

RESEARCH

Open Access



Whole-genome sequencing of Tarim red deer (*Cervus elaphus yarkandensis*) reveals demographic history and adaptations to an arid-desert environment

Buweihailiqiemu Ababaikeri^{1,2†}, Shamshidin Abduriyim^{3,4†} , Yilamujiang Tohetahong¹, Tayerjan Mamat¹, Adil Ahmat¹ and Mahmut Halik^{1*}

Abstract

Background: The initiation of desert conditions in the Tarim Basin in China since the late Miocene has led to the significant genetic structuring of local organisms. Tarim Red Deer (*Cervus elaphus yarkandensis*, TRD) have adapted to the harsh environmental conditions in this basin, including high solar radiation and temperature, aridity, and poor nutritional conditions. However, the underlying genetic basis of this adaptation is poorly understood.

Results: We sequenced the whole genomes of 13 TRD individuals, conducted comparative genomic analyses, and estimated demographic fluctuation. The $\partial a\partial i$ model estimated that the TRD and Tule elk (*Cervus canadensis nannodes*) populations diverged approximately 0.98 Mya. Analyses revealed a substantial influence of the Earth's climate on the effective population size of TRD, associated with glacial advances and retreat, and human activities likely underlie a recent serious decline in population. A marked bottleneck may have profoundly affected the genetic diversity of TRD populations. We detected a set of candidate genes, pathways, and GO categories related to oxidative stress, water reabsorption, immune regulation, energy metabolism, eye protection, heat stress, respiratory system adaptation, prevention of high blood pressure, and DNA damage and repair that may directly or indirectly be involved in the adaptation of TRD to an arid-desert environment.

Conclusions: Our analyses highlight the role of historical global climates in the population dynamics of TRD. In light of ongoing global warming and the increasing incidence of droughts, our study offers insights into the genomic adaptations of animals, especially TRD, to extreme arid-desert environments and provides a valuable resource for future research on conservation design and biological adaptations to environmental change.

Keywords: Arid-desert environment, *Cervus elaphus*, Environmental adaptability, Population demographic history, Tarim red deer, Whole genome sequencing

* Correspondence: mahmuthalic@xju.edu.cn

[†]Buweihailiqiemu Ababaikeri and Shamshidin Abduriyim contributed equally to this work.

¹College of Life Sciences and Technology, Xinjiang University, Urumqi 830046, Xinjiang, China

Full list of author information is available at the end of the article



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Background

Adaptation to diverse and changing environments is a fundamental principle in evolutionary biology, and estimating the potential for adaptive evolution is critical when identifying populations/species at risk of extinction due to environmental change [1]. In wild animals, climate-induced extinctions, distributional and phenological changes, and species' range shifts have been documented at an increasing rate; in particular, the influence of climate change on organisms is exacerbated in extreme environments, such as on plateaus and in desert regions [2]. Adaptation is a complex process that involves many biological systems and quantitative trait loci, each having a small but cumulative effect on the overall expression of the phenotype [3–5]. Identifying genes that are under natural selection and understanding how species have adapted genetically to extreme environments provide information for design of conservation programs to effectively protect important species under scenarios of changing climate [6]. Revealing how populations/species adapt to changing environments thus plays a critically important role in assessing their evolutionary and

ecological dynamics, and predicting population resilience to climate change [7].

Red deer (genus *Cervus*) are widely distributed around most of Holarctic region, from northern Africa, southern Europe (Spain, Sardinia, Corsica) and western Europe (Scotland) across Asia to North America. Species in the genus show a broad range of morphological features (i.e. antler and body size; coat colour) in different climatic conditions, and show considerable biological adaptations for survival under diverse environmental conditions—for example, extreme cold, hot, or moderate temperatures; extremely dry or wet conditions; high altitude; steppe-like habitats, forests, and swampy river plains [8]. The Tarim red deer (TRD; *C. e. yarkandensis*) is a subspecies of *Cervus elaphus* mainly distributed along the Tarim River and its tributaries in the Tarim Basin in Xinjiang Uygur Autonomous Region (XUAR), China (Fig. 1). Aridification and desertification in the basin date back to 5.3 Mya (millions of years ago [9, 10]);. Honored as “the flower of Asia” in the international market owing to its large quantities of velvet antler with good quality [11, 12], and being the main large-

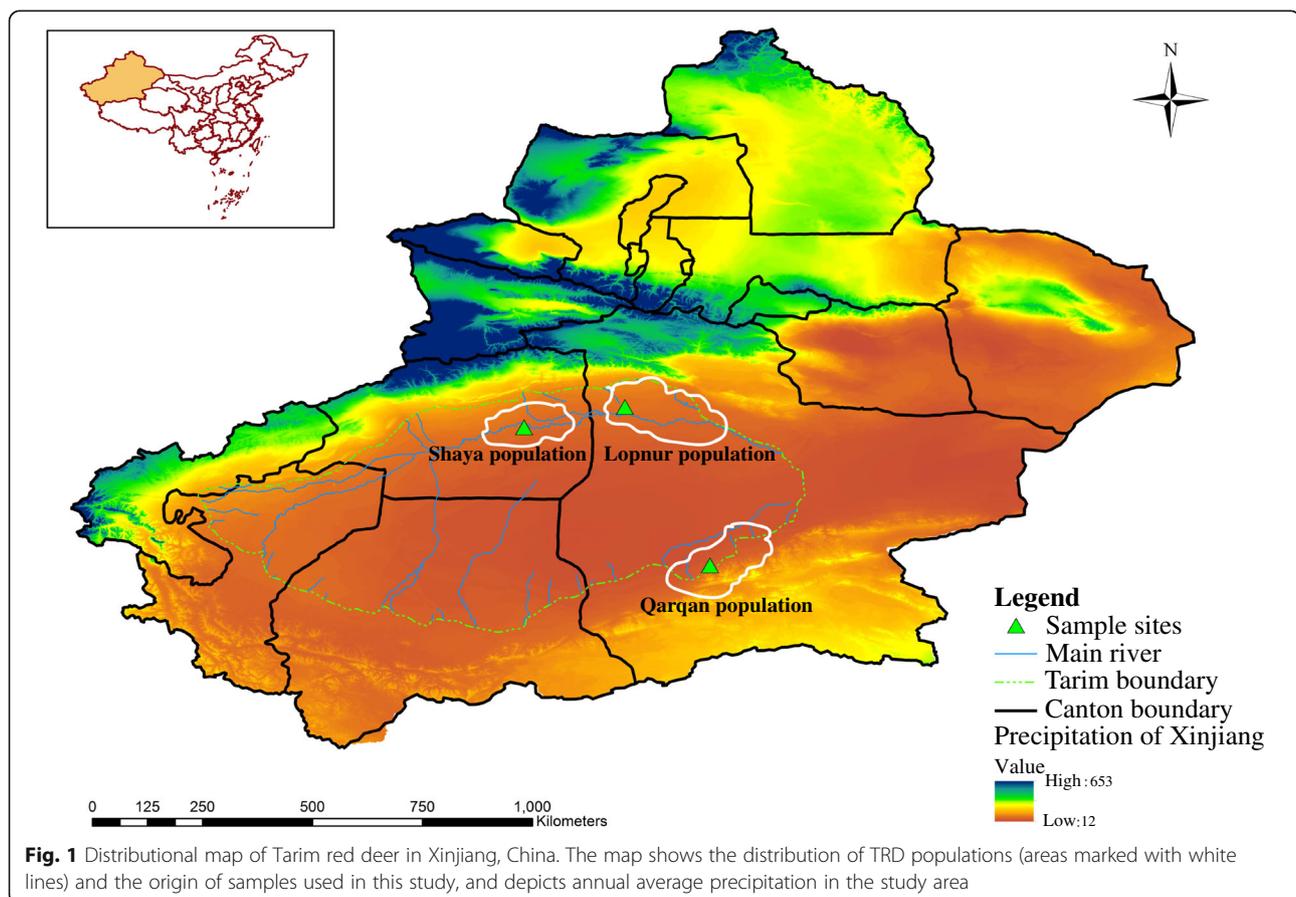


Fig. 1 Distributive map of Tarim red deer in Xinjiang, China. The map shows the distribution of TRD populations (areas marked with white lines) and the origin of samples used in this study, and depicts annual average precipitation in the study area

sized ungulate that plays an important role in maintaining arid and semi-arid forest ecosystems and contributing biodiversity, the Tarim red deer is economically and ecologically the most important animal in Tarim Basin, and has been designated as a Category II protected animal in China [13]. In a long evolutionary process, they have been subjected to strong environmental pressures. The basin is extremely arid, with the average annual precipitation less than 10 mm and the average annual evaporation reaching 1900–2700 mm [14]. There is intense solar radiation and heat; the annual average temperature in the distributional area is 10.5 °C, with temperatures up to 40 °C in summer and surface temperatures reaching over 60 °C [13, 15, 16]. Feeding conditions are poor; TRD forage mainly on *Phragmites communis*, *Tamarix ramosissima*, and *Populus diversifolia*, which are difficult to chew [17], and drink highly mineralized water [13]. Several studies have focused on the adaptive evolution of TRD in this arid-desert environment (e.g. [13, 18–20]), but the mechanisms underlying adaptive traits are poorly understood.

Advances in next-generation sequencing technology now make it possible to sequence individuals' whole genomes and detect selective sweeps, which can provide insights into genome biology and the mechanisms of adaptation to extreme environments. Sequencing cost and time have dramatically decreased [21], allowing whole genome sequencing (WGS) data to be obtained for many individuals, allowing variation to be examined among data sets [22]. Genomic differences can shed light on the genetic basis of adaptation to diverse environments and provide insights into functionally important genetic variants [23]. It is now possible to combine WGS data and population genomic approaches to characterize adaptive variation in an unprecedented way. The integration of genomic information will undoubtedly lead to better management and sustainable utilization of genetic resources.

In recent years, WGS has been widely used to explore adaptive genetic variations of humans and other species living in harsh or extreme environments, for instance, goat (*Capra hircus*) and sheep (*Ovis aries*) in hot, arid environments [3, 24–26]; domestic Bactrian camel (*Camelus actrianus*) in severe desert conditions [27, 28]; and Tibetan yak (*Bos grunniens*), Tibetan antelope (*Pantholops hodgsonii*), Tibetan mastiff (*Canis lupus familiaris*), and Tibetan sheep (*Ovis aries*) at high altitudes on the Tibetan plateau [29–33].

