Genomic insights into zokors’ phylogeny and speciation in China

Xi Liu1,1, Shangzhe Zhang1,1, Zhenyuan Cai1,1, Zhuoran Kuang1,1, Na Wan4, Yinjia Wang4, Leyan Mao5, Xuan An6, Fang Li6,6, Tuo Feng6, Xiaolong Lianga,2, Xi Liang, and Zhenqing Qian

Contributed by Eviatar Nevo; received December 4, 2021; accepted March 30, 2022; reviewed by Hynek Burda and Zhengting Zou

The phylogeny and speciation of subterranean zokors in China are unclear, as previous studies on morphology and limited molecular markers have generated conflicting results. This study unraveled the complex evolutionary history of eight zokor species in China based on de novo assembly at chromosome level and whole-genome sequencing of 23 populations. We found extensive phylogenetic discrepancies between nuclear and mitochondrial phylogenies, and different coalescent phylogenies, which could be explained by introgression and incomplete lineage sorting (ILS). The recent Qinghai-Tibet Plateau uplift (∼3.60 million y ago; Mya) drove Eospalax to speciate into clade A and clade B (∼3.22 Mya), and discordant phylogenies in this node were mainly attributed to introgression rather than ILS. Clade A rapidly divided into three lineages due to geographical isolation and glaciation, while glaciation and C4 plant expansion contributed to the speciation of clade B. ILS contributed to the discordances of two rapidly radiated nodes rather than introgression. The effective population sizes (Ne’s) of all the species of Eospalax were affected by three glaciations. Ancient polymorphisms and divergence hitchhiking contribute to genomic islands of all the species pairs. Positively selected genes putatively related to specific habitation adaptations were identified, such as heart development, neurogenesis, DNA repair, and immune response. Climate, geological tectonism, and C4 vegetation shaped the habitat adaptation and speciation of zokors in China.

Myospalacinae | population genomics | introgression | incomplete lineage sorting | speciation

The genetic basis underlying ecological speciation is one of the key topics in evolutionary biology. The way populations adapt to new environments and the genealogy of closely related species have boosted the interest of ecologists and evolutionary biologists in this field (1). Relationships between species displayed by phylogenetic trees are not always explicit, as many studies have shown the widespread impact of introgression, incomplete lineage sorting (ILS), and the demographic history of tree construction (2, 3). Introgression and ILS leave similar genomic footprints among diversifying species, hence the difficulty distinguishing them across a genome. Therefore, it is always challenging to reconstruct a precise phylogenetic tree representing the true relationships among taxa in the face of combinations of evolutionary processes (4). Recent advances in high-throughput sequencing technology and corresponding analysis have improved our ability to infer phylogenetic relationships. Evolutionary biologists can now elucidate relatively precise relationships between species by integrating different datasets and methods. The development of genomics and genetic models, including the coalescence theory, has contributed to advances in population demography, speciation, and adaptation. These advances allow researchers to study the complex demographic histories, reveal speciation triggers, and identify underlying genetic mechanisms of adaptive evolution (5).

Myospalacinae (Rodentia, Spalacidae), who live and forage underground, is characterized by morphological traits typical for subterranean mammals, like short limbs and tails, reduced auricles and eyes, and prominent burrowing “tools.” Myospalacinae is generally distributed throughout semiarid and semihumid areas with annual precipitation between 200 and 800 mm to the north of the Yangtze River (Fig. 1A). This subfamily consists of two genera: Eospalax and Myospalax; distributed mainly in the Palearctic realm, including China, Mongolia, Far East Russia, Kazakhstan, and North Korea. Zokors’ adaptive evolution and ecological speciation are hypothesized to be driven by rapid environmental shifts and climatic oscillations (6). They originated during the Middle Miocene in Mongolia (6) and then dispersed southward into the rising Qinghai-Tibet Plateau (QTP) (7). Eospalax, distributed only in China (8), contains six species, namely E. fontanierii (distributed from northeast to middle and west China, overlapped with several other species), E. rufescens (distributed throughout the

**Significance**

Phylogenies are the basis of many ecological and evolutionary studies. However, zokor phylogeny and speciation patterns are heavily debated. This study disentangled the phylogeny and speciation of zokors genomically. Six species of the Eospalax were separated into high-altitude E. baileyi and E. smithi and the rest four low-altitude species by recent Qinghai-Tibet Plateau uplift 3.6 million y ago. E. rothschildi and E. smithi speciated south of the Qinling–Huaihe Line, where refuges were supplied during glaciation. Introgression and incomplete lineage sorting led to the complex phylogeny of zokors. Genomic islands were formed due to ancient polymorphisms and divergence hitchhiking. This study concluded that climatic, geological, and tectonic events shaped the phylogeny and speciation of zokors in China.

PNAS 2022 Vol. 119 No. 19 e2121819119
https://doi.org/10.1073/pnas.2121819119
Published May 5, 2022.
Qinling Mountains), *E. cansus* (distributed throughout the Loess Plateau, north of the Qinling Mountains), *E. rothschildi* (found south of *E. rufescens*), *E. baileyi* (scattered in the QTP, with two populations isolated by the Yellow River), and *E. smithi* (found mainly south of *E. baileyi*) (Fig. 1A). *Myospalax*, consisting of *M. aspalax* and *M. psilurus*, was found in northeast China.

The Qinling–Huaihe Line (QHL) climatically divides north and south China by temperature and precipitation. The north of the line has a temperate continental climate with snow in winter, whereas the south of the line has a tropical and subtropical monsoon climate, hot and humid. The QHL classifies all eight species into three groups. *E. smithi* and *E. rothschildi* are distributed to the south of the QHL, and *E. rufescens* and *E. baileyi* are distributed throughout the Qinling Mountains. The distribution of *E. baileyi* is divided by the Yellow River (Fig. 1A), a southern region with annual precipitation of ~400 to 800 mm and a northern region with ~200- to 400-mm annual precipitation. *M. aspalax* and *M. psilurus* occurred only in northeast China (Fig. 1A).

