Genome Biology

RESEARCH

Open Access

Centromeric tandem repeats expansions in plateau zokor enhance chromosome stability for high-altitude adaptation

Zhen-Long Wang (王振龙) [1*†], Lu-Ye Shi (施禄也) [2†], Xiu-Yun Liu (刘秀云) [2,3†], Xiao-Qing Yao (姚晓晴) [2,3], Meng-Yang Li (李孟阳) [1], Qiang-Hui Wang (王强辉) [4], Hao Wu (吴昊) [5], Hui-Sheng Gong (巩会生) [6], Liu-Yang Fu (付留洋) [1], Gang Chang (常罡) [6], Li-Xin Wei (魏立新) [7], Wen-Hui Nie (佴文惠) [8], Wei-Ting Su (苏伟婷) [8], Bo Zhao (赵博) [2], Tao Zhang (张涛) [2*] and Peng Shi (施鹏) [2,9*]

†Zhen-Long Wang, Lu-Ye Shi and Xiu-Yun Liu contributed equally to this work.

*Correspondence: wzl@zzu.edu.cn; zhangtao@mail.kiz.ac.cn; ship@mail.kiz.ac.cn

[1] School of Life Sciences, Zhengzhou University, Zhengzhou, China
[2] State Key Laboratory of Genetic Evolution & Animal Models, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, China
[3] University of Chinese Academy of Sciences, Beijing, China
[4] Novogene Bioinformatics Institute, Beijing, China
[5] Agency for Science, Technology and Research (A*STAR), Singapore, Singapore
[6] Shaanxi Key Laboratory for Animal Conservation, Shaanxi Institute of Zoology, Xian, China
[7] Northwest Institute of Plateau Biology, CAS Key Laboratory of Tibetan Medicine Research, Chinese Academy of Sciences, Xining, China
[8] Kunming Cell Bank, Kunming Institute of Zoology, Chinese Academy of Science, Kunming, China
[9] School of Future Technology, University of Chinese Academy of Sciences, Beijing, China

Abstract

Background: Repetitive elements constitute nearly 50% of mammalian genomes; however, their roles in adaptive evolution remain poorly understood. Their identification and assembly, particularly for tandem repeats, are technically challenging, leading to their frequent omission from downstream analyses.

Results: Here, we employ high-quality long-read sequencing to generate chromosome-level genome assemblies for four zokor species adapted to different altitudes. Comparative analyses reveal that the plateau zokor, the only species primarily inhabiting the high-altitude Qinghai-Tibet Plateau, exhibits an expanded genome size and a distinctive bimodal GC content distribution. These features are largely driven by the extensive expansion of a highly conserved tandem repeat, termed pzTR. Experimental analyses show that pzTR is widely distributed across plateau zokor chromosomes and is likely associated with centromeric regions. Functional assays further demonstrate that depletion of pzTR in plateau zokor cell lines disrupts genomic stability by impairing accurate chromosome segregation. Notably, chromosomes harboring pzTR display reduced mis-segregation rates, particularly under hypoxic conditions.

Conclusions: Together, these findings highlight the unique genomic architecture of the plateau zokor and suggest a critical role for centromeric tandem repeats in safeguarding genome stability under hypoxic stress associated with high-altitude environments.

Keywords: Plateau zokor, Adaptive evolution, Tandem repeat, Centromere drive

Background

The subfamily Myospalacinae comprises fully subterranean rodents endemic to East Asia, classified within the family Spalacidae and exhibiting close phylogenetic affinity with the subfamilies Rhizomyinae and Spalacinae [1–4]. Among these taxa, the plateau zokor (*Eospalax baileyi*) stands out due to its exceptional adaptations to the harsh conditions of high-altitude environments, predominantly inhabiting the Qinghai-Tibet Plateau and surrounding areas [5, 6]. This species demonstrates remarkable adaptations to extreme hypoxia, low temperatures, and elevated ultraviolet radiation [7, 8]. While the subterranean habitat of the plateau zokor mitigates some challenges associated with cold and UV stress, it exacerbates the difficulties posed by hypoxia [9, 10].

Recent comparative genomic analyses across various Myospalacinae taxa have provided insights into potential genes and structural variants associated with hypoxia adaptation in the plateau zokor [7, 10–12]. However, the functional implications of repetitive sequences in these adaptive mechanisms remain insufficiently explored. Historically regarded as "junk DNA", repetitive sequences—including tandem repeats (TRs) and transposable elements (TEs)—are now recognized as important contributors to genome function and evolution [13, 14]. TRs include satellite sequences such as microsatellites and macrosatellites, which are frequently enriched near centromeres and telomeres and can contribute to chromosomal organization and stability [15]. Satellite DNAs are among the most dynamic components of eukaryotic genomes, often showing rapid turnover in sequence composition and copy number even between closely related species [16, 17]. TEs include diverse mobile elements, such as long interspersed nuclear elements (LINEs), short interspersed nuclear elements (SINEs), and long terminal repeats (LTRs), which can shape genome evolution through insertion, deletion, and recombination [18]. Within the plateau zokor genome, over 50% of sequences are composed of repetitive elements. Their high similarity and inherent complexity present substantial challenges to conventional genomic assembly methods, complicating the identification and functional characterization of these sequences [19].

Advances in sequencing technologies, particularly long-read platforms such as Oxford Nanopore Technology (ONT) and PacBio high-fidelity (HiFi) sequencing, have facilitated the large-scale identification and assembly of these repetitive sequences, thereby allowing for a deeper investigation into their contributions to adaptive evolution. In this study, we employed a comprehensive sequencing strategy combining ONT, HiFi, Illumina paired-end reads, and Hi-C data, achieving approximately 300 × coverage across four zokor species from diverse altitudinal gradients within Myospalacinae. The newly assembled genome of the plateau zokor (3.25 Gb) significantly surpasses the prior assembly (2.47 Gb) by nearly 800 Mb [7]. Notably, the plateau zokor genome exhibits an expansion of 200–350 Mb compared to other zokor species, a phenomenon uncommon among mammalian taxa with divergence times less than 4 million years. This genomic expansion is primarily attributed to a highly conserved segment of tandem repeats, designated pzTR, encompassing 372 Mb, and is predominantly localized to the centromeric regions. Our findings suggest that pzTR is crucial for maintaining chromosomal integrity, particularly under hypoxic conditions, underscoring its vital contribution to the adaptive evolution of the plateau zokor.

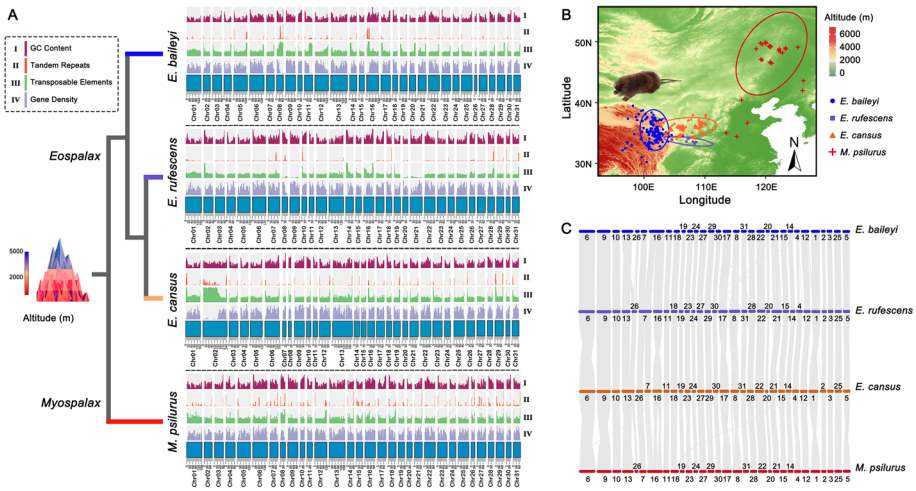

Fig. 1 Chromosome-level genome assemblies of the plateau zokor and three related zokor species across elevational gradients. **A** Genomic features of four zokor species across elevational gradients. **B** Geographic distribution of the plateau zokor, Qinling zokor, Gansu zokor, and Transbaikal zokor. Latitude and longitude data for all sampling points of the four zokor species were obtained from the GBIF website (https://www.gbif.org/) and zokor-related references. **C** Chromosome synteny among the four zokor genome assemblies. Horizontal bars represent chromosome models, with syntenic regions connected by gray lines

