 teleost fish genomes (including *O. ascotanensis*). A phylogenetic tree based on these 509 SCO was built by Orthofinder and was used as input for CODEML. To maximize the number of genes analyzed, we selected all ortholog families that presented a SCO for *O. ascotanensis* and at least for 9 additional species. Hence, we finally performed the analysis with 6,739 SCOs shared by at least 10 species including *O. ascotanensis*. The proteins encoded by these SCOs were aligned using MAFFT [74] using default settings and the nucleotide sequences were codon-aligned using PAL2NAL [75]. We took advantage of the branch-site model implemented in CODEML (part of PAML package) that is designed to detect positive selection that affects only a few sites on the specified branches of a given phylogeny. On the species tree of the 16 fish genomes under analysis, *O.ascotanensis* lineage was marked as 'foreground' and the rest of the fish species as 'background' lineages [35]. Because we evaluated SCO shared by at least 10 species, the number of species in each orthologue families varied. Hence, the Newick utilities [76] were used to adequate the phylogeny accordingly but maintaining the topology of the tree.

We performed a likelihood ratio test between model M2a_null (model = 2, NSsites = 2; fix_omega = 0) and M2a_selection (model = 2, NSsites = 2; fix_omega = 1, omega = 1) as recommende [35,77]. To assess the significance of the likelihood ratio test a *p*-value was calculated using the Chi-Square test for twice the difference of likelihood from model M2a_selection *versus* likelihood M2a_null, with one degree of freedom. Finally, we subjected the p-value of the likelihood ratio test to multiple hypothesis corrections in R using the Benjamini–Hochberg method, thus identifying the genes with a significant overall p-value at 5% FDR ($p < 0.05$).

For gene ontology (GO) analysis of PSG we retrieved the *Danio rerio* orthologue gene IDs based on *O. ascotanensis* gene names. These *D. rerio* gene IDs were then used as input for David [78] and Metascape [36] to perform KEGG (Kyoto Encyclopedia of Genes and Genomes) [79], Pathway and Process, and Protein-protein Interaction enrichment analyses.

4.4. Gene identification and functional annotation

Gene identification analyses were carried out by integrating results from the *ab initio* prediction programs as well as from protein and transcript alignments using the PASA [80] and Evidence Modeler [81] algorithms. The *ab initio* gene prediction was obtained using the Augustus (v2.5.5) [82] and GlimmerHMM (v3.0.1) [83] predictors. First, Augustus was trained with the complete transcript data from the RNA-Seq studies to build a custom model for *Orestias*. Three predictions were run using Augustus (Trained model and zebrafish model) and GlimmerHMM (zebrafish model) in the repeat masked genome assembly. A total number of 88,069 *ab initio* predictions were obtained; 26,278 from Augustus-Trained, 24,058 from Augustus-zebrafish and 37,733 from GlimmerHMM. Protein-coding genes were confirmed using BLAST+, GenBlast and GeneWise [84–86]. Sets of matching homologous proteins were identified by aligning all *ab initio* predictions against the UniProt vertebrate database by BLAST+ [85] (*E*-value, 1e-5). A total number of 30,458 different UniProt proteins were detected as putative homologous proteins (targets). These target proteins were then aligned back to the genome using GenBlast [86] to obtain its most likely gene structure and the boundary positions within the masked genome. A total of 34,788 putative gene structures were found. Putative proteins that were supported by these gene structures were subsequently refined using the GeneWise program [84]. We defined as *bona fide* proteins those supported by gene structures that resulted in at least 80% of their hit length aligned. A total number of 27,646 of the putative proteins passed the coverage filter.

Transcript models generated by Cufflinks and Trinity were then integrated using PASA software. PASA models include complete and partial gene structures based on assembled spliced alignments (GMAP aligner) [87]. First, full-length transcripts were determined by aligning transcript models against the UniProt vertebrate database using BLASTX [85]. Thus, a total of 25,542 transcripts were identified as full-length (covering for at least 98% of the hit protein). PASA was then run using the transcript information obtained from RNA-Seq studies and this list of full-length transcripts as input. A total of 104,382 redundant transcripts were obtained from PASA including alternative splicing transcripts, of which 38,225 were defined as complete structures by PASA.

A consensus gene model for each locus was subsequently produced by source-weighted integration using EVidenceModeler (EVM). Weights of 10, 5, 1, 10 and 6 were assigned to PASA transcripts, GeneWise, *ab initio*, PASA full-length and GeneWise full-length alignment predictions, respectively. A total of 20,952 gene models were consolidated by EVM. The gene models were updated by performing two rounds of PASA to include potentially missing genes, splicing-isoforms and UTR tails. Lastly, the predicted proteins were annotated using BLAST (*E*-value, 1e-5) against Swiss-prot [88], NCBI non-redundant protein (NR) [89], KEGG [79] and UniProt protein databases [90]. Protein domains and GO numbers were assigned using Interpro [91]. Gene names were assigned using the best hits to Swiss-prot database.