Here, we Illumina-sequenced the whole genomes of multiple TRD, compared them to previously published Tule elk (*Cervus canadensis nannodes*) genomic data [34], revealed the divergence times and demographic history of these species, as well as identified a set of

candidate genes and genomic regions under direct or indirect selection in an extremely arid environment. Our findings provide a more complete picture of TRD evolutionary history, making it possible to investigate features unique to this species and how these might reflect adaptation to a hot, arid environment, all of which are imperative for a better understanding of possible responses to current and future climatic changes. Our results are important for their potential application to functional genomics and in the design of programs to protect genetic resources in view of ongoing global warming and the increasing incidence of droughts.

Materials and methods

Sampling

Samples were collected from 13 TRD for sequencing. Muscle samples from three free-ranging individuals that had been poached or died of natural causes were provided by the Local Forestry Bureaus of Qarqan (two individuals) and Shaya (one individual) counties in XUAR, China. Blood samples were obtained in 2018 from 10 captive individuals by means of jugular venipuncture into EDTA-coated vacutainer tubes; these individuals were from a farmed population in Lopnur, XUAR, that had undergone no known hybridization with any other red deer lineages (Fig. 1, Additional file 1: Table S1). Blood collection was conducted in strict accordance with the Animal Ethics Procedures and Guidelines of the People's Republic of China.

The domestication of TRD in Xinjiang began in 1958 with the capture of about 2000 wild individuals, mainly juveniles, from their natural habitat [35]. The capture of wild individuals continued until the TRD was listed as a National class II key protected species in 1982 and capture became prohibited [35]. The average mutation rate per base per generation (μ) for a similar species, Milu (Père David's deer, *Elaphurus davidianus*), was found to be 1.5×10^{-8} [5], and the generation time (g) of red deer is 6.3 years [36]. Thus, the impact of 36 years of captive history has likely had a negligible effect on this species' genome. Importantly, the blood samples we used in this study were collected from farms in Lopnur County, with a dry climate typical of temperate continental plains (average annual precipitation 45.2 mm).

Tule elk (*Cervus canadensis nannodes*) are typified by having lower body masses, lighter pelage, and the longest tooth rows of any North American subspecies [37]. They are far more widespread in North America, occupying much of California's extensive lower elevation oak woodland and have unique morphological adaptation to mild Mediterranean grasslands [38, 39]. They are sensitive to variation in climate, especially precipitation [38], and a hot, dry climate, or droughts, appear to have had substantial negative effects on elk herds independent of

hunting pressure [40]. For these reasons, previously published genomic data of four Tule elk that were sampled from four geographically distinct populations across northern California [34] with average annual precipitation more than 324 mm were retrieved from NCBI (BioProject ID PRJNA345218) to represent deer from a non-arid-desert habitat for comparison.

Sequencing, read mapping, and quality control

The whole genome of each of the 13 TRD samples was Illumina-sequenced. For each sample, a library was constructed with an insert size of 350 bp from high-quality DNA and then sequenced on an Illumina HiSeq Xten platform using paired-end (PE) 150 bp reads following standard procedures. The average sequencing depth was ~8.57X per sample. Adapter and low-quality read pairs, which included reads with >10% unidentified nucleotides (nt) and reads containing >50% low-quality bases ($Q < 5$) in paired reads, >10 nt aligned to the adaptor, allowing $\leq 10\%$ mismatches, were filtered from the sequenced reads (raw reads) to obtain high-quality reads (clean reads). High-quality reads were then mapped to the CerEl1.0 red deer (*Cervus elaphus hippelaphus*) reference genome [41] available from NCBI, using Burrows-Wheeler Aligner software Version: 0.7.8 [42] under the parameters 'mem -t 4 -k 32 -M.' To reduce mismatches generated by PCR amplification before sequencing, duplicate reads were removed by SAMtools Version: 0.1.19 [43] with the parameter '-type rmdup.'

The genome was divided into segments and analyzed in parallel. After alignment, we performed SNP calling on a population scale by using a Bayesian approach implemented in the package SAMtools v0.1.19. 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. To identify SNPs, we used the 'mpileup' command and parameters '-q 1 -C 50 -t SP -t DP -m 2 -F 0.002.' (See Additional file 2: Table S2 for parameter illustration). High-quality variants were selected among raw SNPs based on the following quality scores: coverage depth ≥ 3 , RMS (root mean square) mapping quality ≥ 20 (i.e. SNPs with a sequencing error rate >1% were filtered out); miss ratio ≤ 0.2 ; and MAF (minimum allele frequency) ≥ 0.01 . If two SNPs were less than 5 bp apart, both were removed.

Positional annotation of genetic variants

SNPs were annotated against the CerEl1.0 genome (CerEl1.0.gff file) by using the package ANNOVAR version 2013-05-20 [44]. Based on the genomic annotation, SNPs were categorized as occurring in exons, in introns, at splice sites (within 2 bp of a splice junction), upstream or downstream regions (within 1 kb upstream or downstream from a transcription start site) and intergenic

regions. SNPs in exons were further categorized as synonymous (caused no amino acid change) or nonsynonymous (caused an amino acid change, or caused the gain or loss of a stop codon).

Linkage-disequilibrium analyses

To assess the linkage disequilibrium (LD) pattern between the TRD and Tule elk populations, the coefficient of determination between any two loci in each population was calculated by using PopLDdecay Version 3.4 [45] under the parameters '-MaxDist 500 -MAF 0.05 -Miss 0.1 -Het 1.' For plotting LD decay curves, the average r^2 value (average squared correlations of allele frequencies) was calculated for pairwise markers in 500-kb windows and averaged across the whole genome; this was repeated 10 times. The LD decay rate was measured as the distance at which r^2 dropped to half its maximum value.

Demographic history and population admixture

To better understand historical changes in ancestral TRD and Tule elk population size, demographic history was analyzed with SNP data by applying the pairwise sequentially Markovian coalescent (PSMC) model [46]. This method infers ancestral effective population sizes (N_e) over time, based on the SNP distribution in an individual diploid genome. Parameters were set at -g 6.3, -u 1.5×10^{-8} , -p $4 + 25 \times 2 + 4 + 6$, -b 100, -x 10,000, -X 10000000, including the average mutation rate (μ) of 1.5×10^{-8} per base per generation in Milu [5] and a generation time (g) of 6.3 years [36].

The demographic history was reconstructed and the pattern of gene flow was elucidated for the TRD and Tule elk populations subsequent to their divergence by using the diffusion approximation for demographic inference ($\partial\text{a}\partial\text{i}$) approach [47], which infers demographic parameters based on a diffusion approximation to the site frequency spectrum. Because reduction in genetic variation in the sex chromosomes will affect estimates of the population genetic parameters, and to ensure selective neutrality, SNPs located on the sex chromosomes were eliminated, and only autosomal SNPs were used for the two analyses of demographic history.

Genome-wide selective sweep test

To identify genome-wide selective sweeps associated with adaptation to arid environments, the genome-wide distribution of fixation index (F_{ST}) values [48] and θ_π ratios were calculated for the TRD and Tule elk populations. The F_{ST} values for sliding windows were also calculated using VCFtools [49], for windows 40,000 bp wide sliding in 20,000 bp steps. The F_{ST} values were Z-transformed as follows: $Z(F_{ST}) = (F_{ST} - \mu F_{ST}) / \sigma F_{ST}$, where μF_{ST} is the mean F_{ST} and σF_{ST} is the standard

deviation of F_{ST} . The θ_{π} ratios were \log_2 -transformed ($\pi_{\text{Tule elk population (control group)}/\pi_{\text{TRD population (arid-desert environment group)}}$). We then estimated and ranked the empirical percentiles of $Z(F_{ST})$ and $\log_2(\theta_{\pi}$ ratio) in each window and considered the windows with the top 5% $Z(F_{ST})$ and $\log_2(\theta_{\pi}$ ratio) values simultaneously as candidate outliers under strong selective sweeps. All outlier windows were assigned to corresponding SNPs and genes.

Since the cross-population extended haplotype homozygosity statistic (XP-EHH [50];) maintains power for sample sizes as low as ten [51], XP-EHH was estimated for the TRD and Tule elk populations with XP-EHH version 2012-12-12 (<http://hgdp.uchicago.edu/Software/>), using windows 10,000 bp wide sliding in 5000 bp steps. The genetic map was assumed to be 2.53 cM/Mb for the CerElai1.0 deer genome [41]. Z scores were calculated using a burn-in of 100 and 1000 sweeps. The threshold for identifying candidate genes in the XP-EHH analyses was set to the top 5% of outlier regions.

We compared the $Z(F_{ST})$, $\log_2(\theta_{\pi}$ ratio), and Ztrans (XP-EHH) values of the selective genomic regions with those at the whole-genome scale for TRD and Tule elk. The CerElai1.0 deer reference genome assembly was used to identify the coordinates of nucleotide sequences. Functional information for each gene was obtained by using Blastp to compare the gene set against protein databases, including SwissProt (<http://www.uniprot.org>), TrEMBL (<http://www.uniprot.org/>), KEGG (<http://www.genome.jp/kegg/>), and InterProScan (<https://www.ebi.ac.uk/interpro/>). As enrichment analyses, the Gene Ontology (GO) term and functional pathway were analyzed for the candidate genes, i.e., all of the annotated genes in the outlier windows with the top 5% $Z(F_{ST})$, $\log_2(\theta_{\pi}$ ratio), and Ztrans (XP-EHH) values, using the Go-seq package in R. Specifically, the genes were compared to this classification for functional enrichment analysis of GO biological processes, molecular function, and cellular component terminologies and pathways. Only significantly ($p < 0.05$) over-represented GO terms were considered to be biologically meaningful. The candidate genes were also positioned on known Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways (<http://www.kegg.jp>) by using KOBAS (KEGG Orthology Based Annotation System [52], selecting cow (*Bos taurus*) genome as orthology).

Results

Genome sequencing and mapping

By re-sequencing the whole genomes of 13 TRD individuals at an average depth of 8.57X, we generated 328.167 Gb of raw sequences, which yielded 327.803 Gb of high-quality sequences (Additional file 3: Table S3). We retrieved 113.394 Gb of sequence for the Tule elk from

the NCBI Sequence Read Archive (SRA; BioProject ID PRJNA345218), with 88% of the data showing Phred quality scores higher than 30 and the aligned high-quality genomic data at an overall effective sequence depth of 40 for subsequent analyses [34]. Based on the high-quality sequence data, the genomic GC content averaged 44.03% for the TRD (Additional file 3: Table S3) and 41.55% for the Tule elk [34], which are typical values comparable to those observed in the genomes of other mammals, including ungulates [25, 28–30], indicating that our sequencing data were not affected by GC bias. The mapping rates of our samples to the CerElai1.0 reference genome assembly ranged from 88.77 to 90.35% (Additional file 4: Table S4).