The phylogeny of zokors is still controversial. Based on phylogenetic analysis using mitochondrial DNA data, Su et al. suggested that *E. cansus* was a basal clade, and *E. rufescens*, *E. smithi*, and *E. baileyi* were clustered together in another clade (8). Li and Chen suggested that *Eospalax* originated from the Qinling Mountains with *E. fontanierii* as its ancestor. One clade living in the low-altitude subtropical regions evolved into *E. rothschildi*, and the other clade moved to the cold and high altitude, which later speciated to *E. smithi* (9). Li and Wang suggested that *Eospalax* originated from the Qinling Mountains with *E. fontanierii* as its ancestor, splitting first into *E. rufescens* and then into *E. cansus* after spreading northward. *E. smithi* shows a close relationship with *E. rothschildi* (10), dispersing westward and speciating to *E. baileyi*. Zhang et al. claimed that *E. cansus* could be distinguished from species in the subgenus *Eospalax* and that *E. smithi* is closely related to *E. rothschildi*, suggesting that *E. smithi* could have split from *E. rothschildi* (11). These previous studies were based on antiquated morphologies or the limited number of available genes. Nevertheless, gene trees and species trees are different (12). Some evolutionary histories, like convergent evolution, could lead evolutionary or ecological biologists to the wrong taxonomy based on morphology alone (13). A phylogenetic tree based on one or more genes (14) or...
only the mitochondrial genome (15) creates misleading results due to ILS, gene introgression, hybridization (16), horizontal gene transfer, gene duplication, and gene deletion (17), which would make phylogeny inference more complex and lead to gene tree discordance (16).

This study sequenced genomes of the representative populations across the distributional range of eight zokor species to clarify the relationship among them. We explored the divergence order, speciation patterns, introgression, and effective population size (Ne) fluctuation of eight zokor species based on population-scale genomic data and coalescent modeling. Interspecific divergence timescales and Ne of each species were associated with climatic and geological changes. Genomic islands with elevated differentiation were detected for 12 species pairs, and genes under positive selection were identified for each species. Genomic results highlighted the phylogeny and speciation of zokors across China. High-throughput sequencing technology and population genetic analysis methods generated more accurate phylogenies for zokor species than previous phylogenies based solely on morphological phenotypes.

Results

De Novo Genome Assembly, Scaffolding, and Annotation. We collected 90 individuals from eight zokor species, each from two to five populations (Fig. 1A and SI Appendix, Fig. S17 and Table S1). A chromosome-level genome of zokor (E. fontanierii) was generated through de novo assembly based on 99.2-Gb (~35x) HiFi reads and polished with 125-Gb (~42x) Illumina paired-end (PE) reads (SI Appendix, Table S3). The assembled genome size was 2.70 Gb (Table 1), approximately the same size as predicted (SI Appendix, Fig. S1). The contig N50 was 59.2 Mb and GC content was 41.5% (Table 1 and SI Appendix, Figs. S2 and S3 and Table S4). We anchored 91.6% of contigs to 31 chromosomes (2n = 62) (Fig. 1D, Table 1, and SI Appendix, Fig. S4 and Table S7) using Hi-C data (SI Appendix, Tables S5 and S6). The karyotype number of E. fontanierii was consistent with the previously reported karyotype (18).

The genome assembly covered 3,967 (~97.0%) of the 4,104 total orthologs in the Benchmarking Universal Single-Copy Orthologs (BUSCO) database, indicating completeness of the genome (SI Appendix, Figs. S5 and S6 and Table S8). The mapping rate of PE reads against the assembly was 99.9%, suggesting the high precision of this assembly (SI Appendix, Table S9). There were 21,656 protein-coding genes, 0.08% simple sequence repeats (SSRs), and 42.6% transposable elements (TEs) predicted across the whole genome (SI Appendix, Fig. S7 and Tables S10 and S11). All genome assemblies and annotated gene characters are similar to those reported for E. baileyi (19) at the same time.

| Table 1. Statistics for E. fontanierii genome assembly |
|-------------------------------------------------------|
| **Assembly** | E. fontanierii 3.0 |
| **Sequencing platform** | PacBio Sequel 2,699,719,537 |
| **Genome size** | 227,920,291 |
| **Longest sequence length, bp** | 59,204,535 |

Myospalacinae Has a Closer Relationship with Rhizomyinae Compared with Spalacinae. Phylogenetic trees based on different markers showed conflicting phylogeny among Myospalacinae, Spalacinae, and Rhizomyinae (SI Appendix, Fig. S14). A concatenated maximum-likelihood (ML) tree was reconstructed to clarify the phylogeny of these three subfamilies, based on 13 protein-coding mitochondrial genes, revealing that Myospalacinae was more closely related to Rhizomyinae (SI Appendix, Fig. S15). The coalescent topologies based on the nuclear genome had 40.7% of windows supporting a closer relationship between zokors and bamboo rats. The rest of the two discordant topologies, supported by 30.0 and 29.3% of windows (SI Appendix, Fig. S16), were attributed to introgression (SI Appendix, Table S12). The blind mole rats initially split from zokors about 29.1 million y ago (Mya), while the divergence of zokors and bamboo rats occurred around 26.9 Mya (Fig. 1C).

Phylogenetic Relationships among Zokor Species. Principal-component analysis (PCA) of 19 climatic factors (SI Appendix, Table S1) could distinguish these species, except that E. fontanierii overlapped with E. canus, M. psilurus, and M. aspalax (Fig. 1B). Temperature seasonality and annual precipitation were identified as two main climatic factors contributing to speciation (SI Appendix, Fig. S18). A significant positive correlation between 11 out of 19 climatic factors and corresponding genetic distances (Mantel test, $R^2 = -0.215$ to 0.496, **$P < 0.01$; SI Appendix, Tables S1 and S2) was detected.

