Introgression drives adaptation to the plateau environment in a subterranean rodent

Yukun Kang[1,2], Zhicheng Wang[1,2], Kang An[1,2], Qiqi Hou[1,2], Zhiming Zhang[1,2] and Junhu Su[1,2,3*]

Abstract

Background Introgression has repeatedly been shown to play an important role in the adaptation of species to extreme environments, yet how introgression enables rodents with specialized subterranean lifestyle to acclimatize to high altitudes is still unclear. Myospalacinae is a group of subterranean rodents, among which the high-altitude plateau zokors (*Eospalax baileyi*) and the low-altitude Gansu zokors (*E. cansus*) are sympatrically distributed in the grassland ecosystems of the Qinghai-Tibet Plateau (QTP). Together, they provide a model for the study of the role of introgression in the adaptation of low-altitude subterranean rodents to high altitudes.

Results Applying low-coverage whole-genome resequencing and population genetics analyses, we identified evidence of adaptive introgression from plateau zokors into Gansu zokors, which likely facilitated the adaptation of the latter to the high-altitude environment of the QTP. We identified positively selected genes with functions related to energy metabolism, cardiovascular system development, calcium ion transport, and response to hypoxia which likely made critical contributions to adaptation to the plateau environment in both plateau zokors and high-altitude populations of Gansu zokors.

Conclusions Introgression of genes associated with hypoxia adaptation from plateau zokors may have played a role in the adaptation of Gansu zokors to the plateau environment. Our study provides new insights into the understanding of adaptive evolution of species on the QTP and the importance of introgression in the adaptation of species to high-altitude environments.

Keywords Myospalacinae, Low-coverage whole-genome resequencing, Population genomics, Adaptive evolution, Introgression

*Correspondence:
Junhu Su
sujh@gsau.edu.cn
[1] College of Grassland Science, Key Laboratory of Grassland Ecosystem (Ministry of Education), Gansu Agricultural University, Lanzhou 730070, China
[2] Gansu Agricultural University-Massey University Research Centre for Grassland Biodiversity, Gansu Agricultural University, Lanzhou 730070, China
[3] Gansu Qilianshan Grassland Ecosystem Observation and Research Station, Wuwei 733200, China

Background

Under rapid climate change, species may adapt to their environments by generating novel advantageous mutations, but as the spontaneous generation of these mutations is rare, this scenario is often insufficient to cope with the various environmental challenges [1, 2]. The acquisition of advantageous variants via gene introgression can become a key way for species to improve their fitness, especially in a rapidly changing environment, and the rates of acquiring new genetic alleles or haplotypes via introgression are much faster than those of acquiring beneficial variants through spontaneous mutation [3].

Kang *et al. BMC Biology*   (2024) 22:187

Page 2 of 23

Thus, when a species is not at its optimal fitness due to range expansion or the strong impacts of environmental changes, its adaptation to new environmental challenges can be accelerated through introgressive hybridization with a closely related native species that has already evolved adaptive strategies to cope with similar environmental challenges [4]. This process of rapid adaptive evolution of species by moving adaptive alleles between species through gene introgression is called adaptive introgression. In contrast to de novo mutations, alleles from other species have been pre-tested by selection [5], and through adaptive introgression, recipient species can directly benefit from environmental adaptation strategies that have evolved over a longer period of time in the donor species [4]. Thus, as an important evolutionary mechanism [6], adaptive introgression is currently recognized as an important source of adaptation across diverse taxa [7].

More and more evidence has underscored the importance of gene introgression from other species for adaptation to new environmental challenges in recipient species [8–12]. At present, due to the unique environment and rich species diversity on the Qinghai-Tibet Plateau (QTP), a large number of studies have focused on adaptation to the high-altitude and low oxygen environment in this region via adaptive introgression. These include adaptive introgressions occurring from Neanderthals and Denisovans, both ancient humans, to Tibetans [13, 14], from yak to Tibetan cattle [15, 16], and from Tibetan wolves to Mastiff dogs [4, 17]. These studies all highlight the critical role of gene introgression in the adaptation of species to high altitude and low oxygen.

Myospalacinae, also known as zokors, are subterranean rodents that inhabit the agropastoral ecotone with an annual precipitation of 400 mm in China [18, 19]. Because of their specialized subterranean lifestyle, zokors make excellent subjects for the study of hypoxic adaptation. Among the eight species of Myospalacinae distributed across China, the plateau zokor (*Eospalax baileyi*) is known as a "bioengineer" of grassland ecosystems on the QTP because its excavation of underground tunnels and moderate feeding on grassland vegetation can promote material cycles and energy flows in the ecosystem. Nevertheless, the current increase in its population density under the influence of human activities has accelerated the degradation of grasslands in this region [20]. Plateau zokors have attracted extensive attention due to their prominent ecological functions and great ecological value. Intriguingly, in Myospalacinae, the Gansu zokor (*E. cansus*), a low-altitude species which originated and is currently distributed on the Loess Plateau and in the northern Qinling Mountains in the northwest of China [18], lives across a wide range of altitudes within its distribution and is sympatric with the plateau zokor in the grassland ecosystem along the eastern edge of the QTP. These two species are sister species and diverged at approximately 3.22 Mya [18, 21, 22]. The dual nature of plateau species and subterranean species, combined with the unique distribution pattern of plateau zokor and Gansu zokor, provides an ideal model for the study of high-altitude and hypoxic adaptation and interspecific gene introgression.

What are the molecular mechanisms by which plateau zokors and high-altitude Gansu zokors have adapted to the plateau environment? Considering their sympatric distributions along the eastern edge of the QTP, has there been any interspecific adaptive introgression from plateau zokors to Gansu zokors, thus facilitating the adaptation of Gansu zokors to the plateau environment? Previous studies have shown that *EPAS*1 (endothelial PAS domain protein 1) is a key gene in the adaptation of Tibetans to the plateau environment [23, 24], and the blunted regulation of NO facilitates vasodilatation and thus promotes oxygen transport, which is also one of the important mechanisms for the adaptation of Tibetans to the low oxygen environment of the plateau [25]; Zhang et al. studied the plateau environment adaptation mechanism of plateau zokor and found that genes were significantly enriched in GO entries related to the cardiovascular system [26]. In addition, a comparative transcriptomic study of the hearts of Tibetan and plains pigs showed that the differentially expressed genes in Tibetan and plains pigs were enriched in the Hif-1 signaling pathway, the VEGF signaling pathway, and the hypoxia-related process [27]. Therefore, we hypothesized that if adaptive introgression existed between plateau zokors and Gansu zokors, adaptive introgressed genes from plateau zokors in Gansu zokors would be significantly enriched in the entries related to the cardiovascular system as well as in pathways related to hypoxia.

In the present study, we used low-coverage whole-genome resequencing (LcWGR) to study the population genomics, high-altitude adaptation mechanism, and interspecific gene introgression of plateau zokors and Gansu zokors based on 19 populations across their distributions in China. The aim of this study was to clarify the adaptive evolutionary patterns of the plateau zokor and the Gansu zokor at the molecular level. The results are of great scientific significance to the understanding of biological adaptation models and the formation and maintenance of mechanisms of biodiversity on the QTP as well as for furthering the integrated control and management of zokors.

Results

Raw data quality control

In this study, 129 individuals were sampled from 13 populations of plateau zokors, and 101 individuals were sampled from 6 populations of Gansu zokors (Fig. 1). LcWGR was performed on 184 individuals from 12 zokor populations sampled in the field. A total of 876.23 Gb of raw sequence reads was obtained, and 865.64 Gb of clean sequence reads remained after filtering, accounting for 98.81% of all bases. On average, 97.83% of clean reads met the Q20 quality threshold, and 93.13% met the Q30 threshold. The average GC content of clean reads was 42.16% (Additional file 1: Table S3). The proportion of high-quality clean reads in all individuals exceeded 99% (Additional file 1: Fig. S1), and the base confidences and sequencing data qualities were very high (Additional file 1: Fig. S2, S3), indicating that this data was suitable for further analysis.

Reference genome alignment

We were able to align 99.39% of clean reads from the LcWGR samples to the 2.57 Gb plateau zokor reference genome, with an average sequencing depth of 1.73×. Other alignment information is detailed in Table S4 (Additional file 1).

Mutation detection and SNP annotation

After combining the data from 47 individuals (including one outgroup) downloaded from public online databases with the data from 184 individuals obtained via LcWGR in this study, a total of 44,735,823 SNPs were identified after quality filtering. After annotation of these SNPs, 324,125 were shown to be missense mutations, accounting for 39.03% of all mutations in protein-coding regions; 5482 were shown to be nonsense mutations, accounting for 0.66% of all coding mutations, and 500,915 were silent mutations, accounting for 60.31%. SNPs were most likely to be located within intronic regions, accounting for 54.12% of all SNPs, followed by intergenic regions with 32.17% of all SNPs. Within gene regions, 5.14% of all SNPs were located in downstream regions, 4.99% were in upstream regions, and 1.79% of all SNPs were located in 3′ untranslated regions (3′ UTRs). Most of the remaining SNPS, accounting for less than 1.5% of all SNPs, were located in exonic regions, 5′ untranslated regions (5′ UTRs), and splice site regions (Additional file 1: Table S5). In addition, the transition/transversion (Ts/Tv) ratio for all SNPs was 2.51 (Additional file 1: Table S6).

Population structure

Phylogenetic analysis indicated that most individuals clustered within their respective populations, which were generally distinct. However, individuals from the two
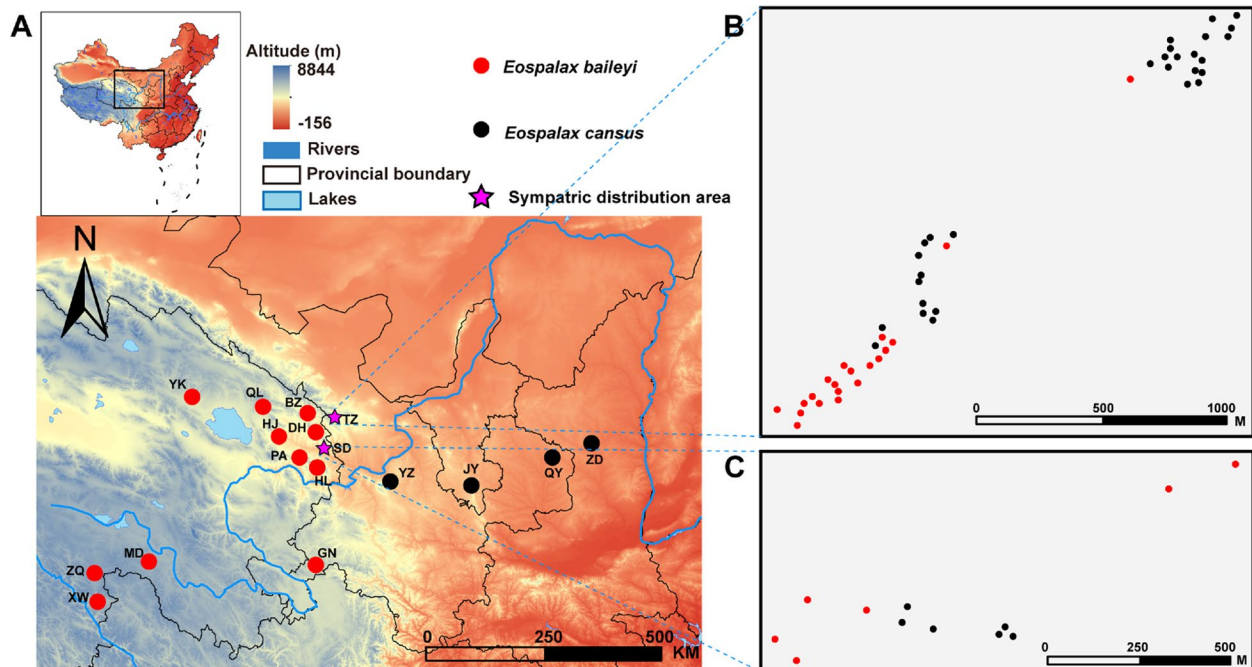

Fig. 1 Sampling sites of plateau zokor and Gansu zokor populations. **A** Geographic distribution of each species and sampling localities. **B** Enlargement of TZ sympatric distribution area. **C** Enlargement of SD sympatric distribution area. Red point: plateau zokor individuals; black point: Gansu zokor individuals

highest altitude populations of plateau zokors, ZQ (plateau zokor, altitude: 4342 m) and XW (plateau zokor, altitude: 4420 m), were mixed together, and a few individuals of the QY (Gansu zokor, altitude: 1300 m) population clustered with the ZD (Gansu zokor, altitude: 1230 m) population. Only three individuals, JY006 (Gansu zokor, altitude: 2000 m), ZD5 (Gansu zokor, altitude: 1230 m), and SDG3 (Gansu zokor, altitude: 2800 m), did not

