RESEARCH ARTICLE

Chromosome-level genome assembly and population genomics of Mongolian racerunner (*Eremias argus*) provide insights into high-altitude adaptation in lizards

Weiming Li[1,2†], Juan Du[1,2†], Lingyun Yang[3†], Qiqi Liang[3], Mengyuan Yang[1,2], Xuming Zhou[1*] and Weiguo Du[1,4*]

Abstract

Background Although the extreme environmental adaptation of organisms is a hot topic in evolutionary biology, genetic adaptation to high-altitude environment remains poorly characterized in ectothermic animals. Squamates are among the most diverse terrestrial vertebrates, with tremendous ecological plasticity and karyotype diversity, and are a unique model system to investigate the genetic footprints of adaptation.

Results We report the first chromosome-level assembly of the Mongolian racerunner (*Eremias argus*) and our comparative genomics analyses found that multiple chromosome fissions/fusions events are unique to lizards. We further sequenced the genomes of 61 Mongolian racerunner individuals that were collected from altitudes ranging from ~80 to ~2600 m above sea level (m.a.s.l.). Population genomic analyses revealed many novel genomic regions under strong selective sweeps in populations endemic to high altitudes. Genes embedded in those genomic regions are mainly associated with energy metabolism and DNA damage repair pathways. Moreover, we identified and validated two substitutions of *PHF14* that may enhance the lizards' tolerance to hypoxia at high altitudes.

Conclusions Our study reveals the molecular mechanism of high-altitude adaptation in ectothermic animal using lizard as a research subject and provides a high-quality lizard genomic resource for future research.

Keywords Lizard, Whole-genome sequencing, Population genomics, High-altitude adaptation

†Weiming Li, Juan Du, and Lingyun Yang contributed equally to this work.

*Correspondence:
Xuming Zhou
zhouxuming@ioz.ac.cn
Weiguo Du
duweiguo@ioz.ac.cn
¹ Key Laboratory of Animal Ecology and Conservation Biology, Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, China
² University of Chinese Academic of Sciences, Beijing, China
³ Novogene Bioinformatics Institute, Beijing 100083, China
⁴ Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences, Kunming 650223, China

Background

An important area in evolutionary biology is understanding the way in which animals have adapted to extreme environment. For example, high altitudes bring multiple challenges, including low oxygen, high ultraviolet radiation, low pressure, and cold temperatures for organisms. However, species that have adapted to such harsh environments usually exhibit numerous physiological capabilities, such as increased heart rates, augmentation of pulmonary ventilation, and increased affinity of hemoglobin for oxygen [1–3]. Previous studies have focused on mammals and birds [4–14] and several genes have been

Li *et al. BMC Biology*   (2023) 21:40

Page 2 of 14

identified by genomic studies, such as *EPAS1*, *MTHFR*, and *EGLN1* in humans [5, 6, 12], *EPAS1* in dogs and horses [9], *TRPC1*, *KCNMA1*, and *SHEEPV1R18* in pigs [8], *SOCS2* in sheep [11], and *SLC35F1* in Tibetan chickens [10]. These genes are the targets of natural selection in high-altitude environments, however, we have a limited understanding of the way in which ectothermic animals such as amphibians and reptiles respond to high altitudes [15, 16].

Squamates, comprising lizards and snakes, have diverse morphological characters and lifestyles, and therefore, have been widely used as research models in evolution and adaptation studies [17–21]. While physiological ecologists have explained and described the biophysical and biochemical patterns of high-altitude adaptation in squamates, the genetic mechanisms underlying the high-altitude adaptive process are still not widely known [22, 23]. Among squamate species, Mongolian racerunner (*Eremias argus*) is the lizard that is extensively distributed in Russia, Korea, Mongolia, and China over a wide range of elevations. This species is currently the only known lizard species that is distributed from the Eastern Plain to the Qinghai-Tibet Plateau in China. Such species could serve as a model for the investigation of genetic adaptation to high altitude in lizards. Here, we first report a chromosome-level genome assembly of the Mongolian racerunner generated using a combination of the latest sequencing and scaffold technologies, including single-molecule real-time, linked-read, second-generation sequencing, and Hi-C scaffolding. We also sequenced genomes of 61 Mongolian racerunner individuals collected from altitudes ranging from ~80 to ~2600 m above sea level (m.a.s.l.). Using these genome resources, we sought to discover the genome variations that respond to high-altitude adaptation in Mongolian racerunners. Our study provides preliminary evidence for extreme environmental adaptation and expands our knowledge of the genome characteristics in lizards.

Results

Genome assembly and chromosomes evolution

We generated a 1.63-Gb genome assembly of the Mongolian racerunner using a sample from the desert in Hohhot, Inner Mongolia Autonomous Region of China. We used a combination of the four latest sequencing technologies: single-molecule real-time sequencing (PacBio Sequel), paired sequencing (Illumina HiSeq), linked-read sequencing (10X Genomics), and Hi-C (Phase Genomics, Inc.). This new genome assembly produced 19 chromosomes with contig N50 of 3.07 Mb and scaffold N50 of 96.73 Mb (Fig. 1a; Additional file 1: Figure S1−4; Additional file 2: Table S1−2). Karyotypic analysis demonstrated that the Mongolian racerunner

possesses a ZW/ZZ sex determination system, and the two sex chromosomes differ in size and shape [24, 25]. To identify the potential Z chromosome in our chromosome-scale assembly, we compared the sequence coverage between male and female individuals and determined chromosome 15 as the Z chromosome (Additional file 1: Figure S5).

To verify the completeness of our assembly, we checked the Core Eukaryotic Genes (CEGs) using CEGMA [27] and Benchmarking Universal Single-Copy Orthologs (BUSCOs) using BUSCO [28]. This analysis recovered 226 (91.13%) complete CEGs and 2441 (94.39%) the complete single-copy orthologs (Additional file 2: Table S3-4), indicating that the Mongolian racerunner genome has a high coverage of the coding region. We also checked the accuracy of the assembly of this genome using Illumina short reads and obtained a mapping rate of 98.65% and a mapping coverage rate of 99.74% (Additional file 2: Table S5), suggesting that the assembled genome has good consistency. Tandem repeat sequences and interspersed repeat sequences account for approximately 47.37% of the Mongolian racerunner genome assembly, and 46.92% of them are transposable elements (Additional file 1: Figure S6; Additional file 2: Table S6). Long interspersed nuclear elements (LINEs) are the most abundant element, making up approximately 33.18% of the genome assembly. Using a comprehensive annotation strategy, 20,107 protein-coding genes were predicted (Additional file 2: Table S7), with an average length of 28.72 kb, an average CDS length of 1.43 kb, an average of 8.63 exons per gene, and an average intron length of 3.57 kb (Additional file 1: Figure S7). Most of the predicted genes (~99.68%) were functionally annotated against known protein databases (Additional file 2: Table S8).

In order to characterize genome synteny between the Mongolian racerunner genome and its relatives, pairwise synteny comparisons were conducted between the Mongolian racerunner, rattlesnake, European common wall lizard, and chicken genomes. Multiple chromosome fissions/fusions events were observed in autosomes and sex chromosomes (Additional file 1: Figure S8-9). For autosomes, genes located at chromosome breakpoint regions (e.g., *RUNX2*, *TFAP2B*, *OSR1*, *WDPCP*, *TBX5*, and *TBX3*) of the Mongolian racerunner genome assembly are enriched in "forelimb morphogenesis," "embryonic forelimb morphogenesis," "limb morphogenesis," and "hindlimb morphogenesis" (Fig. 1b). These limb morphogenesis related genes are located on Chr1 in the snake genome, while their orthologs in the lizard genome are located at different chromosomes: i.e., four genes (*RUNX2*, *TFAP2B*, *OSR1*, and *WDPCP*) are located on Chr3, and two genes (*TBX5* and *TBX3*) were transferred to Chr17 (Fig. 1c). Genes located at chromosome
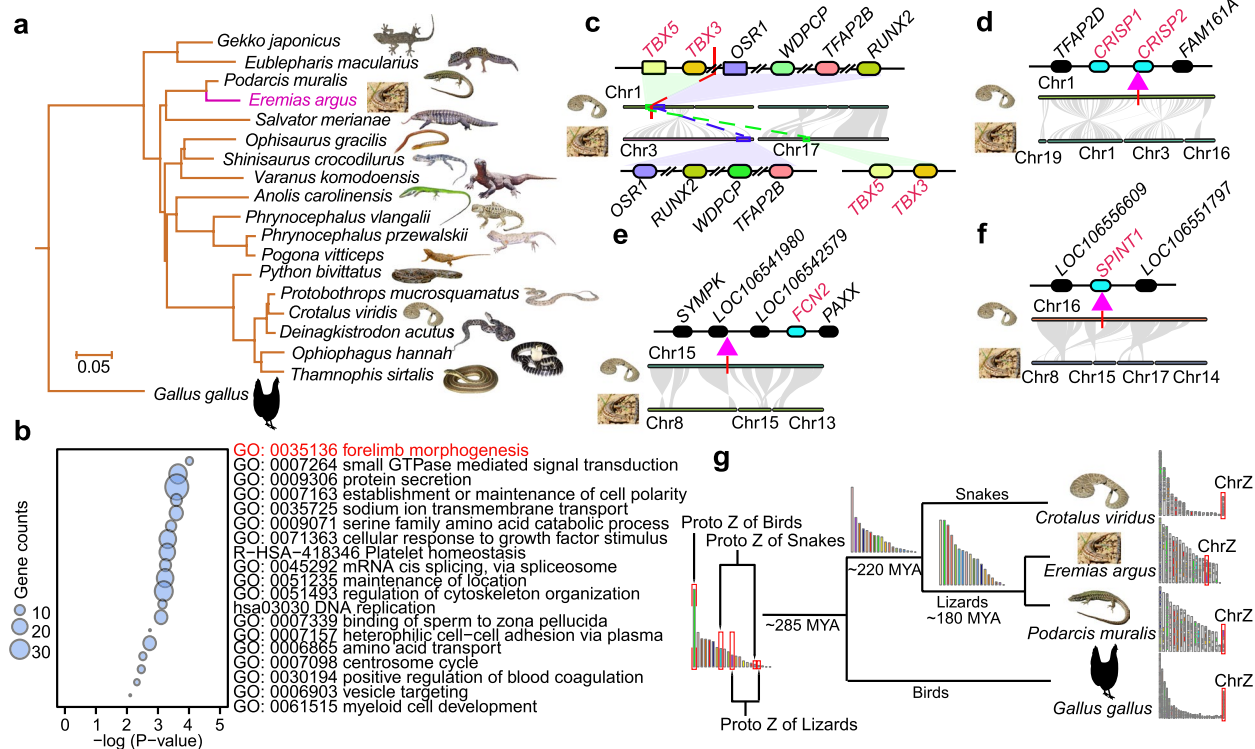
Li *et al. BMC Biology*   (2023) 21:40