4.5. Phylogenetic analysis

12 recently sequenced fish species were used to define a phylogenetic position of *O. ascotanensis*. First, protein coding genes for each species were retrieved from Ensembl v77. Second, BLAST [85] and OrthoMCL [92] (using default options) were used to identify a set of 1:1 orthologous among the selected species. A total of 20,411 orthologous groups were obtained, 1,970 of which fit into the 1:1 orthologous definition. Third, this 1:1 orthologous cluster was aligned using the MUSCLE [93] program, selecting the conserved aligned blocks by using GBlocks [94].

Fourth, the phylogenetic tree was then built using a super sequence of 857,415 residues (57,161 for each specie), obtained from the concatenation of all the conserved blocks. This sequence was utilized as input for the FASTree program [95], using the Whelan-and-Goldman [96] model to estimate the species tree considering a Maximum Likelihood approach.

To estimate the rate of gene expansion and contraction in *Orestias*, we performed a gene family analysis using its four closed phylogenetic species within our tree (see Fig. 2). First, gene families for all teleost fish were obtained from the Ensembl v77 compara database. A total of 10,354 gene families were downloaded and for each family the taxon ID was kept. Protein sequence, CDS sequence and gene ID were thus identified. Second, BLAST was used to assign each *Orestias* gene to a teleost gene family. In brief, we built a multi-fasta file that included a reference gene family for each protein. Then, using blastp (1e-10), we identified the best hit between *Orestias* proteins and these family genes, keeping those that exhibited a minimum identity of 50%. A total of 19,587 *Orestias* genes were thus assigned to 7,474 teleost gene families. Third, we built a matrix were each row corresponded to a gene family and each column to a genome. Each cell in the matrix then included the number of genes that a genome has within the family. The genomes of *O. ascotanensis*, *O. niloticus, O. latipes*, *P. formosa* and *X. maculatus* were considered in these analyses. To diminish errors derived from insufficient assembly and gene annotation completeness, only those gene families that were present in at least three of the five genomes were considered. Therefore, a total number of 7,676 families were finally included in the study. On average, 95% of the genes of each fish species were included in our filtered gene families. Four, the CAFE [97] program was used to determine whether significant family expansions/contractions have occurred among the selected species. CAFE requires as input data a phylogenetic tree with branch lengths expressed in years and a matrix including counts for each gene family. Hence, for this purpose we used the sub-tree constructed for the phylogenetic analysis and branch lengths values that were obtained from the timeTree database [98]. The maximum likelihood value of the birth & death parameter, lambda, were estimated from the data by using the –s option; the lambda value converged to 0.00145. Families with larger size variance were defined as those having a *p*-value below 0.05. Our results indicated that, on average, the genome of *O. ascotanensis* has an expansion/contraction value of −0.12 (see Table S2).

4.6. miRNA expression analysis

Total RNA from the *O. ascotanensis* individuals collected at the Ascotan salt pan was isolated using TRIzol (Invitrogen, Carlsbad, CA, USA), RNA concentration quantified in a Qubit 2.0 Fluorometer (Life Technologies, Carlsbad, CA, USA) and RNA integrity determined using Fragment Analyzer Automated CE System (Advanced Analytical, Ames, IA, USA). RNA samples were extracted from a significant portion of the fish bodies and organs (mostly muscle tissue) to identify as much miRNAs as possible. Sequencing libraries were generated using the Illumina TruSeq Small RNA library preparation kit according to the manufacturer's instructions. Single-read fragments were generated on an Illumina MiSeq sequencer according to the user guide. Raw sequence reads were trimmed by removing Illumina adapter sequences as well as low-quality bases. The raw miR-seq data (see Table S2) was processed and the identification and analysis of putative miRs was carried out using miRDeep2 [99] and the fish miRNA database miRBase (v22.1).

Author contributions

ADG, GN, CH and FG performed most of the bioinformatics analyses during genome annotation and transcriptomic studies. RMA performed PAML analysis, microRNA analyses and wrote most parts of the manuscript. CV and PM carried out fish sampling, DNA and RNA preparation for sequencing analyses and helped during the genomic and transcriptomic studies. RM performed RNA isolation and helped in the transcriptomic analyses. RG, AO, MG, VC, AG, M A Mendez, AM contributed to the genomic and transcriptomic analyses, as well as data analysis and preparation of the manuscript. MLA, MAM, performed the data analyses, oversaw the entire work and wrote most parts of the manuscript.

Availability of supporting data

This Whole Genome Shotgun project has been deposited at DDBJ/ENA/GenBank under the accession number SRP062595/PRJNA293101.

Acknowledgements

The authors thank Dr. Juan C. Opazo for his technical support with the Phylogenetic Analysis by Maximum Likelihood (PAML) analysis.