cluster with their respective populations (Fig. 2A). The two species, plateau zokor and Gansu zokor, were clearly differentiated from each other. With the DH (plateau zokor, altitude: 2900 m) and HL (plateau zokor, altitude: 3000 m) populations as the boundary, the entire upper part of the phylogenetic tree consisted of plateau zokor individuals, and the lower part of the tree consisted of Gansu zokor individuals. Despite potentially serving
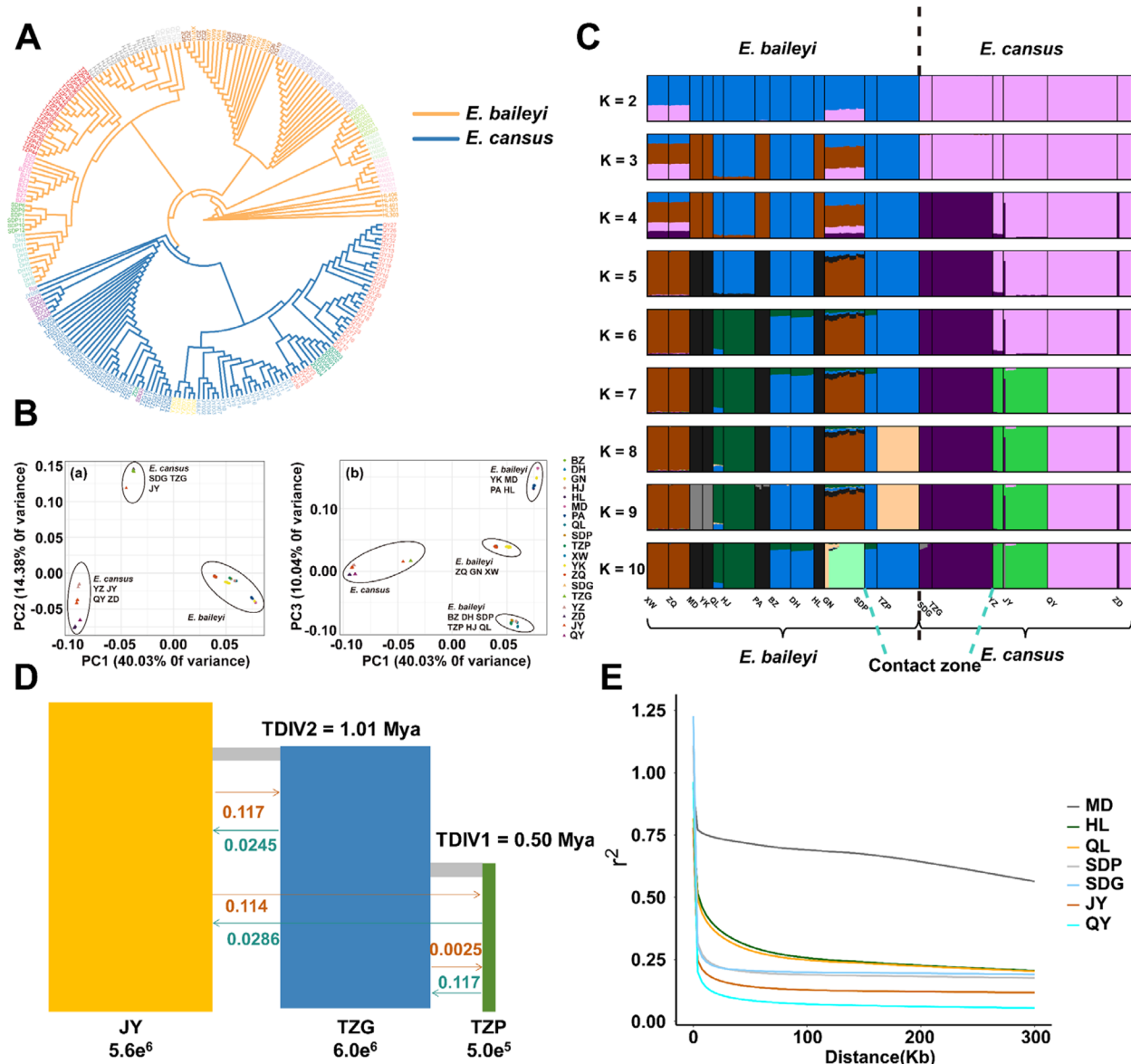
Fig. 2 The phylogenetic analysis, admixture analyses, demographic inferences, and linkage disequilibrium analysis of plateau zokors and Gansu zokors. **A** Phylogenetic tree. **B** Principal component analysis (PCA). (a) PCA with PC1 and PC2; (b) PCA with PC1 and PC3. The circles represent populations of plateau zokors, and the triangles represent populations of Gansu zokors. Each color represents a different population. **C** Population genetic structure. The presence of each color represents new genetic information. **D** Demographic history. Estimated effective population sizes (*Ne*) and divergence times are shown. The numbers next to colored arrows indicate the per generation migration rates between populations. **E** Decay of linkage disequilibrium for different populations of plateau zokors and Gansu zokors. The horizontal axis indicates the distance between two points on the genome, and the vertical axis indicates the linkage unbalance coefficient

Kang *et al. BMC Biology*   (2024) 22:187

Page 5 of 23

as a natural barrier, plateau zokor populations were not strictly divided into two branches separated by the Yellow River. The YK (plateau zokor, altitude: 3700 m) population north of the Yellow River and the MD (plateau zokor, altitude: 4300 m) population south of the Yellow River formed a branch, and the PA (plateau zokor, altitude: 2900 m) and HL (plateau zokor, altitude: 3000 m) populations north of the Yellow River formed a separate branch. Among Gansu zokors, the high-altitude Gansu zokor populations in the sympatric distribution areas, SDG (Gansu zokor, altitude: 2800 m) and TZG (Gansu zokor, altitude: 3000 m), were clearly differentiated from the other low-altitude Gansu zokor populations, YZ (Gansu zokor, altitude: 2200 m), JY (Gansu zokor, altitude: 2000 m), QY (Gansu zokor, altitude: 1300 m), and ZD (Gansu zokor, altitude: 1230 m).

PCA revealed relationships among populations that were very similar to those depicted in the phylogenetic tree (Fig. 2B). PC1 explained 40.3% of the total variance in the dataset. Along PC1, plateau zokors and Gansu zokors were completely distinct, with the Gansu zokor populations on the left and the plateau zokor populations on the right; thus, PC1 mainly reflects interspecific variation. PC2 explained 14.38% of the total variance. Along PC2, the separation between the high-altitude Gansu zokors in the sympatric distribution areas and the rest of the low-altitude Gansu zokors was apparent, which was consistent with the population relationships revealed by the phylogenetic tree, implying genetic differentiation between the low-altitude Gansu zokors and high-altitude Gansu zokors. The plateau zokors did not show separation along PC2. Thus, PC2 mainly reflected altitude differences in Gansu zokors (Fig. 2B). PC3 explained 10.04% of total variance. Along PC3, plateau zokors could be categorized into three groups: a group consisting of the BZ (plateau zokor, altitude: 3000 m), DH (plateau zokor, altitude: 2900 m), SDP (plateau zokor, altitude: 2800 m), TZP (plateau zokor, altitude: 3000 m), HJ (plateau zokor, altitude: 2800 m), and QL (plateau zokor, altitude: 2800 m) populations, a group consisting of the ZQ (plateau zokor, altitude: 4342 m), GN (plateau zokor, altitude: 3600 m), and XW (plateau zokor, altitude: 4420 m) populations, and a group consisting of the YK (plateau zokor, altitude: 3700 m), MD (plateau zokor, altitude: 4300 m), PA (plateau zokor, altitude: 2900 m), and HL (plateau zokor, altitude: 3000 m) populations (Fig. 2B). This was in strong agreement with the population relationships among plateau zokors as reflected in the phylogenetic tree.

Genetic structure analysis revealed the mixing of genetic variation among populations (Fig. 2C). When $K=2$, all populations were clearly divided into two genetic clusters by species, with the plateau zokors on the left and the Gansu zokors on the right. In addition, there were some genetic variants attributed to Gansu zokors mixed into the XW (plateau zokor, altitude: 4420 m), ZQ (plateau zokor, altitude: 4342 m), and GN (plateau zokor, altitude: 3600 m) populations of plateau zokors. When $K=3$, four populations of plateau zokors clustered together (MD, YK, PA, and HL), and the XW (plateau zokor, altitude: 4420 m), ZQ (plateau zokor, altitude: 4342 m), and GN (plateau zokor, altitude: 3600 m) populations still had some genetic variants from the Gansu zokors mixed in. When $K=4$, the high-altitude Gansu zokors in the sympatric distribution areas, SDG (Gansu zokor, altitude: 2800 m) and TZG (Gansu zokor, altitude: 3000 m), clustered distinctly, while genetic variants from high-altitude Gansu zokors were mixed in with the XW (plateau zokor, altitude: 4420 m), ZQ (plateau zokor, altitude: 4342 m), and GN (plateau zokor, altitude: 3600 m) populations of plateau zokors. Thus, when the K value ranged from 2 to 4, the structure analysis indicated mixing of genetic variants between the two species. When the K value ranged from 5 to 10, there was no mixing of genetic components between the plateau zokors and the Gansu zokors. Instead, intraspecific genetic variants appeared to be mixed among the populations of each respective species, accompanied by the distinct clusters for some individual populations. In general, the genetic variants of the following populations were highly similar to each other and remained consistent for all K values: XW (plateau zokor, altitude: 4420 m) and ZQ (plateau zokor, altitude: 4342 m) populations; MD (plateau zokor, altitude: 4300 m) and YK (plateau zokor, altitude: 3700 m) populations; QL (plateau zokor, altitude: 2800 m) and HJ (plateau zokor, altitude: 2800 m) populations; PA (plateau zokor, altitude: 2900 m) and HL (plateau zokor, altitude: 3000 m) populations; BZ (plateau zokor, altitude: 3000 m), DH (plateau zokor, altitude: 2900 m), and SDP (plateau zokor, altitude: 2800 m) populations of plateau zokors; the high-altitude populations of the Gansu zokor, SDG (Gansu zokor, altitude: 2800 m) and TZG (Gansu zokor, altitude: 3000 m) populations; JY (Gansu zokor, altitude: 2000 m) and YZ (Gansu zokor, altitude: 2200 m) populations; and lastly QY (Gansu zokor, altitude: 1300 m) and ZD (Gansu zokor, altitude: 1230 m) populations of Gansu zokors.

Demographic history

The JY (Gansu zokor, altitude: 2000 m) population of Gansu zokor and the TZG (Gansu zokor, altitude: 3000 m) high-altitude population of Gansu zokor diverged into two populations about 1.01 Ma, while the TZP (plateau zokor, altitude: 3000 m) population of plateau zokor diverged from the TZG (Gansu zokor, altitude: 3000 m) population about 0.5 Ma. Gene flow analysis indicated that the level of gene flow from the

TZP (plateau zokor, altitude: 3000 m) population of plateau zokor to the TZG (Gansu zokor, altitude: 3000 m) population of Gansu zokor was higher than the level of gene flow in the opposite direction. The level of gene flow from TZP (plateau zokor, altitude: 3000 m) to TZG (Gansu zokor, altitude: 3000 m) was also higher than that from the TZG (Gansu zokor, altitude: 3000 m) population to the JY (Gansu zokor, altitude: 2000 m) population of Gansu zokor, which was basically equal to that from the JY (Gansu zokor, altitude: 2000 m) to TZG (Gansu zokor, altitude: 3000 m) populations (Fig. 2D).

Analysis of linkage disequilibrium
Linkage disequilibrium in the high-altitude MD (plateau zokor, altitude: 4300 m) population of plateau zokor decayed significantly more slowly than other populations, indicating a high degree of linkage. Thus, this population possessed low population genetic diversity and was likely subject to high selection intensity. The QY (Gansu zokor, altitude: 1300 m) population of Gansu zokor had the fastest decay rate and the highest genetic diversity. The seven populations ranked according to the rate of decay of linkage disequilibrium from slow to fast were as follows: MD (plateau zokor, altitude: 4300 m) < HL (plateau zokor, altitude: 3000 m) < QL (plateau zokor, altitude: 2800 m) < SDP (plateau zokor, altitude: 2800 m) < SDG (Gansu zokor, altitude: 2800 m) < JY (Gansu zokor, altitude: 2000 m) < QY (Gansu zokor, altitude: 1300 m) (Fig. 2E).

Analysis of selection signals in high-altitude populations
The ZQ (plateau zokor, altitude: 4342 m) population of plateau zokor (altitude: 4300 m) and the ZD (Gansu zokor, altitude: 1230 m) population of Gansu zokor (altitude: 1300 m) were examined for evidence of selective sweeps (Additional file 1: Fig. S4A). The purple datapoints in the figure indicate genomic regions under positive selection in the ZQ (plateau zokor, altitude: 4342 m) population. A total of 911 candidate positively selected genes were annotated in these regions. Among all significantly enriched GO terms and KEGG pathways (Additional file 2: Table S7, S8), we identified several significantly enriched GO terms related to hypoxic adaptation, including those related to vascular, red blood cell, heart, and calcium ion transport (Additional file 1: Table S9), and also KEGG pathways related to hypoxic adaptation such as MAPK signaling pathway and Hippo signaling pathway.

Similarly, the high-altitude TZG (Gansu zokor, altitude: 3000 m) population of Gansu zokor and the low-altitude ZD (Gansu zokor, altitude: 1230 m) population of Gansu zokor were examined for evidence of selective sweeps (Additional file 1: Fig. S4B). The purple datapoints in the figure indicate genomic regions under positive selection in the TZG (Gansu zokor, altitude: 3000 m) population. A total of 960 candidate genes were annotated in these regions. Among all significantly enriched GO terms and KEGG pathways (Additional file 2: Table S10, S11), we also identified several significantly enriched terms related to hypoxic adaptation, including those related to hypoxic response, vascularity, erythrocytes, heart, and calcium ion transport (Additional file 1: Table S12), and also KEGG pathways related to hypoxic adaptation such as HIF-1 signaling pathway, PI3K-Akt signaling pathway and MAPK signaling pathway.