Page 3 of 14

Fig. 1 Phylogenetic analysis and chromosome synteny of squamate genomes. **a** Phylogenomic tree constructed using common single-copy orthologous genes from 12 lizards and six snakes, with the chicken as the outgroup. The Mongolian racerunner is highlighted in magenta. All animal pictures are from http://reptile-database.reptarium.cz [26], modified in Adobe Illustrator. **b** Functional enrichment result of Mongolian racerunner genes located at chromosome breakpoint regions caused by chromosome fissions/fusions. **c** Six genes (*RUNX2*, *TFAP2B*, *OSR1*, *WDPCP*, *TBX5*, and *TBX3*) which have a role in limb development were detected in the chromosome breakpoint regions. **d**–**f** Four genes (*CRISP1*, *CRISP2*, *FCN2*, and *SPINT1*) which have a role in rattlesnake venom formation were detected in the chromosome breakpoint regions. **g** The reconstruction of ancestral chromosomes in squamates and chicken genomes. The Chr Z of three species originated from five different autosomes

breakpoint regions of the rattlesnake were enriched in "cellular response to organic cyclic compound" (Additional file 1: Figure S10), and four of them (*CRISP1*, *CRISP2*, *FCN2*, and *SPINT1*) encode proteins that are an important part of rattlesnake venom (Fig. 1d–f). Especially, the *FCN2* gene was gained in the snake genome at Chr 15. This gene encodes ficolin which is an essential snake venom component [29, 30]. For sex chromosomes, we found the chicken, snake, and lizard do not share homolog Z chromosome regions (Additional file 1: Figure S11), suggesting that the sex chromosomes of lizards, snakes, and birds might be derived from different autosomal pairs. To explore the origin of Z chromosomes in the Mongolian racerunner and its relatives, we reconstructed 21 ancestral chromosomes of these four species using GRIMM [31]. We found that the chicken ChrZ evolved from one ancestral autosome, lizard ChrZ evolved from two ancestral autosomes, and ChrZ of the snake evolved from another two ancestral autosomes (Fig. 1g). We further checked the location of sex determination-related genes in the lizard genomes, which shows that *DMRT1*,

DMRT2, and *DMRT3*, the male determining genes in birds [32], are located on the Chr18 of Mongolian racerunner or the Chr17 of European common wall lizard. *SOX3* gene, which is involved in sex determination in mammals [33], is located on ChrZ of the Mongolian racerunner genome and European common wall lizard (Additional file 1: Figure S11). Together, these results suggest that chromosome fissions/fusions occurred frequently and could have contributed to critical evolutionary events in squamates.

Genome adaptation to high altitude

To explore the molecular adaptation to mid altitude and high altitude in the Mongolian racerunner, we collected and sequenced six Mongolian racerunner populations from low-altitude to high-altitude areas (Fig. 2a, b; Additional file 2: Table S9-11). We utilized a combination of four parameters (F_{ST} and θ_π, XP-CLR, and Dxy) to identify genomic regions under selective sweeps. We initially used a top 1% F_{ST} value and a top 1% θ_π ratio cutoff to screen the potential selected region in one mid-altitude
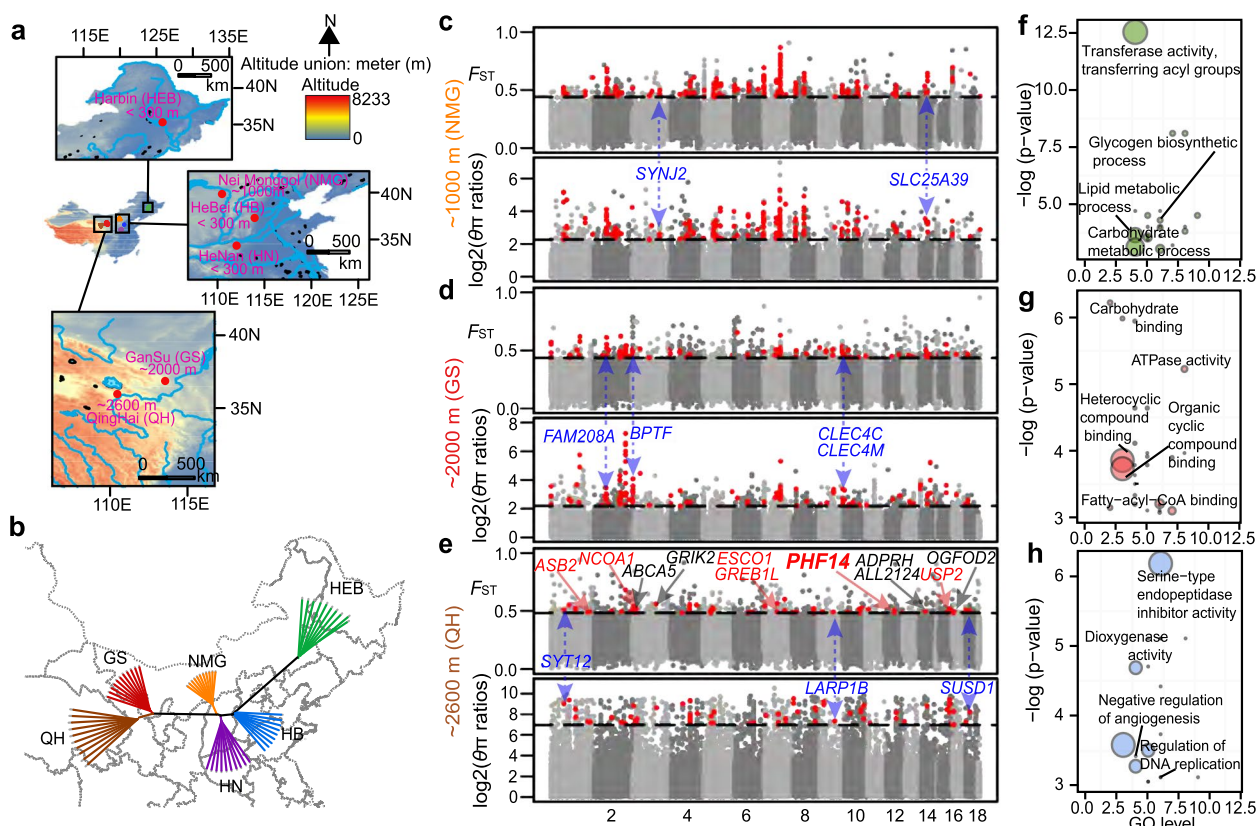
Fig. 2 Sampling altitude and selective sweep analysis of different high-altitude adaptations of the Mongolian racerunner. **a** Geographic distribution of the six representative populations. The altitude (m) of the sampling sites in this study is visualized. **b** Neighbor-joining tree constructed from the allele sharing matrix of variants of 61 Mongolian racerunner individuals. The orientations approximately fit the sample geographic locations. The reliability of the neighbor-joining tree was estimated using 100 bootstrap pseudo-replicates. **c–e** The panels show the distribution of log$_2(\theta_\pi$ ratios) and F_{ST} values that were calculated in 40-kb windows and 20-kb steps in mid- and high-altitude populations compared with low-altitude populations. The points in red (corresponding to the top 1% of the log$_2(\theta_\pi$ ratios) distribution and the top 1% of the F_{ST} value distribution) are genomic regions identified to be under selection in one mid-altitude population (NeiMonggol (NMG, ~ 1000 m)), and two high-altitude populations (GanSu (GS, ~ 2000 m), and Qinghai (QH, ~ 2600 m)). The ~ 1000 m (NMG) population was compared with three low populations (HB, HEB, and HN: altitude under ~ 300 m), the ~ 2000 m (GS) population was compared with three low populations and the NMG population, and the ~ 2600 m (QH) population compared with three low populations, the NMG population, and the GS population. The genes in red functioned by responding to hypoxia or UV, which are important for high-altitude adaptation in lizards. The *PHF14* gene was verified by cell experiments. The explanation of other highlighted genes were shown in Fig. 3 legend. **f–h** The panels show the GO enrichment results (*P* < 0.05 by Fisher exact test) of NMG, GS, and QH positively selected genes. The sizes of the circles represent the number of genes

population (NMG), and two high-altitude populations (GS and QH) compared to low-altitude group (HB, HEB, and HN), respectively (Additional file 1: Figure S12d-f). We identified a total of 179 ($F_{ST} \geq 0.44$; θ_π ratio ≥ 2.27), 115 ($F_{ST} \geq 0.44$; θ_π ratio ≥ 2.19), and 69 ($F_{ST} \geq 0.48$; θ_π ratio ≥ 6.97) genome segments in the NMG, GS, and QH populations relative to the low-altitude group (combined individuals from HB, HEB, and HN populations). These genome segments contained 78, 64, and 55 genes, respectively (Fig. 2c, d). Of these genes, 76.14% (150 out of 197) were also identified under selective sweep using XP-CLR, and 78.68% (155 out of 197) were identified using *D*xy (Additional file 1: Figure S13–15; Additional