This work is part of the Chilean 1000 Genomes Initiative and is carried out within the framework of the Earth Biogenome Project (www.earthbiogenome.org). Funding has been provided by ANID/FONDAP 15090007 and 15200002, ANID/FONDECYT 1180606 (to MLA), ANID/FONDECYT 3190720 (to RMA) and ANID/FONDECYT 1200419 (to M Mendez). This research was partially supported by the supercomputing infrastructure of the NLHPC (ECM-02) and ANID-PIA AFB 170001.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ygeno.2021.12.018.

References

[1] G.A. Collado, M.A. Méndez, Microgeographic differentiation among closely related species of Biomphalaria (Gastropoda: Planorbidae) from the Andean Altiplano, Zool. J. Linnean Soc. (2013), https://doi.org/10.1111/zoj.12073.
[2] G.A. Collado, M.A. Valladares, M.A. Méndez, A new species of Heleobia (Caenogastropoda: Cochliopidae) from the Chilean Altiplano, Zootaxa. (2016), https://doi.org/10.11646/zootaxa.4137.2.8.
[3] P.A. Sáez, P. Fibla, C. Correa, M. Sallaberry, H. Salinas, A. Veloso, J. Mella, P. Iturra, M.A. Méndez, A new endemic lineage of the Andean frog genus Telmatobius (Anura, Telmatobiidae) from the western slopes of the Central Andes, Zool. J. Linnean Soc. (2014), https://doi.org/10.1111/zoj.12152.
[4] L. Parenti, A taxonomic revision of the Andean killifish genus Orestias (Cyprinodontiformes, Cyprinodontidae), Bull. Am. Mus. Nat. Hist. 178 (1984) 107–214.
[5] M. Valqui, S.M. Caziani, O. Rocha-O, E. Rodriguez-R, Abundance and distribution of the south American Altiplano flamingos, Waterbirds. (2000), https://doi.org/10.2307/1522154.
[6] J.C. Wheeler, Evolution and present situation of the South American camelidae, Biol. J. Linn. Soc. (1995), https://doi.org/10.1016/0024-4066(95)90021-7.
[7] Q. Qiu, G. Zhang, T. Ma, W. Qian, Z. Ye, C. Cao, Q. Hu, J. Kim, D.M. Larkin, L. Auvil, B. Capitanu, J. Ma, H.A. Lewin, X. Qian, Y. Lang, R. Zhou, L. Wang, K. Wang, J. Xia, S. Liao, S. Pan, X. Lu, H. Hou, Y. Wang, X. Zang, Y. Yin, H. Ma, J. Zhang, Z. Wang, Y. Zhang, D. Zhang, T. Yonezawa, M. Hasegawa, Y. Zhong, W. Liu, Y. Zhang, Z. Huang, S. Zhang, R. Long, H. Yang, J.A. Lenstra, D.N. Cooper, Y. Wu, J. Wang, P. Shi, J. Wang, J. Liu, J. Wang, The yak genome and adaptation to life at high altitude, Nat. Genet. (2012), https://doi.org/10.1038/ng.2343.
[8] L. Yu, G.D. Wang, J. Ruan, Y. Bin Chen, C.P. Yang, X. Cao, H. Wu, Y.H. Liu, Z.L. Du, X.P. Wang, J. Yang, S.C. Cheng, L. Zhong, L. Wang, X. Wang, J.Y. Hu, L. Fang, B. Bai, K. Le Wang, N. Yuan, S.F. Wu, B.G. Li, J.G. Zhang, Y.Q. Yang, C.L. Zhang, Y. C. Long, H.S. Li, J.Y. Yang, D.M. Irwin, O.A. Ryder, Y. Li, C.I. Wu, Y.P. Zhang, Genomic analysis of snub-nosed monkeys (Rhinopithecus) identifies genes and processes related to high-altitude adaptation, Nat. Genet. (2016), https://doi.org/10.1038/ng.3615.
[9] Q. Zhang, W. Gou, X. Wang, Y. Zhang, J. Ma, H. Zhang, Y. Zhang, H. Zhang, Genome resequencing identifies unique adaptations of Tibetan chickens to hypoxia and high-dose ultraviolet radiation in high-altitude environments, Genome Biol. Evol. (2016), https://doi.org/10.1093/gbe/evw032.
[10] Y. Hao, Y. Xiong, Y. Cheng, G. Song, C. Jia, Y. Qu, F. Lei, Comparative transcriptomics of 3 high-altitude passerine birds and their low-altitude relatives, Proc. Natl. Acad. Sci. U. S. A. (2019), https://doi.org/10.1073/pnas.1819657116.
[11] W. Yang, Y. Qi, J. Fu, Exploring the genetic basis of adaptation to high elevations in reptiles: a comparative transcriptome analysis of two toad-headed agamas (genus Phrynocephalus), PLoS One (2014), https://doi.org/10.1371/journal.pone.0112218.
[12] J.T. Li, Y.D. Gao, L. Xie, C. Deng, P. Shi, M.L. Guan, S. Huang, J.L. Ren, D.D. Wu, L. Ding, Z.Y. Huang, H. Nie, D.P. Humphreys, D.M. Hillis, W.Z. Wang, Y.P. Zhang, Comparative genomic investigation of high-elevation adaptation in ectothermic