Interspecific gene introgression
For populations in the TZ sympatric distribution area, the value of the *D*-statistic was less than 0 and |Z| was greater than 3 regardless of whether P1 was YZ (Gansu zokor, altitude: 2200 m), JY (Gansu zokor, altitude: 2000 m), ZD (Gansu zokor, altitude: 1230 m), or QY (Gansu zokor, altitude: 1300 m), suggesting that there was gene exchange between TZG (Gansu zokor, altitude: 3000 m) and TZP (plateau zokor, altitude: 3000 m). Since the numerator of the formula used in the analysis was BABA-ABBA, the negative *D*-statistics indicated introgression from TZP (P3) to TZG (P2) (Table 1). For populations in the SD sympatric distribution area, the value of the *D*-statistic was less than 0 and |Z| was greater than 3 regardless of whether P1 was YZ (Gansu zokor, altitude: 2200 m), JY (Gansu zokor, altitude: 2000 m), ZD (Gansu zokor, altitude: 1230 m), or QY (Gansu zokor, altitude: 1300 m), suggesting that there was also gene exchange between SDG (Gansu zokor, altitude: 2800 m) and SDP (plateau zokor, altitude: 2800 m) (Additional file 1: Table S13). As observed for the TZ populations,

Table 1 Results of *D*-statistic tests between TZG and TZP populations

P1	P2	P3	Outgroup	*D*-stat	Z	BABA	ABBA
YZ	TZG	TZP	Rpr	-0.1768	-20.869	6167	8816
JY	TZG	TZP	Rpr	-0.168	-20.905	6312	8862
ZD	TZG	TZP	Rpr	-0.2294	-29.775	5728	9139
QY	TZG	TZP	Rpr	-0.1988	-26.684	6779	10,144

Kang *et al. BMC Biology*   (2024) 22:187

Page 7 of 23

Table 2 Results of f_d-statistic (ZD-TZG-TZP-Rpr)

Scaffold	Start	End	Mid	Sites	Sitesused	fd	fdM	Z_score_fd	P_value_fd	Z_score_fdM	P_value_fdM
GWHABJZ00000000030	9,850,001	9,900,000	9,877,987	1284	324	0.2647	0.1827	7.922925	<0.001	7.409773	<0.001
GWHABJZ00000000001	9,050,001	9,100,000	9,077,712	1248	122	0.2325	0.2033	7.23057915	<0.001	7.243370307	<0.001
GWHABJZ00000000263	1,600,001	1,650,000	1,623,773	1512	140	0.2287	0.0538	3.81106021	<0.001	2.10596789	<0.001
GWHABJZ00000000014	15,600,001	15,650,000	15,624,427	939	113	0.1863	0.1445	8.494402	<0.001	6.570579	<0.001
GWHABJZ00000000151	50,001	100,000	78,604	4079	830	0.1861	0.0616	10.54666	<0.001	5.968273	<0.001
GWHABJZ00000000168	4,400,001	4,450,000	4,425,838	1102	340	0.1796	0.1031	3.45407170	<0.001	5.383337537	<0.001
GWHABJZ00000000014	17,150,001	17,200,000	17,175,349	1041	112	0.1792	0.0965	4.41778948	<0.001	6.149553981	<0.001
GWHABJZ00000000192	2,400,001	2,450,000	2,423,362	1285	170	0.1541	0.0962	3.406362	<0.001	3.086847	<0.001
GWHABJZ00000000078	8,250,001	8,300,000	8,268,400	947	110	0.1511	0.066	5.26436596	<0.001	4.898983084	<0.001
GWHABJZ00000000271	950,001	1,000,000	976,562	995	107	0.1511	0.1027	11.6102	<0.001	8.546056	<0.001

Start indicates the starting physical location of scaffold in the genome, and End indicates the physical location of the end of scaffold in the genome. $0 < f_d < 1$, Z-score-$f_d > 3$, and P-value-$f_d < 0.05$ are used as thresholds to indicate significant introgression signals

Kang *et al. BMC Biology*   (2024) 22:187

Page 8 of 23

the negative *D*-statistics for the SD populations indicated introgression from SDP (P3) to SDG (P2).

Proportion of genes that underwent introgression and screening of introgressed genes based on the f_d-statistic

To identify genomic regions that underwent introgression, we calculated the f_d-statistic for genomic regions with positive selection signals, as identified earlier in this study. Regions with $0 < f_d < 1$, *Z*-score-$f_d > 3$, and *P*-value-$f_d < 0.05$ were categorized as introgressed regions. For the TZ sympatric distribution area, we screened 6533, 6462, 5848, and 5761 genomic windows, which were the positively selected regions when P1 was ZD (Gansu zokor, altitude: 1230 m), YZ (Gansu zokor, altitude: 2200 m), QY (Gansu zokor, altitude: 1300 m), and JY (Gansu zokor, altitude: 2000 m), respectively. The size of each window was 50 kb, and the size of the plateau zokor reference genome was 2,569,336,452 bp. The proportions of the genome that were introgressed from TZP (plateau zokor, altitude: 3000 m) to TZG (Gansu zokor, altitude: 3000 m) were estimated to be 0.29%, 0.29%, 0.29%, and 0.27% for the four aforementioned population combinations, with an average proportion of 0.285%.

Due to the large number of windows with significant introgression signals, only the results for the windows with strong signals ($f_d > 0.15$) (Table 2, S14-S16) and the regions (Fig. 4A–D) with strong signals are shown here. We then identified 54 genes located in these putatively introgressed regions with $f_d > 0.15$. The top 30 enriched GO terms for these genes are reported in Fig. S5A (Additional file 1). The top BP (Biological process) terms were cell adhesion, regulation of cell proliferation, miRNA catabolism, glycolysis, and cell differentiation. The top MF (molecular function) terms were calcium binding, enzyme binding, and enzyme activity, and the top CC (Cellular component) terms were organelle membranes, endoplasmic reticulum membranes, plasma membranes, and muscle membranes.

The top 20 enriched KEGG pathways for these introgressed genes are shown in Fig. S5B (Additional file 1). The top Environmental Information Processing pathways were the VEGF signaling pathway and the PI3K-Akt signaling pathway. The top Genetic Information Processing pathway was the nucleotide excision repair pathway. Chemical carcinogenesis and viral infections pathways were the top Human Diseases pathways. The top Metabolism pathways were related to a variety of metabolic processes and interconversion of pentose and glucuronide, and the top Organismal Systems pathways were related to growth hormone synthesis and secretion and olfactory transduction.

We similarly screened genomic windows under positive selection for evidence of introgression using the f_d-statistic in populations in the SD sympatric distribution area, ultimately screening 5712, 5725, 5265, and 5039 windows, which were the positively-selected regions when P1 was ZD (Gansu zokor, altitude: 1230 m), YZ (Gansu zokor, altitude: 2200 m), QY (Gansu zokor, altitude: 1300 m), and JY (Gansu zokor, altitude: 2000 m), respectively. The proportions of the genome that were introgressed from SDP (plateau zokor, altitude: 2800 m) to SDG (Gansu zokor, altitude: 2800 m) were estimated to be 0.33%, 0.32%, 0.34%, and 0.29% for the four aforementioned population combinations, with an average proportion of 0.32%.

Due to the large number of windows with significant introgression signals, only the results for the windows with strong signals ($f_d > 0.15$) (Additional file 1: Table S17-S20) and the regions (Fig. 3E–H) with strong signals are shown here. We then identified 64 genes located in these putatively introgressed regions with $f_d > 0.15$. The top 30 enriched GO terms for these genes, which were highly similar to the top GO terms for introgressed genes in the TZG (Gansu zokor, altitude: 3000 m) population, are reported in Fig. S6A (Additional file 1). The top BP terms were cell adhesion, regulation of cell proliferation, miRNA catabolism, glycolysis, and cell differentiation. The top MF terms were calcium binding, enzyme binding, and enzyme activity, and the top CC terms were presynaptic endocytosis zone membrane, plasma membrane, endoplasmic reticulum membrane, and ruffle membrane.

The top 20 enriched KEGG pathways for these introgressed genes are shown in Fig. S6B (Additional file 1). The top Cellular Processes pathway was the regulation of actin cytoskeleton pathway, and the top Environmental Information Processing pathway was the VEGF signaling pathway. The nucleotide excision repair pathway was the top Genetic Information Processing pathway, and the most significantly enriched Human Diseases pathways were associated with chemical carcinogenesis and viral infection. The top Metabolism pathways were associated with a variety of metabolic processes and steroid hormone biosynthesis. Lastly, the top Organismal Systems pathways were related to growth hormone synthesis and secretion, cortisol synthesis and secretion, thyroid hormone synthesis, and olfactory transduction.

Analysis of adaptive introgression

In the sympatric distribution area TZ, we searched for selective sweeps using the TZP (plateau zokor, altitude: 3000 m)-ZD (Gansu zokor, altitude: 1230 m) species pair and identified 1146 positively selected genes in the TZP (plateau zokor, altitude: 3000 m) population (Fig. 4A).

Kang *et al. BMC Biology*   (2024) 22:187

Page 9 of 23

Similarly, we searched for selective sweeps using the TZG (Gansu zokor, altitude: 3000 m)-ZD (Gansu zokor, altitude: 1230 m) population pair and identified 892 positively selected genes in the TZG (Gansu zokor, altitude: 3000 m) population (Fig. 4B). We then identified 159 genes that were under positive selection in both the TZP (plateau zokor, altitude: 3000 m) and TZG (Gansu zokor, altitude: 3000 m) populations. Of these shared positively selected genes, 99 had positive and significant f_d-statistics, indicating that they likely had undergone introgression as well. Thus, these 99 genes could be characterized as adaptive genes in the TZG (Gansu zokor, altitude: 3000 m) population that were introgressed from the TZP (plateau zokor, altitude: 3000 m) population (Additional file 1: Table S21).

Among all significantly enriched GO terms and KEGG pathways (Additional file 2: Table S22-S23), we identified several significantly enriched terms related to hypoxic adaptation, including some related to the cardiovascular system and response to hypoxia (Table 3), which further suggested a role for some introgressed genes in the high-altitude adaptation process in Gansu zokors.

In the sympatric distribution area SD, we searched for selective sweeps using the SDP (plateau zokor, altitude: 2800 m)-ZD (Gansu zokor, altitude: 1230 m) species pair and identified 629 positively selected genes in the SDP (plateau zokor, altitude: 2800 m) population (Additional file 1: Fig. S7A). Similarly, we searched for selective sweeps using the SDG (Gansu zokor, altitude: 2800 m)-ZD (Gansu zokor, altitude: 1230 m) population pair and identified 488 positively selected genes in the SDG (Gansu zokor, altitude: 2800 m) population (Additional file 1: Fig. S7B). We identified 50 of these genes as under positive selection in both the SDP (plateau zokor, altitude: 2800 m) and SDG (Gansu zokor, altitude: 2800 m) populations. Of these shared positively selected genes, 32 had positive and significant f_d-statistics, indicating that they likely had undergone introgression. Thus, these 32 genes could be characterized as adaptive genes in the SDG (Gansu zokor, altitude: 2800 m) population that were introgressed from the SDP (plateau zokor, altitude: 2800 m) population (Additional file 1: Table S24).

Among all significantly enriched GO terms and KEGG pathways (Additional file 2: Table S25-S26), we also found four significantly enriched terms related to hypoxic adaptation, including those related to the cardiovascular system and calcium ion transport (Additional file 1: Table S27), and KEGG pathway related to hypoxic adaptation such as PI3K-Akt and MAPK signaling pathway. These results suggested a role for some introgressed genes in the process of high-altitude adaptation in Gansu zokors.

Of the adaptive genes introgressed from population TZP (plateau zokor, altitude: 3000 m) into TZG (Gansu zokor, altitude: 3000 m), we visualized haplotype sharing for *Uvrag* (UV radiation resistance-associated gene), *Hpse2* (heparinase 2), *Slit2* (slit guidance ligand 2), *Ece1* (endothelin-converting enzyme 1), and *Nrp1* (neuropilin 1) (Fig. 5). From the degree of haplotype sharing in pairwise comparisons between TZP (plateau zokor, altitude: 3000 m) and TZG (Gansu zokor, altitude: 3000 m), we found that haplotype sharing was much more extensive from the plateau zokors (TZP) than from the Gansu zokors (TZG). Similarly, we visualized haplotype sharing for the genes *Phex* (phosphate regulating endopeptidase X-linked), *Ireb2* (iron-responsive element-binding protein 2), *Cask* (calcium/calmodulin-dependent serine protein kinase), and *Gpc3* (glypican 3), all of which were adaptive genes introgressed from SDP (plateau zokor, altitude: 2800 m) into SDG (Gansu zokor, altitude: 2800 m) (Additional file 1: Fig. S8). We also found that haplotype sharing was more extensive from the plateau zokors (SDP) than from the Gansu zokors (SDG) for these genes.

Discussion

Low-coverage whole-genome resequencing

Routine in-depth whole genome resequencing and reduced-representation genome sequencing are now widely used in various studies of plants and animals. LcWGR is more cost-effective than in-depth whole-genome resequencing, and low-depth sequencing with multiple samples is more accurate in estimating many population parameters than high-depth sequencing