file 2: Table S12–14). The candidate loci identified by F_{ST} and θ_π ratio among NMG, GS, and QH contrasts showed a low overlapping rate (Additional file 1: Figure S16), revealing a uniqueness at molecular adaptation across populations, which also has been observed in amphibians and reptiles [34]. Among these outlier genes, *TMEM68*, a candidate glycerolipid metabolism gene [35], overlapped in the NMG and GS selective sweep sets, and *ABCA8B*, a member of the ABCA transporter subfamily and associated with sphingolipid metabolism [36], overlapped in GS and QH. Genes under selective sweep in lizards endemic to high altitudes (> 2,000 m) are mainly associated with energy metabolism and DNA damage repair
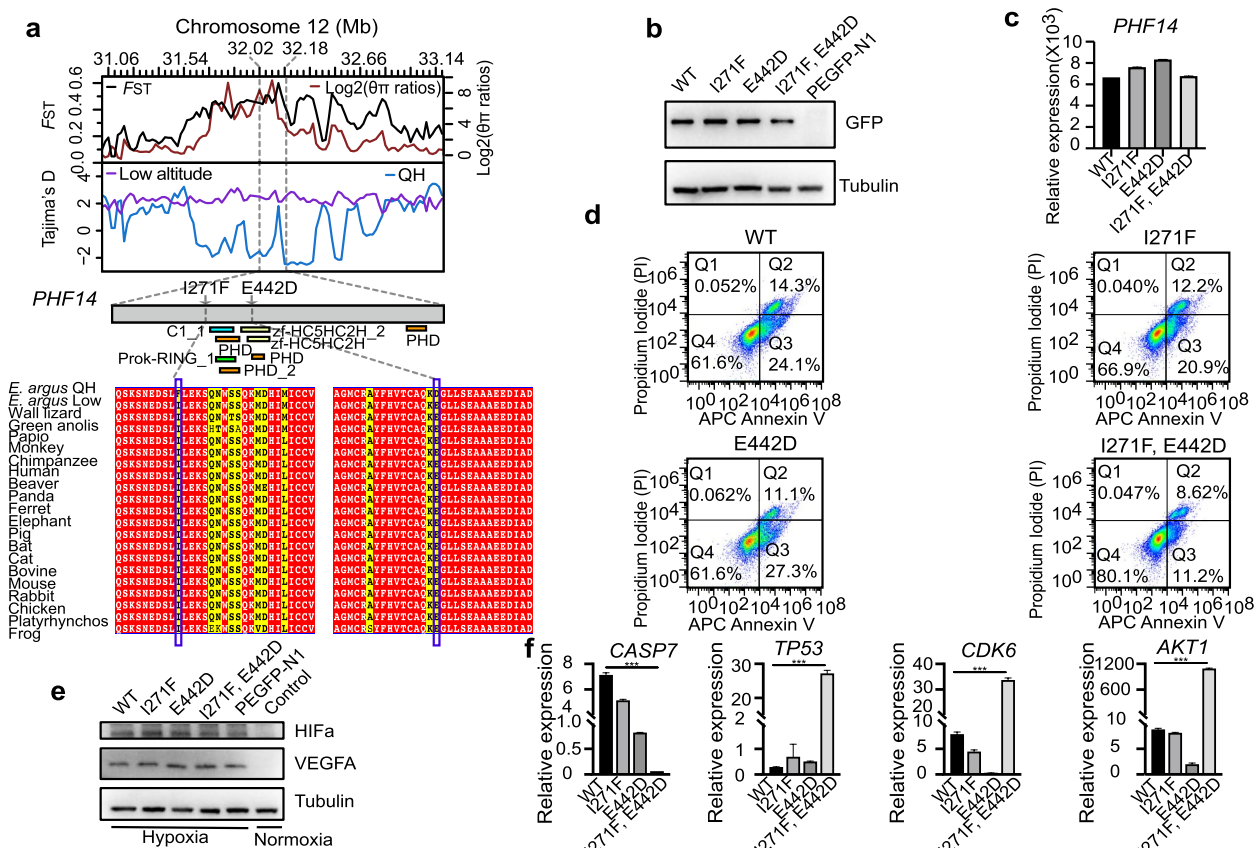
Li *et al. BMC Biology*   (2023) 21:40

Page 5 of 14

Fig. 3 Genotype frequency variation and expression pattern of genes under selective sweeps. **a** The candidate loci that under selective sweep (labeled in blue with boxes shown in Figs. 2c–e) whose genotype frequency increasing with the altitude. The *y*-axis indicates the allele frequency. The populations on the *x*-axis are ordered by altitude from low (left) to high (right). "Low" and "High" mean low altitude and high-altitude, respectively, including in panel **c**. **b** Volcano plot of differentially expressed genes between the lowest (HB) and highest (QH) altitude populations. Dotted line on *y*-axis represents FDR equal 0.01. Dotted lines on-axis represent FoldChange equal − 2 (left) and 2 (right). Top five down and five up differentially expressed genes are also highlighted. The gene symbols in boxes are candidate loci under selective sweep (labeled in black shown in Fig. 2e) in highest altitude population (QH) who also differentially expressed. **c** The expression pattern of five positively selected genes (*OGFOD2*, *GRIK2*, *ABCA5*, *ALL2124*, and *ADPRH*), which were also highlighted in Fig. 2e and Fig. 3b. *OGFOD2*, *ALL2124*, and *ADPRH* are highly expressed, whereas *GRIK2* and *ABCA5* are lowly expressed (FDR < 0.01). The *y*-axis was FPKM (Fragments Per Kilobase of transcript per Million mapped reads)

pathways (Fig. 2e–h; Additional file 2: Table S15–20). Interestingly, the allele frequency in 13 nonsynonymous mutations from nine genes increased with an increase in altitude, suggesting that these genes are important for the high-altitude adaptation of lizards (Fig. 3a). For example, *SYT12*, *CLEC4M*, and *CLEC4C* regulate calcium ion binding, may be involved in hypoxia-induced regulation [37–39]. Given that the QH population is under the highest environmental pressure, we further examined the 55 genes in the F_{ST} and θ_{π} outlier regions between QH and the low-altitude populations. Several genes seem to have helped the Mongolian racerunner adapt to the high altitude based on previous in vivo or in vitro functional experiments: e.g., *ASB2* is related to angiogenesis

[40], and *PITPNM2* and *ARHGAP6* are associated with enzymatic activity binding oxygen. Three genes, namely, *CAMKK2*, *CABP5*, and *MICU2*, have important roles in calcium binding and are required for the expression of hypoxia-inducible genes [41]. One gene, *ESCO1*, ensures the correct replication of DNA in hypoxic environments (Additional file 1: Figure S17) [42, 43].

To investigate how changes in gene expression profiles correlate with changes in altitude of Mongolian racerunner, we performed transcriptomic analysis of lung tissues between low and high populations. The genes with high F_{ST} and θ_{π} values and high XP-CLR score included the upregulated *P2RX4*, *FAM13A*, and *NARF* genes, and the downregulated *NCOA1*, *TMEM168*, and *USP2* genes,

Li *et al. BMC Biology*   (2023) 21:40

Page 6 of 14

Fig. 4 Functional assays of the PHF14 missense SNPs. **a** The $\log_2(\theta_\pi$ ratios), population differentiations (F_{ST}), and Tajima's D of selected regions around gene *PHF14* are shown in the top panel. The middle panel shows the structure of the *PHF14* protein. A multispecies alignment of the amino acid sequence of *PHF14* near the mutation point is provided at the bottom. **b** Western-blotting analysis of the HEK293 cell lysates transfected with the GFP-tagged recombinant plasmid of WT, I271F, E442D, and I271F, E442D, or the empty vector. WT, I271F, E442D and I271F, and E442D have plasmids overexpressing the wild-type PHF14 protein, the I271F mutant, the E442D mutant protein and I271F, and the E442D double mutant protein, respectively. PEGFP-N1 represents the empty vector. Anti-GFP and anti-tubulin antibodies were used to measure the protein expression of PHF14 and the internal reference protein tubulin, respectively. **c** The mRNA expression levels of *PHF14* in HEK293 overexpressing cell lines measured by qPCR. **d** Effect of hypoxic stimulus on the apoptosis of HEK293 cells. **e** HIF-1α and VEGFA expression in the WT, I271F, E442D, and I271F, E442D, or PEGFP-N1 overexpressing HEK293 cells under hypoxic conditions was detected by western blotting. Tubulin was used as the internal loading control. **f** Gene expression of the cell apoptosis genes, including *CASP7*, *TP53*, *CDK6*, and *ATK1* in transfected HEK293 cells, measured by qPCR

which have been suggested to associate with high-altitude adaptation [44–47]. This observation prompted us to explore other genes with expression changes and that may be under selection in Mongolian racerunners. As a result, we identified 493 downregulated and 845 upregulated differentially expressed genes (DEGs) between HB (the lowest altitude population) and QH (the highest altitude population) (Fig. 3b). The DEGs were functionally enriched in small molecule catabolic and organic hydroxy compound metabolic processes (Additional file 1: Figure S18). Of these genes, *OGFOD2*, *GRIK2*, *ABCA5*, *ALL2124*, and *ADPRH* also located in regions with high F_{ST} and θ_π, and low Tajima's D (Additional file 1: Figure S19). Although *OGFOD2*, *ALL2124*, and *ADPRH* were significantly upregulated ($P < 0.01$), *GRIK2* and *ABCA5*

were significantly downregulated ($P < 0.01$) in the high-altitude population (Fig. 3c). Among these five genes, *GRIK2* encodes a cold-sensing receptor [48], and a reduction in its expression may confer cold resistance in high-altitude Mongolian racerunner populations.