To investigate the phylogenetic relationship of zokor species in China, PE reads were mapped against the E. fontanierii genome (SI Appendix, Table S9) and ~7.36 million high-quality single-nucleotide polymorphisms (SNPs) were obtained from the eight species (SI Appendix, Fig. S19). After removing closely related individuals, 51 out of 90 samples remained for subsequent analyses. PCA (Fig. 2A), phylogenetic trees (Fig. 2B), STRUCTURE analysis (Fig. 2C and SI Appendix, Fig. S21), and genetic networking (SI Appendix, Fig. S20) showed separate genetic clusters of eight lineages. E. baileyi was treated as two lineages, E. baileyi1 (north of the Yellow River) and E. baileyi2 (south of the Yellow River), for downstream analysis because of population differentiation within it (Fig. 2C). All species clustered into two clades: E. baileyi1, E. baileyi2, and E. smithi into clade A, and E. fontanierii, E. rufescens, E. canus, and E. rothschildi into clade B. Through STRUCTURE analysis, discrepancies were observed in clustering E. canus and E. rothschildi between K = 5 and 6, and in clustering E. baileyi1, E. baileyi2, and E. smithi between K = 5 and 6 (Fig. 2C). M. psilurus and M. aspalax were observed to separate until K = 9, suggesting they are extremely closely related (Fig. 2C and SI Appendix, Fig. S21).

Genetic Differentiation and Diversity among Zokor Species. The pairwise genetic differentiation was measured by $F_{ST}$. E. baileyi1 showed the largest divergence from all other lineages (Fig. 2D). Intraspecific differentiation ($F_{ST} = 0.314$; Fig. 2D) of E. baileyi was even greater than interspecific differentiation of many other species (SI Appendix, Fig. S22). E. fontanierii, the population with the lowest genetic diversity, showed a high level of genetic differentiation from other lineages. E. baileyi1 and E. baileyi2 showed the highest nucleotide diversity (measured by $\pi$ of $4.1 \times 10^{-3}$ and $3.9 \times 10^{-3}$, respectively, followed by E. rothschildi, E. smithi, E. rufescens, E. canus, and E. fontanierii (Fig. 2D)). The linkage disequilibrium (LD), indicated by $r^2$, decayed to half of its maximum within 100 kb in all the species except E. fontanierii (SI Appendix, Fig. S23).
which was comparable to that of Spalax. *E. fontanierii* dropped to its lowest value at ∼600 kb (SI Appendix, Fig. S23), displaying the highest LD.

**Discordances in Zokor Topology between the Mitochondrial and Nuclear Genome.** The topology based on the mitochondrial genome was discordant with that based on the nuclear genome. *E. fontanierii* was a basal clade in the mitochondrial tree, whereas it was an inner clade in the nuclear genome tree (Fig. 2B). *M. aspalax* and *M. psilurus* were clustered into one branch in the mitochondrial tree. In contrast, *M. aspalax* was a basal clade and *M. psilurus* was clustered with *Eospalax* in the nuclear genome tree (Fig. 2B), possibly because of close distribution leading to introgression. Two *E. rothschildi* individuals in the mitochondrial phylogeny gathered together with *E. fontanierii*, consistent with the haplotype networks of zokors based on the mitochondrial cytochrome oxidase subunit 1 (COI) gene in the mitochondrial DNA (SI Appendix, Fig. S24).

**Discordances between Gene Trees and Species Tree.** We observed 73.6% of gene trees (Fig. 3A) were consistent with the concatenated species tree (SI Appendix, Fig. S25). Discordances existed mainly in two nodes, 7 and 8 (Fig. 3A and B). In the unstable node 7, the proportion of topology (4, 8|5, 6) in red was 58.1%, while the remaining two topologies, (4, 5|6, 8) in blue and (4, 6|5, 8) in purple, were 21.1 and 20.8%, respectively. In node 8, the percentage of the consensus tree in red (12, 13|4, 7) was 63.0%, while the other two topologies, (12, 4|13, 7) and (12, 7|13, 4), were 21.3 and 15.7%, respectively (Fig. 3B–F).

**Extensive Introgression among Zokor Species.** To investigate whether introgression was the primary cause of the phylogenetic discordances, Patterson’s *D* statistic was calculated using the rooted tree expressed as ((P1, (P2, P3)), *Rhizomys pruinatus*). Significant introgression was detected from 50 triplets (***P < 0.01; SI Appendix, Table S13), and the mean value of

---

**Fig. 2.** Phylogeny and genetic diversity of eight zokor species. (A) PCA based on whole-genome SNP. (B) ML tree based on 13 mitochondrial protein-coding genes (Left) and the neighbor-joining tree of nuclear genome SNPs (Right). (C) STRUCTURE analysis of eight zokor species from *K* = 2 to 10, in which *K* = 8 was the best-fitted model (SI Appendix, Fig. S22). (D) Divergence parameter (*F*<sub>ST</sub>) and nucleotide diversity (*π*) of species from *Eospalax*. The clade (Left) comprises three species adapting to high altitude, while the rest consists of four species adapting to low altitude. Values inside the circle are *π* (×10<sup>-3</sup>). Values on the dashed lines connecting two species represent the corresponding *F*<sub>ST</sub>.
absolute $D$ for each species pair was calculated (Fig. 4A). There was no significant evidence for recent hybridization through $F_2$-statistics analysis (SI Appendix, Table S14). The introgression between *Eospalax* and *M. psilurus* was significantly larger than that between *Eospalax* and *M. aspalax* ($***P < 0.0001$; SI Appendix, Table S13). This result illustrated inconsistencies caused by introgression between the mitochondrial and nuclear trees of *M. psilurus* and *M. aspalax* (Fig. 2B). Broad abutting distribution facilitated introgression between *Eospalax* and *M. psilurus* rather than *M. aspalax*. More importantly, the $D$ value of triplet ((*M. psilurus*, *M. aspalax*), *E. fontanieri*) was 0.098, which was larger than that when *E. fontanieri* was replaced by any one of the six species from *Eospalax* ($D$ value around 0.065; SI Appendix, Table S13), suggesting that *E. fontanieri* may be the oldest lineage of *Eospalax*. Introgression occurred between *E. smithi* and all other species except *E. baileyi*. Congruently, introgression was observed between *E. rothschildi* and all species except *E. cansus* (Fig. 4A). These two results indicated extensive migration across the QHL (Fig. 1A). Introgression was restricted between the geographically isolated *E. baileyi* and all the other species, except *E. rothschildi* and *M. psilurus*, suggesting allopatric speciation of *E. baileyi* (Fig. 4A).