(See figure on next page.)
Fig. 3 Some regions with strong introgression signal. **A** JY-TZG-TZP-Rpr. The region shown in the figure contains the genes *Pcdhb* (protocadherin beta) *13*, *Pcdhb20*, *Pcdhb21*, and *Pcdhb22*. **B** QY-TZG-TZP-Rpr. The region shown in the figure contains the genes *Pcdhb13*, *Pcdhb20*, *Pcdhb21*, and *Pcdhb22*. **C** YZ-TZG-TZP-Rpr. The region shown in the figure contains the gene *Ly6m* (lymphocyte antigen 6 family member M). **D** ZD-TZG-TZP-Rpr. The region shown in the figure contains the genes *Pcdhb13*, *Pcdhb20*, *Pcdhb21*, and *Pcdhb22*. **E** JY-SDG-SDP-Rpr. The region shown in the figure contains the genes *Pcdhb13*, *Pcdhb20*, *Pcdhb21*, and *Pcdhb22*. **F** QY-SDG-SDP-Rpr. The region shown in the figure contains the genes *Pcdhb13*, *Pcdhb20*, *Pcdhb21*, and *Pcdhb22*. **G** YZ-SDG-SDP-Rpr. The region shown in the figure contains the genes *Zbtb7a* (zinc finger and BTB domain containing 7a) and *Creb3l3* (cAMP responsive element binding protein 3 like 3). **H** ZD-SDG-SDP-Rpr. The region shown in the figure contains the genes *Pcdhb13*, *Pcdhb20*, *Pcdhb21*, and *Pcdhb22*. Count indicates the number of ABBA or BABA patterns; Position indicates the physical location in the genome
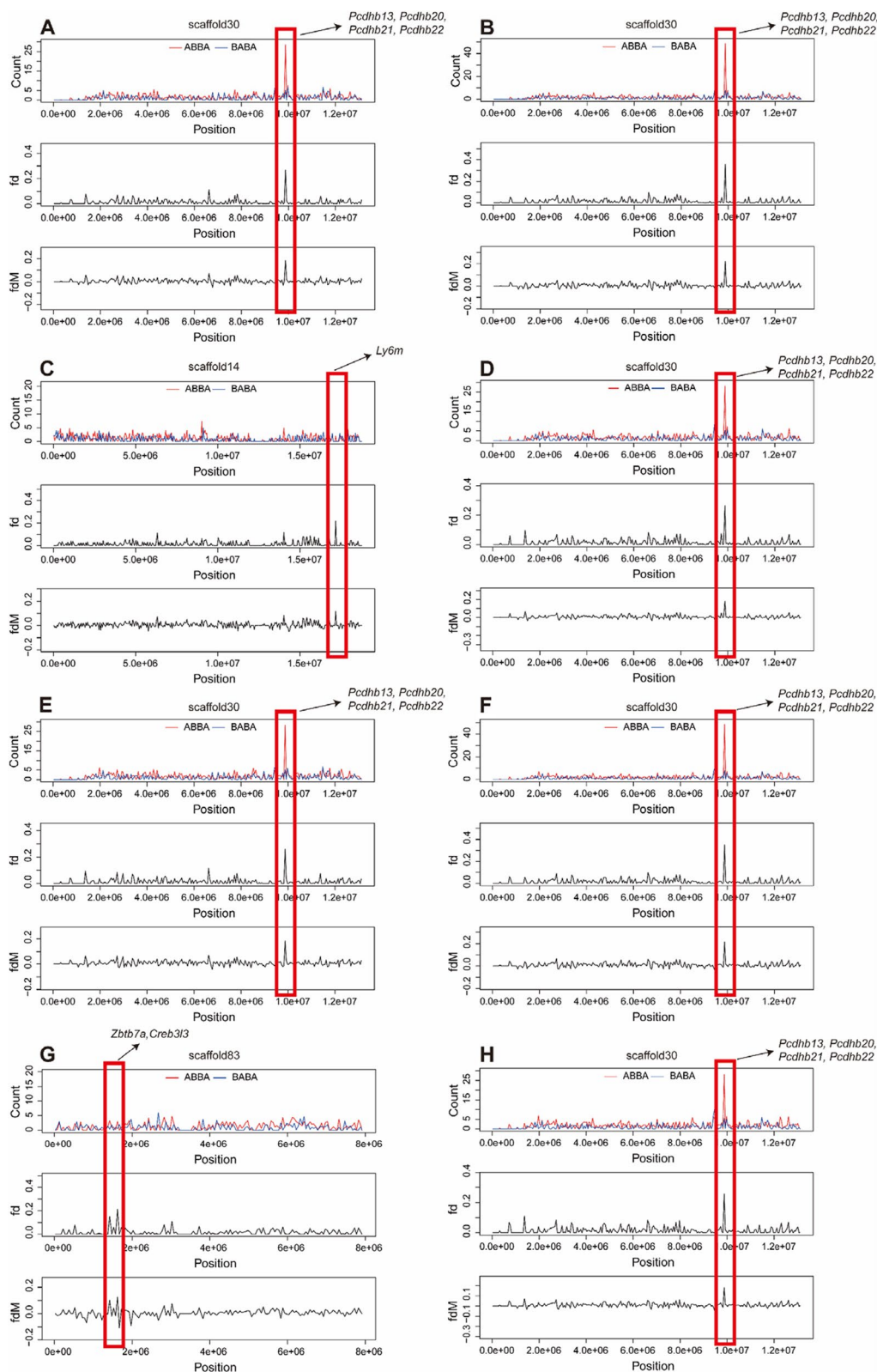
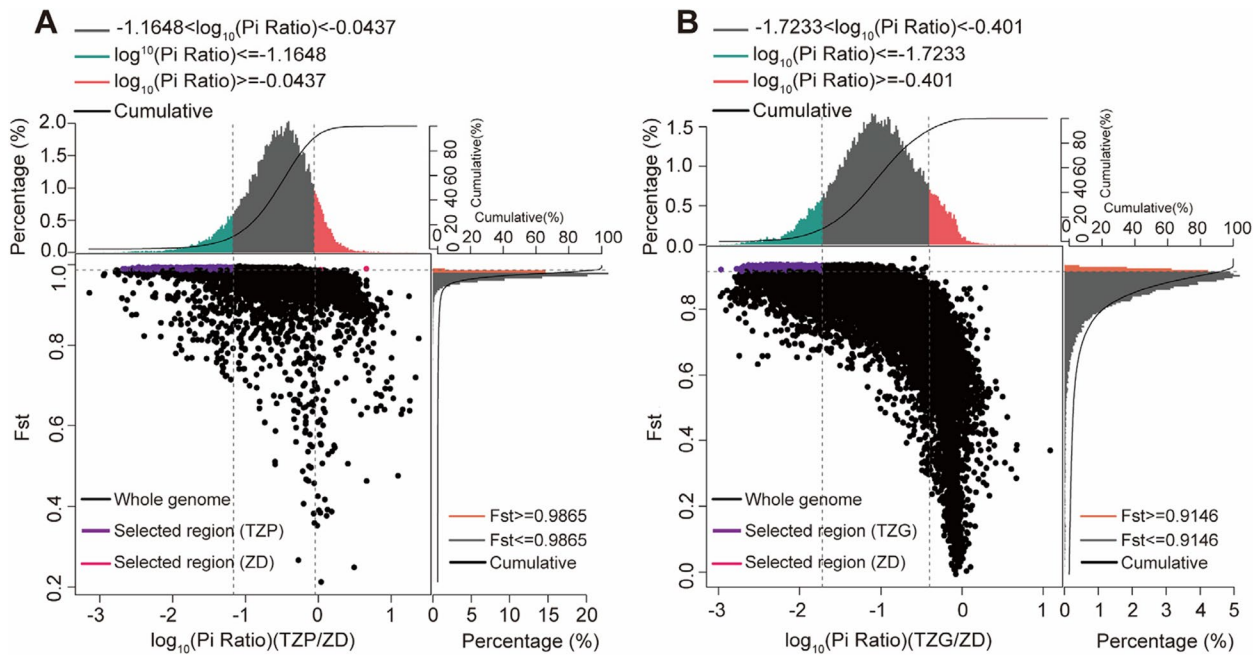**Fig. 3** (See legend on previous page.)

Fig. 4 Genome-wide selection signals. Selection signals for **A** populations TZP and ZD and **B** populations TZG and ZD. The green and red areas at the top of the figures indicate the 5% of genomic regions with lowest and highest *Pi* ratios, respectively, and the orange areas on the right of the figures indicate the 10% of genomic regions with highest *Fst*. The red and purple areas in the center of the figures indicate the intersections of these top *Fst* and *Pi* regions, which constitute candidate loci under positive selection, with the purple areas indicating regions under selection in **A** population TZP and **B** population TZG

Table 3 Significantly enriched GO terms related to cardiovascular system and response to hypoxia for adaptive introgressed genes in the TZG population

Term ID	Term description	*P* value	Genes
GO:0060982	Coronary artery morphogenesis	< 0.001	*Nrp1* (neuropilin 1), *Arid2* (AT-rich interaction domain 2)
GO:0010613	Positive regulation of cardiac muscle hypertrophy	< 0.001	*Ankrd55* (ankyrin repeat domain 55), *Ece1* (endothelin-converting enzyme 1)
GO:0048008	Platelet-derived growth factor receptor signaling pathway	0.01	*Fer* (FER tyrosine kinase), *Nrp1*
GO:0008217	Regulation of blood pressure	0.02	*Corin* (corin, serine peptidase), *Ece1*
GO:1,904,753	Negative regulation of vascular associated smooth muscle cell migration	0.03	*Prkg1* (protein kinase cGMP-dependent 1)
GO:0005021	Vascular endothelial growth factor-activated receptor activity	0.03	*Nrp1*
GO:0035912	Dorsal aorta morphogenesis	0.03	*Rbpj* (recombination signal binding protein for immunoglobulin kappa J region)
GO:0097746	Regulation of blood vessel diameter	0.03	*Ece1*
GO:0060087	Relaxation of vascular smooth muscle	0.04	*Prkg1*
GO:0045822	Negative regulation of heart contraction	0.04	*Prkg1*
GO:0003198	Epithelial to mesenchymal transition involved in endocardial cushion formation	0.05	*Rbpj*
GO:0003214	Cardiac left ventricle morphogenesis	0.05	*Rbpj*
GO:0001666	Response to hypoxia	0.05	*Nox4* (NADPH oxidase 4), *Ece1*, *Pam* (peptidylglycine alpha-amidating monooxygenase)

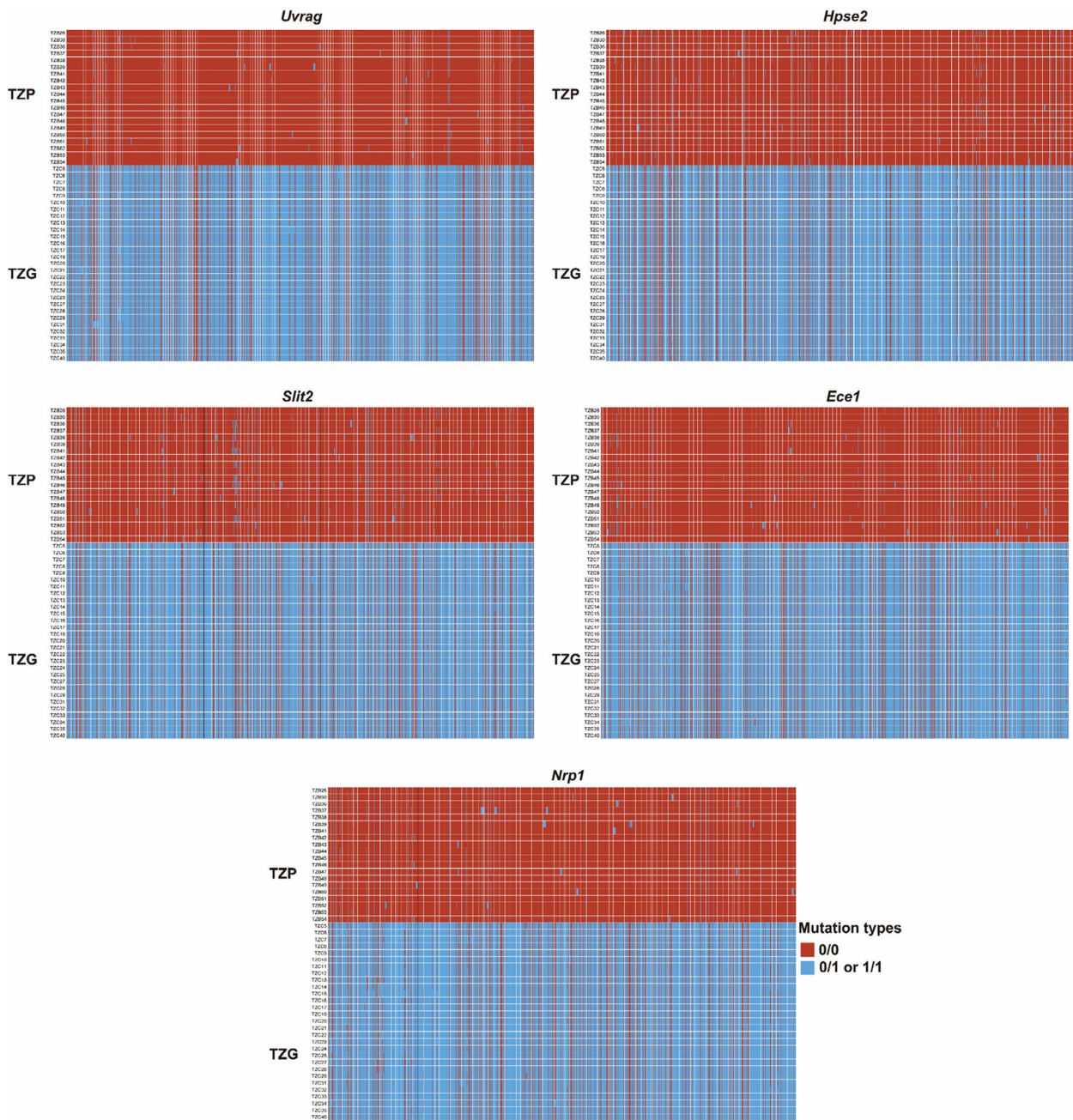*P* ≤ 0.05 was considered statistically significant

Fig. 5 The degree of haplotype sharing in *Uvrag* (UV radiation resistance-associated gene), *Hpse2* (heparinase 2), *Slit2* (slit guidance ligand 2), *Ece1* (endothelin-converting enzyme 1), and *Nrp1* (neuropilin 1) in pairwise comparisons between populations TZP and TZG. The major allele in TZP is indicated in red, and the major allele in TZG is indicated in blue (0/0 means homozygote, 1/1 indicates homozygous mutation, and 0/1 indicates heterozygous mutation)

with fewer samples [28, 29]. Furthermore, reduced-representation genome sequencing only captures 1–10% of the genome, thereby excluding most markers, while low-depth sequencing of multiple samples overcomes this limitation. Therefore, the large-scale application of LcWGR is promising. In this study, 184 samples were sequenced using LcWGR, generating a large amount of data. Combined with previously published high-depth sequencing data of 46 individuals, we were able to identify 44,735,823 SNPs genome-wide, a large number that facilitated downstream population analyses. High percentages of clean reads met Q20 and Q30 quality

thresholds, and GC distributions were within expected ranges without any obvious biases, indicating that library construction and sequence quality were both excellent. Clean sequence reads from individuals of both plateau zokors and Gansu zokors were aligned to the plateau zokor reference genome, with an average alignment rate of 99.39%. These high alignment rates reflect the high quality of the reference genome assembly, the close proximity between the sequenced individuals and the reference genome species, and the high quality of the sequence reads generated in this study. Therefore, the use of LcWGR in this study was effective, and the quality of this sequencing data was more than sufficient for the subsequent population evolution analyses. This study is the largest whole-genome resequencing effort for plateau zokors and Gansu zokors to date. The large amount of sequence data obtained here provides the basis for future research on the evolutionary biology of the two species, greatly expands the catalogue of known genetic variation for plateau zokors and the Gansu zokors, and provides important data support for future research on the molecular mechanisms of rodent outbreaks in the alpine meadow ecosystems on the Qinghai-Tibet Plateau.