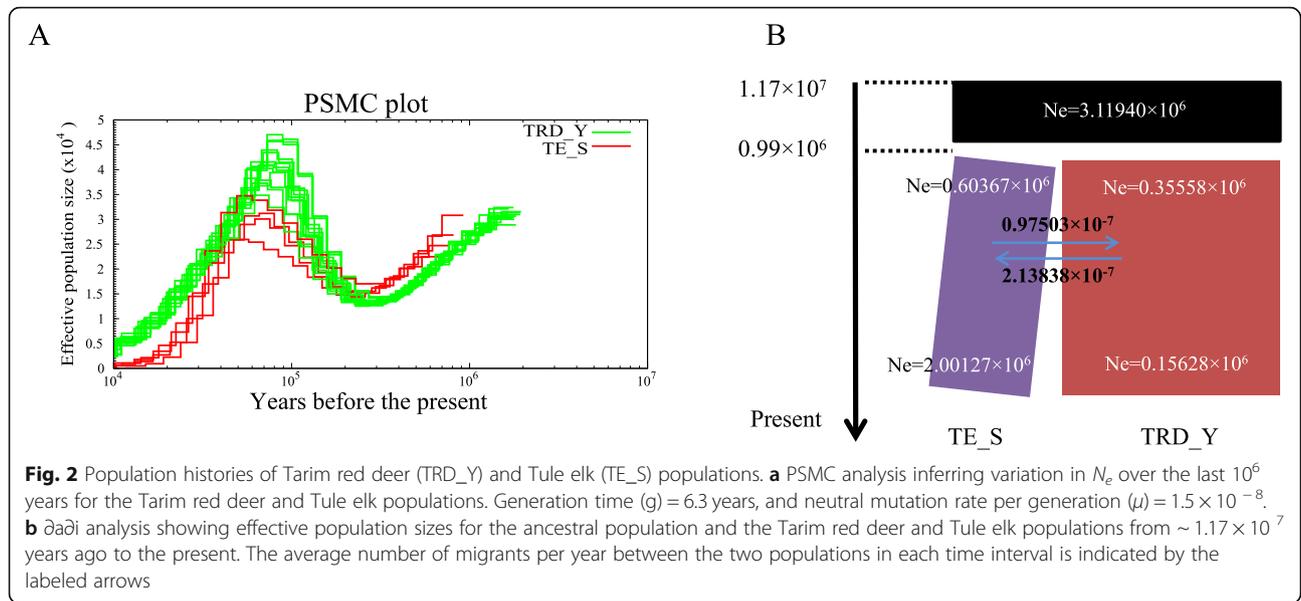
We identified 8,520,805 SNPs after quality control for TRD and 4,896,658 for Tule elk. Most of the high-quality SNPs were located in intergenic (7,002,807) and intronic regions (1,321,254), with only 36,192 synonymous and 27,623 nonsynonymous SNPs located within exons in TRD (Additional file 5: Table S5). Among the SNPs of TRD, heterozygous SNPs for individuals ranged from 1,603,886 to 2,184,617 (Additional file 6: Table S6). As a measure of the quality of our SNP data, the transition-transversion (ts/tv) ratio for TRD and Tule elk was 2.64 and 2.63, respectively, which was consistent with the extensive analysis of mammalian genes by Li [53].

Demographic history

Genome-wide patterns of heterozygosity provide useful information on long-term effective population sizes. We reconstructed historical fluctuations in N_e for TRD and Tule elk with the PSMC model and identified two bottlenecks and one expansion the two populations, which showed highly similar patterns of historical fluctuation in N_e (Fig. 2a). The first continuous decline in N_e was estimated to have taken place between 1.8 and 0.25 Mya, after which the ancestral TRD population expanded until about 0.075 Mya (Fig. 2a), followed by another decline in N_e . The $\partial\text{a}\partial\text{i}$ model, on the other hand, revealed that the TRD and Tule elk populations diverged ~ 0.98 Mya, at an estimated N_e of 0.356×10^6 for TRD and 0.604×10^6 for Tule elk (Fig. 2b, Additional file 7: Table S7). The present N_e for TRD was 0.156×10^6 , an apparent decline of more than two-fold since its initial divergence. Very low levels of gene flow were calculated between TRD and Tule elk (2.13×10^{-7} migrants per year from TRD to Tule elk; 0.975×10^{-7} from Tule elk to TRD), which accords with their geographic distributions and divergence time [8, 13, 54].

Patterns of genomic variation and linkage disequilibrium

The genome-wide average θ_{π} value was 0.561×10^{-3} for TRD and 0.240×10^{-3} for Tule elk. The Tule elk

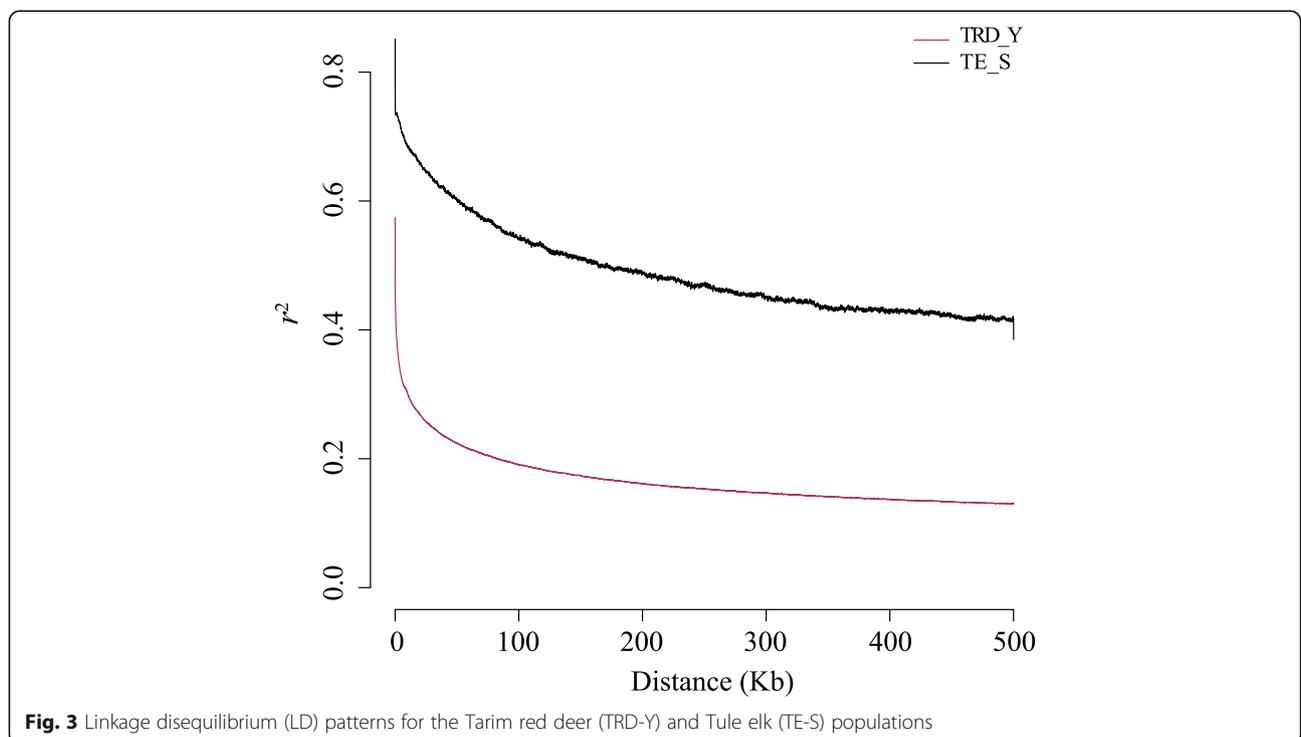


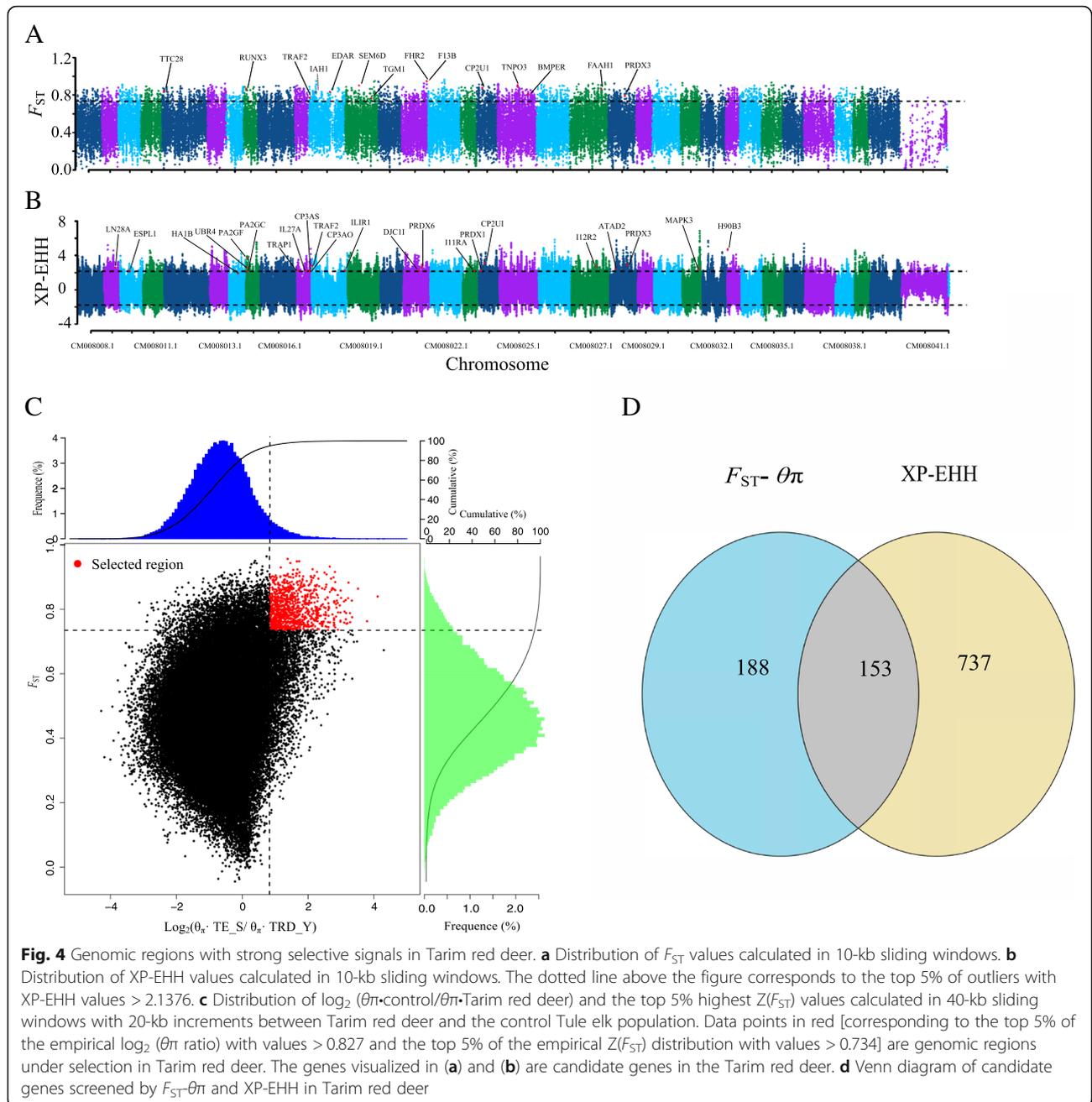
population exhibited an overall slow decay rate and a high level of linkage disequilibrium (LD), whereas the TRD population showed a rapid decay rate and a low level of LD (Fig. 3). For the whole genomes, the LD decay rate was estimated at ~ 437.1 kb for Tule elk and ~ 13 kb for TRD, with r^2 dropping differentially to 0.4257 and 0.2870, respectively. These differences in the genome-wide LD patterns suggest that natural selection differentially affected the population size, structure and

history between the two populations, as LD is a comprehensive reflection of various evolutionary forces.