**Contribution of Introgression and ILS to Discordant Nodes.**

The phylogenetic discordances could be attributed to introgression or ILS. To distinguish introgression from ILS, we performed quantifying introgression via branch length (QuIBL) analysis on 168 topologies with *R. pruinosus* as an outgroup. Of all the tested triplets, 84% showed significant evidence for introgression (141 of 168 triplets, $\Delta$BIC (Bayesian information criterion) > 10; SI Appendix, Table S15). Four discordant topologies (from Fig. 3 C–F) were attributed to 8.1, 7.5, 28.8, and 8.3% introgression loci on average (SI Appendix, Table S15), respectively, suggesting interspecies introgression was the main reason for the conflicting phylogeny. In addition, the similar proportions of discordant topologies (*fo* (ba1/ba2/sm, ro/ru/ca)) ($\sim$28 to 30%) and *(ro/ru/ca, fo, ba1/ba2/sm) ($\sim$6.8 to 9.1%) suggest that ancient introgression possibly occurred between *E. fontanieri* and the ancestor of *E. smithi* and *E. baileyi* (SI Appendix, Table S15).

The discordant topologies of (*E. rothschildi* (*E. cansus*, *E. rufescens*)) and (*E. baileyi2, (E. baileyi1, E. smithi*)) were solely caused by ILS (both $\Delta$BIC $< -10$; SI Appendix, Table S15). Based on the inconsistencies within clade B, we obtained 15 topologies for a rooted quadruple tree, and the major topology was consistent with the phylogenetic tree. The top three discordant topologies were attributed to introgression, ILS, and introgression, respectively (Fig. 4C and SI Appendix, Table S15).

**Ancient Introgression Detection.**

The $D_{COIL}$ statistic was used to detect ancient introgression using a group of quadruple tree from *Eospalax* with *R. pruinosus* as an outgroup. Ancient introgression between *E. fontanieri* and the common ancestor of *E. baileyi* and *E. smithi* was the largest within *Eospalax*, taking up $\sim$50% of windows (Fig. 4B and SI Appendix, Table S16), verifying our inference of ancient introgression based on QuIBL analysis.

To disentangle the problem that the intragenus split of the common ancestor of *E. baileyi* and *E. smithi* was earlier than the intergenus split of *Myospalax* and *Eospalax* in $K = 2$ (Fig. 2C), we tested the ancient introgression between *E. fontanieri* and the recent common ancestor of *M. psilurus* and *M. aspalax*. About 59.5% of windows supported introgression between *E. fontanieri* and *Myospalax*, consistent with PhyloNet analysis (Fig. 4 B and D), with $\sim$10% more windows than supporting *E. fontanieri* and the recent common ancestor of *E. baileyi* and *E. smithi* (Fig. 4B). Therefore, the larger ancient introgression between *E. fontanieri* and *Myospalax* contributed to their clustering from $K = 2$ to 5 (Fig. 2C).

---

**Fig. 3.** Tree structure discordance with the phylogenetic tree. (A) The dominant topology takes up 73.6% of 100-kb neutral loci window trees. The number in black boxes represents each node. Clade A contained *E. baileyi* and *E. smithi*, while clade B contained the remaining four species in *Eospalax*. (B) Frequency of three bipartitions induced in branches 7 and 8 of ASTRAL species trees using the 100-kb neutral site sliding window; headers 7 and 8 correspond to the branch identity in A. The $x$ axis represents the exact definition of each quartet topology. The main topologies are in red, and the other two alternative topologies are in blue and purple. The dashed lines indicate the $1/3$ threshold. (C and D) The discordance tree structures in node 7 correspond to the blue bar and purple bar in B (left), respectively. (E and F) The discordance tree structures in node 8 correspond to the blue bar and purple bar in B (right), respectively.
Demographic History. The fluctuations of Ne for each of the eight lineages were estimated using pairwise sequentially Markovian coalescent (PSMC) models. Three population declines occurred in the five lineages (E. fontanierii, E. cansus, E. rufescens, E. baileyi1, and E. baileyi2) distributed north of the QHL (Fig. 5A), mirroring three glaciation events: Naynayxungla glaciation (NG, ∼0.5 to 0.78 Mya), penultimate glaciation (PG, 135 to ∼194 thousand y ago; kya), and last glacial maximum (LGM, around 20 kya). After the NG ice age retreat, all species except E. baileyi1 and E. rothschildi tended to recover their Ne (Fig. 5A and B). E. fontanierii, E. cansus, and E. rufescens started to recover about ∼600 to 900 kya (Fig. 5A and SI Appendix, Fig. S17). E. baileyi2 and E. cansus showed a small population expansion during ∼0.3 to 0.5 Mya, but neither E. baileyi1 nor E. rothschildi expanded (Fig. 5A and B) during this period. The Ne of all lineages declined during PG and LGM as the sea level increased and atmospheric surface air temperature decreased. In contrast, E. smithi and E. rothschildi, distributed south of the QHL, expanded during the glaciation (Fig. 5A and B) and contracted immediately after the ice age (Fig. 5C).

Eospalax split into clade A and clade B ∼3.22 Mya (95% CI at 3.07 to 3.31 Mya) (Fig. 5C, SI Appendix, Fig. S26, and Dataset S1). Later, E. baileyi1 and E. smithi split from clade A at ∼1.18 Mya (95% CI at 1.17 to 1.34 Mya) and ∼0.75 Mya (95% CI at 0.711 to 0.787 Mya), respectively. For the low-altitude clade B, E. fontanierii first split from their ancestor ∼2.40 Mya (95% CI at 2.21 to 2.40 Mya), confirming that it was the oldest species within Eospalax. E. cansus diverged from E. rufescens ∼1.46 Mya (95% CI at 1.41 to 1.67 Mya) and settled in the Loess Plateau. Finally, the divergence of E. rothschildi from E. rufescens occurred ∼0.44 Mya (95% CI at
0.371 to 0.499 Mya) (Fig. 5 C and SI Appendix, Fig. S26). Introggression was detected between E. fontanieri and the most recent common ancestor of E. rufescens and E. rothschildi, consistent with the D statistics and D_{FST} (Fig. 4 A and B).