Functional validation of genetic variants in the *PHF14* gene

Gene *PHF14* has a high F_{ST} and $\log_2(\theta_\pi)$, as well as a low Tajima's D value, indicating strong signals of selection (Fig. 4a). *PHF14* is reported to be a novel regulator of mesenchyme growth via platelet-derived growth factor (*PDGF*) receptor-alpha [49, 50] and a hypoxia-sensitive epigenetic regulator in cell cycle progression [51]. *PHF14* is highly expressed in lung cancer, and its high expression is associated with poor survival [51, 52]. Silencing

Li *et al. BMC Biology*   (2023) 21:40

Page 7 of 14

of *PHF14* suppresses cancer proliferation [51–54]. This gene harbored two exonic candidate SNPs in the high-altitude population (QH population). In order to test the potential functional differences of the *PHF14* genotypes (I271F and E442D) in high-altitude lizards compared with the lowland species, we assessed the effects of the I271F and E442D mutations in human embryonic kidney cells (HEK293) under hypoxic conditions. Our functional experiments displayed that the wild-type (WT) *PHF14*, which comes from a low-altitude genotype, and its high-altitude variants exhibited comparable expression, as assessed by western blotting and RT-qPCR (Fig. 4b, c; Additional file 2: Table S21). After culturing the cells for 24 h under hypoxic conditions, the HEK293 cell lines with overexpressed PHF14 variants exhibited lower rates of apoptosis relative to wild-type PHF14, with a double mutant PHF14 (I271F and E442D) showing the lowest rate of apoptosis ($P < 0.01$) (Fig. 4d). Hypoxia-inducible factor 1 (HIF-1α) and vascular endothelial growth factor (VEGFA) are the master regulators of homeostasis under hypoxic conditions [55, 56]. Western blotting showed that the expression of the HIF-1α and VEGFA proteins was significantly increased after a hypoxic stimulus in WT, I271F, E442D, double mutant (I271F and E442D) PHF14, and EGFP overexpressing HEK293 cells ($P < 0.05$) (Fig. 4e). We further investigated whether the PHF14 mutant genotypes could alter the transcriptional response to cell apoptosis using RT-qPCR assays. The mRNA expression levels of three out of the nine downstream genes tested, namely, *TP53*, *CDK6*, and *ATK1*, were significantly increased in the double mutant (I271F and E442D) cells ($P < 0.01$) (Fig. 4f). The mRNA expression levels of *CASP7* were significantly decreased in double mutant (I271F and E442D) PHF14 overexpressing HEK293 cells ($P < 0.01$). No significant changes in mRNA expression levels were observed in single mutants (I271F or E442D). These data suggest that the I271F and E442D may collectively amplify the transcriptional activity mediated by PHF14 in response to apoptosis under hypoxic conditions.

Discussion

In this study, we investigated the genetic basis of high-altitude adaptation in lizards based on the high-quality reference genome and whole-genome resequencing. We conducted selective sweeping analysis for the populations from different altitudes using a combination method of F_{ST}, θ_π, XP-CLR, and Dxy and identified candidate genomic regions with exceptionally high levels of population differentiation on a genome scale. A total of 197 genes were found, showing strong evidence of selection in mid- and high-altitude populations. The *PHF14* gene,

displaying reduced genetic diversity and Tajima's *D* in a high-altitude population, was identified and validated as a target gene of natural selection for hypoxic adaptation in lizards. However, it is worth noting that, due to the limitations of the sample size and the genome scanning methods used in the present study, both additional samples and computational methods are required to evidence high-altitude adaptation in lizards. Furthermore, since mutations of the *PHF14* gene are partially validated, further investigation into genes located in genomic regions under selective sweeps would provide greater insight into mechanisms of adaptation.

Conclusions

We sequenced and assembled a high-quality genome of the Mongolian racerunner, the lizard distributed from the central and eastern plain to the Qinghai–Tibet Plateau, to understand how lizards adapt to high-altitude environments. Using this genome as reference, we found that chromosome fissions/fusions might be related to limb degeneration, snake venom, and sex chromosome evolution. Subsequent population genomics of Mongolian racerunners from high and low-altitude populations revealed a variety of novel genes associated with local adaptations to high altitude. Taken together, our study provides a high-quality genome assembly of the lizard and furthers our understanding of the genetic basis of high-altitude adaptation using reptiles as a model system.

Methods

Samples collecting and preparing

For de novo sequencing and assembly, the muscle tissue of an adult female Mongolian racerunner, which was collected from the desert in Hohhot, Inner Mongolia Autonomous Region of China, was harvested for DNA extraction. For genome annotation, RNA libraries were constructed with six tissues (heart, kidney, gonad, lung, liver, and skin) from the same female adult lizard. A total of additional 61 Mongolian racerunner individuals were collected for whole genome resequencing, containing 10 individuals from Harbin of Heilongjiang province, 9 individuals from Xingtai of Hebei province, 10 individuals from Gongyi, Zhengzhou of Henan province, 11 individuals from Hohhot of Inner Mongolia Autonomous Region, 11 individuals from Jingtai, Baiyin of Gansu province, and 10 individuals from Gonghe, Tibetan Autonomous Prefecture of Hainan of Qinghai province. All the lizards used in this study were euthanized by CO_2 and tissue samples were first flash frozen in liquid nitrogen and then stored at $-80℃$ until use. Total genomic DNA was extracted from muscle tissue by using the stand Phenol–chloroform method.

Li *et al. BMC Biology*   (2023) 21:40

Page 8 of 14

Pacific Biosciences (PacBio) library

Genomic DNA was isolated using a Qiagen DNA Genomic kit (Qiagen, Valencia, CA, USA) and the integrity of DNA was checked using agarose gel electrophoresis. DNA purity was checked using the Nano-Photometer® spectrophotometer (IMPLEN, CA, USA). The qualified DNA samples were randomly broken into fragments by g-TUBE (Covaris) and large fragments ($\geq$ 20 kb) of DNA were enriched and purified by magnetic beads. Fragments of DNA were repaired by damage repair and end repair. The stem ring-like adaptors were added at the two ends of DNA fragments, and the fragments that failed to be repaired were removed by exonuclease. Approximately 40-kb SMRTbell libraries were prepared according to the released protocol from PacBio. The single-molecule real-time (SMRT) sequencing was carried out using P6-C4 chemistry on the PacBio Sequel platform at Beijing Novogene Co. Ltd (Beijing, China). An approximate 115.3-fold genome coverage (187.94 Gb) long reads were generated for the assembly.

Illumina sequencing and quality control

A total of 1.5 µg DNA per sample was used for the DNA sample preparations. Sequencing libraries were generated using Truseq Nano DNA HT Sample preparation Kit (Illumina USA) following the manufacturer's recommendations and index codes were added to attribute sequences to each sample. Briefly, the DNA sample was fragmented by sonication to a size of 350 bp, then DNA fragments were end polished, A-tailed, and ligated with the full-length adapter for Illumina sequencing with further PCR amplification. At last, PCR products were purified (AMPure XP system) and libraries were analyzed for size distribution by Agilent2100 Bioanalyzer and quantified using real-time PCR.

For Illumina HiSeq sequencing, paired-end libraries were constructed with insert sizes around 350 bp according to Illumina's protocol. An approximate 65.30-fold genome coverage (106.44 GB) of Illumina paired-end reads was generated for the assembly.

For population genetic analysis, we generated 1399.362 Gb Illumina reads (~ 12.47-average fold per individual). All of the Illumina reads were controlled for quality and sequenced on an Illumina Hiseq sequencer at Beijing Novogene Co. Ltd (Beijing, China). We removed low-quality paired reads as follows: Reads with $\geq$ 10% unidentified nucleotides (N); > 10 nt aligned to the adaptor, allowing $\leq$ 10% mismatches; > 50% bases having phred quality < 5; putative PCR duplicates generated in the library construction process.

10X Genomics library

The Illumina Read1 sequencing primer, 16 bp barcode, and 6 bp random sequence primer were connected to the Gel-Beads sequence. The GelBeads passes through two injection ports. The input DNA is mixed with enzymes and dNTP and other reactants after entering the first sample inlet, and the mixture is wrapped by oil drops when passing through the second sample inlet, then it is collected and placed on a special 96-well plate for the barcode library preparation. The random primers were combined with the random position of genomic DNA fragments in the oil drop reaction system for constant temperature PCR amplification. After the library preparation by adding the Illumina P5 and P7 adaptors into the amplification products, PE150 was sequenced on the Illumina platform at Beijing Novogene Co. Ltd (Beijing, China). An approximate 119.10-fold genome coverage (194.14 Gb) 10X Genomics Linked-Reads were generated for the assembly.

Hi-C library

First, chromatin was digested for 16 h with 400 U HindIII restriction enzyme (NEB) at 37 °C. DNA ends were labeled with biotin and incubated at 37 °C for 45 min, and the enzyme was inactivated with 20% SDS solution. DNA ligation was then performed by the addition of T4 DNA ligase (NEB) and incubation at 16 °C for 4~6 h. After ligation, proteinase K was added to reverse cross-linking during incubation at 65 °C overnight. DNA fragments were purified and dissolved in 86µL of water. Unligated ends were then removed. Purified DNA was next fragmented to a size of 300–500 bp, and DNA ends were then repaired. DNA fragments labeled by biotin were finally separated on Dynabeads® M-280 Streptavidin (Life Technologies). Hi-C libraries were finally controlled for quality and sequenced on an Illumina Hiseq X Ten sequencer. An approximate 143.25-fold genome coverage (233.50 Gb) Hi-C reads were generated for the assembly.