Population evolution characteristics of plateau zokors and Gansu zokors

Using genome-wide data, our results revealed robust and reliable phylogenetic relationships among multiple populations of plateau zokors and Gansu zokors. The two species were completely separated within the phylogenetic tree, with not even a single individual clustering with the other species, thus supporting their independent species status. Based on genome-wide population genetics, previous studies have shown that plateau zokors form two clades, one north of the Yellow River and the other south of the Yellow River, and have concluded that the two clades should be considered as separate species [18]. From the phylogenetic tree in this study, the populations of plateau zokors can also be roughly divided into two clades, north and south of the Yellow River. However, unlike the results of the aforementioned study, in our phylogenetic analysis, the YK (plateau zokor, altitude: 3700 m) population located north of the Yellow River and the MD (plateau zokor, altitude: 4300 m) population located south of the Yellow River clustered together, indicating a relatively close phylogenetic relationship. This was consistent with the results of Zhang et al. [26]. We hypothesize two possibilities for the close phylogenetic relationship between these two populations: [1] the YK (plateau zokor, altitude: 3700 m) population north of the Yellow River was established first during the east-to-west expansion of *Eospalax* species, and then some individuals from the YK (plateau zokor, altitude: 3700 m) population

migrated south to form the MD (plateau zokor, altitude: 4300 m) population, or [2] the MD (plateau zokor, altitude: 4300 m) population south of the Yellow River was established first, and then some individuals from the MD (plateau zokor, altitude: 4300 m) population migrated north to form the YK (plateau zokor, altitude: 3700 m) population. To distinguish between these two possibilities would require a comparison of the divergence times of each of the two populations from the ancestral population. Calculation of divergence times using whole genome data typically requires high sequencing depths. Unfortunately, the low-depth whole genome resequencing technology used in this study precludes this type of analysis, and thus analyses of population divergence and historical dynamics were not performed for all populations. Regardless of which hypothesis is accurate, the impact of river and water system formation on the differentiation of subterranean species populations also needs to be taken into account. The YK (plateau zokor, altitude: 3700 m) population and MD (plateau zokor, altitude: 4300 m) population were located north and south of the Yellow River, respectively. The Yellow River system began to form slowly 1.25 Ma [30], and the YK (plateau zokor, altitude: 3700 m) and MD (plateau zokor, altitude: 4300 m) populations diverged about 8.0–9.4 ka [26]. The extent of formation of the Yellow River in the vicinity of the two populations at the time of their divergence is relevant.

The relationships among populations as depicted in the phylogenetic tree were also validated using PCA, which also showed the obvious interspecific differentiation between plateau zokors and Gansu zokors. Differentiation between populations based on altitude was also apparent using PCA. Gansu zokors clustered into high-altitude and low-altitude groups, indicating significant genetic differences between the high-altitude and low-altitude populations of the Gansu zokor. Population genetic structure analysis revealed moderate intraspecific mixture of genetic variation among plateau zokors. It is worth noting that in the sympatric distribution areas, there was no evidence of genetic mixture between the TZP (plateau zokor, altitude: 3000 m) population of plateau zokor and the TZG (Gansu zokor, altitude: 3000 m) population of Gansu zokor or between the SDP (plateau zokor, altitude: 2800 m) population of plateau zokor and the SDG (Gansu zokor, altitude: 2800 m) population of Gansu zokor. Thus, we did not find any evidence of genetic exchanges between the two species in the sympatric distribution areas from the population structure analysis. However, the analysis of demographic history in the sympatric distribution area using fastsimcoal2 indicated that the highest level of gene flow occurred from population TZP (plateau zokor, altitude: 3000 m)

Kang *et al. BMC Biology*   (2024) 22:187

Page 14 of 23

to population TZG (Gansu zokor, altitude: 3000 m). This was higher than that from population TZP (plateau zokor, altitude: 3000 m) to population JY (Gansu zokor, altitude: 2000 m) of Gansu zokor and was basically of the same magnitude as that from the population JY (Gansu zokor, altitude: 2000 m) of Gansu zokor to the TZG (Gansu zokor, altitude: 3000 m) population. Thus, gene introgression from plateau zokors to Gansu zokors likely occurred in the sympatric distribution areas.

Analysis of selection signals in high-altitude populations
Analysis of selection signals in plateau zokors
The linkage disequilibrium analysis showed that populations at higher altitudes were subjected to stronger selection and exhibited significantly slower rates of decay. For a complete understanding of the mechanisms of high-altitude adaptation in plateau mammals, it is helpful to understand how these animals adapt to low oxygen environments. We first identified positively selected genes in plateau zokors by looking for evidence of selective sweeps using a population from each species, plateau zokor population ZQ (plateau zokor, altitude: 4342 m) and Gansu zokor population ZD (Gansu zokor, altitude: 1230 m). Positively selected genes in plateau zokors were mainly enriched for GO terms related to energy transport and enzyme activity, suggesting that plateau zokors improved their energy metabolism during the process of adaptation to the plateau environment, reflecting the energy demands of species surviving in these harsh environments. Cardiovascular and erythrocyte-associated phenotypic changes can be viewed as typical adaptations to hypoxia, as they are directly related to oxygen uptake and transport [26]. Zhang et al. showed that the ratio of heart mass to body mass, hemoglobin content, red blood cell count, and hematocrit were significantly higher in a high-altitude population of plateau zokors as compared to a low-altitude population. Using the Mouse Genome Informatics (MGI) database, positively selected genes in high-altitude populations were significantly enriched for terms related to physiological and quantitative abnormalities of erythrocytes. Using the GO database, these positively selected genes were significantly enriched for GO terms related to cardiac development, cardiac morphogenesis, angiogenesis, and vascular morphogenesis [26]. Our results were similar to those from Zhang et al. [26]. Positively selected genes in the ZQ (plateau zokor, altitude: 4342 m) population of plateau zokor, located in Yushu, Qinghai, at an altitude of nearly 4400 m, were significantly enriched for cardiovascular terms like vascular-associated smooth muscle cell migration, ventricular compact myocardium morphogenesis, cardiac muscle cell proliferation, coronary artery morphogenesis, and heart contraction. The extensive selection of these cardiovascular-related genes suggested a vital role for the cardiovascular system of the plateau zokor in its adaptation to high-altitude environments.

In addition, positively selected genes were enriched for calcium ion transport-related terms, similar to some previous studies [31–34]. These terms included voltage-gated calcium channel complex, voltage-gated calcium channel activity, regulation of calcium ion transmembrane transport via high voltage-gated calcium channel, and calcium ion transport. When the body is hypoxic, calcium transport and calcium signaling pathways are involved in the stimulation of hypoxia-inducible factor α (*HIF-α*) transcription [35], and these pathways also play a very essential role in cardiac development and cardiac function [36, 37]. Positively-selected genes were enriched for several KEGG pathways related to cell growth, sugar synthesis, metabolism, and, notably, both the classical MAPK and Hippo signaling pathways. MAPK signaling can promote the expression of *HIF-α*, which in turn indirectly regulates the HIF-1 signaling pathway [38]. At the molecular level, cells of higher mammals mainly sense oxygen through *HIF-α*, and through the HIF-1 signaling pathway, a series of responses can be induced when the organism is in a hypoxic state, thus regulating the efficiency of oxygen utilization [39]. The Hippo signaling pathway is a key pathway that controls organ size, such as heart size [26]. The loss of *STK*3 (serine/threonine kinase 3), a central regulator within the Hippo pathway, promotes the proliferation of cardiomyocytes during heart development, leading to cardiomegaly in mice [40]. By interacting with the Wnt signaling pathway, the Hippo signaling pathway can control cardiac size via β-linker proteins [40].

Analysis of selection signals in high-altitude populations of Gansu zokors
Similar to the enrichment results for positively selected genes in plateau zokors, we found that positively selected genes in high-altitude Gansu zokors were significantly enriched for cardiovascular and erythrocyte-related functional terms, such as erythrocyte development, vascular remodeling, negative regulation of blood pressure, branching involved in vascular morphogenesis, and cardiac valve morphogenesis, as well as calcium transport-related terms, suggesting that the cardiovascular system played an important role in the process of adaptation to the plateau environment in this species as well. In addition, positively selected genes in the high-altitude Gansu zokor population, including *HIF-α*, *Ece1*, and *Nos1* (nitric oxide synthase 1), were also significantly enriched for the response to hypoxia GO term. *HIF1* is a heterodimer formed by α subunits (*HIF-α*) and β subunits [41] and is responsible for mediating hypoxia signal transduction, in

which the expression of α subunits is directly regulated by oxygen levels. *Ece1*, a rate-limiting enzyme, plays a key role in endothelin-1 (ET-1) biosynthesis [42], and Li et al. found strong signals of selective sweeps in *Ece1* in relation to hypoxia in Tibetan pigs [43]. Hypoxia can promote the activation of *NOS* [44], and acute hypoxia has been found to cause changes in the expression of *NOS* in rats [45]. Of the KEGG pathways, the positively selected genes in the high-altitude Gansu zokor populations were significantly enriched for classical pathways related to response to hypoxia, such as the HIF-1, MAPK, and PI3K-Akt signaling pathways. PI3K is a phosphatidylinositol 3-kinase. Under hypoxic conditions, PI3K can be activated and phosphorylates Akt upon binding to it, which enhances *HIF-α* activity and increases cell proliferation while reducing cell apoptosis [46]. In summary, genes related to energy metabolism, cardiovascular system development, calcium ion transport, and response to hypoxia were under positive selection in the high-altitude Gansu zokor population, which likely facilitated the adaptation of Gansu zokors to the plateau environment.

Interspecific gene introgression of Myospalacinae
Gene introgression from other species is an important source of new genetic variation that allows species to adapt to new environments. The adaptive introgression resulting from hybridization and backcrossing with another species has the potential to make the recipient populations adapt to their environments at a much faster rate than if they relied solely on de novo mutations [47–49]. Introgression is more likely to occur between closely related species, as incomplete reproductive barriers are easier to overcome, especially if the introgressed allele is already beneficial to the donor population [3, 50]. In the present study, we identified the genomic locations of introgressed genes and then characterized their types and functions. To investigate whether introgressed genes played a role in adaptation of Gansu zokors to the plateau environment, we combined analyses of positive selection analysis with detection of introgression signals to ensure both that introgression signals were significant and that the introgressed alleles were beneficial to both the donor and recipient populations.

Liu et al. [18] studied the phylogeny and speciation of Myospalacinae using whole genome sequencing and identified an ancient introgression between *E. fontanierii* and *Myospalax*, thus providing an explanation for the mixing of *E. fontanierii* genetic variation with that of *M. psilurus* and *M. aspalax*. Based on the *D*-statistic, gene introgression likely occurred between *E. smithii* and all species in Myospalacinae except the branch of plateau zokors north of the Yellow River. Gene introgression was inferred between *E. rothschildi* and other Myospalacinae species except for Gansu zokors. The plateau zokor branch north of the Yellow River only experienced gene introgression with *E. rothschildi* and *M. psilurus*. This study revealed extensive migration of Myospalacinae species along the Qinling-Huaihe Line and also hinted at allopatric speciation in the plateau zokor branch north of the Yellow River. Zhang et al. [51] investigated the interspecific phylogenetic relationships of the genus *Eospalax* using whole genome resequencing. In the phylogeny constructed using mitochondrial genomes, an individual Gansu zokor from a region adjacent to the plateau zokor distribution area clustered with the plateau zokors, which implied possible post-speciation gene flow between plateau zokors and Gansu zokors. Our study is the first to elaborate on the issues related to gene introgression and adaptive introgression in Myospalacinae. The plateau zokors and the Gansu zokors are sympatric at the eastern edge of the QTP, and the unique ecological environment of the QTP as well as the essential roles of the two species in grassland ecosystems of the region provide an excellent model to study the issues related to adaptive evolution. The results of this study also provide vital information about the molecular mechanisms of rodent outbreaks in the alpine grassland ecosystems of the QTP.

Adaptive introgression from plateau zokors to Gansu zokors
In the *D*-statistic analysis, the occurrence of gene introgression between plateau zokors and Gansu zokors in the two sympatric distribution areas was determined by using each of the four populations, ZD (Gansu zokor, altitude: 1230 m), YZ (Gansu zokor, altitude: 2200 m), QY (Gansu zokor, altitude: 1300 m), and JY (Gansu zokor, altitude: 2000 m) as P1. Based on the f_d-statistic, the levels of gene introgression from plateau zokors to Gansu zokors were estimated to be low, which may explain the relative lack of genetic mixture between these two species in the sympatric distribution areas according to the genetic structure analysis. In the sympatric distribution area TZ, based on the f_d-statistic, we identified 99 adaptive genes that were introgressed into the TZG (Gansu zokor, altitude: 3000 m) population of Gansu zokor from the TZP (plateau zokor, altitude: 3000 m) population of plateau zokor. Among these 99 adaptive introgressed genes, several genes are related to cardiovascular development and hypoxic adaptation, such as *Uvrag*, *Hpse2*, and *Rbpj* (recombination signal binding protein for immunoglobulin kappa J region); *Slit2*, *Ece1*, *Nrp1*, and *Tspan4* (tetraspanin 4); and *Sik2* (salt inducible kinase 2). *Uvrag*, the ultraviolet resistance-related gene, is an autophagy gene with multiple biological functions [52, 53]. Mice lacking *Uvrag* developed cardiomyopathy accompanied by decreased cardiac function [54], and the deletion of *Uvrag* further promoted the aging of

mouse heart cells [55]. *Hpse2* encodes a heparanase that can degrade heparin sulfate proteoglycan located in the extracellular matrix and at the cell surface [56], and this protein may participate in angiogenesis by participating in biological processes [57, 58]. The Rbpj protein is an important transcriptional regulatory factor in the Notch signaling pathway as well as a homeostatic suppressor of a variety of angiogenic and vascular inhibitory factor genes in cardiomyocytes, and it can act as a major gene controller to regulate vascular growth in the adult heart [59]. The Slit2 protein is an important multifunctional protein involved in a variety of biological processes, including angiogenesis, neurodevelopment, and immune regulation. It can be a pro-angiogenic factor by inducing the formation of new blood vessels in angiogenic tissues, and it can positively regulate angiogenesis by binding to the Robo1 protein [60]. Ece1, as a rate-limiting enzyme, plays a key role in endothelin-1 (ET-1) biosynthesis [42], and Li et al. found strong evidence of selective sweeps in *Ece*1 associated with hypoxia in Tibetan pigs [43]. By observing pulmonary vascular development in mouse fetuses, Joza et al. found that Sema3 and Nrp1 were jointly involved in the early formation of the bronchial vascular tree as well as in the late development of the alveolar-capillary respiratory membrane [61]. When Sema3 and Nrp1 proteins were absent, the development of mouse lung tissue suffered from severe defects such as pulmonary vein dilatation and reduced capillary density, which ultimately led to the death of newborn mice due to respiratory failure. *Tspan4* is a key gene for migrasome formation in mammalian cells, and knockdown or knockout of *Tspan4* inhibited migrasome formation and impeded angiogenesis [62]. In a study of the role of HIF-1α in tumor cell glycolysis, Sik2 protein was found to upregulate the expression of HIF-1α, the hypoxia-inducible factor, by activating the PI3K-Akt signaling pathway [63]. Evidence of haplotype sharing in the sympatric distribution area TZ also supported the introgression of *Uvrag*, *Hpse2*, *Slit2*, *Ece1*, and *Nrp1* from plateau zokors to Gansu zokors. Our results also showed that adaptive introgressed genes are significantly enriched for multiple GO terms related to the cardiovascular system. Zhao et al. conducted transcriptome sequencing of the lungs of Tibetan sheep at different altitudes and found that some important differentially expressed genes (DEGS) were related to angiogenesis; these differentially expressed genes were significantly enriched in GO entries such as regulation of blood vessel size, regulation of vasoconstriction, cardiovascular system development, and blood vessel development, which was similar to our findings [64]. In the process of revealing the origin of the Tibetans and the genetic basis of their adaptation to the plateau environment, Wang et al. found that the positively selected genes of the Tibetans were significantly enriched in the GO terms such as blood vessel development, vasculature development, and response to hypoxia, which also suggests that the genetic mechanism of the zokor's adaptation to the plateau environment is similar to that of humans [65].