**Genomic Islands.** To further investigate the evolutionary forces of speciation, we identified 8,705 genomic islands (GIs) with the top 1% F_{ST} divergence from 21 species pairs (SI Appendix, Figs. S27 and S28). After combining adjacent islands, a total of 7,788 GIs were detected across the genome. We compared D_{XY}, π, Tajima’s D, LD, and the recombination rate (ρ) of GIs with those of background regions. A significantly elevated level of LD, reduced level of ρ, and declined Tajima’s D were observed in GI regions, indicating the presence of divergent hitchhiking for these regions. The number or size of GIs did not correlate with their population divergence (SI Appendix, Fig. S30).

**Genes under Positive Selection.** Although six species in Eospalax were closely related, phenotypes among them were diverse due to their adaptation to specific habitat. To investigate the genetic basis of zokor’s adaptation and speciation, positively selected genes (PSGs) were identified for each species. A total of 747 PSGs were identified (286 in E. baileyi, 115 in E. smithi, 55 in E. fontanieri, 186 in E. rufescens, 126 in E. cansus, and 91 in E. rothschildi) among 20,580 genes analyzed. Gene ontology (GO) enrichment analyses of those PSGs for E. baileyi revealed that there were 146 enriched GO terms mainly involved in musculature (GO:0060537 and GO:0007517), neurogenesis (GO:0051961 and GO:0021766), heart morphogenesis (GO:0003007, GO:0001947, GO:0061371, and GO:0007507), DNA repair (GO:0006281), autophagy (GO:0005058, GO:0010506, GO:0016239, and GO:0016241), immunology (GO:0030890), and regulation of HMOX1 (heme oxygenase 1) activity (R-MMU-9707587) (Dataset S2). As for E. smithi, there were 43 GO terms enriched, of which eight were involved in DNA modification, dendritic spine development, and blood circulation development (GO:0006305, GO:0030890, and GO:0060999) (Dataset S3). The remaining four species, E. fontanieri, E. cansus, E. rothschildi, and E. rufescens, enriched 25, 21, 37, and 102 GO terms, respectively. For E. fontanieri, seven terms involved in the regulation of neuron projection development (GO:0010975) and apoptotic process (GO:0043065), as well as the behavior of cognition (GO:0050890) and learning or memory (GO:0007611), were enriched (Dataset S4). E. cansus enriched three terms associated with renal and kidney system development (GO:0072001 and GO:0001822) (Dataset S5), whereas E. rothschildi enriched six terms related to limb development (GO:0035108) and more than seven terms involved in embryonic development (GO:0048568) (Dataset S6). E. rufescens, a high-altitude species, enriched terms related to neurogenesis (GO:0051960), immunology (GO:0002366), DNA repair (GO:0006281), ventricular system development (GO:0021591), the respiratory system (GO:0044065), starch and sucrose metabolism (mmu00100), insulin receptor signal (GO:0009141), memory (GO:0007613), and cognition (GO:0050890) (Dataset S7).

**Discussion**

**Zokors’ Distribution in China.** Zokors are herbivorous subterranean mammals that spend most of their lives underground. Organisms living in such habitats generally cope with hypoxia,
darkness, and hypercapnia (20). Oxygen concentration within the burrow systems is dependent on air permeability and barometric pressures (21). If the soil moisture is high, ventilation would be blocked, leading to hypoxia and hypercapnia in the tunnels. In contrast, if the soil is too dry, limited vegetation could survive to supply quantitative and qualitative food resources for the animals. Zokors occupy semihumid and semiarid areas (Fig. 1A), similar to the climatic niche of another subterranean mammal, Spalax in Israel (22). These habitats offer a compromise between vegetation-rich mesic areas with extreme hypoxia and hypercapnia due to high soil moisture and better ventilated, low soil moisture areas with insufficient food resources.

Phylogeny of the Three Subterranean Subfamilies from Spalacidae. Spalacinae, Myospalacinae, and Rhizomyinae are considered as one family, Spalacidae, a sister lineage of Muridae (23). However, the phylogenetic relationship of the three subfamilies remains controversial. Several studies have attempted to elaborate the phylogeny through morphology, nuclear or mitochondrial genes (24–27), and transcriptomes (28), creating conflicting results (SI Appendix, Fig. S14). There are several possibilities for this discordance. First, the phylogeny was constructed using one or only a few concatenated molecular markers, which may be under selection. Second, introgression or ILS could also cause inconsistencies in tree topologies, leading to this discordance. The phylogeny reconstructed in this study based on whole genomes confirmed that zokors are more closely related to bamboo rats than blind mole rats, providing solid evidence for the discordance. Introgresion, rather than ILS, is responsible for 90% of two discordant topologies, (Rattus, Eospalax) Spalax, Rhizomyos in clade 2 and (Rattus, Rhizomyos) Spalax, Eospalax in clade 3 (SI Appendix, Table S12). The mapping rate of Spalax and Rhizomyos against E. fontanieri is 70.7 and 95.4%, respectively, also suggesting a closer relationship between Rhizomyos and Eospalax.

Population Structure of the Eight Zokor Species. Climate, including 19 variables (SI Appendix, Table S1), might affect the adaptation and speciation of zokors in China (Fig. 1B). Precipitation and temperature were the most effective factors shaping the distribution and evolution of zokors (SI Appendix, Fig. S18). The first PC explained 57.0% of variations, separating E. rothschildi from other species (Fig. 1B). E. rothschildi is the only species distributed in areas with annual precipitation over 800 mm (Fig. 1A). The second PC explained 25.2% of variations and separated E. smithi and E. baileyi from all other species because they were distributed at higher elevations (SI Appendix, Fig. S17). The genetic distances were positively correlated with the 11 out of 19 climate factors (SI Appendix, Tables S1 and S2) and the altitude (R² = 0.520; ***P = 1.0e-4), which was consistent with the large divergences observed between high-altitude and low-altitude lineages (Fig. 2D).