Selective sweep and GO enrichment analyses

We considered the top 5% of F_{ST} values and the θ_π ratio cut-off values, $Z(F_{ST}) > 0.734$ and $\log_2(\theta_\pi \text{ ratio}) > 0.827$ (Fig. 4c) for the Tule elk population (control group) / TRD population (arid-desert environment group), to be potentially selected regions associated with adaptation to





an arid-desert environment. We identified 955 putative regions spread across all chromosomes, comprising of 341 candidate genes (Fig. 4a, Additional file 8: Table S8). Among these genes, however, nine did not match any annotated genes, due either to incomplete annotation of the genomes used in this study, or more likely to the selected functional mutations within each of these regions not being located within or close to a protein-coding gene [24]. In addition, the XP-EHH analysis identified 890 genes (top 5% outliers, XP-EHH value > 2.1376) as positive candidates for involvement in adaptation to an

arid-desert environment (Fig. 4b, Additional file 9: Table S9). Overall, using multiple approaches, we identified 153 overlapping candidate genes for adaptation to an arid-desert environment (Fig. 4d, Additional file 10: Table S10).

The candidate genes identified by $F_{ST} - \theta_{\pi}$ method were enriched in various fundamental biological processes and molecular functions, and were highlighted by 90 significant gene ontology (GO) terms ($p < 0.05$, without Bonferroni correction) (Fig. 5a, Additional file 11: Table S11). The top several categories from the GO

database associated with these genes included transferase activity, inorganic anion transport, cellular response to stimulus, metabolic and biosynthetic process (such as RNA metabolic and biosynthetic process, transcription), oxidoreductase activity, hydrolase activity, and regulation of signaling pathway (Additional file 11: Table S11).

We also performed a gene-set enrichment test for KEGG pathways and GO categories for the genes selected by the XP-EHH method and identified significantly enriched categories. Most of the candidate genes were associated with multiple signaling, signal transduction pathways, and various metabolic pathways (Fig. 5). Enrichment tests revealed GO categories that were significantly related to myosin binding, prosthetic group metabolic process, protein secretion, ion binding, galacturonan metabolic process, pectin metabolic-catabolic process, response to inorganic substance, regulation of immune system process, response to water stimulus, and sensory organ development (Additional file 12: Table S12). Among the top KEGG pathways, the candidate genes were significantly enriched in fanconi anemia pathway, lysine degradation, ECM-receptor interaction, Notch signaling pathway, bacterial invasion of epithelial cells, Linoleic acid metabolism, melanogenesis, and others (Additional file 13: Table S13).

Discussion

Demographic history

The first period of decline in N_e (Fig. 2a) for TRD coincided with a shift in the climate cycle 1–0.25 Mya [55]. Decreases in N_e have also been reported for the giant panda (*Ailuropoda melanoleuca* [56]); brown bear (*Ursus arctos* [57]); and snub-nosed monkeys (*Rhinopithecus roxellana*, *R. brelichi* and *R. bieti* [58]) during this period, when global glaciation, cold climate, and frequent fluctuations in sea level [59] had a profound evolutionary effect on the population sizes of some mammalian species.

The timing of ancestral population expansion for TRD (Fig. 2a) coincided with the end of the interglacial period (0.13–0.07 Mya [60]); when environmental conditions became similar to those at present [61]. A recovery of N_e occurred in *R. roxellana* around the same time (~0.07 Mya [58]); but that for giant panda was inferred to have occurred later, around 30,000–40,000 years ago [56]. The population expansion in TRD was likely due to the relatively warm, wet climatic conditions during the last interglacial period, allowing organisms to recolonize the northeast Tarim as glacial melt-water increased and rivers resumed their courses [62]. TRD underwent a subsequent decline in N_e after 0.075 Mya. This time interval coincided with extreme cooling during the last glaciation [59], when the climate became colder and drier, and the meltwater that served as the main source

of water in the Tarim Basin greatly decreased [63]. The northern and eastern parts of the Tarim River dried up, and the resident biota must have either gone locally extinct or retreated to the southwest [62].

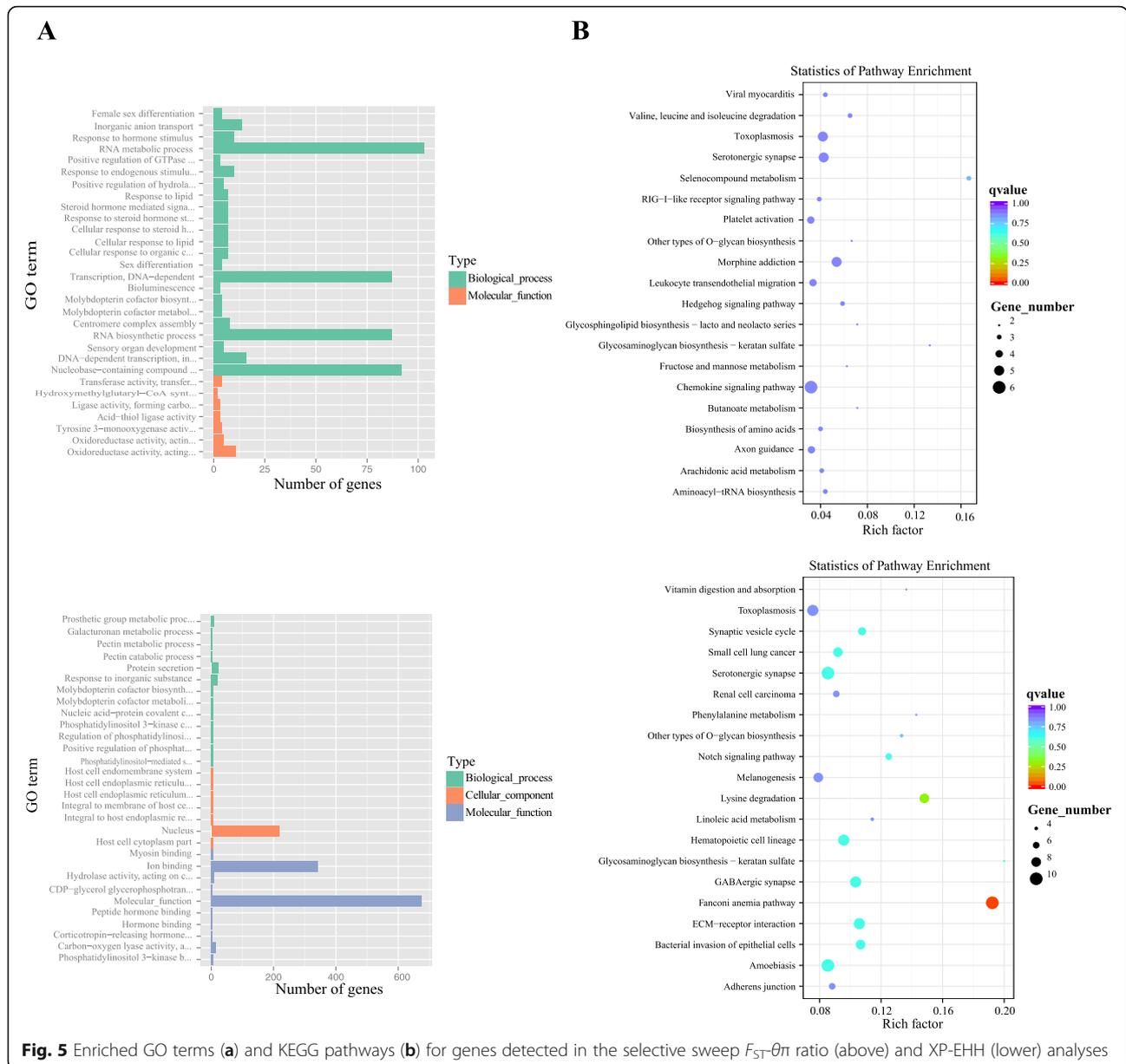
On the other hand, the latter decrease in TRD N_e coincided with the dispersal of modern humans from Africa to Eurasia [64] which caused a similar decrease in N_e in several other mammals [56–58, 65]. Lorenzen et al. [59] suggested that the decline in genetic diversity observed in horse and bison after the last glaciation, and to a lesser extent in reindeer, might reflect the impact of expanding human populations in Europe and Asia. We similarly observed reduced genomic diversity in TRD relative to other mammals (Additional file 14: Table S14). We thus hypothesize that the combination of an extremely cold climate and anthropogenic effects is responsible for the second reduction in the TRD N_e .

Interestingly, the divergence time of ~0.98 Mya between TRD and Tule elk (Fig. 2b, Additional file 7: Table S7) estimated by using the *daði* model differed from that estimated by mitochondrial Cyt *b* and D-loop data [66], in which the most recent common ancestor between Eastern (*C. canadensis* and *C. nippon*) and Western (*C. elaphus* and *C. hanglu*) Clades may have occurred in the late Miocene, at a mean age of approximately 6 Mya, but was consistent with an estimate based on genomic SNP data that the isolation of TRD and *C. canadensis* occurred 0.8–2.2 Mya [67], suggesting that TRD colonized the Tarim Basin more recently than previously thought [66].

Patterns of genomic variation and linkage disequilibrium

In agreement with previous studies based on single or multiple genes [37, 68–70], θ_π values for the TRD and Tule elk populations (0.561×10^{-3} and 0.240×10^{-3} , respectively) showed a lower level of genomic diversity than that of other mammals [28, 30, 71, 72], and was similar to or even lower than that of some endangered mammals (Additional file 14: Table S14). Most of these species have gone through bottleneck events at least once in their evolutionary history, suggesting that historical changes in population size contributed to the loss of genetic variation (e.g. [73, 74]), including that in TRD. Corroborating this conclusion, Broughton et al. [38] provided provisional evidence, based on DNA from archaeological elk bones, that a decline in genetic diversity in the California Tule elk (included in our study) was due to a historical bottleneck event.