Table 1 Summary statistics of the four zokor genome assemblies

	E. baileyi	*E. rufescens*	*E. cansus*	*M. psilurus*
Common name	plateau zokor	Qinling zokor	Gansu zokor	Transbaikel zokor
Assembly size (Gb)	3.250	2.899	3.033	2.993
Scaffold N50 (Mb)	80.26	90.33	89.08	89.52
BUSCO completeness (%)	93.2	94.9	95.2	93.7
Repeat content (%)	54.6	37.0	43.0	44.1
GC content (%)	42.8	41.9	41.7	41.4

Results

Genome sequencing, assembly, and annotation of four zokor species

The plateau zokor (*Eospalax baileyi*) is an intriguing species that has adapted to high-altitude environments, with its core distribution found between 2800 and 4600 m [7]. To explore the genomic mechanisms underlying adaptive evolution in this species, we compared it with three other zokor species representing different altitudinal gradients: *E. rufescens* (Qinling zokor, 2000–2700 m), *E. cansus* (Gansu zokor, 800–2500 m), and *Myospalax psilurus* (Transbaikal zokor, 0–900 m) (Fig. 1A, B). We employed a comprehensive sequencing strategy integrating long ONT reads, long HiFi reads, short paired-end Illumina reads, and short paired-end Hi-C reads, achieving an approximate cumulative coverage of 300 × per species (Additional file 1: Tables S1-S3). Through a multi-round assembly and polishing strategy, high-quality chromosome-level genome assemblies were produced for all four zokor species (Fig. 1A, C, Table 1, Additional file 2: Fig. S1).

Quality assessments of the genome assemblies, using Benchmarking Universal Single-Copy Orthologs (BUSCO) and Core Eukaryotic Genes Mapping Approach (CEGMA)

metrics, yielded scores between 93.20%−95.20% and 91.94%−96.37%, respectively (Additional file 1: Table S4), confirming the high integrity of the assembled genomes. Mapping rates of short paired-end Illumina reads to their respective genomes ranged from 99.38% and 99.69% (Additional file 1: Table S4), further validating the accuracy and consistency of the assemblies. Scaffold N50 values ranged from 80.26 Mb to 90.33 Mb, indicating strong assembly continuity (Additional file 1: Table S5). To further assess the assembly quality, we compared our genomes with 105 publicly available genomes in the UCSC database, including 45 rodent genomes [20] (Additional file 3: Table S6). This comparison revealed that, despite the larger genome sizes in the four zokor species, they maintained exceptionally high contig N50 values (Additional file 2: Fig. S2A), highlighting the robustness of the assemblies. The final genome sizes were documented as 3.25 Gb for plateau zokor, 2.90 Gb for Qinling zokor, 3.03 Gb for Gansu zokor, and 2.99 Gb for Transbaikal zokor (Additional file 1: Table S5), with 81.94%, 89.52%, 90.91%, and 90.39% of the sequences respectively anchored onto 31 pseudochromosomes (2n = 62; Additional file 1: Table S7). Notably, the updated version of the plateau zokor genome (v3.1) expanded the previously published genome size by nearly 800 Mb compared to the v3.0 release (Additional file 2: Fig. S2B). This newer assembly version exhibited an increase in the number of ultra-long contigs (greater than 30 Mb) (Additional file 2: Fig. S2C), further enhancing assembly continuity and reliability in the genomic analysis of high-altitude adaptation.

The plateau zokor presents a noteworthy case of significant genomic expansion, measuring 200 Mb to 350 Mb larger than its zokor relatives, despite utilizing similar long-read sequencing techniques for genome assembly (Additional file 2: Fig. S1D). This was corroborated by *k*-mer analysis, which confirmed the unbiased nature of the genomic expansion (Additional file 1: Table S8). Besides, in comparison to the previously released chromosome-level genome of *E. fontanierii* (Chinese zokor, 500–1700 m) [8], which was also assembled using long HiFi reads, the plateau zokor genome exhibits a substantial expansion of approximately 550 Mb (Additional file 2: Fig. S1D).

We identified 25,137, 25,323, 24,620, and 25,271 protein-coding genes in the plateau zokor, Qinling zokor, Gansu zokor, and Transbaikal zokor, respectively. Functional annotation was achieved for 98.48%, 98.40%, 98.82%, and 98.69% of these genes across at least one database (Additional file 1: Tables S9-S14). Despite the comparability in the number of protein-coding genes among the four zokor species, the plateau zokor stands out due to a notable increase in repetitive sequences. Repetitive DNA constitutes 54.60% of its genome, a significant contrast to the 37.00%, 43.01%, and 44.13% found in the Qinling, Gansu, and Transbaikal zokors, respectively (Additional file 1: Table S15). This suggests that the amplification of repetitive sequences may be a key contributor to the plateau zokor's genomic enlargement.

Distinct tandem repeat (pzTR) expansion in plateau zokor

Large-scale expansion of repetitive sequences with atypical nucleotide composition can alter the overall GC distribution of genomic reads [21, 22]. To determine whether this large-scale expansion is specific to the plateau zokor, we selected five additional zokor species with publicly available Illumina short reads [4], in addition to three zokor species sequenced concurrently with the plateau zokor, to examine their GC distribution

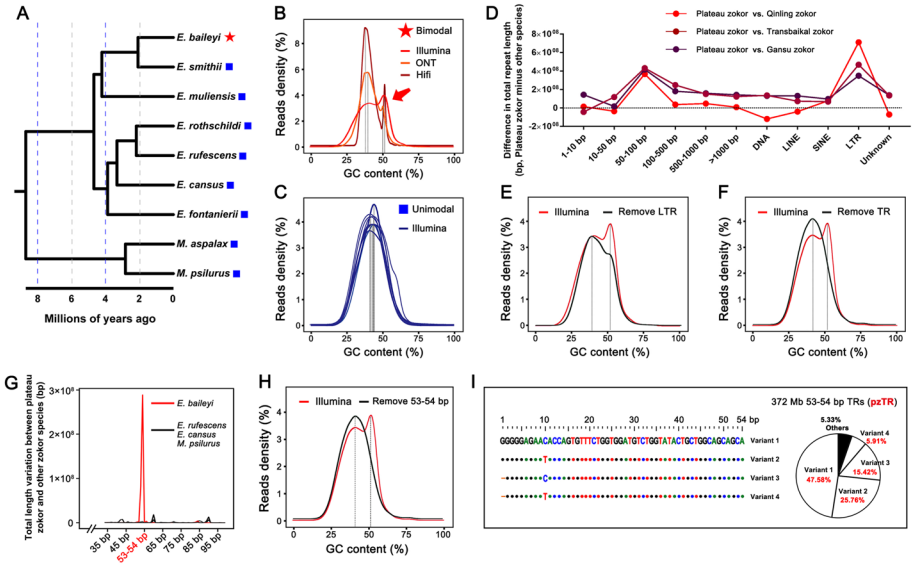

Fig. 2 Identification of pzTR in the plateau zokor. **A** Phylogenetic relationships of nine zokor species used to analyze the GC distribution patterns of sequencing reads, with estimated divergence times indicated at the corresponding nodes. Divergence times for *Eospalax* species were taken from Zhang et al. [4], and those for *Myospalax* species were taken from Wan et al. [23]. **B** GC distribution patterns of Illumina short reads, ONT, and HiFi long reads in the plateau zokor. **C** GC distribution patterns of Illumina short reads in eight zokor species, excluding the plateau zokor. **D** Differences in the total lengths of repeat classes between the plateau zokor and the other three zokor species. The first six categories are tandem repeats (TRs) grouped by unit length, and the remaining categories are transposable element (TE) classes. Values are shown as plateau zokor minus the comparison species. **E** GC distribution patterns of Illumina short reads in the plateau zokor after removal of LTRs. **F** GC distribution patterns of Illumina short reads in the plateau zokor after removal of 50–100 bp TRs. **G** Comparison of the total length of each repeat unit within the 50–100 bp TRs across the plateau zokor and the other three zokor species sequenced in this study. **H** GC distribution patterns of Illumina short reads in the plateau zokor after removal of 53–54 bp TRs. **I** Classification of homologous variants of pzTR in the plateau zokor genome and their proportions in the 53–54 bp TRs

profiles in a phylogenetic framework with estimated divergence times (Fig. 2A). Our findings show that only the plateau zokor genome exhibits the anticipated bimodal GC distribution, while the other eight zokor species present a normal unimodal distribution (Fig. 2B, C). Moreover, to eliminate any potential biases associated with sequencing platforms, we analyzed the GC content distributions of long reads obtained through ONT and HiFi sequencing technologies in the plateau zokor. Both datasets confirmed a consistent bimodal distribution, with two major peaks centered at approximately 41% and 52% GC content, and with peak positions aligning precisely with those observed in the Illumina short read data (Fig. 2B). These results confirm that the extensive expansion of repetitive sequences is a unique characteristic of the plateau zokor genome.