snakes, Proc. Natl. Acad. Sci. U. S. A. (2018), https://doi.org/10.1073/pnas.1805348115.
[13] J. Kang, X. Ma, S. He, Evidence of high-altitude adaptation in the glyptosternoid fish, Creteuchiloglanis macropterus from the Nujiang River obtained through transcriptome analysis, BMC Evol. Biol. (2017), https://doi.org/10.1186/s12862-017-1074-0.
[14] L. Yang, Y. Wang, Z. Zhang, S. He, Comprehensive transcriptome analysis reveals accelerated genic evolution in a Tibet fish, *Gymnodiptychus pachycheilus*, Genome Biol. Evol. (2014), https://doi.org/10.1093/gbe/evu279.
[15] C. Tong, F. Tian, C. Zhang, K. Zhao, The microRNA repertoire of Tibetan naked carp *Gymnocypris przewalskii*: a case study in Schizothoracinae fish on the Tibetan plateau, PLoS One 12 (2017), https://doi.org/10.1371/journal.pone.0174534.
[16] F. Cruz-Jofré, P. Morales, I. Vila, Y. Esquer-Garrigos, B. Hugueny, P. Gaubert, E. Poulin, M.A. Méndez, Geographical isolation and genetic differentiation: the case of *Orestias ascotanensis* (Teleostei: Cyprinodontidae), an Andean killifish inhabiting a highland salt pan, Biol. J. Linn. Soc. (2016), https://doi.org/10.1111/bij.12704.
[17] B. Keller, D. Soto, Hydrogeologic influences on the preservation of *Orestias ascotanensis* (Teleostei: Cyprinodontidae), in Salar de Ascotán, northern Chile, Rev. Chil. Hist. Nat. 71 (1998) 147–156.
[18] C. Macaya, N. Lam, I. Vila, Embryological development of the high-altitude killifish *Orestias ascotanensis* Parenti 1984 (Teleostei: Cyprinodontidae), Environ. Biol. Fish (2019), https://doi.org/10.1007/s10641-019-00859-6.
[19] P. Rojas, S. Scott, I. Tobar, Ú. Romero, I. Vila, Head morphometry of Orestias (Cyprinodontiformes). Response to extreme southern Altiplano systems? Environ. Biol. Fish (2020) https://doi.org/10.1007/s10641-020-00997-2.
[20] S. Scott, P. Rojas, I. Vila, Meristic and morphological differentiation of Orestias species (Teleostei; Cyprinodontiformes) from the southern Altiplano, Environ. Biol. Fish (2020), https://doi.org/10.1007/s10641-020-00995-4.
[21] C. Araya-Jaime, N. Lam, I.V. Pinto, M.A. Méndez, P. Iturra, Chromosomal organization of four classes of repetitive DNA sequences in killifish *Orestias ascotanensis* Parenti, 1984 (Cyprinodontiformes, Cyprinodontidae), Comp. Cytogenet. (2017), https://doi.org/10.3897/compcytogen.v11i3.11729.
[22] F. Meng, J. Chen, Z. Miao, K. Zhu, Y. Huang, Q. Wang, B. Liu, Y. Liu, Phylogenetic analysis of the complete mitochondrial genome of *Orestias ascotanensis* (Cyprinodontiformes, Cyprinodontidae), Mitochondrial DNA Part B Resour. (2019), https://doi.org/10.1080/23802359.2019.1687033.
[23] A. Lüssen, T.M. Falk, W. Villwock, Phylogenetic patterns in populations of Chilean species of the genus Orestias (Teleostei: Cyprinodontidae): results of mitochondrial DNA analysis, Mol. Phylogenet. Evol. (2003), https://doi.org/10.1016/S1055-7903(03)00077-0.
[24] C. Quezada-Romegialli, C.J. Guerrero, D. Véliz, I. Vila, The complete mitochondrial genome of the killifish Orestias sp. (Cyprinodontiformes, Cyprinodontidae) from the high Andean range, Mitochondrial DNA (2015), https://doi.org/10.3109/19401736.2015.1053119.
[25] I. Vila, P. Morales, S. Scott, E. Poulin, D. Véliz, C. Harrod, M.A. Méndez, Phylogenetic and phylogeographic analysis of the genus Orestias (Teleostei: Cyprinodontidae) in the southern Chilean Altiplano: the relevance of ancient and recent divergence processes in speciation, J. Fish Biol. (2013), https://doi.org/10.1111/jfb.12031.
[26] I. Vila, S. Scott, N. Lam, P. Iturra, M.A. Mendez, Karyological and morphological analysis of divergence among species of the killifish genus Orestias (Teleostei: Cyprinodontidae) from the southern Altiplano, Orig. Phylogenetic Interrelat. Teleosts. (2010) 471–480.
[27] S.A. Pavey, H. Collin, P. Nosil, S.M. Rogers, The role of gene expression in ecological speciation, Ann. N. Y. Acad. Sci. (2010), https://doi.org/10.1111/j.1749-6632.2010.05765.x.