Genome assembly

As for the assembly of the genome, the longest PacBio reads were selected as the seed reads, Daligner [57] was used to map all of the PacBio long reads to the seed reads. The pre-assembled reads were obtained by generating a consensus of mapped reads using LASort, LAMerge, and pbdacgon. Unitigs of the pre-assembled reads were used to generate a layout of overlapping reads and the contigs of assembly. The high-quality consensus was obtained after mapping the PacBio long reads to the de novo assembled reference by Quiver [58] (algorithm: arrow). Duplicated assembled haploid contigs were purged using purge_haplotigs [59] with short-insert Illumina reads.

Li *et al. BMC Biology*   (2023) 21:40

Page 9 of 14

The Linked-reads from the 10X Genomics platform were used to construct a scaffold genome by fragScaff [60]. Illumina reads were used to correct the inaccurate bases in the genome introduced by fragScaff using Pilon [61].

We used Hi-C reads to obtain a chromosome-level genome by Lachesis [62], though fragScaff had improved assemblies and produced a scaffold genome. The frequency of long-range interactions decreases rapidly as the linear distance along a chromosome increases, which helps produce a number of high-quality and contiguous assemblies. The final assembly totaled 1.63 Gb of sequence with a contig N50 of 3.07 Mb and a scaffold N50 of 96.73 Mb.

Genome annotation

To annotate the repeat sequences of the genome, we initially used LTR_FINDER (version 1.0.7) [63], RepeatModeler (version 2.1), and RepeatScount (version 1.0.5) to find repeats. Next, RepeatMasker (version 4.0.5) [64] was used to search for known and novel transposable elements (TEs) by mapping sequences against the de novo repeat library and Repbase TE library (v20140131). In addition, we used the ProteinMask software (version open-2.1, with parameters: -no LowSimple *-p* value 0.0001) to identify TE-relevant proteins. Overlapping transposable elements belonging to the same type of repeats were integrated together.

We used a homology-based method to annotate the protein-coding gene structure. First, we download genome and annotation information from online database and built a non-redundant protein database of 8 species: *Anolis carolinensis* (ftp://ftp.ensembl.org/pub/release-100/fasta/anolis_carolinensis/dna/Anolis_carolinensis.AnoCar2.0.dna.toplevel.fa.gz), *Pogona vitticeps* (https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/900/067/755/GCF_900067755.1_pvi1.1/GCF_900067755.1_pvi1.1_genomic.fna.gz), *Shinisaurus crocodilurus* (ftp://parrot.genomics.cn/gigadb/pub/10.5524/100001_101000/100315/Shinisaurus_crocodilurus.fa.gz), *Gekko japonicus* (https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/447/785/GCF_001447785.1_Gekko_japonicus_V1.1/GCF_001447785.1_Gekko_japonicus_V1.1_genomic.fna.gz), *Python molurus bivittatus* (https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/186/305/GCF_000186305.1_Python_molurus_bivittatus-5.0.2/GCF_000186305.1_Python_molurus_bivittatus-5.0.2_genomic.fna.gz), *Protobothrops mucrosquamatus* (https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/527/695/GCF_001527695.2_P.Mucros_1.0/GCF_001527695.2_P.Mucros_1.0_genomic.fna.gz), *Thamnophis sirtalis* (https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/077/635/GCF_001077635.1_Thamnophis_sirtalis-6.0/GCF_001077635.1_Thamnophis_sirtalis-6.0_genomic.fna.gz), and *Deinagkistrodon acutus* (ftp://parrot.genomics.cn/gigadb/pub/10.5524/100001_101000/100196/Deinagkistrodon_acutus.fa.gz). Then the protein sequences were aligned to the genome by using TBlastN [65] with an E-value cutoff by 1E-5. The blast hits were conjoined by solar [66]. For each blast hit, Genewise [67] was used to predict the exact gene structure in the corresponding genomic regions.

For de novo gene prediction, we utilized SNAP (version 2006–07-28) [68], GENSCAN (version 1.0) [69], GlimmerHMM (version 3.0.4) [70], and AUGUSTUS (version 3.1) [71] to analyze the repeat-masked genome. EVidenceModeler software (EVM, version 1.1.1) [72] was used to integrate the genes predicted by the homology and de novo approaches and generate a consensus gene set.

RNA-seq data from 6 issues as mentioned above were mapped to the genome using Tophat (version 2.0.13) [73]. Then cufflinks (version 2.1.1) [74] was used to assemble transcripts to gene models. Transcripts were assembled to expressed sequence tags (ESTs) by Trinity [75] with parameters (-ss 0.5 -jc 0 -minkmercov 2 -minglue 2). We used the transcript information to revise the gene set.

Functional annotation of protein-coding genes was evaluated by BLASTP (*e* value 1E−05) using two integrated protein sequence databases — SwissProt [76] and NR database. Protein domains were annotated by searching InterPro and Pfam databases, using InterProScan and Hmmer, respectively. Gene Ontology (GO) [77] terms for each gene were obtained from the corresponding InterPro or Pfam entry. The pathways, in which the gene might be involved, were assigned by blast against the KEGG database (release53) [78]. The tRNA genes were identified by tRNAscan-SE [79] software. The rRNA fragments were predicted by aligning to the rRNA sequences database using BlastN at *E*-value of 1E−10. The miRNA and snRNA genes were predicted by INFERNAL software against the Rfam database [80].

Phylogenetic tree construction

Gene families were constructed through a hierarchical clustering algorithm and 'all against all' BLASTP with a cutoff of 1E−9 (blast-2.2.26). The alignments with high-scoring segment pairs were conjoined for each gene pair by solar. To identify homologous gene pairs, more than 30% coverage of the aligned regions in both homologous genes was required. Finally, homologous genes were clustered into gene families using the software hcluster_sg.

The phylogenetic tree was reconstructed by using shared single-copy genes. Protein sequences for these single-copy genes were aligned by MUSCLE (version 3.7) [81], then protein sequence alignment was transformed back to CDS alignments. We concatenated the CDS alignments of single-copy genes to a "supermatrix." Using

Li *et al. BMC Biology*   (2023) 21:40

Page 10 of 14

this supermatrix, we constructed the phylogenetic tree using the ML (maximum likelihood) algorithm as implemented in RAxML software (version 8.0.19) [82].

Genome synteny

We used MCscan (version v0.8.12) [83] to get synteny blocks among five chromosome-level genomes: Mongolian racerunner, European common wall lizard (*Podarcis muralis*), chicken, green anole lizard, and rattlesnake (*Crotalus viridis*) genomes. We converted the gff files to BED format files, then the CDS sequences were extracted from five genomes and translated into protein sequences as input files. The parameters used for the genome alignment were "−minspan = 10." The alignments of syntenic chromosomes were visualized by MCscan. The parameters used for visualizing were "−notex." The chromosome breakpoint regions are defined as the interval between two syntenic blocks from different chromosomes with more than 10 genes in size, which were detected between the Mongolian racerunner and rattlesnake genomes. Ancestral genome reconstruction was performed using GRIMM-Synteny [84] and MGR [31].

SNP calling and filtering

The high-quality paired-end reads were mapped to the reference genome using BWA (Burrows-Wheeler Aligner) (Version: 0.7.8) [85] with the command "mem -t 4 -k 32 −M." After alignment, we performed SNP calling on a population scale using a Bayesian approach as implemented in the package SAMtools [86]. We then calculated genotype likelihoods from reads for each individual at each genomic location and the allele frequencies in the sample with a Bayesian approach. The "mpileup" command was used to identify SNPs with the parameters as "-q 1 -C 50 -t SP -t DP -m 2 -F 0.002." Then, to exclude SNP calling errors caused by incorrect mapping or InDels, only high-quality SNPs (coverage depth ≥ 6 and ≤ 100, RMS mapping quality ≥ 20, maf ≥ 0.1, miss ≤ 0.1) were kept for subsequent analysis. Consequently, 6,303,245 SNPs were left after filtering from 78,513,343 raw SNPs (Additional file 2: Table S11-12). SNP annotation was performed according to the Mongolian racerunner genome using the package ANNOVAR (Version: 2013−05−20) [87].

Selective sweep and enrichment analysis

We calculated genome-wide distribution of F_{ST} values and $\theta\pi$ using Vcftools (version 0.1.14) [88] in 40-kb windows with a step size of 20 kb. The $\theta\pi$ ratios ($\theta\pi$–*group1*/$\theta\pi$–*group2*) were calculated for three extreme-control group pairs and log2-transformed. The windows

with top 1% F_{ST} and -$\log_2(\theta\pi$ ratio), simultaneously were considered as candidate outliers under strong selective sweeps. We also used XP-CLR [89] as a supplement to verify the reliability of the above two selective sweep methods for this project. All outlier regions were assigned to corresponding SNPs and annotated genes. Except above relative divergence methods, we used the absolute divergence (*D*xy) approach to further validate higher divergence regions.

For genes annotated in outlier regions, we used KOBA2.0 [90] software to test the statistical enrichment in KEGG pathways. Gene Ontology (GO) enrichment analysis was implemented by the GOseq R package, in which gene length bias was corrected. GO terms with corrected *P*-value less than 0.05 were considered significant.

The divergent levels (F_{ST}) of between two subpopulations were simulated by the coalescent simulator ms [91]. We performed the simulations under three isolation with migration (IWM) models. The between-population split time (-ej parameter) and migration rate (-I parameter) were adjusted to match the overall observed F_{ST} distributions. The F_{ST} values from simulations were calculated by the *evo* program (https://github.com/millanek/evo) with "fst –ms" parameter [92]. The result showed that the top 1% of observed F_{ST} (approximately $F_{ST} \geq 0.42$) are always higher than the corresponding neutral F_{ST} values, which is consistent with divergent selection (F_{ST}) acting on approximately this top 1% of selective sweeping regions.