Genetic Diversity and Differentiation of Zokors. The primary cause of variation in genetic diversity may be associated with demographic histories and ecological stresses. The π of E. baileyi was the highest among all species because of its largest Nₑ (9.4 × 10⁻⁷) and environmental heterogeneity (SI Appendix, Fig. S26). The current E. baileyi was reported to recover from four refugia during the LGM (29), and the environmental heterogeneity enhanced the maintenance of polymorphisms (30). E. rothschildi and E. smithi, distributed in the warmer climate south of the QHL, could have received many migrants during the glaciation because of warm shelter. Extensive introgression accounted for their second- (E. rothschildi) and third- (E. smithi) highest π (Figs. 2D and 4A). Larger π in E. rothschildi than in E. smithi could have resulted from the larger Nₑ and the stronger stress exerted by more precipitation (Figs. 1A and 2D and SI Appendix, Fig. S26). E. fontanieri, although being a basal clade species, showed the lowest π (Fig. 2D), the slowest LD decay rate (SI Appendix, Fig. S23), but not the smallest Nₑ, a result that requires further investigation in the future. The intraspecific divergence between E. baileyi and E. baileyi is even larger than between different species pairs (Fig. 2B), suggesting that E. baileyi split into two distinct species due to the Yellow River barrier. E. baileyi was subjected to drought stress with 200 to 400 mm of annual precipitation (Fig. 1A). By contrast, all other species were distributed in areas with annual precipitation above 400 mm. A previous study reported that E. baileyi consisted of four subpopulations (29), likely due to more individuals and the larger distribution areas in that study. We suggest that E. baileyi and E. baileyi were two species because of their genetic and phenotypic differences. The divergence between E. baileyi and E. rufescens was minimal (Fig. 2E), likely due to similar geologies (Fig. 1A) and introgression occurring between them (Fig. 4A and SI Appendix, Fig. S26).

Discordances in Phylogeny. There are limitations in using the concatenated multilocus approach in the face of introgression and ILS, which could create difficulties when inferring the phylogeny, and may not provide the true history of the species (2). The discordances that occurred on nodes 7 and 8 (Fig. 3B) were mainly due to ILS [SI Appendix, Table S15; topologies (ba2, (ba1, sm)) and (ca, (ro, ru))]. The inconsistency in the clustering of E. rufescens, E. cansus, and E. rothschildi from K = 4 to 6 and the clustering of E. smithi, E. baileyi, and E. baileyi from K = 5 to 6 in STRUCTURE analysis (Fig. 2C) could be attributed to ILS. ILS is ubiquitous during rapidly diverging lineages, such as adaptive radiation, occurring due to the short time between speciation events (31–34). Therefore, the comparatively short divergence time on these two nodes for zokors might have caused this discordance. The larger Nₑ’s, 7.9 × 10⁷ (ancestor of E. baileyi and E. smithi) and 6.3 × 10⁷ (ancestor of E. cansus, E. rufescens, and E. rothschildi), allowed for a high degree of ILS in these two clades (SI Appendix, Fig. S26) (35).

Population Demographic History. Climate shaped the adaptation and speciation of zokors (Fig. 2B) and blind mole rats (22). Although both of them live underground with comparatively constant temperatures, they are affected by the aboveground temperature (36). Glaciation affected zokors by freezing them to death and diminishing vegetation, which supplied food to the herbivorous zokors. Limited vegetation during glaciation explains the decline of Nₑ of the five species north of the QHL. However, in the warmer and wetter climate south of the QHL, the vegetation was expected to be better, which could supply shelter for zokors. The Nₑ declines of E. smithi and E. rothschildi after PG and LGM were likely due to heavy rain in south and east China (37, 38). Introgresion from E. baileyi to E. smithi and from E. rufescens to E. rothschildi was detected (Fig. 3D), which is consistent with the hypothesis that zokors migrated to warmer shelters during the glaciation increasing the Nₑ (Fig. 5B) and π (Fig. 2D) of E. smithi and E. rothschildi (Fig. 2E). The speciation of these two species may have been triggered by the southward migration of zokors driven by glaciation. The temperature and vegetation within the three species (E. rufescens, E. fontanieri, and E. cansus) distribution areas were different because of the different altitudes. E. rufescens was distributed at the highest altitude and, correspondingly, its Nₑ
was the last to recover. On the other hand, *E. fontanieri* occurred at the lowest altitude of all three species (*SI Appendix, Fig. S17*) and, correspondingly, its *Ne* expanded first (Fig. 5A and *SI Appendix, Fig. S17*).

Compared with *E. bailey1*, *E. bailey2* showed a small population expansion after the NG, because *E. bailey2* was distributed in areas with annual precipitation of more than 400 mm. Population expansion was observed in *E. cansus* and *E. bailey2* after NG glaciation (Fig. 5 A and B), consistent with C4 plant expansion in the Loess Plateau 0.43 Mya when the summer rainfall increased although the total precipitation decreased (39). C4 plants are adapted to hot, dry environments, and zokors mainly feed on C4 plants (40), which allows for their synchroic expansion around ~0.40 to 0.50 Mya.