Other factors could also have contributed to the low genomic diversity to some extent, such increased human activities and the degraded and fragmented habitat of the TRD population [75, 76]. Higher LD in the Tule elk population (Fig. 3) may have contributed to reduced genomic variation, since selection in cases of relatively



high LD can reduce neutral nucleotide diversity by strong selection against deleterious alleles or by substitutions of beneficial alleles at linked loci [77–79]. Finally, genetic variation responds to experimental manipulations of population size [80, 81], and we cannot rule out that low genetic variation in TRD and Tule elk might be due to small population sizes and/or inbreeding effects [75, 82].

Adaptive signatures to an arid-desert environment

The $Z(F_{ST})$ and $\log_2(\theta_{\pi}$ ratio) values we estimated for selected genomic regions of TRD from an arid-

desert zone were much higher than the overall genomic level (Additional file 15: Fig. S1). Relative to the overlap expected by chance, we observed a marked excess of overlapping selective signals shared between the candidate genes identified in this study and candidate genes published previously for other mammals in the same or a similar environment (Additional file 16: Fig. S2). Likewise, a number of the candidate genes overlapped among the F_{ST} , θ_{π} , and XP-EHH methods (Fig. 4d, Additional file 10: Table S10). Hence, it is evident that the candidate genes identified in this study were reliable.

Several overlapping selective signals were shared between the candidate genes identified by the $F_{ST}-\theta_{\pi}$ ratio here and the predefined gene panel for previously published candidate genes in other mammalian species in a similarly extreme environment (Additional file 16: Fig. S2). For example, several genes with high selective sweeps in TRD have also been identified as candidate genes in the adaptation of Bactrian camels (*TRAF2*, *PRDX3*, *EDAR*, *TGMI*, *F13B*, *IAHI*) and sheep breeds (*RUNX3*, *TNPO3* and *TTC28*) to desert environments [25, 28]. *EDAR* (Tumor necrosis factor receptor superfamily member EDAR) functions in the repair of skin tissue damage, maintains skin microenvironment homeostasis [83], and is involved in the formation and regulation of hair curl, which helps in temperature homeostasis and protects against heat damage [84]. Furthermore, Bornert et al. [85] found that the more the function of mouse *EDAR* was preserved, the shorter the mandible was. This parallels the significantly shorter and wider skull morphology of TRD compared to congeners like the Tianshan red deer (*C. e. songaricus*) and Altai red deer (*C. e. sibiricus*) in Xinjiang, China; this morphology is presumably adaptive to eating food hardened by dryness and heat [86]. We hypothesize that the *EDAR* gene in TRD is involved in the development of hair and sweat glands, which likely play important roles in regulating body temperature and maintaining water-salt balance, and is related to adaptive skull morphology, but these functions need to be confirmed.

The functions of the GO terms for candidate genes selected by $F_{ST}-\theta_{\pi}$ ratio were quite consistent with adaptive features of TRD, such as the relatively compact, medium-sized body and high salt-alkali tolerance. Our results agree well with the conclusions of Yang et al. [25] that the nervous system development, metabolic process, and response to stimulus terms are related to adaptations of sheep breeds to desert environments. The KEGG pathway analysis indicated that these selected genes were enriched in several biological pathways, including selenocompound metabolism, toxoplasmosis, the hedgehog signaling pathway involved in ovarian activity [87], valine, leucine and isoleucine degradation, and others (Fig. 5, Additional file 17: Table S15).

Some candidate genes (Additional file 16: Fig. S2) in TRD located in the selective sweep regions identified by the XP-EHH analysis were related to the adaptation of Bactrian camels to a desert environment (*TRAF2*, *UBR5*, *DBX1*, *PRDX3*, *SLX4*, *TPO* and *NLE1* [28]); and sheep breeds to the Taklimakan Desert (*EHBPI*, *TTC28*, *IRF5*, *RNF24*, *TSHB* and *KCP* [25]); and to arid environments (*CCL22*, *RUNX3*, *STK40*, *HECW2*, *NDOR1* *ATAD2* and *SMC1B* [25]). TRD and sheep breeds inhabit same environment in the Tarim Basin, and overlapping candidate genes indicate that species in the basin have faced

similar selective pressures and have undergone convergent evolution. Apart from these candidate genes, none of our candidate regions overlapped with those identified in previous studies, perhaps due to different study designs, sample sizes, sequencing coverage, and statistical methods, as well as species-specific differences. We found 16 candidate genes (*TGMI*, *FHR2*, *RPN2*, *GPR98*, *PK3CB*, *I12R2*, *IL1R1*, *TBA*, *CABP8*, *TCPA*, *VINC*, *CASP3*, *LAMB2*, *LAMB1*, *ITAM* and *CO4A4*) were also identified by the XP-EHH analysis to be significantly over-represented in GO categories and KEGG pathways (Additional file 12: Table S12 and Additional file 13: Table S13), including genes related to myosin binding (GO:0017022, $p < 0.01$), regulation of immune system process (GO:0002682, $p < 0.05$) and immune response (GO:0050776, GO:0050778, $p < 0.05$), bacterial invasion of epithelial cells (bta05100, $p < 0.05$), ECM-receptor interaction (bta04512, $p < 0.05$), amoebiasis (bta05146, $p < 0.05$), small cell lung cancer (bta05222, $p < 0.05$), renal cell carcinoma (bta05211), toxoplasmosis (bta05145), HTLV-I infection (bta05166), NF-kappa B signaling pathway (bta04064), and apoptosis (bta04210). These clusters and pathways are immunologically relevant to adaptation to arid-desert conditions, because many of them are involved in disease and pathogen resistance. TRD suffer diseases such as tuberculosis [88, 89], pasteurellosis [90], and cisticercustenuicollis [91]; in addition, parasites like hard ticks, widely distributed in the Gobi Desert within the Tarim Basin, are a main threat during the hot season [92]. It is thus plausible that genes related to the immune system would be targeted by natural selection in TRD. Previous studies for other mammalian species have similarly reported genes related to the immune system to be involved in adaptation to hot, arid environments [3, 93].

The Tarim Basin has high solar radiation, and long-term exposure to ultraviolet radiation can lead to a number of ophthalmic conditions in animals. Five candidate genes (*LAMB1*, *LAMB2*, *CYC*, *FANCF* and *GPR98*) were positively selected by XP-EHH analysis in TRD are functionally involved in ocular development, visual protection, and photoreceptor cell synapses [94–96]. Two of these (*CYC* and *FANCF*) evolved more rapidly in the Bactrian camel and dromedary exposed to long-term ultraviolet radiation than in the alpaca [28]. These genes may play a central role in visual perception in TRD, and selection on eye physiology may be an important feature of adaptation to arid-desert conditions.

Relevant to the adaptation of TRD to arid conditions, ten selected genes (including *SLX4*, *FANCF*, *FANCG*, *FANCI*, *ATR*, and *POLH*) were related to the fanconi anemia pathway (bta03460, $p < 0.01$) (Additional file 13: Table S13) involved in DNA repair, DNA replication fork stability, and other cellular processes [97]. In

addition, the peroxiredoxins (*PRDX1*, *PRDX3* and *PRDX6*) we identified in the high selective sweep regions belong to a family of peroxidases that regulate the level of reactive oxygen species (ROS) in cells, are involved in antioxidant protection and cell signaling, and are associated with increased resistance to radiation [98]. *PRDX3* was enriched in the oxidoreductase activity (GO: 0016628, GO: 0016627, $p < 0.05$), and response to various stimulus (GO: 0009725, GO: 0009719, GO: 0033993, GO: 0048545, GO: 0071383, $p < 0.05$, Additional file 11: Table S11) GO categories. These candidate genes are likely important for the survival of TRD in an arid environment by playing significant roles in prevention of DNA damage, DNA repair, oxidative stress, and stress responses.

The TRD inhabits the edge of Taklimakan Desert, where sandstorms are frequent and airborne dust can initiate respiratory diseases such as asthma. The candidate gene *TRAF2* (TNF receptor-associated factor 2 protein) was identified for TRD in both the $F_{ST}-\theta_{\pi}$ and XP-EHH approaches was also a critical gene in the adaptation of Bactrian camels to a desert environment [28]. Involved in the signal transduction of immune function related receptors (e.g. CD40, CD30, CD27, 4-1BB and RANK) and combined with RIP, GCK, ASK1 and NIK, *TRAF2* plays an important role in the regulation of most kinases and their corresponding transcription factors [99, 100]. In addition, high level of *TRAF2* expression in asthmatic rats and humans showed that it is more likely involved in the inflammatory mechanism of asthma possibly through airway epithelial cells and inflammatory cells [101]. Meanwhile, *IL1R1* (Interleukin-1 receptor type 1), positively identified under XP-EHH selection in TRD, is the signal-transducing receptor for IL-1 and encodes cytokine receptor that participate in host defense mechanisms, including immune and inflammatory responses [102]. Playing a crucial role in IL-1 signaling pathway, *IL1R1* can increase the effect of IL-1, which is critically involved in the regulation of inflammation and pathobiology of immune and inflammatory conditions, such as asthma and allergic diseases [103, 104]. We thus speculate that *TRAF2* and *IL1R1* may help defend against airborne dusts in TDR, and that selection on respiratory physiology helps TRD adapt to its arid-desert environment.

Among the candidate genes for arid-desert adaptation we obtained by XP-EHH analysis, *CP2U1*, *PGH1*, *PTGD S*, *HYES*, *PA2GF* and *PA2GC* are functionally involved in the arachidonic acid metabolic pathway (bta00590, Additional file 13: Table S13), which was likewise identified as highly functionally important in desert adaptation in the Bactrian camel [27] and sheep breeds [25], indicating convergent evolution. Four candidate genes (*PA2GF*, *PA2GC*, *CP3AS* and *CP3AO*) are involved in linoleic acid

metabolism (bta00591, Additional file 13: Table S13), where *PA2GF* and *PA2GC* play a vital role in converting lecithin into linoleic acid. *CP2U1* (cytochrome P450 2 U1), *CP3AS* (cytochrome P450 3A28), and *CP3AO* (cytochrome P450 3A24) belong to the cytochrome P450 (CYP) family, members of which display a wide variety of functions in the metabolism of many xenobiotic species, including the endogenous biosynthesis of fatty acids, bile acids, steroids, or vitamins [105, 106]. The *CP2U1* (bta00590, Additional file 13: Table S13) gene product, in particular, can help to transform arachidonic acid into 19(S)-HETE and 20-HETE. 19(S)-HETE is a potent vasodilator of the renal preglomerular vessels that stimulate water reabsorption [107], and is potentially useful for the survival of camels in the desert [27]. 20-HETE is a potent vasoconstrictor in the kidney and brain and is involved in the pathogenesis of hypertension [108, 109], which is consistent with the consumption of highly mineralized water and salty foods (i.e. *Phragmites communis* and *Tamarix ramosissima*) by TRD [5, 13, 110]. It appears, then, that arachidonic acid, linoleic acid, and fatty acid metabolisms, and particularly the *CP2U1* gene, are important factors for the survival of TRD in its arid environment. As in Onzima et al. [26], the candidate gene *MAPK3* (mitogen-activated protein kinase 3) may also play role in desert adaptation; this gene is related to encoding the Na^+/K^+ -ATPase and the epithelial Na^+ channel and is thus involved in the reabsorption of sodium in the kidney [28].