[28] K.L. Mack, P. Campbell, M.W. Nachman, Gene regulation and speciation in house mice, Genome Res. (2016), https://doi.org/10.1101/gr.195743.115.
[29] J.H. Davidson, C.N. Balakrishnan, Gene regulatory evolution during speciation in a songbird, G3 genes, genomes, Genet. (2016), https://doi.org/10.1534/g3.116.027946.
[30] G. Parra, K. Bradnam, I. Korf, CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes, Bioinformatics. 23 (2007) 1061–1067, https://doi.org/10.1093/bioinformatics/btm071.
[31] P.H.F. Lucinda, R.E. Reis, Systematics of the subfamily Poeciliinae Bonaparte (Cyprinodontiformes: Poeciliidae), with an emphasis on the tribe Cnesterodontini Hubbs, Neotrop. Ichthyol. (2005), https://doi.org/10.1590/s1679-62252005000100001.
[32] J. Zhang, Evolution by gene duplication: an update, Trends Ecol. Evol. (2003), https://doi.org/10.1016/S0169-5347(03)00033-8.
[33] J.P. Demuth, M.W. Hahn, The life and death of gene families, BioEssays. (2009), https://doi.org/10.1002/bies.080085.
[34] F.A. Kondrashov, Gene duplication as a mechanism of genomic adaptation to a changing environment, Proc. R. Soc. B Biol. Sci. (2012), https://doi.org/10.1098/rspb.2012.1108.
[35] Z. Yang, PAML 4: phylogenetic analysis by maximum likelihood, Mol. Biol. Evol. (2007), https://doi.org/10.1093/molbev/msm088.
[36] Y. Zhou, B. Zhou, L. Pache, M. Chang, A.H. Khodabakhshi, O. Tanaseichuk, C. Benner, S.K. Chanda, Metascape provides a biologist-oriented resource for the analysis of systems-level datasets, Nat. Commun. 10 (2019) 1–10.
[37] A.R. Chaudhuri, A. Nussenzweig, The multifaceted roles of PARP1 in DNA repair and chromatin remodelling, Nat. Rev. Mol. Cell Biol. 18 (2017) 610–621.
[38] E. Fanning, V. Klimovich, A.R. Nager, A dynamic model for replication protein a (RPA) function in DNA processing pathways, Nucleic Acids Res. 34 (2006) 4126–4137.
[39] J.-H. Yoon, P.M. Swiderski, B.E. Kaplan, M. Takao, A. Yasui, B. Shen, G.P. Pfeifer, Processing of UV damage in vitro by FEN-1 proteins as part of an alternative DNA excision repair pathway, Biochemistry. 38 (1999) 4809–4817.
[40] E. Tumini, S. Barroso, C. Pérez-Calero, A. Aguilera, Roles of human POLD1 and POLD3 in genome stability, Sci. Rep. 6 (2016) 1–13.
[41] D. Mu, D.S. Hsu, A. Sancar, Reaction mechanism of human DNA repair excision nuclease, J. Biol. Chem. 271 (1996) 8285–8294.
[42] K. Bebenek, A. Tissier, E.G. Frank, J.P. McDonald, R. Prasad, S.H. Wilson, R. Woodgate, T.A. Kunkel, 5'-deoxyribose phosphate Lyase activity of human DNA polymerase ι in vitro, Science (80) 291 (2001) 2156–2159.
[43] J.P. De Winter, M.A. Rooimans, L. Van Der Weel, C.G.M. Van Berkel, N. Alon, L. Bosnoyan-Collins, J. De Groot, Y. Zhi, Q. Waisfisz, J.C. Pronk, The Fanconi anaemia gene FANCF encodes a novel protein with homology to ROM, Nat. Genet. 24 (2000) 15–16.
[44] P.R. Andreassen, A.D. D'Andrea, T. Taniguchi, ATR couples FANCD2 monoubiquitination to the DNA-damage response, Genes Dev. 18 (2004) 1958–1963.
[45] Y.J. Machida, Y. Machida, Y. Chen, A.M. Gurtan, G.M. Kupfer, A.D. D'Andrea, A. Dutta, UBE2T is the E2 in the Fanconi anemia pathway and undergoes negative autoregulation, Mol. Cell 23 (2006) 589–596.
[46] G.-L. Moldovan, A.D. D'Andrea, How the fanconi anemia pathway guards the genome, Annu. Rev. Genet. 43 (2009) 223–249.
[47] V. Fernández Zenoff, F. Siñeriz, M.E. Farías, Diverse responses to UV-B radiation and repair mechanisms of bacteria isolated from high-altitude aquatic environments, Appl. Environ. Microbiol. (2006), https://doi.org/10.1128/AEM.01333-06.
[48] P. Xiong, R.F. Schneider, C.D. Hulsey, A. Meyer, P. Franchini, Conservation and novelty in the microRNA genomic landscape of hyperdiverse cichlid fishes, Sci. Rep. (2019), https://doi.org/10.1038/s41598-019-50124-0.