Transcriptome sequencing and analysis

For the gene profile of lizard at high altitude, the lungs of four individuals from Qinghai (highest altitude) and four individuals from Hebei (lowest altitude) were harvested locally and preserved in RNA later immediately. A total of 3 μg RNA per sample was used as input material for the RNA sample preparations. Sequencing libraries were generated using NEBNext® UltraTM RNA Library Prep Kit for Illumina® (NEB, USA) following the manufacturer's recommendations and index codes were added to attribute sequences to each sample. We generated 40.13 Gb RNA-seq reads of six different organ tissues for genome annotation and 55.02 Gb RNA-seq reads of eight lung tissues for transcriptomic analysis of high-altitude adaptation. Index of the reference genome was built using HISAT2-BUILD and the high-quality RNA-seq reads were aligned to the reference genome by the HISAT2 v2.0.4 [93] program with default parameters. HTSeq v0.6.1 [94] was used to count the reads numbers mapped to each gene. And then FPKM (Fragments Per Kilobase of transcript per Million mapped reads) of each gene was calculated based on the length of the gene

and reads count numbers. FPKM considers the effect of sequencing depth and gene length for the reads count at the same time, and is currently the most commonly used method for estimating gene expression levels. Differential expression analysis of two conditions (four biological replicates per condition) was performed by using the edgeR [95].

Cell maintenance and plasmid construct

HEK293 cells were cultivated in DMEM (C11995500BT; Gibco) with 10% FBS (10099141; Gibco) and penicillin (0.2 U/ml)/streptomycin (0.2 lg/ml). The wild-type *PHF14* gene was synthesized in Sangon Company, China. The mutagenesis of the PHF14 sequence (I271F and E442D) was achieved through site-directed mutagenesis and verified by Sanger sequencing. Rapid Recombinant Clone Kit (VI201, TIANGEN) was used to subclone the wildtype PHF14 cDNA, the SNP1 I271F sequence (A to T), or the SNP2 E442D sequence (A to T) into the pEGFP-N1 plasmid for the overexpression analysis.

Transfection, Western blotting, and qPCR quantification

Transfection of the recombinant plasmids into HEK293 cells was carried out by using Lipofectamine3000 (Invitrogen, America) according to the manufacturer's instructions. First, 4 μL of Lipofectamine3000 Reagent (Invitrogen, America) and 5 μg of Endofree plasmid were diluted in 125 μL of DMEM (C11995500BT; Gibco). This solution was then mixed with Lipofectamine plasmids Reagent (1:1 ratio) and incubated for 15 min at room temperature, they were then added to 6-well plates (Costar, USA) with 50% cell confluence. These cells were incubated at 37℃ with 5% CO_2 for 24 h.

Cells were hypoxic stimulus with 1% O_2 for 24 h after 24 h after transfection and then were washed with cold PBS, harvested, and immediately lysed (RIPA Lysis Buffer, P0013B; Beyotime and PMSF, ST506; Beyotime) for 30 min on ice, cell lysates were centrifuged at 14,000 g for 20 min at 4℃, and supernatants were collected. The total amount of proteins was measured with the BCA method. 10ug of cell lysates supernatants were subjected to western blot with anti-GFP (AE012; Abclonal), anti-HIFα (a16873; ABclonal) anti-VEGFA (ab1316; Abcam), and anti-Tubulin (AC008; Abclonal) antibodies. Antigen–antibody complexes were visualized by enhanced chemiluminescence detection (mageQuantTM LAS 4000).

After the cells were hypoxic stimulus with 1% O_2 for 24 h, the cells were washed in PBS, harvested, and total RNA was extracted using TransZol Up Plus RNA Kit (ER501-01, TRANS) and RNA was reverse transcribed with GOScriptTM Reverse Transcription System (A5001,

Promega) according to the manufacturer's instructions. Real-time qPCR was performed using SYBR™ Universal PCR Master Mix (439,155; Invitrogen) with first-strand cDNA to evaluate gene expression. Quantitative Real-Time PCR analysis of the TP53, CDK6, AKT1, and CASP7 was carried out using the ABI7500 sequence detection system (Applied Biosystems by Life Technologies, Darmstadt, Germany). The ubiquitous Tubulin gene served as a reference control. PCR primers are shown in Additional file 2: Table S21 and qPCR conditions were as follows: 50 °C for 2 min, initial denaturation at 95 °C for 2 min, 40 cycles of denaturation at 95 °C for 15 s, with combined annealing and extension at 58 °C for 1 min. Three biological replicates were performed per sample.

Cell apoptosis assays

The HEK293 cells overexpressed WT, I271F, E442D, and I271F, E442D, or the PEGFP-N1 were treated on hypoxic conditions for 24 h. The apoptosis rate was evaluated using the Annexin V-AF647/PI Apoptosis Detection kit (P-CA-203, Procell) according to the instructions from the manufacturer. The cells were seeded into 6-well plates (3×10^5 cells/well). Following hypoxic conditions, the cells were collected, washed with cold PBS, and resuspended in 500 μL binding buffer. Then, 5 μL Annexin V-AF647 and 5 μL PI were added to the buffer and incubated at room temperature for 15 min. Cells were analyzed by flow cytometry (BD FACSCanto) within 1 h.

Supplementary Information

Additional file 1: Figure S1. Distribution of 17-mer frequency in Mongolian racerunner genome. **Figure S2.** GC content against the sequencing depths of Mongolian racerunner genome. **Figure S3.** Hi-C interactome plot among Mongolian racerunner chromosomes (Chr1-Chr19). **Figure S4.** Circos plot of the Mongolian racerunner genome assembly showing (from outermost to innermost) Mongolian racerunner chromosomes (Mb), gene density, GC content (%) and TE density (%). **Figure S5.** The female to male depth ratio (log2) for each chromosome. **Figure S6.** Divergence distribution of classified families in Mongolian racerunner genome. **Figure S7.** A comparison of gene parameters among the five lizard genomes. **Figure S8.** Synteny between chicken and Mongolian racerunner genomes (left), and green anole lizard and Mongolian racerunner genomes (right). **Figure S9.** Comparative synteny of chromosomes among European wall common lizard (P. muralis), Mongolian racerunner and rattlesnake (C. viridis) genomes. **Figure S10.** The top 20 functional enrichment results of genes located at rattlesnake genome chromosome breakpoint regions using Metascape. **Figure S11.** Synteny of the Z chromosome among four species. **Figure S12.** The isolation with migration (IWM) models of species formation used for coalescent simulations. **Figure S13.** This figure showed the distribution of XP-CLR values which calculated in 40-kb window and 20-kb step between high and low altitude populations. **Figure S14.** This figure showed the distribution of absolute divergence (Dxy) values which calculated in 40-kb window and 20-kb step between high and low altitude populations. **Figure S15.** The absolute divergence (Dxy) comparison between the top 1% FST. **Figure**

Li *et al. BMC Biology*   (2023) 21:40

Page 12 of 14

S16. Venn plots were shown the global overlaps of the top 1% candidate loci between NMG/GS/QH. **Figure S17.** A concise map of positively selected genes labeled with blue that may relate to hypoxia and UV light response. **Figure S18.** The functional enrichment results of differentially expressed genes (DEGs) using Metascape. **Figure S19.** The FST and θπ, and Tajima'D for the upstream and downstream 1Mb genomic regions of these five genes were calculated in 40kb window. **Figure S20.** The original and uncropped gels for cell experiments.

Additional file 2: Table S1. Reads and libraries used in *de novo* Mongolian racerunner genome sequencing. **Table S2.** Assembly statistics of Mongolian racerunner genome. **Table S3.** The CEGMA (Core Eukaryotic Genes Mapping Approach) result of Mongolian racerunner genome. **Table S4.** The BUSCO (Benchmarking Universal Single-Copy Orthologs) result of Mongolian racerunner genome. **Table S5.** The statistics on genome coverage and depth of Mongolian racerunner genome. **Table S6.** The statistics of repeat sequences in Mongolian racerunner genome. **Table S7.** Gene prediction of Mongolian racerunner (*Eremias argus*) genome. **Table S8.** The statistics on functional gene annotation of Mongolian racerunner genome. **Table S9.** Geographical position and sex of sampled Mongolian racerunner individuals. **Table S10.** Mapping statistics of re-sequenced individuals. The genome-wide resequencing data for 61 samples were mapped to the de novo Mongolian racerunner genome. **Table S11.** Summary of SNPs in 61 re-sequenced individuals of Mongolian racerunner. **Table S12.** The top 1% genes under select sweep region in the Inner-Mongolia (NMG) population. **Table S13.** The top 1% genes under select sweep region in the GanSu (GS) population. **Table S14.** The top 1% genes under select sweep region in the Qinghai (QH) population. **Table S15.** Biological Process (BP) GO term enrichment results of selected genes in the NMG populations. **Table S16.** KEGG enrichment results of selected genes in the NMG population. **Table S17.** Biological Process (BP) GO term enrichment results of selected genes in the GS populations. **Table S18.** KEGG enrichment results of selected genes in the GS populations. **Table S19.** Biological Process (BP) GO term enrichment results of selected genes in the QH populations. **Table S20.** KEGG enrichment results of selected genes in the QH populations. **Table S21.** The primers for PCR.