To determine which classes of repetitive sequences contributed to the genome expansion in the plateau zokor, we conducted a comprehensive classification of repetitive sequences across the assembled genomes of four zokor species. TRs were categorized into six classes based on repeat unit length, while TEs were classified into five types: DNA, LINE, SINE, LTR, and Unknown. Comparative analysis revealed that the plateau zokor exhibited significantly higher total lengths of TRs with repeat units of 50–100 bp and LTR-type TEs compared to the other zokor species (Fig. 2D, Additional file 2: Fig. S3).

To confirm these findings, we systematically removed short reads containing these two types of repetitive sequences and reassessed the GC distribution of the remaining genomic short reads. After removing LTRs, the GC distribution remained bimodal, although the profile was altered (Fig. 2E), whereas removal of 50–100 bp TRs shifted the GC distribution to a unimodal pattern, consistent with other zokor species (Fig. 2F). Additional analyses suggested that the apparent LTR-associated high-GC signal mainly reflected overlap with pzTR-rich regions rather than an independent contribution from LTRs (Additional file 2: Fig. S4). Further removal of other repetitive sequences did not produce similar effects. These results suggest that the expansion of 50–100 bp TRs is primarily responsible for both the genomic expansion and the bimodal GC distribution in the plateau zokor.

In our detailed comparison of TRs within the 50–100 bp range in the plateau zokor, we observed that the expanded repeats predominantly consist of 53–54 bp units (Fig. 2G). Notably, the removal of these 53–54 bp TRs from the plateau zokor short reads also resulted in a unimodal GC distribution, similar to that observed in other zokor species (Fig. 2H). This prompted us to investigate the interrelationship among these expanding 53–54 bp TRs through homology analysis. To conduct this analysis, we identify tandem repeats with identical motifs by extracting consensus sequences from the TRF output. Repeats were classified as having identical motifs if one consensus sequence, matching in length, was found within the duplicated sequence of another (or its reverse complement) (see Methods for details). Surprisingly, we discovered that the 53–54 bp TRs were primarily composed of four highly homologous variants, which together accounted for approximately 95% of all such TRs in the plateau zokor. These variants differed by only 1–2 base pairs (Fig. 2I). Consistent with our previous findings, removal of these four highly homologous TR variants also resulted in a shift toward a unimodal GC distribution. The cumulative length of these four TR variants was approximately 372 Mb, which strongly correlates with the observed genomic expansion in the plateau zokor compared to other zokor species. Thus, our findings suggest that the expansion of these homologous 53–54 bp TRs, referred to as plateau zokor specific tandem repeats (pzTR), is a decisive factor driving both genome expansion and the bimodal GC distribution pattern in the plateau zokor. Consistent with this conclusion, TAREAN analysis identified the most abundant satellite sequence in the plateau zokor genome, accounting for approximately 11% of the genome, and its consensus sequence was identical to pzTR variant 1 (Additional file 3: Table S16).

We further compared pzTR against tandem repeat consensus sequences extracted from other published chromosome-level zokor genomes [8, 23, 24], but detected no homologous tandem repeats in any of these species. Thus, given the estimated divergence times among zokor species (Fig. 2A), the high sequence homogeneity of pzTR is more consistent with recent amplification than with long-term sequence conservation. In the reference assembly, pzTR was detected on approximately 70% of anchored chromosomes and was mainly located near chromosome ends (Additional file 2: Fig. S5A). However, only a small proportion of total pzTR was anchored to chromosomes, whereas most copies were located on unanchored sequences (Additional file 2: Fig. S5B, S5C), indicating that pzTR-rich regions remain difficult to fully resolve and place in the current assembly.

The pzTRs we identified are notably GC-rich, with a GC content exceeding 53%. Recent mammalian telomere-to-telomere (T2T) assemblies have revealed diverse centromeric repeat families, and some centromeric repeat units identified in other mammals, such as sheep and pig, also appear to be relatively GC-rich [25, 26]. Furthermore, pzTRs are devoid of CpG sites, thereby eliminating potentials for high methylation level, a characteristic they share with active centromeric repeats, which also exhibit low methylation level [27]. These unique features suggest that pzTRs may not only contribute to genome expansion but could also serve a structural role, potentially in the centromeric regions of plateau zokor chromosomes.

Distribution of pzTR on the chromosomes of plateau zokor

To confirm the presence of pzTR in plateau zokor tissues, we designed probes based on variant 1, which predominates in pzTR, and conducted fluorescence in situ hybridization (FISH). Our analysis revealed a widespread and substantial distribution of pzTR in skin tissues isolated from the plateau zokor (Fig. 3A). Building on the hypothesis that pzTRs may play a structural role in chromosome stability, particularly in centromeric

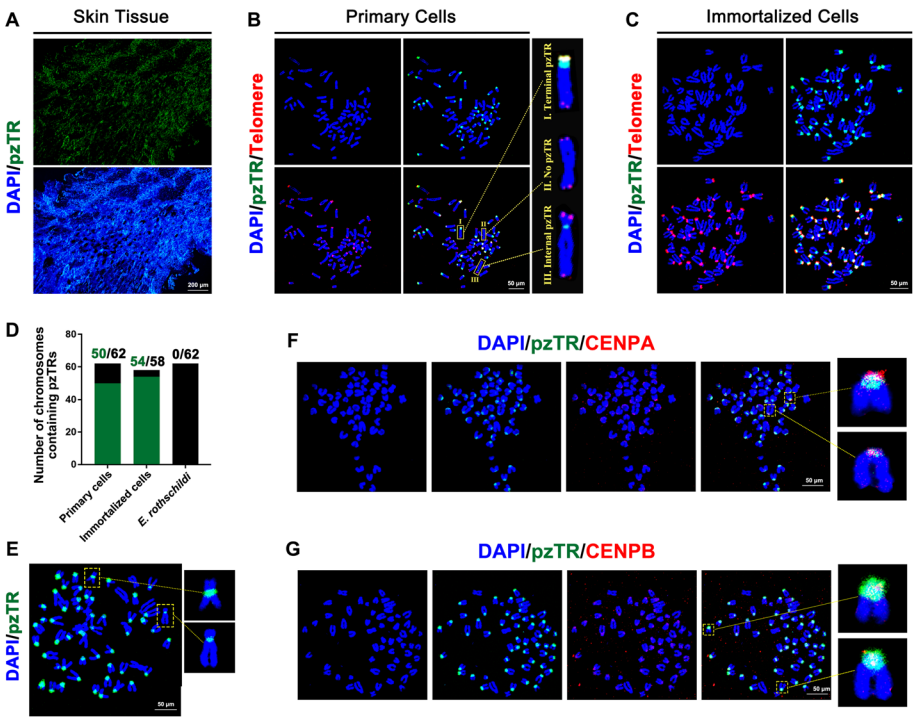

Fig. 3 Distribution of pzTR on the chromosomes of plateau zokor. **A** Fluorescence in situ hybridization (FISH) of pzTR in the skin tissue of the plateau zokor. **B** FISH of pzTR and telomeres in primary fibroblasts of the plateau zokor. Enlarged representative chromosomes illustrate three signal patterns: I, terminal pzTR, with pzTR localized near one chromosome end; II, no pzTR, with no detectable pzTR signal; and III, internal pzTR, with pzTR localized at a more internal position. **C** FISH of pzTR and telomeres in immortalized fibroblasts of the plateau zokor. **D** Quantification of the number of chromosomes containing pzTRs in primary fibroblasts and immortalized fibroblasts of the plateau zokor, and fibroblasts from *E. rothschildi*. **E** Confirmation of pzTR distribution on chromosomes based on the positions of centromere constrictions in M-type and SM-type chromosomes of the plateau zokor. **F** Co-localization of pzTR with centromere protein CENP-A in immortalized fibroblasts of the plateau zokor. **G** Co-localization of pzTR with centromere protein CENP-B in immortalized fibroblasts of the plateau zokor