The candidate genes *H90B3* (putative heat shock protein HSP 90-beta-3, known as *HSP90AB3P* in humans) and *TRAP1* (mitochondrial heat shock protein 75 kDa) also identified by XP-EHH analysis are involved in ion binding (GO:0043167, $p < 0.01$), transition metal ion binding (GO: 0046914, $p < 0.05$), oxidoreductase activity (GO: 0016652, $p < 0.05$), and metabolic process (GO: 0008152, $p < 0.05$) (Additional file 12: Table S12), which may be related to heat stress response [111]. *DJCI1* (DNAJ homolog subfamily C member 11, known as *DNAJC11* in humans) and *AUX1* (putative tyrosine-protein phosphatase auxilin, known as *DNAJC6* in *Bos taurus*) are associated with the heat shock family of proteins [112]. The identification of multiple genes associated with heat stress suggests that these genes are involved in the adaptation TRD to a hot environment. Moreover, we found enrichment for eight candidate genes (*GNAI2*, *FZD4*, *MP2K2*, *CREB3*, *CBP*, *GNAO*, *TF7L2* and *GNAO*) in the melanogenesis pathway (bta04916, Additional file 13: Table S13). This is consistent with the dark gray-brown coat color of TRD in summer turning to light gray-brown in winter, and the obvious black-brown back line. Melanogenesis associated with heat-tolerance appears to be common among species adapted to desert and desert-like environments

[3]. Accordingly, we assumed that this KEGG pathway is likely involved in the coat color and thermal tolerance traits of this species.

We found intriguing over-represented biological pathways based on raw p -values, though the results of the more stringent Bonferroni correction were not statistically significant. This is likely attributable to the small sample size and relatively low coverage of sequencing depth in this study and may also due to the genomic differences between the red deer and the Tule elk. However, our results provide a useful indication of the mechanisms involved in arid-desert adaptation in TRD. We found putative signatures containing a complex of genes, pathways, and GO terms directly or indirectly related to oxidative stress, water reabsorption, immune-regulation, energy metabolism, eye protection, heat stress, the respiratory system, prevention of high blood pressure, and prevention and repair of DNA damage. As with previous studies [3, 25, 28, 113], our study underscores that arid-desert adaptation is complex, involving various biological processes and quantitative trait loci, each contributing a small but cumulative effect to the overall phenotype.

Conclusions

The genomic sequencing data for TRD could be valuable and useful for both conservations of wild populations and utilization of domestic ones as a resource for further study, providing baseline sequence and annotation data for the species. Meanwhile the selective sweep results advance our understanding of the genetic mechanisms underlying the adaptations of TRD and other species in similar extreme environments. However, as the samples we analyzed were minimal and to some extent the Tule elk we use as comparison was somehow distantly related to Tarim red deer, more detailed studies are necessary to further confirm and refine our results by integrating comprehensive genomic data (for example, candidate gene sequencing, high density SNP genotyping, gene expression profiling) with environmental and physiological data, and to identify underlying genomic mechanisms.

Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s12983-020-00379-5>.

Additional file 1: Table S1. Identity, location, and sample type for *Cervus elaphus yarkandensis* and *Cervus canadensis nannodes* samples used in this study.

Additional file 2: Table S2. The explanation of parameters in samtools 'mpileup' command used for SNP calling.

Additional file 3: Table S3. Summary of sequence data generated in this study.

Additional file 4: Table S4. Statistics for mapping rate and coverage of samples analyzed in this study.

Additional file 5: Table S5. SNP annotation statistics of Tarim red deer and Tule elk.

Additional file 6: Table S6. Individual SNP statistics of Tarim red deer.

Additional file 7: Table S7. Parameters inferred from the diffusion approximation for demographic inference ($\partial\text{a}\partial\text{i}$) approach.

Additional file 8: Table S8. List of genes in the overlapping regions selected by the top 5% highest $\log_2(\theta_{\pi\text{-control}}/\theta_{\pi\text{-Tarim red deer}})$ and top 5% highest $Z(F_{ST})$ scores for Tarim red deer.

Additional file 9: Table S9. List of genes in the overlapping regions selected by the top 5% highest XP-EHH values for Tarim red deer in XP-EHH analyses.

Additional file 10: Table S10. List of genes in the selected overlapping regions by top 5% highest $\log_2(\theta_{\pi\text{-control}}/\theta_{\pi\text{-Tarim red deer}})$ and top 5% highest $Z(F_{ST})$ for Tarim red deer as well as the significance of these genes in XP-EHH (XP-EHH value) analyses.

Additional file 11: Table S11. GO functional enrichment of the candidate genes obtained by $F_{ST}-\theta_{\pi}$ ratio analysis that are associated with adaptation to arid environments.

Additional file 12: Table S12. GO functional enrichment of the candidate genes obtained by XP-EHH analysis associated with the drought environment adaptations.

Additional file 13: Table S13. The top 30 KEGG pathway enrichment of the candidate genes obtained by XP-EHH analysis associated with the drought environment adaptations.

Additional file 14: Table S14. Summary statistics for genomic nucleotide diversity in different mammalian species.

Additional file 15: Figure S1. Differences in $Z(F_{ST})$ and θ_{π} ratio values between the selected regions and the whole-genome scale for Tarim red deer (TRD). (A) Comparison between $Z(F_{ST})$ values for genomic regions that have undergone selective sweeps and $Z(F_{ST})$ values at the whole-genome scale in TRD. The upper, middle (within the box), and lower boundary lines for the boxes represent 25, 50% (median value), and 75% of the $Z(F_{ST})$ and θ_{π} ratio values, respectively. (B) Comparison between θ_{π} ratio values for genomic regions that have undergone selective sweeps and θ_{π} ratio values at the whole-genome scale.

Additional file 16: Figure S2. Venn diagrams showing the overlap of candidate genes identified by the F_{ST} & θ_{π} (A) and XP-EHH (B) analyses in Tarim red deer. Numbers in the intersecting regions are the observed overlapping genes among the candidate genes in Tarim red deer, and predefined gene panel, i.e., previously published candidate genes in other mammalian species in arid environments, including the Bactrian camel, sheep breeds from the Taklimakan Desert region, and sheep breeds from arid regions.

Additional file 17: Table S15. Top 30 KEGG pathway enrichment of the candidate genes obtained by $F_{ST}-\theta_{\pi}$ ratio analysis associated with the drought environment adaptations.

Acknowledgements

We thank Dr. Matthew H. Dick (Hokkaido University) for English editing and comments on our manuscript, and the Xinjiang Uyghur Autonomous Regional Forestry Bureau for sampling authorization. This work was funded by the National Natural Science Foundation of China (No. 31560600) and in part by China Postdoctoral Science Foundation (2019 M652701).

Authors' contributions

M.H. conceived the study; B.A. and M.H. designed the research; I.T., T.M., A.A. and M.H. collected the samples; B.A. with help of I.T. generated the data; B.A. and S.A. analyzed the data and interpreted the results; B.A. and S.A. wrote and edited the manuscript; all authors read and approved the final version of manuscript.

Funding

The authors acknowledge the financial support from the National Natural Science Foundation of China (No. 31560600). S. A. was funded by China Postdoctoral Science Foundation (2019 M652701).

Availability of data and materials

The datasets analyzed during the current study are available from the corresponding author on reasonable request.

Ethics approval and consent to participate

The samples used in this study were conducted in strict accordance with the Animal Ethics Procedures and Guidelines of the People's Republic of China.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹College of Life Sciences and Technology, Xinjiang University, Urumqi 830046, Xinjiang, China. ²College of Xinjiang Uyghur Medicine, Hoten 848000, Xinjiang, China. ³College of Life Science, Shihezi University, Shihezi 832003, Xinjiang, China. ⁴Department of Ecology, Hubei Key Laboratory of Cell Homeostasis, College of Life Science, Wuhan University, Wuhan 430072, Hubei, China.