**Positively Selected Genes.** Genes are always selected during adaptation and speciation to new environments. All zokors living underground cope with darkness, hypoxia, hypercapnia, and environments full of pathogens (41). However, zokors’ environments differ in altitude (*SI Appendix, Fig. S17*), temperature, precipitation (*SI Appendix, Table S1*), and vegetation. *E. bailey* lives at an altitude between 2,800 and 4,600 m, which is characterized by low temperature, low oxygen caused by low barometric pressure, and strong UV (ultraviolet) light. Genes related to immunology were selected since pathogens in humid underground tunnels threaten the mammal’s health (20). The brain and nervous system of zokors are expected to be damaged under hypoxia, so a developed neurological system could facilitate their adaptation to hypoxia, accounting for the PSGs enriched in neurogenesis (42). PSGs enriched in the cell cycle and DNA repair underlie their resistance to oxidative injury and strong UV at high altitudes (43). Additionally, *E. bailey* could also well protect itself from superoxide caused by hypoxia by regulating the expression and activity of HMOX1 (44). Genes related to autophagy could increase the energy utilization for lack of food at high altitudes (45). As expected, multiple genes were enriched in the functions of heart development and determination of heart left/right asymmetry conducive to oxygen delivery (46), contributing to hypoxia adaptation. Vegetation at high altitude is scarce and of low quality and therefore, in order to find more food sources, zokors must explore much larger areas (47). Correspondingly, genes related to the development of the hippocampus and memory behavior were under selection. PSGs of *E. bailey* were also enriched in musculature, potentially due to their extensive digging activities (Dataset S2). *E. smithi*, located in the second-highest altitude, enriched many GO terms similar to *E. bailey*, like nervous system development and DNA modification. In addition, genes related to blood circulation that would facilitate oxygen delivery could have been selected to cope with hypoxia. *E. rufescens*, located in the Qinling Mountains with a relatively high altitude, enriched genes in ventricular system development and respiratory system development, which are responsible for oxygen inhalation and delivery, facilitating their adaptation to hypoxia. Starch and sucrose metabolic pathway transformation could promote the accumulation of lipids, helping zokors survive the long-term winter without a new food supply (48). Genes related to insulin receptor signaling could potentially increase glucose absorption for heat production (49), which could help zokors cope with low temperatures at high altitudes. PSGs related to memory were beneficial in remembering complex underground tunnels. For *E. cansus*, genes related to the renal system were selected to increase water absorption because of the arid environment in the Loess Plateau (50). Genes related to apoptosis were identified in *E. fontanieri*, reflecting antitumor potential (51). Typically, *E. rothschildi* has a smaller body size and is located south of the QHHL, where the average temperature is high. In contrast, *E. bailey* normally has a larger body size and is located in the high-altitude QTP, where the average temperature is low. Both of them enriched terms of growth regulation and limb development, which is consistent with Bergmann’s rule (52).

**Genome Islands.** The diversifications between each zokor species pair across the genome are heterogeneous and could generate GIs (*SI Appendix, Fig. S28*). GIs frequently occur across the zokor genome, especially between closely related species. Local adaptation (53), reproductive isolation (54), selection (55–57), and sorting of ancient balanced polymorphisms (53, 58) could lead to GIs. In the present study, both *FST* and *DX* within islands are larger than those of background regions (*SI Appendix, Table S17*), which could be explained by divergent sorting of ancient polymorphisms (53, 58, 59) or speciation with introgression (60). If the latter model induced the GIs, more robust islands with increasing divergence time and introgression would be expected (61, 62). However, the divergence of GIs from different species pairs was not significantly different, and the size or number of GIs did not correlate with introgression and divergence time (*SI Appendix, Fig. S30*), rejecting the speciation-with-gene-flow model.

The GIs of all the species pairs were characterized by elevated LD, reduced level of π, and decreased recombination rates (*SI Appendix, Table S17*). We observed a significant correlation between the divergence and recombination rate from GIs between most species pairs (*SI Appendix, Fig. S28*), suggesting the possible occurrence of divergence hitchhiking or background selection (53, 58, 60). *DX* within island regions would not increase if the GIs were initiated by linked selection (53, 60). In contrast, *DX* within island regions did increase in our results (*SI Appendix, Table S17*), rejecting the hypothesis of linked selection leading to GIs. Elevated *DX* indicates that haplotypes containing GIs were likely to become genetically isolated earlier than the rest of the genomes of all species pairs under comparison (*SI Appendix, Table S17*; Mann–Whitney *U* test, ***P < 2.2e-16***). These results suggest that divergent sorting of ancient polymorphisms produced the GIs (53, 58, 60, 63).

**Zokor Divergence and Speciation Momentum.** Intrinsic (genetics, demographic history) and extrinsic (climate, habitat) factors could produce speciation momentum (64). *Eospalax* separated from *Myospalax* more than 4 Mya, proved by fossil data (8), and *Eospalax* speciated southwestward to *E. bailey* in the QTP about 3.22 Mya (Fig. 5D), when the Kun–Huang tectonism occurred and the QTP had been uplifted (65). The separation between *E. bailey1* and *E. bailey2* occurred 1.18 Mya (Fig. 5C), which is consistent with the tectonism causing the Yellow River penetration to the QTP about 1.2 Mya (66), separating these two populations (Figs. 1A and 5C). *E. bailey2* speciated southward to *E. smithi* 0.75 Mya (Fig. 5C), around the NG, and the *Ne* of *E. smithi* expanded during both the PG and LGM (Fig. 5B). Therefore, glaciation is considered the speciation momentum that drove animals southward to the warm refuge south of the QHHL. *E. rufescens* speciated northward to *E. cansus* 1.46 Mya and southward to *E. rothschildi* 0.44 Mya (Fig. 5C). The northward dispersal of *E. cansus* was probably triggered by heat-adapted C4 plant expansion that occurred 1.6 Mya in the south Loess Plateau (39). *E. rothschildi* speciated southward during the NG (Fig. 5B and C), and its population...
expanded during the LGM, suggesting glacialiation drove zokors to the warm southern shelter (Fig. 3B), causing speciation there. The migration from E. baijii toward to E. smithi, and from E. rufescens toward to E. rathbuni, is consistent with the detected introgression (Fig. 4A and D) and higher genetic diversity (Fig. 2D). Although the distribution of E. baijii abutted with E. canus (Fig. 1A) (67), it was secondary contact with restricted recent introgression. Climate (SI Appendix, Fig. S18), teocim, the uplift of the QTP, and C4 plant expansion triggered the speciation of zokors in China.

Materials and Methods

Long-read was performed with the PacBio Sequel platform, while short-read sequencing was performed with the Illumina NovaSeq 6000 and DNBSEQ T7 platform. Clean reads were mapped against the chromosome-level reference genome by BWA, and SNPs were called with GATK. Genetic differentiation parameters, such as Fst and ω, were calculated by VCFTools. Distant tree analysis was performed with RAxML, ASTRAL, and DiscoVista. Recent introgression, incomplete lineage sorting, and ancient introgression were analyzed by D statistics, QuIBL, and δOL. Demographic history was inferred using PSMC and fastsimcoal v2.7. Detailed information is available in SI Appendix, Materials and Methods.