regions, we aimed to further investigate their chromosomal localization within the plateau zokor genome. To this end, we performed FISH analysis on pzTR and telomeres in both primary and spontaneously immortalized fibroblasts at metaphase (see Methods for details). Enlarged representative chromosomes showed three signal patterns: chromosomes with pzTR localized near one chromosome end (terminal pzTR), chromosomes with pzTR at a more internal position (internal pzTR), and chromosomes lacking detectable pzTR signal. Among pzTR-positive chromosomes, pzTR was typically concentrated at a single locus located near one end of the chromosome (Fig. 3B, C). This one-locus-per-chromosome pattern is consistent with pzTR being associated with the centromeric region rather than representing a broadly dispersed tandem repeat family. Notably, in chromosomes carrying pzTR, the telomeric signal was often stronger at the pzTR-adjacent end than at the non-pzTR end. This asymmetric pattern was observed in both primary and immortalized fibroblasts and was further supported by quantitative analysis of telomeric fluorescence intensity (Additional file 2: Fig. S6). Besides, pzTR was detected on the majority of chromosomes in both cell types, exceeding 80% across all chromosomes, while it was absent in the chromosomes of *E. rothschildi* (Fig. 3D, Additional file 2: Fig. S7).

The FISH results suggest that pzTR may co-localize with telomeres at one end of the plateau zokor chromosomes. However, unlike the predominantly metacentric (M-type) or submetacentric (SM-type) chromosomes found in most zokor species, the chromosomes of the plateau zokor are primarily acrocentric (A-type) or subtelocentric (ST-type) (Additional file 2: Fig. S8), raising the possibility of co-localization with centromeres as well. To verify this, we identified a limited number of metacentric chromosomes in the plateau zokor, which have clearer centromeric positions (primary constrictions) compared to the acrocentric chromosomes. Using the positions of the primary constrictions of these chromosomes, we preliminarily assessed centromeric distribution and examined for overlap with pzTR. The results confirmed a high degree of overlap between pzTR and the locations of the primary constrictions (Fig. 3E). To further validate these findings, we co-stained pzTR with centromere proteins CENP-A and CENP-B, which are well-established markers that specifically localize to centromeric regions and are essential for proper chromosome segregation during cell division [28, 29]. Co-localization of pzTR with CENP-A and CENP-B signals was observed on some chromosomes (Fig. 3F, G), providing additional support for the association of pzTR with centromeric regions.

Impact of pzTR on chromosomal stability in plateau zokor

Given the remarkable conservation, high abundance, and preferential distribution of pzTR in the centromeric regions, we hypothesize that pzTR plays a significant role in maintaining chromosomal and genomic stability in the plateau zokor. To test this hypothesis, we employed CRISPR-Cas9 technology, utilizing five single-guide RNAs (sgRNAs) to deplete pzTR in spontaneously immortalized plateau zokor fibroblasts (Fig. 4A, B). The efficiency of the pzTR depletion was confirmed through FISH experiments at both the cellular and chromosomal levels, revealing a substantial reduction in pzTR levels in the depletion group compared to the wild-type (WT) group (Fig. 4C-E).

Following the establishment of pzTR-depleted cell lines, we assessed the proliferative capacity of both pzTR-depleted and WT cells to determine the impact of pzTR depletion

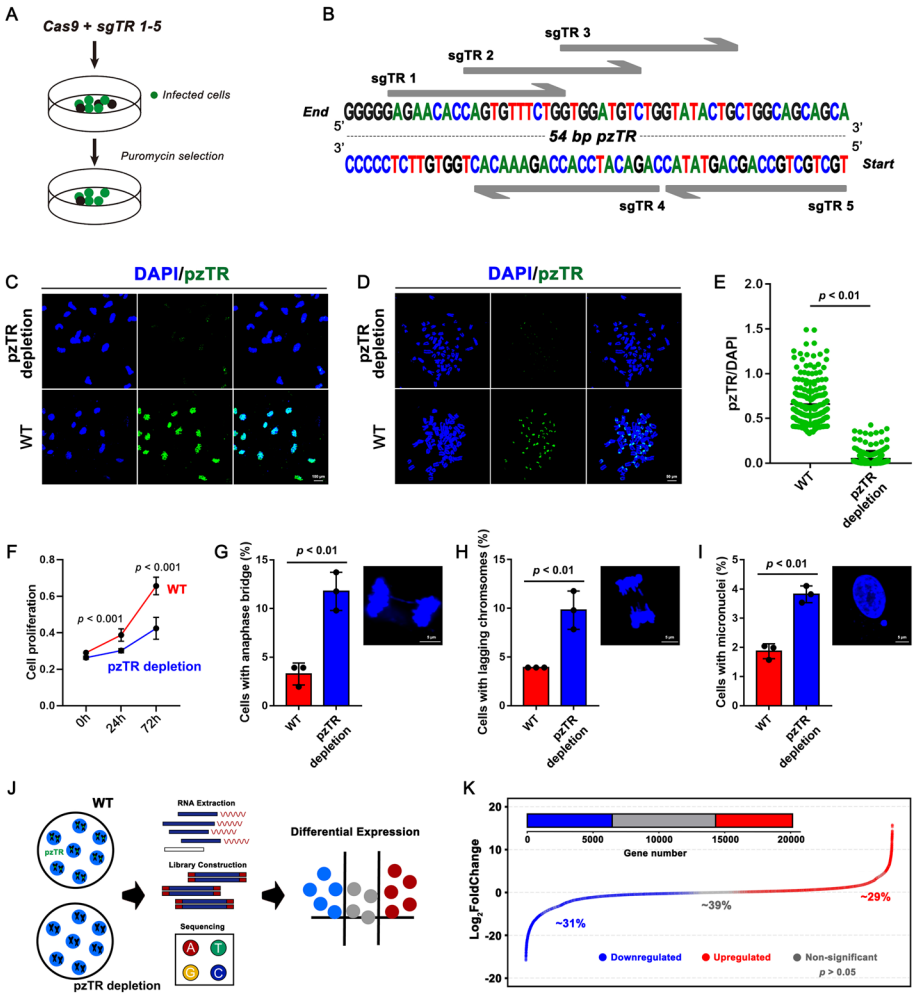

Fig. 4 Impact of pzTR on chromosomal stability in plateau zokor. **A** Flowchart of the CRISPR-Cas9-mediated pzTR depletion strategy. **B** Detailed information on the five single-guide RNAs used for targeting pzTR depletion. **C** FISH-based evaluation of the efficiency of pzTR depletion at the cellular level. **D** FISH-based evaluation of the efficiency of pzTR depletion at the chromosomal level. **E** Statistical analysis of pzTR content in pzTR-depleted and wild-type (WT) fibroblasts based on cellular FISH results. **F** Statistical comparison of proliferation rates between pzTR-depleted and WT fibroblasts. **G-I** Statistical comparison of anaphase bridges, chromosome analysis, and micronuclei formation in pzTR-depleted and WT fibroblasts during mitosis. **J** Workflow for transcriptome sequencing of pzTR-depleted and WT fibroblasts. **K** More than 60% of coding genes in pzTR-depleted fibroblasts exhibited significant differential expression compared to WT fibroblasts. Significant differences are indicated in the figure ($p < 0.01$; unpaired two-tailed Student's *t* test)

on normal cell proliferation. Our results indicated a significant reduction in the proliferative ability of pzTR-depleted cells compared to WT cells (Fig. 4F). To elucidate the underlying mechanisms behind this reduced proliferative capacity, we investigated whether pzTR influences genomic or chromosomal stability. We quantitatively assessed three pivotal indicators of chromosomal instability: anaphase bridges, lagging chromosomes, and micronuclei across different stages of cell division. All three indicators were significantly elevated in pzTR-depleted cells compared to WT cells, consistently indicating a detrimental effect on chromosomal stability following pzTR depletion (Fig. 4G-I). Consistently, depletion of the centromeric satellite in laboratory rat (*Rattus norvegicus*) fibroblast cells also reduced cell proliferation and increased the frequencies of anaphase

bridges, lagging chromosomes, and micronuclei (Additional file 2: Fig. S9), suggesting that the role of centromeric satellite sequences in maintaining chromosomal stability is evolutionarily conserved rather than specific to plateau zokor pzTR. Additionally, we performed transcriptome sequencing on both pzTR-depleted and WT cells to explore the impact of pzTR depletion on gene expression stability in plateau zokor cells (Fig. 4J). The results revealed that over 60% of gene expressions in pzTR-depleted cells exhibited significant fluctuations compared to WT cells, with nearly 31% significantly downregulated and 29% significantly upregulated ($p < 0.05$, Fig. 4K, Additional file 3: Table S17). This suggests that pzTR depletion substantially disrupts existing gene expression patterns within plateau zokor cells. Collectively, these findings highlight the indispensable role of pzTR in maintaining chromosomal stability and regulating gene expression in plateau zokor cells.