Received: 16 June 2020 Accepted: 2 October 2020

Published online: 16 October 2020

References

- Thullier W, Münkemüller T, Lavergne S, Mouillot D, Mouquet N, Schiffers K, et al. A road map for integrating eco-evolutionary processes into biodiversity models. *Ecol Lett*. 2013;16:94–105.
- Easterling DR, Meehl GA, Parmesan C, Changnon SA, Karl TR, et al. Climate extremes: observations, modeling, and impacts. *Science*. 2000;289:2068–74.
- Kim ES, Elbeltagy AR, Aboul-Naga AM, Rischkowsky B, Sayre B, Mwacharo JM, et al. Multiple genomic signatures of selection in goats and sheep indigenous to a hot arid environment. *Heredity*. 2016;116:255–64.
- Liu X, Zhang Y, Li Y, Pan J, Wang D, Chen W, et al. EPAS1 gain-of-function mutation contributes to high-altitude adaptation in Tibetan horses. *Mol Biol Evol*. 2019;11:11.
- Zhu L, Deng C, Zhao X, Ding J, Huang H, Zhu S, et al. Endangered Père David's deer genome provides insights into population recovering. *Evol Appl*. 2018;11:2040–53.
- Skelly DK, Joseph LN, Possingham HP, Freidenburg LK, Farrugia TJ, Kinnison MT, et al. Evolutionary responses to climate change. *Conserv Biol*. 2007;21:1353–5.
- Oostra V, Saastamoinen M, Zwaan BJ, Wheat CW. Strong phenotypic plasticity limits potential for evolutionary responses to climate change. *Nat Commun*. 2018;9:1005.
- Grubb P. Nomenclature, subspeciation and affinities among elaphine deer (*Cervus elaphus sensu lato*). In: Oswald C, editor. *Symposium on red deer taxonomy*. Engleng: Bavaria; 2004. p. 13–31.
- Sun JM, Liu TS. The age of the Taklimakan Desert. *Science*. 2006;312:1621.
- Sun J, Zhang L, Deng C, Zhu R. Evidence for enhanced aridity in the Tarim Basin of China since 5.3 Ma. *Quat Sci Rev*. 2008;27:1012–23.
- Tao D, Zhao J, Deng G, Jiao J. Relationship between velvet antler ossification and PTH and androgen serum levels in Tarim red deer (*Cervus elaphus*). *J Exp Zool*. 2015;323A:696–703.
- Zhang XW, Ma CH, Bai GB. Identification of Tarim red deer (*C. elaphus yarkandensis*) by allele-specific diagnostic PCR and PCR-RFLP based on sequences of 12S ribosomal RNA. *China J Tradit Chin Med Pharm*. 2011;26:570–3 (In Chinese).
- Mahmut H, Anwar T, Ohtaishi N. Tarim red deer of Xinjiang in China. Urumqi: Xinjiang University Press; 2012. (In Chinese).
- Piao S, Ciais P, Huang Y, Shen Z, Peng S, Li J, et al. The impacts of climate change on water resources and agriculture in China. *Nature*. 2010;467:43–51.
- Gu J, Gao X, Xiang L. Red deer in western Lopnor district—scientific survey and research in Lopnor. Beijing: Science Press; 1987. p. 226–34. (In Chinese).
- Luo N, Gu J. Status of Yarkand deer and strategy for its conservation and utilization. *Acta Zool Arid Inland*. 1993;1:38–41 (In Chinese).
- Qiao JF, Yang WK, Gao XY. Analysis of natural diet and food habitat selection of the Tarim red deer, *Cervus elaphus yarkandensis*. *Chin Sci Bull*. 2006;51(Suppl):147–52 (In Chinese).
- Qian WX, Ao WP, Yusup A. Feed intake and digestibility of Tarim red deer. *China Herbivore Sci*. 2014;32:31–2 (In Chinese).
- Qian WX, Ao WP, Hui XH, Wu JP. Lower dietary concentrate level increases bacterial diversity in the rumen of *Cervus elaphus yarkandensis*. *Can J Microbiol*. 2018;64(7):501–9.
- Qian WX, Ao WP, Jia CH, Li ZP. Bacterial colonisation of reeds and cottonseed hulls in the rumen of Tarim red deer (*Cervus elaphus yarkandensis*). *Antonie Van Leeuwenhoek*. 2019;112(9):1283–96.
- Snyder M, Du J, Gerstein M. Personal genome sequencing: current approaches and challenges. *Genes Dev*. 2010;24:423–31.
- Altshuler DM, Durbin RM, Abecasis GR, Bentley DR, Chakravarti A, Clark A, et al. An integrated map of genetic variation from 1,092 human genomes. *Nature*. 2012;491:56–65.
- Andersson L, Georges M. Domestic-animal genomics: deciphering the genetics of complex traits. *Nat Rev Genet*. 2004;5:202–12.
- Benjelloun B, Alberto FJ, Streeter I, Boyer F, Coissac E, Stucki S, et al. Characterizing neutral genomic diversity and selection signatures in indigenous populations of Moroccan goats (*Capra hircus*) using WGS data. *Front Genet*. 2015;6:107.
- Yang J, Li WR, Lv FH, He SG, Tian SL, Peng WF, et al. Whole-genome sequencing of native sheep provides insights into rapid adaptations to extreme environments. *Mol Biol Evol*. 2016;33:2576–92.
- Onzima RB, Upadhyay MR, Doeke HP, Brito LF, Bosse M, Kanis E, et al. Genome-wide characterization of selection signatures and runs of homozygosity in Ugandan goat breeds. *Front Genet*. 2018;9:318.
- Jirimutu, Wang Z, Ding G, Chen G, Sun Y, Sun Z, et al. Genome sequences of wild and domestic bactrian camels. *Nat Commun*. 2012;3:1202.
- Wu H, Guang X, Al-Fageeh MB, Cao J, Pan S, Zhou H, et al. Camelid genomes reveal evolution and adaptation to desert environments. *Nat Commun*. 2014;5:5188.
- Ge RL, Cai Q, Shen YY, San A, Ma L, Zhang Y, et al. Draft genome sequence of the Tibetan antelope. *Nat Commun*. 2013;4:1858.
- Qiu Q, Zhang G, Ma T, Qian W, Ye Z, Cao C, et al. The yak genome and adaptation to life at high altitude. *Nat Genet*. 2012;44:946–9.
- Gou X, Wang Z, Li N, Qiu F, Xu Z, Yan D, et al. Whole-genome sequencing of six dog breeds from continuous altitudes reveals adaptation to high-altitude hypoxia. *Genome Res*. 2014;24:1308–15.
- Wang X, Liu J, Zhou G, Guo J, Yan H, Niu Y, et al. Whole-genome sequencing of eight goat populations for the detection of selection signatures underlying production and adaptive traits. *Sci Rep*. 2016;6:38932.
- Yang J, Jin ZB, Chen J, Huang XF, Li XM, Liang YB, et al. Genetic signatures of high-altitude adaptation in Tibetans. *P Natl Acad Sci USA*. 2017;114:4189–94.
- Mizzi JE, Lounsberry ZT, Brown CT, Sacks BN. Draft genome of Tule elk *Cervus elaphus nannodes*. *F1000Res*. 2017;6:1691.
- Jiang Y. Domestication history of Tarim red deer. *Anim Husbandry Vet Sci Technol Inf*. 2014;8:125–6 (In Chinese).
- Zheng XT, Zhao M, Li S, Chang ZJ. Statistical analysis of genetic parameters of quantity characters of different deer species, varieties and strains. *J Econ Anim*. 2001;5:25–9 (In Chinese).
- Meredith EP, Rodzen JA, Banks JD, Schaefer R, Ernest HB, Famula TR, et al. Microsatellite analysis of three subspecies of Elk (*Cervus elaphus*) in California. *J Mammal*. 2007;88:801–8.
- Broughton JM, Beck RK, Coltrain JB, O'Rourke DH, Rogers AR. A late Holocene population bottleneck in California Tule Elk (*Cervus elaphus nannodes*): provisional support from ancient DNA. *J Archaeol Method Th*. 2012;20:495–524.
- McCullough DR. The tule elk: its history, behavior, and ecology. Berkeley: California University Press; 1971.
- Howell JA, Brooks GC, Semenov-Irving M, Greene C. Population dynamics of Tule elk at Point Reyes National Seashore, California. *J Wildlife Manage*. 2002;66:478–90.
- Bana N, Nyiri A, Nagy J, Frank K, Nagy T, Stéger V, et al. The red deer *Cervus elaphus* genome CerEl1.0: sequencing, annotating, genes, and chromosomes. *Mol Gen Genomics*. 2018;293:665–84.
- Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25:1754–60.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. *Bioinformatics*. 2009;25:2078–9.
- Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from next-generation sequencing data. *Nucleic Acids Res*. 2010;38:e164.