Similarly, in the sympatric distribution area SD, we identified 33 adaptive genes introgressed into the SDG (Gansu zokor, altitude: 2800 m) population of the Gansu zokor from the SDP (plateau zokor, altitude: 2800 m) population of the plateau zokor. Among these, *Phex*, *Ireb2*, *Cask*, and *Gpc3* were significantly enriched for terms related to hypoxic adaptation, including lung development, erythrocyte homeostasis, calcium ion transport, and coronary vasculature development. Haplotype sharing between the two species in the sympatric distribution area SD also supported the introgression of *Phex*, *Ireb2*, *Cask*, and *Gpc3* from plateau zokors to Gansu zokors. At present, the direct relationships between these four genes and lung development, erythrocyte homeostasis, calcium ion transport, and coronary vasculature development is not clear, and relevant research is limited. Thus, further functional research into these four genes is necessary to help elucidate their contributions to high-altitude adaptation. PI3K can be activated and phosphorylates Akt upon binding to it, enhancing HIF-α activity under hypoxic conditions. Jia et al. found that the most significantly enriched KEGG pathways of differentially expressed genes between Tibetan pig and Duroc pig were the PI3K-Akt signaling pathway [66], and this pathway still appeared in the TOP 20 pathways in our results despite its insignificance.

In conclusion, our results support the hypothesis proposed in this study, and introgressed genes from plateau zokors may have played pivotal roles in the adaptation of Gansu zokors to the plateau environment.

Methods

Sample collection
Details for individuals from all 19 populations are shown in Table S1 (Additional file 1). Of the 230 individuals involved in this study, 184 were collected directly from the field, and deep sequencing datasets for the remaining 46 individuals were obtained from online databases. Among these, the sequencing data from MD (plateau zokor, altitude: 4300 m), YK (plateau zokor, altitude: 3700 m), PA (plateau zokor, altitude: 2900 m), HL (plateau zokor, altitude: 3000 m), and QL (plateau zokor, altitude: 2800 m) populations were generated by Zhang et al. [26], the sequencing data from JY (Gansu zokor, altitude: 2000 m) (5 individuals) were generated by Zhang et al. [51], and sequencing data for ZD (Gansu zokor, altitude: 1230 m) and YZ (Gansu zokor, altitude:

2200 m) populations were generated by Liu et al. [18] (Additional file 1: Table S2). We sampled two sympatric distribution areas of plateau zokors and Gansu zokors to detect interspecific gene introgression: Tianzhu Tibetan Autonomous County (TZ) in Wuwei City, Gansu Province, and Songduo Township (SD) in Huzhu Tu Autonomous County of Haidong City, Qinghai Province. Twenty plateau zokors (TZP, plateau zokor, altitude: 3000 m) and 29 Gansu zokors (TZG, Gansu zokor, altitude: 3000 m) were collected from the TZ; 6 plateau zokors (SDP, plateau zokor, altitude: 2800 m) and 6 Gansu zokors (SDG, Gansu zokor, altitude: 2800 m) were collected from the SD. For each individual, 5 g of thigh muscle was collected and soaked in 75% alcohol and stored in a freezer at − 80 °C, with periodic changes of the alcohol storage solution.

Library construction
In this study, DNA extraction, library construction, and sequencing work were performed by Qingdao Ouyi Biological Co., LTD. The library was constructed and sequencing was conducted on Illumina NovaSeq 6000 Platform with 150 bp pair-end (PE) library size. The requirements for DNA quality control were as follows: DNA appeared as a single band with no degradation via gel electrophoresis; the concentration was > 25 ng/μL as determined via spectrophotometry, the total amount was > 500 ng, the 260/280 absorbance ratio was 1.8–1.9, and the 260/230 absorbance ratio was 1.9–2.3. The initial amount of DNA used for library construction was 100–200 ng.

Sequencing data analysis pipeline
The process for analyzing LcWGR data was as follows: (1) data quality control: fastp 0.20.0 [67] was used to perform quality control on raw reads to obtain clean reads; (2) data alignment: clean reads were aligned to the plateau zokors reference genome [26] using BWA 0.7.17 [68]; (3) genotyping: SNP (single-nucleotide polymorphism) detection and imputation were performed based on alignment results and a reference panel using loimpute 0.1.4 [69]; (4) information analysis: a series of subsequent evolutionary analyses were conducted using SNP data.

Data quality control and filtering
Incorrect base calls are always generated in sequencing data for various reasons, including biases in sequencing instruments and poor sample quality. To eliminate the influence of sequencing errors on the results, clean reads were obtained by quality filtering of raw reads. fastp 0.20.0 was used with the following filtering criteria: (1) joint sequences were removed; (2) reads with five or more N (non-AGCT) bases were removed; (3) average base mass values were calculated using a 4-base sliding window, and reads with average base mass values less than 20 were removed.

Reference genome alignment
Clean reads were aligned to the plateau zokor reference genome [26] using the mem algorithm in BWA 0.7.17. After converting the format in SAMtools 1.9 [70], the MarkDuplicates module in GATK 4.1.3.0 [71] was used to remove redundancies, and Qualimap 2.2.2d [72] was used to analyze the results of the alignment.

Mutation detection and SNP annotation
For LcWGR, whole-genome low-depth resequencing and mutation identification is performed for all individuals in the population. Because the low-coverage nature of the data will result in gaps, inference and imputation of missing genotypes based on linkage disequilibrium among SNP loci is then performed, ultimately resulting in high-density genetic markers at the whole-genome level for large-scale samples. Genotype imputation is a process wherein missing genotypes are estimated and filled in target samples based on haplotypes and genotypes in reference panels, effectively increasing the density of single-nucleotide polymorphisms (SNPs). SNP detection and imputation were performed using loimpute 0.1.4 based on the alignments of the samples to the reference genome as well as the species panel [73, 74] information. After merging the high and low depth data samples, we used vcftools 0.1.16 [75] for further filtering: (1) loci with minimum allele frequency (MAF) less than 0.05 were deleted; (2) loci with genotype data in greater than 80% of individuals were retained; (3) only the second most frequent allele at each site was retained. Using the gff annotation file of the zokor reference genome, SnpEff 4.1 g [76] was used to determine whether newly identified SNPs were located in gene elements and whether they resulted in changes at the amino acid level.

Population structure
In this study, PLINK 1.90p [77] was applied to calculate the IBS matrix, and then the phylogenetic tree of 230 individuals and 1 outgroup individual was constructed using the Neighbor-Joining method in PHYLIP 3.697 [78]. Sequencing data from an individual of *Rhizomys pruinosus* was obtained from the National Center for Biotechnology Information (NCBI) database (accession number SRX4501361) and used as the outgroup. We performed principal component analysis (PCA) on all SNP markers using GCTA 1.26.0 [79] to obtain the eigenvector with the largest eigenvalue. Before the genetic structure analysis, we first filtered SNPs obtained from the whole genome, removed loci with minimum allele frequencies less than 0.05, and then used PLINK 1.90p to

screen the non-tightly linked loci. ADMIXTURE 1.3.0 [80] was used to analyze the population structure, using a range of *K* values from 2 to 10 and repeated 10 times using 10 different seeds. The best *K* value was determined according to the cross-validation error (CVE). Due to the large population size of this study, involving a total of 230 individuals from 19 populations, pong [81] was used to cluster the results of the 10 repetitions under each *K* value and visualize the population structure.

Demographic history

The TZP (plateau zokor, altitude: 3000 m) population of plateau zokor and the TZG (Gansu zokor, altitude: 3000 m) population of Gansu zokor in the sympatric distribution area and the JY (Gansu zokor, altitude: 2000 m) population of Gansu zokor were used to analyze demographic history, and fastsimcoal2 [82] was used to infer the differentiation times among the three populations. We analyzed the genome data using the get4foldSites tool (https://github.com/brunonevado/get4foldSites) to identify all fourfold degenerate sites. SNPs that significantly deviated from Hardy–Weinberg equilibrium ($P < 0.05$) were excluded using PLINK 1.90p. The mutation rate was set to 3.03×10^{-9} based on previous studies [18]. Gene flow between populations was calculated using Treemix [83].

Analysis of linkage disequilibrium

We used the Plot_MultiPop.pl script in PopLDdecay 3.4 [84] to calculate the r^2 values using the MeanBin method for each population. Based on the generated r^2 values, we used the ggplot2 package in R to plot the chain decay plot. Due to the large number of populations, we selected the MD (plateau zokor, altitude: 4300 m), HL (plateau zokor, altitude: 3000 m), QL (plateau zokor, altitude: 2800 m), SDP (plateau zokor, altitude: 2800 m), SDG (Gansu zokor, altitude: 2800 m), JY (Gansu zokor, altitude: 2000 m), and QY (Gansu zokor, altitude: 1300 m) populations as representative populations across the altitude gradient.

Analysis of selection signals in high-altitude populations
Genome-wide selective sweep

We identified regions under positive selection by combining the genetic differentiation coefficient *Fst* and the nucleotide diversity estimate *Pi*. We selected a plateau zokor/Gansu zokor species pair (ZQ-ZD) and a high-altitude population/low-altitude population pair (TZG-ZD) to analyze the molecular mechanism of high-altitude adaptation in high-altitude populations. The genome was scanned using vcftools 0.1.16 to calculate *Fst* between two populations in each group and *Pi* for each population, with a window size of 50 kb and a step size of 50 kb. Areas with extremely high or extremely low population

Pi ratios (log10) and high *Fst* (top 10%) were identified as areas under positive selection.

Analysis of gene functional enrichment in selected regions

After screening the regions subject to positive selection, GO (Gene ontology) functional enrichment analysis and KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway enrichment analysis of the genes in these regions were performed using the R package ClusterProfiler [85].

Interspecific gene introgression
D-statistic detection of interspecific gene introgression

In this study, the *D*-statistic [86] was used to detect gene introgression between plateau zokors and Gansu zokors. The *D*-statistic was calculated as follows:

$$(P_1, P_2, P_3, O) = \frac{\sum_{i=1}^{n}(C_{ABBA}(i) - C_{BABA}(i))}{\sum_{i=1}^{n}(C_{ABBA}(i) + C_{BABA}(i))}$$

where C_{ABBA} (*i*) is the mode value of ABBA at the ith locus and C_{BABA} (*i*) is the mode value of BABA at the ith locus, both of which have a magnitude between 0 and 1. If the *D*-statistic deviates significantly from 0, then introgression is inferred to have occurred. When ABBA is greater than BABA, the *D*-statistic is positive, indicating that there was gene exchange between P2 and P3. When ABBA is less than BABA, the *D*-statistic is negative, indicating that there was gene exchange between P1 and P3. The significance of the *D*-statistic is represented by the *Z*-value, which is generally considered to be statistically significant when $|Z| > 3$. Since P3 is defined as the introgression source population, a positive *D*-statistic is usually interpreted as introgression in the direction of P3 to P2, and a negative one is usually interpreted as introgression in the direction of P3 to P1. In the sympatric distribution area TZ, we used four combinations of populations to adequately assess whether introgression occurred between TZP (plateau zokor, altitude: 3000 m) and TZG (Gansu zokor, altitude: 3000 m). In all four population combinations, P2 was the TZG (Gansu zokor, altitude: 3000 m) population, P3 was the TZP (plateau zokor, altitude: 3000 m) population, and the outgroup was *Rhizomys pruinosus* (Rpr); P1 was represented by either the YZ (Gansu zokor, altitude: 2200 m), JY (Gansu zokor, altitude: 2000 m), ZD (Gansu zokor, altitude: 1230 m), or QY (Gansu zokor, altitude: 1300 m) populations, respectively. In the sympatric distribution area SD, we used four combinations of populations to adequately assess whether introgression occurred between SDP (plateau zokor, altitude: 2800 m) and SDG (Gansu zokor, altitude: 2800 m). In all four population combinations, P2 was the SDG (Gansu zokor, altitude: 2800 m) population, P3 was the SDP (plateau zokor, altitude: 2800 m) population, and

Kang *et al. BMC Biology*   (2024) 22:187

Page 19 of 23

the outgroup was Rpr; P1 was represented by either the YZ (Gansu zokor, altitude: 2200 m), JY (Gansu zokor, altitude: 2000 m), ZD (Gansu zokor, altitude: 1230 m), or QY (Gansu zokor, altitude: 1300 m) populations, respectively. The qpDstat module in ADMIXTOOLS 7.0.2 [86] was used to calculate the *D*-statistics of the four populations in all combinations. Since the numerator of the formula used in this software was BABA-ABBA, a negative *D*-statistic in the results of this study indicated introgression in the direction of P3 to P2.