Positive contribution of pzTR to chromosomal stability under hypoxic conditions

In comparison with other zokor species that lack pzTR expansion in their genomes, the plateau zokor occupies an extremely high-altitude distribution unique to the species, while also living almost exclusively in closed or semi-closed underground burrows [7]. This combination of ecological factors likely drives the plateau zokor to develop a range of unique adaptive strategies for surviving in low-oxygen environments. Having observed a positive contribution of pzTR to chromosomal stability in the plateau zokor, we hypothesized that pzTR may also play a role in the plateau zokor's extreme hypoxic adaptation.

To test whether hypoxia affects genomic stability in plateau zokor fibroblasts, we first quantified the overall micronucleus rate in plateau zokor-WT cells under normoxic and extreme hypoxic (0.3% O_2) conditions. The percentage of cells with micronuclei increased significantly under hypoxia, from approximately 1.7% under normoxia to approximately 2.4% under hypoxia (Additional file 2: Fig. S10). We then examined the frequency of pzTR-positive micronuclei under the same conditions. Micronuclei arise from chromosomal instability during cell division and typically contain mis-segregated chromosomes. If pzTR contributes to chromosomal stability under hypoxic conditions, we would expect chromosomes containing pzTR to exhibit greater stability, while those lacking pzTR would be more prone to mis-segregation, leading to micronuclei formation. Given that over 80% of plateau zokor chromosomes are pzTR-positive (Fig. 3D), the proportion of pzTR-positive micronuclei should align closely with this figure, around 80%, if pzTR has a neutral effect on micronuclei formation.

Under normoxic conditions, however, we observed that the proportion of pzTR-positive micronuclei was approximately 60%, which was significantly lower than the expected value of 80% based on the proportion of pzTR-positive chromosomes ($p < 0.01$). This discrepancy became even more pronounced under hypoxic conditions, where the proportion of pzTR-positive micronuclei further declined to between 40 and 50%, concurrent with an increase in pzTR-negative micronuclei (Fig. 5A, B). The significant reduction of pzTR-positive micronuclei in hypoxia, coupled with the rise of pzTR-negative micronuclei, suggests that chromosomes harboring pzTR manifest enhanced stability in low-oxygen environments. These findings support the hypothesis that pzTR may play a role in

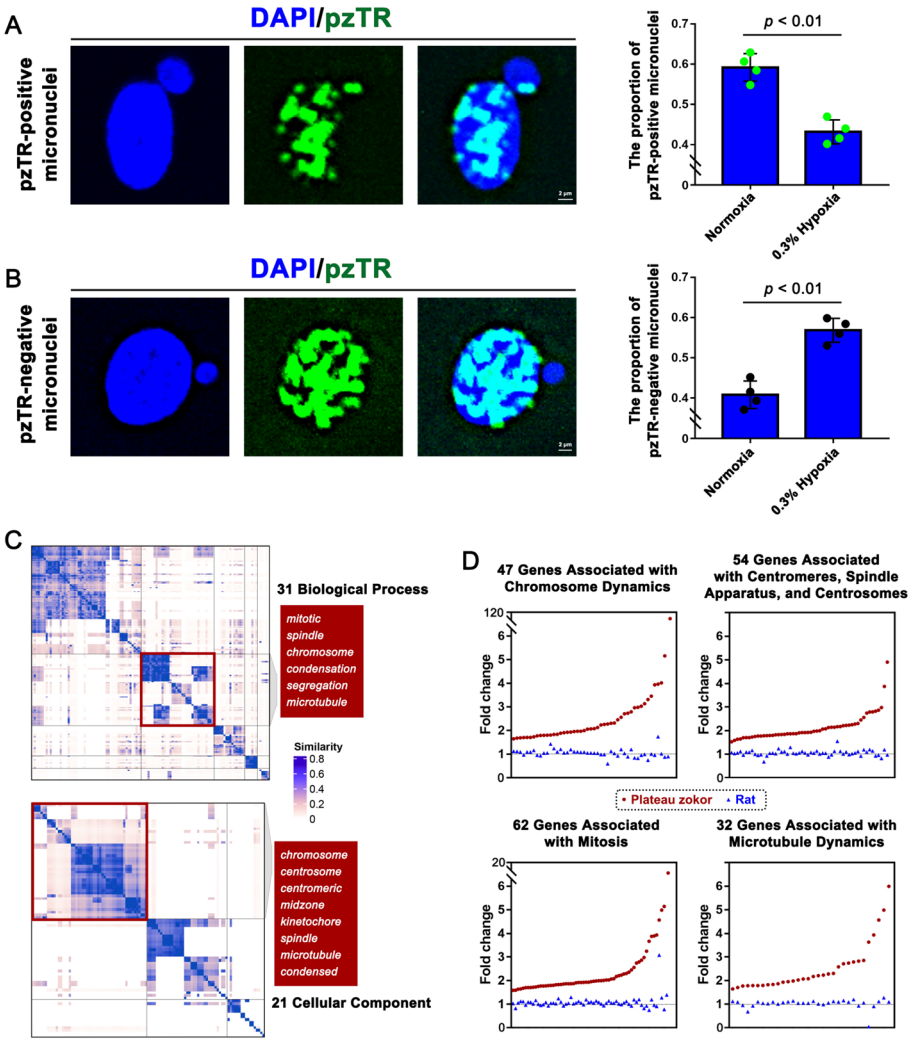

Fig. 5 Positive contribution of pzTR to chromosome stability under hypoxic conditions. **A**, **B** Statistical analysis of the positive and negative rates of pzTR in micronuclei formed during mitosis of plateau zokor fibroblasts under ambient oxygen (21%) and hypoxia (0.3%). **C** Under hypoxic (0.3% O_2) conditions, a large number of differentially expressed genes related to mitosis (such as those associated with centromeres, centrosomes, spindles, and kinetochores) were identified in plateau zokor fibroblasts. The results represent semantic enrichment of GO terms for these differentially expressed genes, with key terms highlighted in red boxes on the right-hand side. **D** In contrast to ambient oxygen conditions, most of the differentially expressed genes related to mitosis under hypoxia in plateau zokor fibroblasts showed no significant differential expression in rat fibroblasts. Differentially expressed genes were classified according to GO terms. Significant differences are indicated in the figure ($p < 0.01$; unpaired two-tailed Student's *t* test)

mitigating chromosomal damage and preserving mitotic integrity in the plateau zokor, particularly under extreme hypoxic stress.

To further investigate whether the maintenance of centromere stability or other mitotic processes contributes to the plateau zokor's unique adaptation to extreme hypoxia, we performed transcriptomic sequencing on plateau zokor-WT and rat fibroblasts (RAT-WT) exposed to normoxic and hypoxic (0.3% O_2) conditions for 48 h. Although both species showed a similar number of differentially expressed genes (565 in plateau zokor and 457 in rat, with $|\text{Log}_2(\text{Fold change})| > 0.5$, adj. $p < 0.05$), yet only 22

genes were commonly differentially expressed (Additional file 3: Table S18). Gene Ontology (GO) enrichment analysis revealed that, under hypoxia, plateau zokor-WT cells showed significant upregulation of genes involved in chromosome dynamics, spindle assembly, centrosome regulation, and mitosis (Fig. 5C, Additional file 3: Table S19). In stark contrast, these genes were largely unchanged in RAT-WT cells, indicating a distinct hypoxic response in the plateau zokor (Fig. 5D, Additional file 3: Table S20). These findings underscore the critical role of pzTR in maintaining chromosomal stability under hypoxia and highlight the importance of mitotic integrity in the plateau zokor's adaptation to low-oxygen environments.