Data Availability. Genome and Sequence Read Archive data were deposited in the National Center for Biotechnology Information (NCBI), BioProject: PRJNA776391. All study data downloaded from NCBI are listed in SI Appendix.

ACKNOWLEDGMENTS.

This project was supported by the National Natural Science Foundation of China (32071487 and 32000305), National Key Research and Development Programs (2021YFA1200900), Science Fund for Creative Research Groups of Gansu Province (21JR7RA533), Lanzhou University’s ‘Double First Class’ Guided Project-Team Building Funding Research Startup Fee (to K.L.), Chang Jiang Scholars Program, the Fundamental Research Funds for Excellent Youth from the Central Universities (Lanzhou University) (1Uzubky-2021 e7y7 and Uzubky-2020 c007), and State Key Laboratory of Grassland Agro-Ecosystems (Lanzhou University; Grants SKLGE-202001, 202009, and 202010). We thank Dr. Nawan Shrestha for English editing. We are also grateful for the Ancell-Teicher Research Foundation for Genetic and Molecular Evolution for its continuous financial support for the Spalax Research Program. We received support for computational work from the Big Data Computing Platform for Western Ecological Environment and Regional Development and Supercomputing Center of Lanzhou University.

Author affiliations: 1State Key Laboratory of Grassland Agro-Ecosystems, College of Ecology, Lanzhou University, Lanzhou 730000, China; 2Key Laboratory of Adaptation and Evolution of Plateau Biology, Northwest Institute of Plateau Biology, Chinese Academy of Sciences, Xining 810001, China; 3Qinghai Provincial Key Laboratory of Animal Industrial Genomics, Lanzhou 730000, China; 4College of Life Sciences and Technology, Mudanjiang Normal University, Mudanjiang 157001, China; 5Shanxi Key Laboratory of Animal Conservation, Shaanxi Institute of Zoology, Xian 710032, China; and 6Institute of Evolution, University of Haifa, Haifa 3498838, Israel.
46. T. Zhang et al., Phenotypic and genomic adaptations to the extremely high elevation in plateau zokor (Myospalax baileyi). Mol. Ecol. 30, 5765-5779 (2021).
47. K. Li et al., Sympatric speciation revealed by genome-wide divergence in the blind mole rat Spalax. Proc. Natl. Acad. Sci. U.S.A. 112, 11905–11910 (2015).
48. J. Yang et al., Whole-genome sequencing of native sheep provides insights into rapid adaptations to extreme environments. Mol. Biol. Evol. 33, 2576-2592 (2016).
49. M. Li et al., Genomic analyses identify distinct patterns of selection in domesticated pigs and Tibetan wild boars. Nat. Genet. 45, 1431–1438 (2013).
50. F. Li, Z. Qiao, Q. Duan, E. Nevo, Adaptation of mammals to hypoxia. Animal Model. Exp. Med. 4, 311–318 (2021).
51. B. A. Carneiro, W. S. El-Deiry, Targeting apoptosis in cancer therapy. Nat. Rev. Clin. Oncol. 17, 395–417 (2020).
52. M. Rodríguez et al., Bergmann’s rule and the geography of mammal body size in the Western Hemisphere. Glob. Ecol. Biogeogr. 17, 274–283 (2008).
53. F. Han et al., Gene flow, ancient polymorphism, and ecological adaptation shape the genomic landscape of divergence among Darwin’s finches. Genome Res. 27, 1004–1015 (2017).
54. M. Malinsky et al., Genomic islands of speciation separate cichlid ecomorphs in an East African crater lake. Science 350, 1493–1498 (2015).
55. R. Burni et al., Linked selection and recombination rate variation drive the evolution of the genomic landscape of differentiation across the speciation continuum of Ficedula flycatchers. Genome Res. 25, 1656–1665 (2015).
56. S. Renaut et al., Genomic islands of divergence are not affected by geography of speciation in sunflowers. Nat. Commun. 4, 1827 (2013).
57. N. Vijay et al., Evolution of heterogeneous genome differentiation across multiple contact zones in a crow species complex. Nat. Commun. 7, 13195 (2016).
58. T. Ma et al., Ancient polymorphisms and divergence hitchhiking contribute to genomic islands of divergence within a poplar species complex. Proc. Natl. Acad. Sci. U.S.A. 115, E236-E243 (2018).
59. B. Wang et al., Ancient polymorphisms contribute to genome-wide variation by long-term balancing selection and divergent sorting in Boechera stricta. Genome Biol. 20, 126 (2019).
60. T. E. Cruzickbank, M. W. Hahn, Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. Mol. Ecol. 23, 3133–3157 (2014).
61. F. Noriil, D. J. Funk, D. Drito-Barrantos, Divergent selection and heterogeneous genomic divergence. Mol. Ecol. 18, 375–402 (2009).
62. J. L. Feder, P. Nosil, The efficacy of divergence hitchhiking in generating genomic islands during ecological speciation. Evolution 64, 1729–1747 (2010).
63. J. Wang, N. R. Street, D. G. Scofield, P. K. Ingvarsson, Variation in linked selection and recombination drive genomic divergence during allopatric speciation of European and American aspens. Mol. Biol. Evol. 33, 1754–1767 (2016).
64. H. F. Gante, W. Salzburger, Evolution: Cichlid models on the runaway to speciation. Curr. Biol. 22, R956-R958 (2012).
65. J. Li, Y. Shi, B. Li, Uplift of the Qinghai-Xizang (Tibet) Plateau and Global Change (Lanzhou University Press, Lanzhou, China, 1995).
66. Z. Hu et al., Rapid fluvial incision and headward erosion by the Yellow River along the Jinshaan gorge during the past 1.2 Ma as a result of tectonic extension. Quat. Sci. Rev. 133, 1–14 (2016).
67. L. Liu et al., Classification of two zokor species based on mitochondrial gene, morphological and habitat indices. Shou Lei Xue Bao 38, 402 (2018).