45. Zhang C, Dong SS, Xu JY, He WM, Yang TL. PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. *Bioinformatics*. 2019;35:1786–8.
46. Li H, Durbin R. Inference of human population history from individual whole-genome sequences. *Nature*. 2011;475:493–6.
47. Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genet*. 2009;5:e1000695.
48. Weir BS, Cockerham CC. Estimating *F*-statistics for the analysis of population structure. *Evolution*. 1984;38:1358–70.
49. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. *Bioinformatics*. 2011;27:2156–8.
50. Sabeti PC, Varilly P, Fry B, Lohmueller J, Hostetter E, Cotsapas C, et al. Genome-wide detection and characterization of positive selection in human populations. *Nature*. 2007;449:913–8.
51. Pickrell JK, Coop G, Novembre J, Kudaravalli S, Li JZ, Absher D, et al. Signals of recent positive selection in a worldwide sample of human populations. *Genome Res*. 2009;19:826–37.
52. Mao X, Cai T, Olyarchuk JG, Wei L. Automated genome annotation and pathway identification using the KEGG orthology (KO) as a controlled vocabulary. *Bioinformatics*. 2005;21:3787–93.
53. Li WH. *Molecular Evolution*. Sunderland: Sinauer Associates; 1997. p. 487.
54. Ludt JC, Schroeder W, Rottmann O, Kuehn R. Mitochondrial DNA phylogeography of red deer (*Cervus elaphus*). *Mol Phylogenet Evol*. 2004;31:1064–83.
55. Lisiecki LE, Raymo ME. A Pliocene-Pleistocene stack of 57 globally distributed benthic $\delta^{18}\text{O}$ records. *Paleoceanography*. 2005;20:PA103.
56. Zhao S, Zheng P, Dong S, Zhan X, Wu Q, Guo X, et al. Whole-genome sequencing of giant pandas provides insights into demographic history and local adaptation. *Nat Genet*. 2013;45:67–71/U99.
57. Miller W, Schuster SC, Welch AJ, Ratan A, Bedoya-Reinao OC, Zhao F, et al. Polar and brown bear genomes reveal ancient admixture and demographic footprints of past climate change. *Proc Natl Acad Sci USA*. 2012;109:E2382–90.
58. Zhou X, Wang B, Pan Q, Zhang J, Kumar S, Sun X, et al. Whole-genome sequencing of the snub-nosed monkey provides insights into folivory and evolutionary history. *Nat Genet*. 2014;46:1303–10.
59. Lorenzen ED, Nogués-Bravo D, Orlando L, Weinstock J, Binladen J, Marske KA, et al. Species-specific responses of Late Quaternary megafauna to climate and humans. *Nature*. 2011;479:359–64.
60. Ehlers J, Gibbard PL. The extent and chronology of Cenozoic global glaciation. *Quatern Int*. 2007;164/165(Apr):6–20.
61. Orlando L, Ginolhac A, Zhang G, Froese D, Albrechtsen A, Stiller M, et al. Recalibrating equus evolution using the genome sequence of an early Middle Pleistocene horse. *Nature*. 2013;499:74–8.
62. Shan WJ, Liu J, Yu L, Robert WM, Mahmut H, Zhang YP. Genetic consequences of postglacial colonization by the endemic Yarkand hare (*Lepus yarkandensis*) of the arid Tarim Basin. *Chin Sci Bull*. 2011;56:1370–82.
63. Yang X, Zhu Z, Jaekel D, Owen LA, Han J. Late Quaternary palaeoenvironment change and landscape evolution along the Keriya River, Xinjiang, China: the relationship between high mountain glaciation and landscape evolution in foreland desert regions. *Quatern Int*. 2002;97–98:0–166.
64. Mellars P. Why did modern human populations disperse from Africa ca. 60,000 years ago? A new model. *Proc Natl Acad Sci USA*. 2006;103:9381–6.
65. Prado-Martinez J, Sudmant PH, Kidd JM, Li H, Kelley JL, Lorente-Galdos B, et al. Great ape genetic diversity and population history. *Nature*. 2013;499:471–5.
66. Lorenzini R, Garofalo L. Insights into the evolutionary history of *Cervus* (*Cervidae*, tribe *Cervini*) based on Bayesian analysis of mitochondrial marker sequences, with first indications for a new species. *J Zool Syst Evol Res*. 2015;53:340–9.
67. Hu PP, Shao YC, Xu JP, Wang TJ, Li YQ, Liu HM, et al. Genome-wide study on genetic diversity and phylogeny of five species in the genus *Cervus*. *BMC Genomics*. 2019;20:384.
68. Zhang SY, Zhong L, Jia B, Xi JF, Peng LZ, Zeng XC, et al. Microsatellite analysis of genetic diversity of Tarim red deer in Xinjiang. *Anim Husband Vet Med*. 2006;38:17–20 (In Chinese).
69. Yuan GW, Xi JF, Guo QH, Yuan QB, Zhong L, Peng LZ, et al. Relationships between genetic diversity of microsatellites and antler production capacity of Tarim red deer. *J Econ Anim*. 2008;12:6–12 (In Chinese).
70. Dong XY, Shan WJ, Yu LJ, Mahmut H. Structure of the mitochondrial DNA control region and population genetic diversity analysis of Tarim red deer (*C.e.yarkandensis*). *Biotechnology*. 2010;20:16–20 (In Chinese).
71. Carneiro M, Rubin CJ, Di Palma F, Albert FW, Alföldi J, Barrio AM, et al. Rabbit genome analysis reveals a polygenic basis for phenotypic change during domestication. *Science*. 2014;345:1074–9.
72. Liu S, Lorenzen ED, Fumagalli M, Li B, Harris K, Xiong Z, et al. Population genomics reveal recent speciation and rapid evolutionary adaptation in polar bears. *Cell*. 2014;157:785–94.
73. Hoelzel AR, Fleischer RC, Campagna C, Le Boeuf BJ, Alvord G. Impact of a population bottleneck on symmetry and genetic diversity in the northern elephant seal. *J Evol Biol*. 2002;15:567–75.
74. Leonard JA. Ancient DNA applications for wildlife conservation. *Mol Ecol*. 2008;17:4186–96.
75. Mahmut H, Ganzorig S, Onuma M, Masuda R, Suzuki M, Ohtaishi N. A preliminary study of the genetic diversity of Xinjiang Tarim red deer (*Cervus elaphus yarkandensis*) using the microsatellite DNA method. *Jpn J Vet Res*. 2001;49:231–7.
76. Liu YH. Genetic diversity and phylogenetic evolution of red deer population. Harbin: PhD thesis, Northeast Forestry University; 2011. (In Chinese).
77. Corbett-Detig RB, Hartl DL, Sackton TB. Natural selection constrains neutral diversity across a wide range of species. *PLoS Biol*. 2015;13:e1002112.
78. Filatov DA. Extreme Lewontin's paradox in ubiquitous marine phytoplankton species. *Mol Biol Evol*. 2019;36:4–14.
79. Stephan W. Genetic hitchhiking versus background selection: the controversy and its implications. *Philos Trans R Soc B Biol Sci*. 2010;365:1245–53.
80. Frankham R. Relationship of genetic variation to population size in wildlife. *Conserv Biol*. 1996;10:1500–8.
81. Montgomery ME, Woodworth LM, Nurthen RK, Gilligan DM, Briscoe DA, Frankham R. Relationships between population size and loss of genetic diversity: comparisons of experimental results with theoretical predictions. *Conserv Genet*. 2000;1:33–43.
82. Ba H, Jia B, Wang G, Yang Y, Kedem G, Li C. Genome-wide SNP discovery and analysis of genetic diversity in farmed sika deer (*Cervus nippon*) in northeast China using double-digest restriction site-associated DNA sequencing. *G3*. 2017;7:3169–76.
83. Garcin CL, Huttner KM, Kirby N, Schneider P, Hardman MJ. Ectodysplasin A pathway contributes to human and murine skin repair. *J Invest Dermatol*. 2016;136:102210–30.
84. Jablonski NG, Chaplin G. The evolution of skin pigmentation and hair texture in people of African ancestry. *Dermatol Clin*. 2014;32:113–21.
85. Bornert F, Choquet P, Gros CI, Aubertin G, Perrin-Schmitt F, Claus F, et al. Subtle morphological changes in the mandible of tabby mice revealed by micro-CT imaging and elliptical fourier quantification. *Front Physiol*. 2011;2:15.
86. Mahmut H, Anwar T, Omar A, Ohtaishi N. Geographical variations of skull morphology in red deer (*Cervus elaphus*) in Xinjiang, China. *Acta Theriologica Sinica*. 2005;25:313–8.
87. Chen C, Liu Z, Pan Q, Chen X, Wang H, Guo H, et al. Genomic analyses reveal demographic history and temperate adaptation of the newly discovered honey bee subspecies *Apis mellifera sinixinyuan* n. ssp. *Mol Biol Evol*. 2016;33:1337–48.
88. Zhang YC, Deng HJ. Diagnosis and cure of Tarim red deer tuberculosis. *J Econ Anim*. 2009;13:121–2 (In Chinese).
89. Xu JP. Epidemiological status and prevention and control of para-tuberculosis in red deer in Tarim area. *Chin J Vet Med*. 2010;46:91–2 (In Chinese).
90. Yang XC. Diagnosis and treatment of pasteurellosis of Tarim red deer. *Rural Sci Technol*. 2010;3:59–60 (In Chinese).
91. Jiang XM, Xu JP, Jiao HH, Zhao AY. Cases of cysticercosis infection in the Tarim red deer. *Contemp Anim Husband*. 2017;398:23. <http://www.cqvip.com/QK/93041X/201710/673947647.html>.
92. Liang FX. Integrated control of in vitro parasitics of hard ticks in Tarim red deer. *Spec Econ Anim Plants*. 2009;1:16–7 (In Chinese).
93. Kim J, Hanotte O, Mwai OA, Dessie T, Salim B, Diallo B, et al. The genome landscape of indigenous African cattle. *Genome Biol*. 2017;18:34.
94. Hunter DD, Shah V, Merlie JP, Sanes JR. A laminin-like adhesive protein concentrated in the synaptic cleft of the neuromuscular junction. *Nature*. 1989;338:229–34.
95. Khan SH, Javed MR, Qasim M, Shahzadi S, Rehman SU. Domain analyses of Usher syndrome causing Clarin-1 and GPR98 protein models. *Bioinformation*. 2014;10:491–5.

96. Zenker M, Tralau T, Lennert T, Pitz S, Mark K, Madlon H, et al. Congenital nephrosis, mesangial sclerosis, and distinct eye abnormalities with microcoria: an autosomal recessive syndrome. *Am J Med Genet A*. 2004; 130A:138–45.
97. Rodríguez A, D'Andrea A. Fanconi anemia pathway. *Curr Biol*. 2017;27:R986–8.
98. Yoo YD, Chung YM, Park JK, Ahn CM, Kim SK, Kim HJ. Synergistic effect of peroxiredoxin II antisense on cisplatin-induced cell death. *Exp Mol Med*. 2002;34:273–7.
99. Graham JP, Moore CR, Bishop GA. Roles of the TRAF2/3 binding site in differential B cell signaling by CD40 and its viral oncogenic mimic, LMPI. *J Immunol*. 2009;183(5):2966–73.
100. Dupoux A, Cartier J, Cathelin S, Filomenko R, Solary E, Laurence DD. cIAP1-dependent TRAF2 degradation regulates the differentiation of monocytes into macrophages and their response to CD40 ligand. *Blood*. 2009;113(1): 175–85.
101. Jiang DF, Tong XS, Luo DJ. Regulation with methylprednisolone on the expression of toll like receptor 1, FasL, and tumor necrosis factor receptor associated factor 2 in asthmatic rats. *Zhejiang J Integr Tradit Chin West Med*. 2012;22(6):422–8 (In Chinese).
102. Parrish-Novak J, Dillon SR, Nelson A, Hammond A, Sprecher C, Gross JA, et al. Interleukin 21 and its receptor are involved in NK cell expansion and regulation of lymphocyte function. *Nature*. 2000;408:57–63.
103. Schmitz N, Kurrer M, Kopf M. The IL-1 receptor 1 is critical for Th2 cell type airway immune responses in a mild but not in a more severe asthma model. *Eur J Immunol*. 2010;33(4):991–1000.
104. Madore AM, Vaillancourt VT, Bouzigon E, Sarnowski C, Monier F, Dizier MH, et al. Genes involved in Interleukin-1 receptor type II activities are associated with asthmatic phenotypes. *Allergy Asthma Immun*. 2016;8:466–70.
105. Nebert DW, Russell DW. Clinical importance of the cytochromes P450. *Lancet*. 2002;360:1155–62.
106. Nelson DR, Koymans L, Kamataki T, Stegeman JJ, Feyereisen R, Waxman DJ, et al. P450 superfamily: update on new sequences, gene mapping, accession numbers and nomenclature. *Pharmacogenetics*. 1996;6:1–42.
107. Carroll MA, Balazy M, Margiotta P, Huang DD, Falck JR, McGiff JC. Cytochrome P-450-dependent HETEs: profile of biological activity and stimulation by vasoactive peptides. *Am J Physiol-Reg I*. 1996;271:R863–9.
108. Chuang SS, Helvig C, Taimi M, Ramshaw HA, Collop AH, Amad M, et al. CYP2U1, a novel human yhmus- and brain-specific cytochrome P450, catalyzes ω - and (ω -1)-hydroxylation of fatty acids. *J Biol Chem*. 2004;279: 6305–14.
109. Miyata N, Roman RJ. Role of 20-hydroxyeicosatetraenoic acid (20-HETE) in vascular system. *J Smooth Muscle Res*. 2005;41:175–93.
110. Bai XFF, Zhu JJ, Wang ZL, Tan YQ, Liu LD. Relationship between the salt accumulation and the drought resistance in several woody plants in arid zone. *Scientia Silvae Sinicae*. 2012;48:45–9 (In Chinese).
111. Felts SJ, Owen BAL, Nguyen PM, Trepel J, Donner DB, Toft DO. The hsp90-related protein TRAP1 is a mitochondrial protein with distinct functional properties. *J Biol Chem*. 2000;275:3305–12.
112. Shi Y, Manley JL. A complex signaling pathway regulates SRp38 phosphorylation and pre-mRNA splicing in response to heat shock. *Mol Cell*. 2007;28:79–90.
113. Lv FH, Agha S, Kantanen J, Colli L, Stucki S, Kijas JW, et al. Adaptations to climate-mediated selective pressures in sheep. *Mol Biol Evol*. 2014;31:3324–43.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