Discussion

Comparative genomics has traditionally emphasized protein-coding genes and regulatory elements in the study of adaptive evolution [30–32]. However, the functional significance of repetitive elements in this context has been insufficiently addressed. These elements, dispersed across genomes and interacting with other genomic features, present significant challenges in identification and assembly [33]. While short-read sequencing technologies have been pivotal in genomic analysis, they are inherently limited in resolving repetitive regions. In contrast, the advent of long-read sequencing platforms has proven invaluable, providing higher resolution and yielding critical insights into the variability of repetitive sequences and the evolution of essential genomic structures, such as centromeres [14, 34, 35].

In this study, we expand the understanding of TRs in mammalian adaptive evolution by leveraging high-quality ONT and HiFi data. Specifically, we show that the expansion of pzTRs, which are likely associated with centromeric regions in the plateau zokor genome, is closely linked to the genomic stability required by its hypoxic environment. Notably, homologous pzTRs are absent in other zokor species, indicating that their presence in plateau zokors likely arose from a recent and rapid expansion event. The formation and fixation of these putative centromere-associated pzTRs can be explained by the centromere drive model. In this model, novel centromeric variants preferentially orient toward the egg rather than the polar body during asymmetric female meiosis, resulting in non-Mendelian inheritance [36, 37]. These variants are hypothesized to exhibit increased strength, driven by enhanced kinetochore protein-binding capabilities and elevated repeat abundance [38, 39]. Although direct evidence for increased kinetochore protein binding is lacking, the high conservation and substantial abundance of pzTRs suggest they contribute to a stronger centromeric variant. The centromere drive model traditionally posits that stronger centromeric variants act as selfish elements without conferring adaptive benefit [40]. However, a recent study in humans have demonstrated a negative correlation between centromere strength and chromosome mis-segregation rates during mitosis [41]. In plateau zokors, chromosomes containing pzTRs exhibit significantly lower mis-segregation frequencies under normoxic conditions and even further reductions under hypoxic conditions. These findings suggest that, beyond their role in the centromere drive model, pzTR-containing chromosomes may enhance chromosome segregation fidelity, potentially offering an adaptive advantage in plateau zokors. Thus, our findings provide new evidence for the role of centromeric TR expansion in the

adaptive evolution of mammals, offering insights into strategies for coping with extreme environments.

The evolution of the centromere is often associated with chromosomal rearrangements, which play a critical role in establishing reproductive isolation and speciation [42, 43]. In our study, the contribution of pzTR expansion to speciation of plateau zokor remains an open question. Notably, the chromosomal karyotype of the plateau zokor is predominantly of the A or ST type, which contrasts with the M or SM types found in other zokor species. This difference in karyotype may influence the dynamics of centromeric sequence exchange among nonhomologous chromosomes, potentially facilitating the spread and conservation of pzTRs within the genome [44, 45]. The association between centromere drive, karyotype variation, and the expansion of pzTRs raises the possibility that these genomic features could be linked to processes of reproductive isolation and speciation. Therefore, future investigations that integrate karyotype variation with pzTR analysis could provide valuable insights into whether these genomic phenomena contribute to the speciation events of the plateau zokor.

Despite providing valuable insights into the role of centromeric pzTRs in the adaptive evolution of the plateau zokor, our study has several limitations. First, although our cytological results support the association of pzTR with centromeric regions, the mouse antibodies used to detect centromeric proteins such as CENP-A and CENP-B exhibited limited specificity, which constrains our capacity to thoroughly explore the molecular mechanisms governing pzTR functionality. Second, while our CRISPR-Cas9 experiments suggest that pzTRs contribute to chromosomal stability, the precise molecular pathways remain to be elucidated. We postulate that pzTRs may modulate the binding of centromere-associated proteins, such as CENP-A or CENP-B, in a manner similar to the role of centromeric satellite DNA in other species. However, studying these processes in the extreme hypoxic environment of the plateau zokor presents considerable technical challenges. Finally, the driving forces behind the expansion of pzTRs, whether primarily influenced by hypoxic selective pressures or indicative of a broader evolutionary trend, remain unclear. Future research investigating the ecological factors affecting pzTR copy number variation across different altitudes will be crucial for elucidating the role of TR expansion in adaptive evolution.

Conclusions

By utilizing high-quality long-read sequencing to assemble chromosome-level genomes, this study reveals the adaptive role of tandem repeats (pzTRs) in the genome of the plateau zokor (*Eospalax baileyi*). We show that pzTRs contribute to genome expansion, a unique bimodal GC content distribution, and chromosome stability through their association with centromeric regions, especially under hypoxic conditions at high altitudes. We provide new insights into the functional significance of repetitive sequences, which have often been overlooked in evolutionary genomics. This research not only challenges traditional views of repetitive sequences as neutral elements but also offers a novel perspective on the role of non-coding regions in evolutionary adaptation. Furthermore, our findings provide new insights into the centromere drive model, shedding light on the selective pressures shaping centromere evolution. Together, these contributions significantly advance our understanding of genome evolution and adaptation in extreme environments.

Methods

Animals

Four zokor species (*Eospalax* and *Myospalax*) were live-trapped between 2019 and 2020 in China: *E. baileyi* (plateau zokor) from the Laji Mountains, Qinghai Province; *E. rufescens* (Qinling zokor) and *E. cansus* (Gansu zokor) from the Qinling Mountains, Shanxi Province; and *M. psilurus* (Transbaikal zokor) from the outskirts of Kaifeng City, Henan Province. All subjects were male. Following euthanasia under mass anesthesia, fresh hearts, livers, brains, spleens, lungs, kidneys, testicles, and thigh muscles were collected and immediately frozen in liquid nitrogen before storage at −80 ℃. Ethical approval for these experiments was obtained from the Life Science Ethics Review Committee of Zhengzhou University (ZZUIRB-2022–44) and the Ethics Committee of the Kunming Institute of Zoology, Chinese Academy of Sciences (IACUC-RE-2022–08–014). The distribution map of each species was generated using the R packages maps, raster, and ggplot2 [46], with 95% confidence interval ellipses added to the scatter points for each species using the stat_ellipse function.

Genome sequencing

Genomic DNA was extracted from muscle tissues of four zokors using the sodium dodecyl sulfate (SDS) method. The concentration, purity, and integrity of the genomic DNA were evaluated via 1% agarose gel electrophoresis (180 V for 20 min), 1% pulsed-field gel electrophoresis (17 h), Qubit 3.0 fluorometer, and NanoDrop 2000c spectrophotometer.

For the preparation and sequencing of small fragment libraries, DNA was subjected to random fragmentation, end repair, A-tail addition, adapter ligation, purification, and PCR amplification, producing libraries with a 350 bp insert size. These libraries were sequenced on the Illumina HiSeq X Ten and NovaSeq 6000 platforms. Raw sequencing data were filtered to remove reads with adapter sequences, more than 10% N bases, or over 20% low-quality bases [47].

Simultaneously, Oxford Nanopore Technology (ONT) libraries were prepared by repairing DNA damage, adding A-tails, ligating adapters, and constructing 20 kbp libraries. Sequencing was conducted on the PromethION 48 platform (R9.4), with basecalling performed using Guppy 5.0.122 (https://nanoporetech.com/) to convert fast5 files to fastq format, followed by removal of reads with low quality or errors.

For high fidelity (HiFi) reads, DNA from plateau zokor and Transbaikal zokor underwent damage repair, end repair, adapter ligation, and purification to generate 20 kbp SMRT bell libraries. This method was chosen specifically for these species due to challenges in genome assembly. Sequencing was performed on the PacBio Sequel II platform, and the resulting data were analyzed with the CCS module in SMRTlink v13.1 (https://www.pacb.com/support/software-downloads/) to produce high-precision HiFi reads for further analysis.

Genome size estimation and chromosomal assembly

Based on *k*-mers analysis, short paired-end Illumina reads were utilized to estimate the genomic characteristics of four zokors, including genome size, heterozygosity ratio, and repeat rate. We employed Kmerfreq program in GCE-1.0.2 [48] to estimate genome size

(Genome size = *k*-mer number/*k*-mer depth) by analyzing the frequency and depth of 17-mers, and revised the genome size by correcting errors from incorrect *k*-mers.