[49] E. Mor, N. Shomron, Species-specific microRNA regulation influences phenotypic variability: perspectives on species-specific microRNA regulation Prospects & Overviews E. Mor and N. Shomron, BioEssays (2013), https://doi.org/10.1002/bies.201200157.
[50] A. Hershko, A. Ciechanover, The ubiquitin system for protein degradation, Annu. Rev. Biochem. 61 (1992) 761–807.
[51] A. Al-Hakim, C. Escribano-Diaz, M.C. Landry, L. Donnell, S. Panier, R.K. Szilard, D. Durocher, The ubiquitous role of ubiquitin in the DNA damage response, DNA Repair (Amst) (2010), https://doi.org/10.1016/j.dnarep.2010.09.011.
[52] R.B. Vallee, J.C. Williams, D. Varma, L.E. Barnhart, Dynein: an ancient motor protein involved in multiple modes of transport, J. Neurobiol. 58 (2004) 189–200.
[53] Y. Li, D.D. Wu, A.R. Boyko, G.D. Wang, S.F. Wu, D.M. Irwin, Y.P. Zhang, Population variation revealed high-altitude adaptation of tibetan mastiffs, Mol. Biol. Evol. (2014), https://doi.org/10.1093/molbev/msu070.
[54] C. Kielbassa, L. Roza, B. Epe, Wavelength dependence of oxidative DNA damage induced by UV and visible light, Carcinogenesis. (1997), https://doi.org/10.1093/carcin/18.4.811.
[55] N.A. Cabrol, E.A. Grin, G. Chong, E. Minkley, A.N. Hock, Y. Yu, L. Bebout, E. Fleming, D.P. Häder, C. Demergasso, J. Gibson, L. Escudero, C. Dorador, D. Lim, C. Woosley, R.L. Morris, C. Tambley, V. Gaete, M.E. Galvez, E. Smith, I. Uskin-Peate, C. Salazar, G. Dawidowicz, J. Majerowicz, The high-lakes project, J. Geophys. Res. Biogeosci. 114 (2009), https://doi.org/10.1029/2008JG000818.
[56] C. Demergasso, C. Dorador, D. Meneses, J. Blamey, N. Cabrol, L. Escudero, G. Chong, Prokaryotic diversity pattern in high-altitude ecosystems of the Chilean Altiplano, J. Geophys. Res. Biogeosci. (2010), https://doi.org/10.1029/2008JG000836.
[57] X. Zhang, K. Horibata, M. Saijo, C. Ishigami, A. Ukai, S. Kanno, H. Tahara, E. G. Neilan, M. Honma, T. Nohmi, Mutations in UVSSA cause UV-sensitive syndrome and destabilize ERCC6 in transcription-coupled DNA repair, Nat. Genet. 44 (2012) 593–597.
[58] R.-L. Ge, Q. Cai, Y.-Y. Shen, A. San, L. Ma, Y. Zhang, X. Yi, Y. Chen, L. Yang, Y. Huang, Draft genome sequence of the Tibetan antelope, Nat. Commun. 4 (2013) 1–7.
[59] T.G. Evans, D. Kültz, The cellular stress response in fish exposed to salinity fluctuations, J. Exp. Zool. Part A Ecol. Integr. Physiol. (2020), https://doi.org/10.1002/jez.2350.
[60] J.C. Balasch, L. Tort, Netting the stress responses in fish, Front. Endocrinol. (Lausanne). (2019), https://doi.org/10.3389/fendo.2019.00062.
[61] M. Li, S. Tian, L. Jin, G. Zhou, Y. Li, Y. Zhang, T. Wang, C.K.L. Yeung, L. Chen, J. Ma, J. Zhang, A. Jiang, J. Li, C. Zhou, J. Zhang, Y. Liu, X. Sun, H. Zhao, Z. Niu, P. Lou, L. Xian, X. Shen, S. Liu, S. Zhang, M. Zhang, L. Zhu, S. Shuai, L. Bai, G. Tang, H. Liu, Y. Jiang, M. Mai, J. Xiao, X. Wang, Q. Zhou, Z. Wang, P. Stothard, M. Xue, X. Gao, Z. Luo, Y. Gu, H. Zhu, X. Hu, Y. Zhao, G.S. Plastow, J. Wang, Z. Jiang, K. Li, N. Li, X. Li, R. Li, Genomic analyses identify distinct patterns of selection in domesticated pigs and Tibetan wild boars, Nat. Genet. (2013), https://doi.org/10.1038/ng.2811.
[62] Y. Qu, H. Zhao, N. Han, G. Zhou, G. Song, B. Gao, S. Tian, J. Zhang, R. Zhang, X. Meng, Y. Zhang, Y. Zhang, X. Zhu, W. Wang, D. Lambert, P.G.P. Ericson, S. Subramanian, C. Yeung, H. Zhu, Z. Jiang, R. Li, F. Lei, Ground tit genome reveals avian adaptation to living at high altitudes in the Tibetan plateau, Nat. Commun. (2013), https://doi.org/10.1038/ncomms3071.
[63] L. Yi, M. Dalai, R. Su, W. Lin, M. Erdenedalai, B. Luvsantseren, C. Chimedtseren, Z. Wang, S. Hasi, Whole-genome sequencing of wild Siberian musk deer (*Moschus moschiferus*) provides insights into its genetic features, BMC Genomics (2020), https://doi.org/10.1186/s12864-020-6495-2.
[64] T.T. Bizuayehu, I. Babiak, MicroRNA in teleost fish, Genome Biol. Evol. 6 (2014) 1911–1937.