Proportion of genes that underwent introgression and screening of introgressed genes based on the f_d-statistic

Since the *D*-statistic only detects gene introgression events among the four populations, it cannot detect the exact location of the genome where gene introgression occurs. Therefore, the f_d-statistic [87] was further used to calculate the proportion of genes that underwent introgression from plateau zokors (TZP, SDP) to Gansu zokors (TZG, SDG) and identify the locations of introgressed genes. f_d is a statistic derived from the *D*-statistic, which detects the location of gene introgression across the entire genome in the form of a sliding window. Due to the directional bias of f_d in detecting gene introgression, it works better when detecting introgression from P3 to P2 than when detecting introgression from P2 to P3. Therefore, in order to analyze gene introgression from plateau zokors (TZP, SDP) in Gansu zokors (TZG, SDG), the selection of P1, P2, P3, and outgroups in the sympatric distribution areas TZ and SD was the same as indicated in the previous section. By using the python script "ABBABABAwindows.py" [87], we computed f_d values for non-overlapping windows of 50 kb throughout the genomes while excluding windows containing fewer than 100 SNPs. The significance of each f_d value was determined by calculating the *Z*-score and *P*-value for each window using Jackknife resampling [88, 89]. Jackknife resampling was performed by dividing each window into 20 blocks, removing one of the blocks each time (leave-one-out), and re-counting the f_d values using the remaining 19 blocks. This was performed 20 times to obtain the re-sampled f_d values, and then the original f_d values of the window and the 20 re-sampled f_d values were used to determine the *Z*-scores and *P*-values of the f_d values. Finally, $0 < f_d < 1$, *Z*-score-$f_d > 3$, and *P*-value-$f_d < 0.05$ were used as thresholds to indicate significant introgression signals, and the regions meeting the above criteria were selected as the genomic regions where plateau zokors (TZP, SDP) had introgressed into Gansu zokors (TZG, SDG).

The proportion of introgression across the genome (PIG) was calculated using the following formula [11]:

$$PIG = \frac{\sum f_{di} W_i}{G}$$

where f_{di} represents the fraction of introgression within the ith window, W_i is the window size, and G refers to the genome size. Genes located within the genomic regions of Gansu zokors that were introgressed from plateau zokors in the sympatric distribution areas were identified using the genome gff annotation file. GO functional enrichment analysis and KEGG pathway enrichment analysis were performed using the R software package ClusterProfiler to identify the functions of these introgressed genes [85].

Analysis of adaptive introgression

Genes that had significant introgression signals and were under positive selection in both donor and recipient populations were considered to be genes associated with adaptive introgression [9]. We identified regions under positive selection in donor and recipient populations by combining *Fst* and *Pi*. The genome was scanned using vcftools 0.1.16 to calculate *Fst* between two populations in each group and *Pi* for each population, with a window size of 50 kb and a step size of 50 kb. Areas with extremely high or extremely low population *Pi* ratios (log10) and high *Fst* (top 10%) were identified as areas under positive selection. In the sympatric distribution area TZ, we searched genome-wide for evidence of selective sweeps in the TZP (plateau zokor, altitude: 3000 m)-ZD (Gansu zokor, altitude: 1230 m) species pair and the TZG (Gansu zokor, altitude: 3000 m)-ZD (Gansu zokor, altitude: 1230 m) population pair to identify positively selected genes in the TZP (plateau zokor, altitude: 3000 m) population and TZG (Gansu zokor, altitude: 3000 m) population, respectively. We then identified genes that were under positive selection in both TZP (plateau zokor, altitude: 3000 m) and TZG (Gansu zokor, altitude: 3000 m) populations. Among this set, those with significant introgression signals were considered as genes associated with adaptive introgression. Similarly, in the sympatric distribution area SD, we searched genome-wide for evidence of selective sweeps in the SDP (plateau zokor, altitude: 2800 m)-ZD (Gansu zokor, altitude: 1230 m) species pair and the SDG (Gansu zokor, altitude: 2800 m)-ZD (Gansu zokor, altitude: 1230 m) population pair to identify positively selected genes in the SDP (plateau zokor, altitude: 2800 m) population and the SDG (Gansu zokor, altitude: 2800 m) population, respectively. We then identified genes that were under positive selection in both SDP (plateau zokor, altitude: 2800 m) and SDG (Gansu zokor, altitude: 2800 m) populations. Among this set, those with significant introgression signals were considered as genes associated with adaptive introgression.

After identification of genes associated with adaptive introgression, GO functional enrichment analysis and KEGG pathway enrichment analysis were performed using the R software package ClusterProfiler to identify their functions [85].

Based on these results, adaptive and introgressed genes that were associated with hypoxia adaptation were selected to visualize haplotype sharing between plateau zokors and Gansu zokors in the sympatric distribution areas.

Conclusions

The molecular mechanism of adaptation of the high-altitude populations of Gansu zokors to the plateau environment was similar to that of plateau zokors. Genes related to energy metabolism, cardiovascular system development, calcium ion transport, and response to hypoxia were under positive selection, which likely drove adaptation to the plateau environments. Gene introgression occurred between plateau zokors and Gansu zokors, and a number of genes associated with cardiovascular system development, lung development, and calcium ion transport were found among the positively selected genes introgressed into Gansu zokors from plateau zokors. These genes, which are highly relevant to hypoxic adaptation, may have played pivotal roles in the adaptation of the Gansu zokor to the plateau environment.

Abbreviations

XW	Xiewu town
ZQ	Zhenqin town
MD	Maduo county
YK	Yangkang village
BZ	Bazha village
HJ	Huangjiazhai town
PA	Ping'an district
HL	Hualong county
DH	Donghe village
QL	Qinglin town
GN	Gannan prefecture
TZ	Tianzhu county
SD	Songduo town
TZP	Tianzhu county, plateau zokor
SDP	Songduo town, plateau zokor
ZD	Zhidan county
QY	Qingyang city
JY	Jingyuan county
YZ	Yuzhong county
TZG	Tianzhu county, Gansu zokor
SDG	Songduo town, Gansu zokor
QTP	Qinghai-Tibet Plateau
LcWGR	Low-coverage whole-genome resequencing
SNP	Single-nucleotide polymorphism
PCA	Principal component analysis
GO	Gene ontology
KEGG	Kyoto Encyclopedia of Genes and Genomes
NCBI	National Center for Biotechnology Information
Rpr	*Rhizomys pruinosus*
BP	Biological process
MF	Molecular function
CC	Cellular component
Pcdhb	Protocadherin beta

Ly6m	Lymphocyte antigen 6 family member M
Zbtb7a	Zinc finger and BTB domain containing 7a
Creb3l3	CAMP responsive element-binding protein 3 like 3
Nrp1	Neuropilin 1
Arid2	AT-rich interaction domain 2
Ankrd55	Ankyrin repeat domain 55
Ece1	Endothelin-converting enzyme 1
Fer	FER tyrosine kinase
Corin	Corin, serine peptidase
Prkg1	Protein kinase cGMP-dependent 1
Rbpj	Recombination signal binding protein for immunoglobulin kappa J region
Nox4	NADPH oxidase 4
Pam	Peptidylglycine alpha-amidating monooxygenase
Uvrag	UV radiation resistance-associated gene
Hpse2	Heparinase 2
Slit2	Slit guidance ligand 2
Phex	Phosphate regulating endopeptidase X-linked
Ireb2	Iron-responsive element-binding protein 2
Cask	Calcium/calmodulin-dependent serine protein kinase
Gpc3	Glypican 3
STK3	Serine/threonine kinase 3
Nos1	Nitric oxide synthase 1
Tspan4	Tetraspanin 4
Sik2	Salt inducible kinase 2

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12915-024-01986-y.

Additional file 1: Fig. S1-S8. Table S1-S6, S9, S12-21, S24, S27. Fig. S1 Statistics of filtering results. Fig. S2 Distribution of base quality. Fig. S3 Distribution of base proportion. Fig. S4 Genome-wide selection signals forpopulations ZQ and ZD and populations TZG and ZD. Fig. S5 Top 30 enriched GO terms and top 20 enriched KEGG pathways for candidate genes introgressed from population TZP to TZG. Fig. S6 Top 30 enriched GO terms and top 20 enriched KEGG pathways for candidate genes introgressed from population SDP to SDG. Fig. S7 Genome-wide selection signals for populations SDP and ZD and populations SDG and ZD. Fig. S8 The degree of haplotype sharing in *Phex*, *Ireb2*, *Cask*, and *Gpc3* in pairwise comparisons between populations SDP and SDG. Table S2 Information on 46 individual data downloaded from the database. Table S3 Sample sequencing data size and quality. Table S4 Comparison information of resequenced samples. Table S5 Regional distribution of SNPs. Table S6 Transversions and transitions of SNPs. Table S9 Significantly enriched GO terms related to vascular, cardiac, and calcium ion transport for positively selected genes in the ZQ population. Table S12 Significantly enriched GO terms related to hypoxia adaptation, vascular, cardiac, and calcium ion transport for positively selected genes in the TZG population. Table S13 Results of D-statistic tests between SDG and SDP populations. Table S14 Results of fd-statistic. Table S15 Results of fd-statistic. Table S16 Results of fd-statistic. Table S17 Results of fd-statistic. Table S18 Results of fd-statistic. Table S19 Results of fd-statistic. Table S20 Results of fd-statistic. Table S21 Adaptive genes in population TZG introgressed from population TZP. Table S24 Adaptive genes in population SDG introgressed from population SDP. Table S27 Significantly enriched GO terms related to lung development, cardiovascular system, and calcium ion transport for adaptive introgressed genes in the SDG population.

Additional file 2: Table S7, S8, S10, S11, S22, S23, S25, S26. Table S7 Significantly enriched Gene Ontology terms for candidate genes from the ZQ population. Table S8 Significantly enriched KEGG pathways for candidate genes from the ZQ population. Table S10 Significantly enriched Gene Ontology terms for candidate genes from the TZG population. Table S11 Significantly enriched KEGG pathways for candidate genes from the TZG population. Table S22 Significantly enriched Gene Ontology terms for adaptive introgressed genes in the TZG population. Table S23 Significantly enriched KEGG pathways for adaptive introgressed genes in the TZG population. Table S25 Significantly enriched Gene Ontology terms for

Kang *et al. BMC Biology*   (2024) 22:187

Page 21 of 23

adaptive introgressed genes in the SDG population. Table S26 Significantly enriched KEGG pathways for adaptive introgressed genes in the SDG population.

Authors' contributions
J. H. S. conceived and designed the study. Y. K. K., Z. C. W., K. A., Q. Q. H., and Z. M. Z. collected samples. Y. K. K. and J. H. S. made all analyses and revised the manuscript. Y. K. K. wrote the manuscript draft. All authors read, revised, and approved the final manuscript.

Availability of data and materials
The genome assembly of plateau zokor used in this study were deposited in the Genome Warehouse in National Genomics Data Center, under accession number GWHABJZ00000000 [90]; the raw PacBio sequence and Illumina re-sequencing data of MD, YK, PA, HL, and QL populations used in this study have been deposited in the Genome Sequence Archive, under accession numbers CRA002263 [91]; the raw PacBio sequence and Illumina re-sequencing data of JY (5 individuals) populations used in this study have been deposited in the Genome Sequence Archive, under accession numbers CRA005933 [92]. All sequences of other individuals reported in this study were deposited in the National Center for Biotechnology Information (NCBI), BioProject: PRJNA1067323 [93].

Declarations

Ethics approval and consent to participate
This study was approved by the Animal Ethics Committee of Gansu Agricultural University and supported by the local government (GAU-LC-2020–014).