References
1. Okuzaki Y, Sota T. Factors related to altitudinal body size variation in the earthworm-eating ground beetle carabus japonicus. Zoolog Sci. 2017;34(3):229–34.
2. Tufts DM, Revsbech IG, Cheviron ZA, Weber RE, Fago A, Storz JF. Phenotypic plasticity in blood-oxygen transport in highland and lowland deer mice. J Exp Biol. 2013;216(Pt 7):1167–73.
3. Naeije R. Physiological adaptation of the cardiovascular system to high altitude. Prog Cardiovasc Dis. 2010;52(6):456–66.
4. Witt KE, Huerta-Sanchez E. Convergent evolution in human and domesticate adaptation to high-altitude environments. Philos Trans Royal Soc B Biol Sci. 2019;374(1777):20180235.
5. Peng Y, Yang Z, Zhang H, Cui C, Qi X, Luo X, Tao X, Wu T, Ouzhuluobu, Basang, et al. Genetic variations in Tibetan populations and high-altitude adaptation at the Himalayas. Mol Biol Evol. 2011;28(2):1075–81.
6. Xu S, Li S, Yang Y, Tan J, Lou H, Jin W, Yang L, Pan X, Wang J, Shen Y, et al. A genome-wide search for signals of high-altitude adaptation in Tibetans. Mol Biol Evol. 2011;28(2):1003–11.
7. Hendrickson SL. A genome wide study of genetic adaptation to high altitude in feral Andean Horses of the paramo. BMC Evol Biol. 2013;13:273.
8. Li M, Tian S, Jin L, Zhou G, Li Y, Zhang Y, Wang T, Yeung CK, Chen L, Ma J, et al. Genomic analyses identify distinct patterns of selection in domesticated pigs and Tibetan wild boars. Nat Genet. 2013;45(12):1431–8.
9. Gou X, Wang Z, Li N, Qiu F, Xu Z, Yan D, Yang S, Jia J, Kong X, Wei Z, et al. Whole-genome sequencing of six dog breeds from continuous altitudes reveals adaptation to high-altitude hypoxia. Genome Res. 2014;24(8):1308–15.
10. Wang MS, Li Y, Peng MS, Zhong L, Wang ZJ, Li QY, Tu XL, Dong Y, Zhu CL, Wang L, et al. Genomic analyses reveal potential independent adaptation to high altitude in Tibetan chickens. Mol Biol Evol. 2015;32(7):1880–9.
11. Yang J, Li WR, Lv FH, He SG, Tian SL, Peng WF, Sun YW, Zhao YX, Tu XL, Zhang M, et al. Whole-genome sequencing of native sheep provides insights into rapid adaptations to extreme environments. Mol Biol Evol. 2016;33(10):2576–92.
12. Yang J, Jin ZB, Chen J, Huang XF, Li XM, Liang YB, Mao JY, Chen X, Zheng Z, Bakshi A, et al. Genetic signatures of high-altitude adaptation in Tibetans. Proc Natl Acad Sci. 2017;114(16):4189–94.
13. Lai YT, Yeung CKL, Omland KE, Pang EL, Hao Y, Liao BY, Cao HF, Zhang BW, Yeh CF, Hung CM, et al. Standing genetic variation as the predominant source for adaptation of a songbird. Proc Natl Acad Sci. 2019;116(6):2152–7.
14. Liu X, Zhang Y, Li Y, Pan J, Wang D, Chen W, Zheng Z, He X, Zhao Q, Pu Y et al. EPAS1 gain-of-function mutation contributes to high-altitude adaptation in Tibetan horses. Mol Biol Evol. 2019;36(11):2591–603.
15. Li JT, Gao YD, Xie L, Deng C, Shi P, Guan ML, Huang S, Ren JL, Wu DD, Ding L, et al. Comparative genomic investigation of

Li *et al. BMC Biology*   (2023) 21:40

Page 13 of 14

high-elevation adaptation in ectothermic snakes. Proc Natl Acad Sci. 2018;115(33):8406–11.

16. Sun YB, Fu TT, Jin JQ, Murphy RW, Hillis DM, Zhang YP, Che J. Species groups distributed across elevational gradients reveal convergent and continuous genetic adaptation to high elevations. Proc Natl Acad Sci. 2018;115(45):E10634–41.

17. Pasquesi GIM, Adams RH, Card DC, Schield DR, Corbin AB, Perry BW, Reyes-Velasco J, Ruggiero RP, Vandewege MW, Shortt JA, et al. Squamate reptiles challenge paradigms of genomic repeat element evolution set by birds and mammals. Nat Commun. 2018;9(1):2774.

18. Kvon EZ, Kamneva OK, Melo US, Barozzi I, Osterwalder M, Mannion BJ, Tissieres V, Pickle CS, Plajzer-Frick I, Lee EA, et al. Progressive loss of function in a limb enhancer during snake evolution. Cell. 2016;167(3):633-642 e611.

19. Roscito JG, Sameith K, Parra G, Langer BE, Petzold A, Moebius C, Bickle M, Rodrigues MT, Hiller M. Phenotype loss is associated with widespread divergence of the gene regulatory landscape in evolution. Nat Commun. 2018;9(1):4737.

20. Vallender EJ, Lahn BT. Multiple independent origins of sex chromosomes in amniotes. Proc Natl Acad Sci. 2006;103(48):18031–2.

21. Pie MR, Campos LLF, Meyer ALS, Duran A. The evolution of climatic niches in squamate reptiles. Proc Biol Sci. 1858;2017:284.

22. Navas CA. Herpetological diversity along Andean elevational gradients: links with physiological ecology and evolutionary physiology. Comp Biochem Physiol A Mol Integr Physiol. 2002;133(3):469–85.

23. He J, Xiu M, Tang X, Yue F, Wang N, Yang S, Chen Q. The different mechanisms of hypoxic acclimatization and adaptation in Lizard Phrynocephalus vlangalii living on Qinghai-Tibet Plateau. J Exp Zool A Ecol Genet Physiol. 2013;319(3):117–23.

24. Dai X, Zeng XM, Chen B, Wang YZ. The research on the karyotypes of six species in the genus Eremias from China. Hereditas. 2004;26(5):669–75.

25. Qu A, Li Z. Study on the karyotype and C-banding pattern of chromosomes of Eremias Argus Argus. Journal of Xuzhou Normal University. 1992;4(1):30–33.

26. Uetz P, Stylianou A. The original descriptions of reptiles and their subspecies. Zootaxa. 2018;4375(2):257–64.

27. Parra G, Korf BI. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. Bioinformatics. 2007;23(9):1061–7.

28. Simao FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015;31(19):3210–2.

29. Schield DR, Card DC, Hales NR, Perry BW, Pasquesi GM, Blackmon H, Adams RH, Corbin AB, Smith CF, Ramesh B, et al. The origins and evolution of chromosomes, dosage compensation, and mechanisms underlying venom regulation in snakes. Genome Res. 2019;29(4):590–601.

30. Suryamohan K, Krishnankutty SP, Guillory J, Jevit M, Schroder MS, Wu M, Kuriakose B, Mathew OK, Perumal RC, Koludarov I, et al. The Indian cobra reference genome and transcriptome enables comprehensive identification of venom toxins. Nat Genet. 2020;52(1):106–17.

31. Bourque G, Pevzner PA. Genome-scale evolution: reconstructing gene orders in the ancestral species. Genome Res. 2002;12(1):26–36.

32. Smith CA, Roeszler KN, Ohnesorg T, Cummins DM, Farlie PG, Doran TJ, Sinclair AH. The avian Z-linked gene DMRT1 is required for male sex determination in the chicken. Nature. 2009;461(7261):267–71.

33. Koopman P, Gubbay J, Vivian N, Goodfellow P, Lovell-Badge R. Male development of chromosomally female mice transgenic for Sry. Nature. 1991;351(6322):117–21.

34. Sun YB, Fu TT, Jin JQ, Murphy RW, Hillis DM, Zhang YP, Che J. Species groups distributed across elevational gradients reveal convergent and continuous genetic adaptation to high elevations. Proc Natl Acad Sci U S A. 2018;115(45):E10634–41.

35. Chang P, Heier C, Qin W, Han L, Huang F, Sun Q. Molecular identification of transmembrane protein 68 as an endoplasmic reticulum-anchored and brain-specific protein. PLoS One. 2017;12(5):e0176980.

36. Yang C, Yuan H, Gu J, Xu D, Wang M, Qiao J, Yang X, Zhang J, Yao M, Gu J, et al. ABCA8-mediated efflux of taurocholic acid contributes to gemcitabine insensitivity in human pancreatic cancer via the S1PR2-ERK pathway. Cell Death Discov. 2021;7(1):6.

37. Eizuka K, Nakashima D, Oka N, Wagai S, Takahara T, Saito T, Koike K, Kasamatsu A, Shiiba M, Tanzawa H, et al. SYT12 plays a critical role in oral cancer and may be a novel therapeutic target. J Cancer. 2019;10(20):4913–20.

38. Probert F, Mitchell DA, Dixon AM. NMR evidence for oligosaccharide release from the dendritic-cell specific intercellular adhesion molecule 3-grabbing non-integrin-related (CLEC4M) carbohydrate recognition domain at low pH. FEBS J. 2014;281(16):3739–50.

39. Seta KA, Yuan Y, Spicer Z, Lu G, Bedard J, Ferguson TK, Pathrose P, Cole-Strauss A, Kaufhold A, Millhorn DE. The role of calcium in hypoxia-induced signal transduction and gene expression. Cell Calcium. 2004;36(3–4):331–40.

40. Okumura F, Joo-Okumura A, Nakatsukasa K, Kamura T. The role of cullin 5-containing ubiquitin ligases. Cell Div. 2016;11:1.

41. Salnikow K, Kluz T, Costa M, Piquemal D, Demidenko ZN, Xie K, Blagosklonny MV. The regulation of hypoxic genes by calcium involves c-Jun/AP-1, which cooperates with hypoxia-inducible factor 1 in response to hypoxia. Mol Cell Biol. 2002;22(6):1734–41.

42. Lu Y, Li S, Cui Z, Dai X, Zhang M, Miao Y, Zhou C, Ou X, Xiong B. The cohesion establishment factor Esco1 acetylates alpha-tubulin to ensure proper spindle assembly in oocyte meiosis. Nucleic Acids Res. 2018;46(5):2335–46.