[65] M. Baumgart, E. Barth, A. Savino, M. Groth, P. Koch, A. Petzold, I. Arisi, M. Platzer, M. Marz, A. Cellerino, A miRNA catalogue and ncRNA annotation of the short-living fish *Nothobranchius furzeri*, BMC Genomics 18 (2017) 693.
[66] S. Gnerre, I. MacCallum, D. Przybylski, F.J. Ribeiro, J.N. Burton, B.J. Walker, T. Sharpe, G. Hall, T.P. Shea, S. Sykes, A.M. Berlin, D. Aird, M. Costello, R. Daza, L. Williams, R. Nicol, A. Gnirke, C. Nusbaum, E.S. Lander, D.B. Jaffe, High-quality draft assemblies of mammalian genomes from massively parallel sequence data, Proc. Natl. Acad. Sci. U. S. A. (2011), https://doi.org/10.1073/pnas.1017351108.
[67] A.F.A. Smit, R. Hubley, P. Green, RepeatModeler Open-1.0. 2008–2015 1, 2015, p. 2018. http://www.repeatmasker.org.
[68] A.F.A. Smit, R. Hubley, P. Green, RepeatMasker Open-4.0. 2013–2015. http://www.repeatmasker.org, 2021.
[69] C. Trapnell, L. Pachter, S.L. Salzberg, TopHat: discovering splice junctions with RNA-Seq, Bioinformatics. (2009), https://doi.org/10.1093/bioinformatics/btp120.
[70] C. Trapnell, A. Roberts, L. Goff, G. Pertea, D. Kim, D.R. Kelley, H. Pimentel, S. L. Salzberg, J.L. Rinn, L. Pachter, Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and cufflinks, Nat. Protoc. (2012), https://doi.org/10.1038/nprot.2012.016.
[71] B.J. Haas, A. Papanicolaou, M. Yassour, M. Grabherr, P.D. Blood, J. Bowden, M. B. Couger, D. Eccles, B. Li, M. Lieber, M.D. Macmanes, M. Ott, J. Orvis, N. Pochet, F. Strozzi, N. Weeks, R. Westerman, T. William, C.N. Dewey, R. Henschel, R. D. Leduc, N. Friedman, A. Regev, De novo transcript sequence reconstruction from RNA-seq using the trinity platform for reference generation and analysis, Nat. Protoc. (2013), https://doi.org/10.1038/nprot.2013.084.
[72] D.L. Rabosky, J. Chang, P.O. Title, P.F. Cowman, L. Sallan, M. Friedman, K. Kaschner, C. Garilao, T.J. Near, M. Coll, An inverse latitudinal gradient in speciation rate for marine fishes, Nature. 559 (2018) 392–395.
[73] D.M. Emms, S. Kelly, OrthoFinder: phylogenetic orthology inference for comparative genomics, Genome Biol. 20 (2019) 1–14.
[74] K. Katoh, H. Toh, Recent developments in the MAFFT multiple sequence alignment program, Brief. Bioinform. 9 (2008) 286–298.
[75] M. Suyama, D. Torrents, P. Bork, PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments, Nucleic Acids Res. 34 (2006) W609–W612.
[76] T. Junier, E.M. Zdobnov, The Newick utilities: high-throughput phylogenetic tree processing in the UNIX shell, Bioinformatics. 26 (2010) 1669–1670.
[77] D.R. Valenzano, B.A. Benayoun, P.P. Singh, E. Zhang, P.D. Etter, C.-K. Hu, M. Clément-Ziza, D. Willemsen, R. Cui, I. Harel, The African turquoise killifish genome provides insights into evolution and genetic architecture of lifespan, Cell. 163 (2015) 1539–1554.
[78] D.W. Huang, B.T. Sherman, R.A. Lempicki, Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources, Nat. Protoc. 4 (2009) 44–57.
[79] M. Kanehisa, KEGG: Kyoto encyclopedia of genes and genomes, Nucleic Acids Res. (2000), https://doi.org/10.1093/nar/28.1.27.
[80] B.J. Haas, A.L. Delcher, S.M. Mount, J.R. Wortman, R.K. Smith, L.I. Hannick, R. Maiti, C.M. Ronning, D.B. Rusch, C.D. Town, S.L. Salzberg, O. White, Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies, Nucleic Acids Res. (2003), https://doi.org/10.1093/nar/gkg770.
[81] B.J. Haas, S.L. Salzberg, W. Zhu, M. Pertea, J.E. Allen, J. Orvis, O. White, C. R. Robin, J.R. Wortman, Automated eukaryotic gene structure annotation using EVidenceModeler and the program to assemble spliced alignments, Genome Biol. (2008), https://doi.org/10.1186/gb-2008-9-1-r7.