To optimize genome assembly for each zokor species, we adopted distinct assembly strategies. For Qinling and Gansu zokors, initial contigs v1 were assembled from long ONT reads using Canu v1.9 [49]. These contigs were then polished through three rounds with NextPolish v1.3.1 [50], utilizing both long ONT and short paired-end Illumina reads. Chromosome anchoring was performed with ALLHiC v0.9.8 [51], resulting in assembly v1. Subsequently, long ONT reads were mapped to assembly v1 with Minimap2 v2.20 [52], followed by local assembly with NextDenovo v2.4.0 (https://github.com/Nextomics/NextDenovo) to produce contigs v2. These contigs were further polished with NextPolish v1.3.1, and the final chromosome-level assembly v2 was generated using ALLHiC v0.9.8.

For plateau and Transbaikal zokors, the initial contig sizes obtained with Canu v1.9 from long ONT reads were substantially smaller than anticipated. Consequently, we employed a HiFi sequencing strategy. Contigs v1 were assembled from long HiFi reads using Hifiasm 0.13-r308 [53], and chromosome anchoring was performed with 3D-DNA v180922 [54] using short paired-end Hi-C reads to produce assembly v1. Long ONT reads were mapped to this assembly with Minimap2 v2.20, followed by local assembly with NextDenovo v2.4.0 to generate contigs v2. Gaps were closed using assembly v1, and contigs v2 were polished through three rounds with NextPolish v1.3.1 using long HiFi and short paired-end Illumina reads. Finally, chromosome-level assembly v2 was achieved with ALLHiC v0.9.8, based on short paired-end Hi-C reads.

Genome assessment

The completeness of the assembled genomes for the four zokor species was assessed using Benchmarking Universal Single-Copy Orthologs (BUSCO v4.1.2; http://busco.ezlab.org/) [55] and the Core Eukaryotic Genes Mapping Approach (CEGMA v2.5; http://korflab.ucdavis.edu/datasets/cegma/) [56]. BUSCO analysis was conducted against the mammalian single-copy orthologs dataset (mammalia_odb10, $n = 9226$), while CEGMA was used to evaluate the presence of conserved genes across eukaryotic model organisms ($n = 248$). To further validate the genome assemblies, short paired-end Illumina reads were aligned to the assembled genomes using BWA-MEM 0.7.8 [57]. SNP calling was subsequently performed on the BWA alignment results with Samtools 0.1.19 [58]. The consistency of the genome assemblies was evaluated based on read mapping rate, coverage, and the ratio of homology SNPs.

Transcriptome sequencing

We generated transcriptomic data for four zokor species from eight different tissues: heart, liver, brain, spleen, lung, kidney, muscle, and testis, to assist with gene prediction. All tissues were obtained from the same adult male specimen used for whole-genome DNA sequencing. After harvest, the tissues were rapidly frozen in liquid nitrogen and stored at −80 ℃. RNA was extracted from the tissues using Trizol (Invitrogen) reagent according to the manufacturer's instructions. The concentration,

purity, and integrity of the RNA were assessed using 1% agarose gel electrophoresis (180 V, 16 min), an Agilent 2100 Bioanalyzer, a Qubit 3.0 fluorometer, and a NanoDrop 2000c spectrophotometer. Following the preparation of the 350 bp cDNA library, paired-end sequencing was performed on the Illumina HiSeq X Ten platform.

Annotation of gene structure

For repeat annotation of the four zokor genomes, we employed both homology-based annotation and de novo prediction strategies. Homology-based annotation involved aligning the genomes to RepBase (http://www.girinst.org/repbase) using RepeatMasker v4.0.5 and RepeatProteinMask v4.0.5 (http://www.repeatmasker.org/), based on sequence similarity. De novo prediction was performed using LTR_FINDER v1.0.7 [59], RepeatScout v1.0.5 [60], and RepeatModeler v1.0.8 (http://www.repeatmasker.org) to construct a repeat sequence library, which was then used for prediction with RepeatMasker. Additionally, tandem repeat sequences were identified using TRF v4.07b [61]. For independent evaluation of LTR annotation in the plateau zokor genome, LTRs were also re-annotated using HiTE v3.3.3 with default parameters, except that the animal genome mode was specified [62].

For non-coding RNA annotation, miRNA and snRNA were predicted using INFERNAL [63], tRNA was predicted with tRNAscan-SE v1.3.1 [64], and rRNA was identified using the BLAST program against a mouse rRNA sequence library [65].

Coding gene structure annotation for the four zokor genomes was achieved through a combination of homology-based, ab initio, and transcriptome-based prediction methods. Homology-based prediction involved aligning the zokor genomes to the protein-coding sequences of five species: mouse (*Mus musculus*, GCA_000001635.8), plateau zokor (*Eospalax baileyi*, GWHABJZ00000000), Chinese hamster (*Cricetulus griseus*, GCF_000223135.1), naked mole-rat (*Heterocephalus glaber*, GCF_000247695.1), and blind mole-rat (*Nannospalax galili*, GCF_000622305.1) using Tblastn v2.2.26 (E-value $\leq$ 1e-5). The alignment results were integrated using Solar v0.9.6 [66], and structural prediction was performed with GeneWise v2.4.1 [67]. Ab initio predictions were conducted using five approaches: AUGUSTUS v3.2.3 [68], GlimmerHMM v3.0.4 [69], SNAP v2013.11.29 [70], geneid v1.4 [71], and GENSCAN v1.0 [72]. Transcriptome-based prediction was conducted using two complementary approaches. First, transcriptome sequencing data from eight tissues were aligned to the genome with TopHat v2.0.13 [73] to identify exon regions and splice sites, followed by gene structure prediction with Cufflinks v2.1.1 [74]. Second, a de novo assembly of the transcriptome data from the same eight tissues was performed using Trinity v2.1.1 [75], with subsequent gene structure prediction carried out using PASA v2.0.2 [76]. Integration of these structural annotations was achieved using EvidenceModeler (EVM) v1.1.1 [77], and the annotations were further refined with PASA, resulting in a comprehensive, non-redundant set of coding genes.

Functional annotation of protein-coding genes

To functionally annotate coding genes, we utilized two protein sequence databases: Swiss-Prot (http://web.expasy.org/docs/swiss-prot_guideline.html) and the NR Protein

Sequence Database (https://www.ncbi.nlm.nih.gov/protein). Protein domains were predicted using InterProScan v.4.8 and HMMER v.3.1 with reference to the InterPro v.32.0 (http://www.ebi.ac.uk/interpro/) and Pfam v.27.0 (https://pfam-legacy.xfam.org/) databases [78–81]. Both databases also provide access to Gene Ontology (GO) terms (http://geneontology.org/) [82] for functional annotation. Additionally, gene pathways were identified through BLAST searches against the KEGG database v.53 (http://www.kegg.jp/kegg/kegg1.html) [83] with an E-value threshold of 1e-5.

pzTR identification

Tandem repeats with identical motifs can exhibit different consensus sequences due to variations in starting positions. For instance, two tandem repeats with consensus sequences "ATCC" and "TCCA" share the same motif, with the second being a one-base pair shift from the first. To facilitate this analysis, we identify tandem repeats with identical motifs by extracting consensus sequences from the TRF output. If two consensus sequences matched in length and one was found within the duplicated sequence of the other (or its reverse complement), they were classified as having identical motifs. Based on this criterion, the major 53–54 bp tandem repeats in the plateau zokor genome were grouped into four highly homologous sequence variants (pzTRs).

To further identify pzTR-like repeats in the plateau zokor and other zokor genomes, we used pzTR variant 1 as the query sequence for similarity-based screening. Following the strategy of previous work on the inference of overall similarity between tandem repeat consensus sequences [22], TR consensus sequences were extracted from the TRF output of each species and used as subject sequences for comparison. Because the boundaries of tandem-repeat monomers are often arbitrary, related repeat units may align in a staggered manner and therefore show only partial similarity in a standard local alignment. To reduce this boundary effect, each subject consensus sequence was duplicated in tandem prior to sequence comparison, and the duplicated sequences were used to build a local BLASTN database. The pzTR sequence was then searched against this database using BLASTN. Because BLASTN reports local high-scoring segment pairs (HSPs), whereas our aim was to estimate overall similarity at the monomer level, all valid HSPs from the same query-subject pair were integrated rather than relying on a single best hit. Subject coordinates obtained from duplicated repeats were folded back to the original monomer coordinate system, and HSPs were stitched together along the longer-sequence backbone. Overlapping aligned regions were counted only once, whereas short non-overlapping HSPs were retained when they contributed additional unique coverage. Based on the stitched alignment, an overall similarity score was calculated for each query-subject pair. For the plateau zokor genome, TR consensus sequences showing overall similarity $\geq 80\%$ to pzTR were retained as candidate hits. In contrast, BLASTN returned no valid hits from any of the other zokor species, indicating that no TR consensus sequences with detectable similarity to pzTR were present in those genomes under this search strategy.