43. Alomer RM, da Silva EML, Chen J, Piekarz KM, McDonald K, Sansam CG, Sansam CL, Rankin S. Esco1 and Esco2 regulate distinct cohesin functions during cell cycle progression. Proc Natl Acad Sci. 2017;114(37):9906–11.

44. Smith SM, Mitchell GS, Friedle SA, Sibigtroth CM, Vinit S, Watters JJ. Hypoxia Attenuates Purinergic P2X Receptor-Induced Inflammatory Gene Expression in Brainstem Microglia. Hypoxia (Auckl). 2013;2013(1):1–11.

45. Ziółkowska-Suchanek I, Mosor M, Podralska M, Iżykowska K, Gabryel P, Dyszkiewicz W, Słomski R, Nowak J. FAM13A as a novel hypoxia-induced gene in non-small cell lung cancer. J Cancer. 2017;8(19):3933.

46. Ji S. Asahina N, Kitano S, Kino Y: Bioinformatics data mining approach indicates the expression of chromatin immunoprecipitation followed by deep sequencing (ChIP-Seq)-based hypoxia-inducible factor-1α target genes in periplaque lesions of multiple sclerosis. Clin Exp Neuroimmunol. 2015;6(2):159–69.

47. Ghorbel MT, Cherif M, Jenkins E, Mokhtari A, Kenny D, Angelini GD, Caputo M. Transcriptomic analysis of patients with tetralogy of Fallot reveals the effect of chronic hypoxia on myocardial gene expression. J Thorac Cardiovasc Surg. 2010;140(2):337-345 e326.

48. Gong J, Liu J, Ronan EA, He F, Cai W, Fatima M, Zhang W, Lee H, Li Z, Kim GH, et al. A cold-sensing receptor encoded by a glutamate receptor gene. Cell. 2019;178(6):1375-13861 e311.

49. Kitagawa M, Takebe A, Ono Y, Imai T, Nakao K, Nishikawa S, Era T. Phf14, a novel regulator of mesenchyme growth via platelet-derived growth factor (PDGF) receptor-alpha. J Biol Chem. 2012;287(33):27983–96.

50. Huang Q, Zhang L, Wang Y, Zhang C, Zhou S, Yang G, Li Z, Gao X, Chen Z, Zhang Z. Depletion of PHF14, a novel histone-binding protein gene, causes neonatal lethality in mice due to respiratory failure. Acta Biochim Biophys Sin (Shanghai). 2013;45(8):622–33.

51. Park JE, Tse SW, Xue G, Assisi C, Maqueda AS, Ramon GPX, Low JK, Kon OL, Tay CY, Tam JP, et al. Pulsed SILAC-based proteomic analysis unveils hypoxia- and serum starvation-induced de novo protein synthesis with PHD finger protein 14 (PHF14) as a hypoxia sensitive epigenetic regulator in cell cycle progression. Oncotarget. 2019;10(22):2136–50.

52. Zhang L, Huang Q, Lou J, Zou L, Wang Y, Zhang P, Yang G, Zhang J, Yu L, Yan D, et al. A novel PHD-finger protein 14/KIF4A complex overexpressed in lung cancer is involved in cell mitosis regulation and tumorigenesis. Oncotarget. 2017;8(12):19684–98.

53. Cheng M, Michalski S, Kommagani R. Role for Growth Regulation by Estrogen in Breast Cancer 1 (GREB1) in Hormone-Dependent Cancers. Int J Mol Sci. 2018;19(9):2543.

54. Rae JM, Johnson MD, Cordero KE, Scheys JO, Larios JM, Gottardis MM, Pienta KJ, Lippman ME. GREB1 is a novel androgen-regulated gene required for prostate cancer cell growth. Prostate. 2006;66(8):886–94.

55. Mazure NM, Chen EY, Laderoute KR, Giaccia AJ. Induction of vascular endothelial growth factor by hypoxia is modulated by a phosphatidylinositol 3-kinase/Akt signaling pathway in Ha-ras-transformed cells through a hypoxia inducible factor-1 transcriptional element. Blood. 1997;90(9):3322–31.

56. Pugh CW, Ratcliffe PJ. Regulation of angiogenesis by hypoxia: role of the HIF system. Nat Med. 2003;9(6):677–84.

57. Myers G. Efficient Local Alignment Discovery amongst Noisy Long Reads. Berlin, Heidelberg: Springer Berlin Heidelberg; 2014. p. 52–67.

58. Chin CS, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler EE, et al. Nonhybrid, finished microbial

Li *et al. BMC Biology*   (2023) 21:40

Page 14 of 14

genome assemblies from long-read SMRT sequencing data. Nat Methods. 2013;10(6):563–9.

59. Roach MJ, Schmidt SA, Borneman AR. Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. BMC Bioinformatics. 2018;19(1):460.

60. Adey A, Kitzman JO, Burton JN, Daza R, Kumar A, Christiansen L, Ronaghi M, Amini S, Gunderson KL, Steemers FJ, et al. In vitro, long-range sequence information for de novo genome assembly via transposase contiguity. Genome Res. 2014;24(12):2041–9.

61. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One. 2014;9(11):e112963.

62. Burton JN, Adey A, Patwardhan RP, Qiu R, Kitzman JO, Shendure J. Chromosome-scale scaffolding of de novo genome assemblies based on chromatin interactions. Nat Biotechnol. 2013;31(12):1119–25.

63. Xu Z, Wang H. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. Nucleic Acids Res. 2007;35(Web Server issue):W265-268.

64. Bergman CM, Quesneville H. Discovering and detecting transposable elements in genome sequences. Brief Bioinform. 2007;8(6):382–92.

65. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215(3):403–10.

66. Yu X-J, Zheng H-K, Wang J, Wang W, Su B. Detecting lineage-specific adaptive evolution of brain-expressed genes in human using rhesus macaque as outgroup. Genomics. 2006;88(6):745–51.

67. Birney E, Clamp M, Durbin R. GeneWise and Genomewise. Genome Res. 2004;14(5):988–95.

68. Korf I. Gene finding in novel genomes. BMC Bioinformatics. 2004;5:59.

69. Burge C, Karlin S. Prediction of complete gene structures in human genomic DNA. J Mol Biol. 1997;268(1):78–94.

70. Majoros WH, Pertea M, Salzberg SL. TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. Bioinformatics. 2004;20(16):2878–9.

71. Stanke M, Waack S. Gene prediction with a hidden Markov model and a new intron submodel. Bioinformatics. 2003;19(Suppl 2):ii215-225.

72. Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, White O, Buell CR, Wortman JR. Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. Genome Biol. 2008;9(1):R7.

73. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. 2013;14(4):R36.

74. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc. 2012;7(3):562–78.

75. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol. 2011;29(7):644–52.

76. UniProt Consortium T. UniProt: the universal protein knowledgebase. Nucleic Acids Res. 2018;46(5):2699.

77. The Gene Ontology C. Expansion of the Gene Ontology knowledgebase and resources. Nucleic Acids Res. 2017;45(D1):D331–8.

78. Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M, Tanabe M. Data, information, knowledge and principle: back to metabolism in KEGG. Nucleic Acids Res. 2014;42(Database issue):D199-205.

79. Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 1997;25(5):955–64.

80. Griffiths-Jones S, Moxon S, Marshall M, Khanna A, Eddy SR, Bateman A. Rfam: annotating non-coding RNAs in complete genomes. Nucleic Acids Res. 2005;33(Database issue):D121-124.

81. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004;32(5):1792–7.

82. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30(9):1312–3.

83. Wang Y, Tang H, Debarry JD, Tan X, Li J, Wang X, Lee TH, Jin H, Marler B, Guo H, et al. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res. 2012;40(7):e49.

84. Pevzner P, Tesler G. Genome rearrangements in mammalian evolution: lessons from human and mouse genomes. Genome Res. 2003;13(1):37–45.

85. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25(14):1754–60.

86. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. Genome Project Data Processing S: The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25(16):2078–9.

87. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010;38(16):e164.

88. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, et al. The variant call format and VCFtools. Bioinformatics. 2011;27(15):2156–8.

89. Chen H, Patterson N, Reich D. Population differentiation as a test for selective sweeps. Genome Res. 2010;20(3):393–402.

90. Xie C, Mao X, Huang J, Ding Y, Wu J, Dong S, Kong L, Gao G, Li CY, Wei L. KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. Nucleic Acids Res. 2011;39(Web Server issue):W316-322.

91. Hudson RR. Generating samples under a Wright-Fisher neutral model of genetic variation. Bioinformatics. 2002;18(2):337–8.

92. Malinsky M, Challis RJ, Tyers AM, Schiffels S, Terai Y, Ngatunga BP, Miska EA, Durbin R, Genner MJ, Turner GF. Genomic islands of speciation separate cichlid ecomorphs in an East African crater lake. Science. 2015;350(6267):1493–8.

93. Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. Nat Protoc. 2016;11(9):1650–67.

94. Anders S, Pyl PT, Huber W. HTSeq–a Python framework to work with high-throughput sequencing data. Bioinformatics. 2015;31(2):166–9.

95. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26(1):139–40.

96. Li WM, Du J, Yang LY, Liang QQ, Yang MY, Zhou XM, Du WG. Chromosome-level genome assembly and population genomics of Mongolian racerunner (*Eremias argus*) provide insights into high-altitude adaptation in lizards. 2022. NCBI https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA659114.

97. Li WM, Du J, Yang LY, Liang QQ, Yang MY, Zhou XM, Du WG. Chromosome-level genome assembly and population genomics of Mongolian racerunner (*Eremias argus*) provide insights into high-altitude adaptation in lizards. 2022. Figshare. https://doi.org/10.6084/m9.figshare.21098470.

Publisher's Note