[82] M. Stanke, O. Keller, I. Gunduz, A. Hayes, S. Waack, B. Morgenstern, AUGUSTUS: a b initio prediction of alternative transcripts, Nucleic Acids Res. (2006), https://doi.org/10.1093/nar/gkl200.
[83] W.H. Majoros, M. Pertea, S.L. Salzberg, TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders, Bioinformatics. (2004), https://doi.org/10.1093/bioinformatics/bth315.
[84] E. Birney, M. Clamp, R. Durbin, GeneWise and Genomewise, Genome Res. (2004), https://doi.org/10.1101/gr.1865504.
[85] C. Camacho, G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, T. L. Madden, BLAST+: architecture and applications, BMC Bioinformatics. (2009), https://doi.org/10.1186/1471-2105-10-421.
[86] R. She, J.S.C. Chu, B. Uyar, J. Wang, K. Wang, N. Chen, genBlastG: using BLAST searches to build homologous gene models, Bioinformatics. (2011), https://doi.org/10.1093/bioinformatics/btr342.
[87] T.D. Wu, C.K. Watanabe, GMAP: a genomic mapping and alignment program for mRNA and EST sequences, Bioinformatics. (2005), https://doi.org/10.1093/bioinformatics/bti310.
[88] A. Bairoch, The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000, Nucleic Acids Res. (2000), https://doi.org/10.1093/nar/28.1.45.
[89] K.D. Pruitt, T. Tatusova, D.R. Maglott, NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins, Nucleic Acids Res. (2007), https://doi.org/10.1093/nar/gkl842.
[90] Uniprot Consortium, UniProt: a hub for protein information, Nucleic Acids Res. (2015), https://doi.org/10.1093/nar/gku989.
[91] R.D. Finn, T.K. Attwood, P.C. Babbitt, A. Bateman, P. Bork, A.J. Bridge, H. Y. Chang, Z. Dosztanyi, S. El-Gebali, M. Fraser, J. Gough, D. Haft, G.L. Holliday, H. Huang, X. Huang, I. Letunic, R. Lopez, S. Lu, A. Marchler-Bauer, H. Mi, J. Mistry, D.A. Natale, M. Necci, G. Nuka, C.A. Orengo, Y. Park, S. Pesseat, D. Piovesan, S. C. Potter, N.D. Rawlings, N. Redaschi, L. Richardson, C. Rivoire, A. Sangrador-Vegas, C. Sigrist, I. Sillitoe, B. Smithers, S. Squizzato, G. Sutton, N. Thanki, P. D. Thomas, S.C.E. Tosatto, C.H. Wu, I. Xenarios, L.S. Yeh, S.Y. Young, A.L. Mitchell, InterPro in 2017-beyond protein family and domain annotations, Nucleic Acids Res. (2017), https://doi.org/10.1093/nar/gkw1107.
[92] L. Li, C.J. Stoeckert, D.S. Roos, OrthoMCL: identification of ortholog groups for eukaryotic genomes, Genome Res. (2003), https://doi.org/10.1101/gr.1224503.
[93] R.C. Edgar, MUSCLE: a multiple sequence alignment method with reduced time and space complexity, BMC Bioinformatics. (2004), https://doi.org/10.1186/1471-2105-5-113.
[94] G. Talavera, J. Castresana, Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments, Syst. Biol. 56 (2007) 564–577.
[95] M.N. Price, P.S. Dehal, A.P. Arkin, FastTree 2 - approximately maximum-likelihood trees for large alignments, PLoS One (2010), https://doi.org/10.1371/journal.pone.0009490.
[96] S. Whelan, N. Goldman, A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach, Mol. Biol. Evol. (2001), https://doi.org/10.1093/oxfordjournals.molbev.a003851.
[97] T. De Bie, N. Cristianini, J.P. Demuth, M.W. Hahn, CAFE: a computational tool for the study of gene family evolution, Bioinformatics. (2006), https://doi.org/10.1093/bioinformatics/btl097.
[98] S. Kumar, G. Stecher, M. Suleski, S.B. Hedges, TimeTree: a resource for timelines, Timetrees, and divergence times, Mol. Biol. Evol. (2017), https://doi.org/10.1093/molbev/msx116.
[99] M.R. Friedländer, S.D. MacKowiak, N. Li, W. Chen, N. Rajewsky, MiRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades, Nucleic Acids Res. (2012), https://doi.org/10.1093/nar/gkr688.