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

References
1. Orr HA, Unckless RL. The population genetics of evolutionary rescue. PLos Genet. 2014;10:e1004551.
2. Hamilton JA, Miller JM. Adaptive introgression as a resource for management and genetic conservation in a changing climate. Conserv Biol. 2016;30:33–41.
3. Harrison RG, Larson EL. Hybridization, introgression, and the nature of species boundaries. J Hered. 2014;105:795–809.
4. Signore AV, Yang YZ, Yang QY, Qin G, Moriyama H, Ge RL, et al. Adaptive changes in hemoglobin function in high-altitude Tibetan canids were derived via gene conversion and introgression. Mol Biol Evol. 2019;36:2227–37.
5. Hedrick PW. Adaptive introgression in animals: examples and comparison to new mutation and standing variation as sources of adaptive variation. Mol Ecol. 2013;22:4606–18.
6. Anderson E. Introgressive Hybridization. Hoboken: Wiley; 1949.
7. Arnold ML, Kunte K. Adaptive genetic exchange: a tangled history of admixture and evolutionary innovation. Trends Ecol Evol. 2017;32:601–11.
8. Jones MR, Mills LS, Alves PC, Callahan CM, Alves JM, Lafferty DJR, et al. Adaptive introgression underlies polymorphic seasonal camouflage in snowshoe hares. Science. 2018;360:1355–8.
9. Ma Y, Wang J, Hu Q, Li J, Sun Y, Zhang L, et al. Ancient introgression drives adaptation to cooler and drier mountain habitats in a cypress species complex. Commun Biol. 2019;2:213.
10. Li C, Wu YJ, Chen B, Cai Y, Guo J, Leonard AS, et al. Markhor-derived introgression of a genomic region encompassing *PAPSS2* confers high-altitude adaptability in Tibetan goats. Mol Biol Evol. 2022;39:253.
11. Fu R, Zhu Y, Liu Y, Feng Y, Lu RS, Li Y, et al. Genome-wide analyses of introgression between two sympatric Asian oak species. Nat Ecol Evol. 2022;6:924–35.
12. Nanaei HA, Cai Y, Alshawi A, Wen J, Hussain T, Fu WW, et al. Genomic analysis of indigenous goats in Southwest Asia reveals evidence of ancient adaptive introgression related to desert climate. Zool Res. 2023;44:20–9.
13. Huerta-Sánchez E, Jin X, Asan, Bianba Z, Peter BM, Vinckenbosch N, et al. Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. Nature. 2014;512:194–7.
14. Quan C, Li Y, Liu X, Wang Y, Ping J, Lu Y, et al. Characterization of structural variation in Tibetans reveals new evidence of high-altitude adaptation and introgression. Genome Biol. 2021;22:159.
15. Chen N, Cai Y, Chen Q, Li R, Wang K, Huang Y, et al. Whole-genome resequencing reveals world-wide ancestry and adaptive introgression events of domesticated cattle in East Asia. Nat Commun. 2018;9:2337.
16. Wu DD, Ding XD, Wang S, Wójcik JM, Zhang Y, Tokarska M, et al. Pervasive introgression facilitated domestication and adaptation in the *Bos* species complex. Nat Ecol Evol. 2018;2:1139–45.
17. Wang MS, Wang S, Li Y, Jhala Y, Thakur M, Otecko NO, et al. Ancient hybridization with an unknown population facilitated high-altitude adaptation of canids. Mol Biol Evol. 2020;37:2616–29.
18. Liu X, Zhang S, Cai Z, Kuang Z, Wan N, Wang Y, et al. Genomic insights into zokors' phylogeny and speciation in China. Proc Natl Acad Sci U S A. 2022;119:e2121819119.
19. Kang Y, Wang Z, Yao B, An K, Pu Q, Zhang C, et al. Environmental and climatic drivers of phenotypic evolution and distribution changes in a widely distributed subfamily of subterranean mammals. Sci Total Environ. 2023;878:163177.
20. Kang Y, Su J, Yao B, Wang C, Zhang D, Ji W. Interspecific skull variation at a small scale: the genus *Eospalax* exhibits functional morphological variations related to the exploitation of ecological niche. J Zool Syst Evol Res. 2021;59:902–17.
21. Fan N, Shi Y. A revision of the zokors of subgenus *Eospalax*. Acta Theriol Sin. 1982;2:183–99.
22. Zhang Y, Liu J. Effects of plateau zokor (*Myospalax fontanierii*) on plant community and soil in an alpine meadow. J Mammal. 2003;84:644–51.
23. Yi X, Liang Y, Huerta-Sanchez E, Jin X, Xi Ping Cuo Z, Pool JE, et al. Sequencing of 50 human exomes reveals adaptation to high altitude. Science. 2010;329:75–8.
24. Peng Y, Yang Z, Zhang H, Cui C, Qi X, Luo X, et al. Genetic variations in Tibetan populations and high-altitude adaptation at the Himalayas. Mol Biol Evol. 2011;28:1075–81.
25. He Y, Qi X, Ouzhuluobu, Liu S, Li J, Zhang H, et al. Blunted nitric oxide regulation in Tibetans under high-altitude hypoxia. Natl Sci Rev. 2018;5:516–29.
26. Zhang T, Chen J, Zhang J, Guo YT, Zhou X, Li MW, et al. Phenotypic and genomic adaptations to the extremely high elevation in plateau zokor (*Myospalax baileyi*). Mol Ecol. 2021;30:5765–79.
27. Zhang B, Chamba Y, Shang P, Wang Z, Ma J, Wang L, et al. Comparative transcriptomic and proteomic analyses provide insights into the key genes involved in high-altitude adaptation in the Tibetan pig. Sci Rep. 2017;7:3654.
28. Alex BC, Gompert Z. Population genomics based on low coverage sequencing: how low should we go? Mol Ecol. 2013;22:3028–35.
29. Nevado B, Ramos-Onsins SE, Perez-Enciso M. Resequencing studies of nonmodel organisms using closely related reference genomes: optimal experimental designs and bioinformatics approaches for population genomics. Mol Ecol. 2014;23:1764–79.
30. Wang X, Hu G, Saito Y, Ni G, Hu H, Yu Z, et al. Did the modern Yellow River form at the Mid-Pleistocene transition? Sci Bull. 2022;67:1603–10.
31. Li Y, Wu DD, Boyko AR, Wang GD, Wu SF, Irwin DM, et al. Population variation revealed high-altitude adaptation of Tibetan mastiffs. Mol Biol Evol. 2014;31:1200–5.

32. Scheinfeldt LB, Soi S, Thompson S, Ranciaro A, Woldemeskel D, Beggs W, et al. Genetic adaptation to high altitude in the Ethiopian highlands. Genome Biol. 2012;13:R1.

33. Wang MS, Li Y, Peng MS, Zhong L, Wang ZJ, Li QY, et al. Genomic analyses reveal potential independent adaptation to high altitude in Tibetan chickens. Mol Biol Evol. 2015;32:1880–9.

34. Zhang W, Fan Z, Han E, Hou R, Zhang L, Galaverni M, et al. Hypoxia adaptations in the grey wolf (*Canis lupus chanco*) from Qinghai-Tibet Plateau. PLos Genet. 2014;10:e1004466.

35. Hui AS, Bauer AL, Striet JB, Schnell PO, Czyzyk-Krzeska MF. Calcium signaling stimulates translation of HIF-α during hypoxia. FASEB J. 2006;20:466–75.

36. Fearnley CJ, Roderick HL, Bootman MD. Calcium signaling in cardiac myocytes. Cold Spring Harbor Perspect Biol. 2011;3:a004242.

37. Paudel S, Sindelar R, Saha M. Calcium signaling in vertebrate development and its role in disease. Int J Mol Sci. 2018;19:3390.

38. Sang N, Stiehl DP, Bohensky J, Leshchinsky I, Srinivas V, Caro J. MAPK signaling up-regulates the activity of hypoxia-inducible factors by its effects on p300. J Biol Chem. 2003;278:14013–9.

39. Majmundar AJ, Wong WJ, Simon MC. Hypoxia-inducible factors and the response to hypoxic stress. Mol Cell. 2011;40:294–309.

40. Heallen T, Zhang M, Wang J, Bonilla-Claudio M, Klysik E, Johnson RL, et al. Hippo pathway inhibits Wnt signaling to restrain cardiomyocyte proliferation and heart size. Science. 2011;332:458–61.

41. Wang GL, Jiang BH, Rue EA, Semenza GL. Hypoxia-inducible factor 1 is a basic-helix-loop-helix-PAS heterodimer regulated by cellular O_2 tension. Proc Natl Acad Sci U S A. 1995;92:5510–4.

42. Whyteside AR, Turner AJ, Lambert DW. Endothelin-converting enzyme-1 (ECE-1) is post-transcriptionally regulated by alternative polyadenylation. PLoS ONE. 2014;9:e83260.

43. Li M, Tian S, Jin L, Zhou G, Li Y, Zhang Y, et al. Genomic analyses identify distinct patterns of selection in domesticated pigs and Tibetan wild boars. Nat Genet. 2013;45:1431–8.

44. Bunn HF, Poyton RO. Oxygen sensing and molecular adaptation to hypoxia. Physiol Rev. 1996;76:839–85.

45. Gess B, Schricker K, Pfeifer M, Kurtz A. Acute hypoxia upregulates NOS gene expression in rats. Am J Physiol. 1997;273:R905–10.

46. Chen J, Bai M, Ning C, Xie B, Zhang J, Liao H, et al. Gankyrin facilitates follicle-stimulating hormone-driven ovarian cancer cell proliferation through the PI3K/AKT/HIF-1α/cyclin D1 pathway. Oncogene. 2015;35:2506–17.

47. Arnold ML. Transfer and origin of adaptations through natural hybridization: were Anderson and Stebbins right? Plant Cell. 2004;16:562–70.

48. Barton NH. The role of hybridization in evolution. Mol Ecol. 2001;10:551–68.

49. Suarez-Gonzalez A, Lexer C, Cronk QCB. Adaptive introgression: a plant perspective. Biol Lett. 2018;14:20170688.

50. Mallet J. Hybridization as an invasion of the genome. Trends Ecol Evol. 2005;20:229–37.

51. Zhang T, Lei ML, Zhou H, Chen ZZ, Shi P. Phylogenetic relationships of the zokor genus *Eospalax* (Mammalia, Rodentia, Spalacidae) inferred from whole-genome analyses, with description of a new species endemic to Hengduan Mountains. Zool Res. 2022;43:331–42.

52. Zhu H, He L. UVRAG: A multifunctional protein with essential roles in the heart. Journal of Cardiology & Current Research. 2014;1:13–6.

53. Song Y, Quach C, Liang C. UVRAG in autophagy, inflammation, and cancer. Autophagy. 2020;16:387–8.

54. Song Z, An L, Ye Y, Wu J, Zou Y, He L, et al. Essential role for UVRAG in autophagy and maintenance of cardiac function. Cardiovasc Res. 2014;101:48–56.

55. Lai S, Zhang S, Liu X, Mazhar H, Naz A. Ultraviolet resistance-associated gene (*Uvrag*) deficiency promotes cellular senescence in the heart. Chinese J Tissue Eng Res. 2020;25:2241–6. (In Chinese).

56. Oskarsson T. Extracellular matrix components in breast cancer progression and metastasis. Breast. 2013;22:S66–72.

57. Poltavets V, Kochetkova M, Pitson SM, Samuel MS. The role of the extracellular matrix and its molecular and cellular regulators in cancer cell plasticity. Front Oncol. 2018;8:431.

58. Zcharia E, Metzger S, Chajek-Shaul T, Friedmann Y, Pappo O, Aviv A, et al. Molecular properties and involvement of heparanase in cancer progression and mammary gland morphogenesis. J Mammary Gland Biol. 2001;6:311–22.

59. Díaz-Trelles R, Scimia MC, Bushway P, Tran D, Monosov A, Monosov E, et al. Notch-independent RBPJ controls angiogenesis in the adult heart. Nat Commun. 2016;7:12088.

60. Coll M, Ariño S, Martínez-Sánchez C, Garcia-Pras E, Gallego J, Moles A, et al. Ductular reaction promotes intrahepatic angiogenesis through Slit2-Roundabout 1 signaling. Hepatology. 2022;75:353–68.

61. Joza S, Wang J, Fox E, Hillman V, Ackerley C, Post M. Loss of semaphorin-neuropilin-1 signaling causes dysmorphic vascularization reminiscent of alveolar capillary dysplasia. Am J Pathol. 2012;181:2003–17.

62. Zhang C, Li T, Yin S, Gao M, He H, Li Y, et al. Monocytes deposit migrasomes to promote embryonic angiogenesis. Nat Cell Biol. 2022;24:1726–38.

63. Gao T, Zhang X, Zhao J, Zhou F, Wang Y, Zhao Z, et al. SIK2 promotes reprogramming of glucose metabolism through PI3K/AKT/HIF-1α pathway and Drp1-mediated mitochondrial fission in ovarian cancer. Cancer Lett. 2020;469:89–101.

64. Zhao P, Zhao F, Hu J, Wang J, Liu X, Zhao Z, et al. Physiology and transcriptomics analysis reveal the contribution of lungs on high-altitude hypoxia adaptation in Tibetan sheep. Front Physiol. 2022;13:885444.

65. Wang B, Zhang YB, Zhang F, Lin H, Wang X, Wan N, et al. On the origin of Tibetans and their genetic basis in adapting high-altitude environments. PLoS ONE. 2011;6:e17002.

66. Jia C, Kong X, Koltes JE, Gou X, Yang S, Yan D, et al. Gene co-expression network analysis unraveling transcriptional regulation of high-altitude adaptation of Tibetan pig. PLoS ONE. 2016;11:e0168161.

67. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 2018;2018(34):884–90.

68. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at arXiv. 2013. https://arxiv.org/abs/1303.39.

69. Wasik K, Berisa T, Pickrell JK, Li JH, Fraser DJ, King K, et al. Comparing low-pass sequencing and genotyping for trait mapping in pharmacogenetics. BMC Genomics. 2021;22:197.

70. 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.

71. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The genome analysis toolkit: a mapreduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20:1297–303.

72. Okonechnikov K, Conesa A, García-Alcalde F. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. Bioinformatics. 2015;32:292–4.

73. Yang W, Yang Y, Zhao C, Yang K, Wang D, Yang J, et al. Animal-ImputeDB: a comprehensive database with multiple animal reference panels for genotype imputation. Nucleic Acids Res. 2019;48:659–67.

74. Gao Y, Yang Z, Yang W, Yang Y, Gong J, Yang QY, et al. Plant-ImputeDB: an integrated multiple plant reference panel database for genotype imputation. Nucleic Acids Res. 2020;49:1480–8.

75. 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.

76. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms. SnpEff Fly. 2012;6:80–92.

77. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007;81:559–75.

78. Felsenstein J. PHYLIP: Phylogeny inference package - v3.2. Cladistics. 1989;64:164-166.

79. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet. 2011;88:76–82.

80. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 2009;19:1655–64.

81. Behr AA, Liu KZ, Liu-Fang G, Nakka P, Ramachandran S. Pong: fast analysis and visualization of latent clusters in population genetic data. Bioinformatics. 2016;32:2817–23.

82. Excoffier L, Dupanloup I, Huerta-Sánchez E, Sousa VC, Foll M. Robust demographic inference from genomic and SNP data. PLos Genet. 2013;9:e1003905.

83. Pickrell JK, Pritchard JK. Inference of population splits and mixtures from genome-wide allele frequency data. PLos Genet. 2012;8:e1002967.

84. 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.
85. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS. 2012;16:284–7.
86. Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, Zhan Y, et al. Ancient admixture in human history. Genetics. 2012;192:1065–93.
87. Martin SH, Davey JW, Jiggins CD. Evaluating the use of ABBA-BABA statistics to locate introgressed loci. Mol Biol Evol. 2015;32:244–57.
88. Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, et al. A draft sequence of the Neandertal genome. Science. 2010;328:710–22.
89. Zhang W, Dasmahapatra KK, Mallet J, Moreira GRP, Kronforst MR. Genome-wide introgression among distantly related *Heliconius* butterfly species. Genome Biol. 2016;17:25.
90. The genome assembly data. National Genomics Data Center. 2020. https://ngdc.cncb.ac.cn/gwh/Assembly/941/show. Accessed 16 July 2024.
91. Zhang T. High-altitude adaptation of plateau zokor. National Genomics Data Center. 2021. https://ngdc.cncb.ac.cn/gsa/search?searchTerm=CRA002263.
92. Zhang T. Whole-genome resequence of *Eospalax* species. National Genomics Data Center. 2022. https://ngdc.cncb.ac.cn/gsa/search?searchTerm=%22CRA005933%22.
93. Kang Y, Wang Z, An K, Hou Q, Zhang Z, Su J. Introgression drives adaptation to the plateau environment in a subterranean rodent. GenBank. 2024. https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1067323/.

Publisher's Note