TAREAN analysis of candidate satellite repeats

To identify satellite repeats in plateau zokor and laboratory rat (*Rattus norvegicus*), we performed TAREAN analysis based on Illumina paired-end sequencing data [84, 85]. For

the plateau zokor, raw Illumina reads were obtained from the same individual used for genome sequencing in this study. For laboratory rat, publicly available Illumina sequencing data were downloaded from the NCBI Sequence Read Archive (SRA; accession SRR7755593). For each dataset, 2 million read pairs were randomly selected for analysis. The selected read pairs were analyzed using the RepeatExplorer2/TAREAN pipeline implemented on the public Galaxy server (https://repeatexplorer-elixir.cerit-sc.cz/galaxy/). Prior to clustering, read quality control and preprocessing were performed with the "Preprocessing of FASTQ paired-end reads" tool available on the same platform. TAREAN was then run with default settings, except that the queue option was set to "long".

Cell culture

Primary fibroblasts were isolated from the dorsal skin of plateau zokors captured in the Laji Mountains (3,700 m above sea level, Qinghai Province, China) and Rothschild's zokors (*Eospalax rothschildi*) from the Daba Mountains (1,100 m above sea level, Shaanxi Province, China). Laboratory rat fibroblast cell line (cat. no. CCC-REPF-1) was purchased from the Chinese Academy of Medical Sciences (Beijing, China); no additional authentication was performed in our laboratory. The cells were cultured in DMEM (Gibco, cat. no. 11965) supplemented with 10% fetal bovine serum (Gibco, cat. no. 10099141 C) and maintained in a humidified incubator at 37 ℃ with 5% CO_2. While most primary fibroblast cells underwent senescence after 5–7 population doublings, one fibroblast line derived from the plateau zokor exhibited notably rapid growth and achieved spontaneous immortalization. The identity of this cell line as plateau zokor fibroblast was confirmed through transcriptome sequencing. All primary fibroblasts and cell lines were regularly tested for mycoplasma contamination every month and were consistently found negative throughout the study.

Karyotyping, fluorescence in situ hybridization (FISH) and Immunofluorescence combined with FISH

For karyotyping analysis and FISH of metaphase cells, fibroblasts were cultured to approximately 80% confluency and subsequently treated with 0.1 μg/mL colchicine (Sigma, cat. no. C3915) for 4 h to induce metaphase arrest. The cells were then subjected to a hypotonic treatment by slowly adding 1 mL of pre-warmed hypotonic solution (0.35% KCl) to the resuspended cells and incubating at 37 °C for 30 min. Following this, the cell suspensions were fixed in fresh methanol acid (3:1) solution for 10 min and subsequently dropped onto glass slides. The slides were dehydrated through an ethanol series (70%, 80%, 100%; 2 min each) and air-dried. Karyotypes were analyzed after staining with Giemsa, as previously described [86]. Chromosome types were categorized based on arm ratio and classified as metacentric (M), submetacentric (SM), subtelocentric (ST), and acrocentric (A) [87].

For the preparation of pzTR and rat satellite I probes, oligonucleotides corresponding to the pzTR, rat satellite I were synthesized and labeled with FAM. Telomere sequences were synthesized and labeled with TAMRA. The sequences used were GGGGGAGAA CACCAGTGTTTCTGGTGGATGTCTGGTATACTGCTGGCAGCAGCA for pzTR, AGCACACTCATACAAGCATGTCCCATTGGGAACTCACTGAATTCGCCTAG for

rat satellite I and $(TTAGGG)_7$ for telomeres. Probes were diluted in hybridization buffer (20 mM Na_2HPO_4; 20 mM Tris; 60% formamide; 2 × saline sodium citrate (SSC)) and applied to the slides. The slides were then denatured at 85 °C for 5 min and incubated overnight in a humidified chamber at room temperature. Following hybridization, slides were washed twice in 2 × SSC/0.1% Tween-20 at 60 °C and once at room temperature for 10 min each.

For co-staining pzTR with CENP-A or CENP-B, after FISH were performed, the slides were fixed with 4% paraformaldehyde (PFA) for 20 min after hybridization and permeabilized with phosphate-buffered saline (PBS) containing 0.3% Triton X-100 for 15 min, then incubated with blocking buffer (PBS containing 1% bovine serum albumin) for 1 h at room temperature. Rabbit anti-CENP-A (Cell signaling technology, Cat. no. 2048, 1:500 for immunofluorescence) or mouse anti-CENP-B (Santa Cruz, Cat. no. sc-376283, 1:200 for immunofluorescence) primary antibody were added and incubated overnight at 4 °C. Following primary antibody incubation, slides were washed three times with PBST (PBS containing 0.2% Tween-20) and then incubated for 1 h at room temperature with the corresponding secondary antibodies: donkey anti-rabbit IgG conjugated to Alexa Fluor 555(Invitrogen, Cat. no. A32732, 1:1000 for immunofluorescence) or donkey anti-mouse IgG conjugated to Alexa Fluor 555 (Invitrogen, Cat. no. A31570, 1:1000 for immunofluorescence). Then, the slides were washed three times with PBST. After counterstaining with DAPI (4',6-diamidino-2-phenylindole), slides were examined using an Olympus FV1000 fluorescent microscope.

CRISPR-Cas9-mediated depletion of repetitive DNAs in plateau zokor and rat fibroblast

The guide RNA (gRNA) sequences targeting pzTR or rat satellite I were designed using the Broad Institute's online tool (https://portals.broadinstitute.org/gpp/public/) and cloned into the lentiCRISPR v2 plasmid (Addgene plasmid #52,961) via Esp3I digestion (ThermoFisher, cat. no. ER0451). Lentiviruses were produced by transfecting 293 T cells with the packaging plasmids pCMVΔ8.9 and pMD2.G at a ratio of 5:2.5:1 using Lipofectamine 3000 (Invitrogen, cat. no. L3000015). Forty-eight hours post-transfection, the medium containing lentiviruses was collected and concentrated using Centrifugal Filters (Millipore, cat. no. UFC910096). Lentiviruses were then added to plateau zokor or rat fibroblasts in the presence of 8 μg/ml Polybrene (Millipore, cat. no. TR-1003-G). Puromycin (5 μg/mL) selection began 72 h after infection. To assess the depletion efficiency, FISH for pzTR and rat satellite I were performed as previously described. The fluorescence intensity ratios of pzTR to DAPI and rat satellite I to DAPI were calculated using ImageJ, with more than 200 cells analyzed per group.

Cell proliferation assay

The cell proliferation assay was conducted following the manufacturer's protocol for the Enhanced Cell Counting Kit-8 (Beyotime, Cat. no. C0043). In brief, 5,000 cells were seeded per well in a 96-well plate. For each well, 10 μL of CCK-8 reagent was added to 100 μL of culture medium and gently mixed. The plate was incubated at 37 °C with 5% CO_2 for 2 h, and absorbance was measured at 450 nm using a microplate reader.

Micronucleus, anaphase bridges, and lagging chromosome analysis

Cells were seeded into 24-well plates containing round coverslips at a density of 1×10^5 cells per well and cultured under standard conditions for 24 h. The cells were then fixed with 4% formaldehyde in PBS for 30 min and washed three times with PBS, with each wash lasting 5 min. After fixation, the cells were counterstained with DAPI for 10 min, followed by three additional washes. Coverslips were subsequently removed and imaged using an Olympus FV1000 fluorescent microscope. For each sample, a minimum of fifty mitotic cells were analyzed to assess anaphase bridges and lagging chromosomes, while at least 300 nuclei were examined to identify micronuclei [88, 89]. Each experiment was conducted in triplicate.

Transcriptomic analysis

To explore the role of pzTR in maintaining chromosome stability and regulating gene expression in plateau zokor, we conducted RNA sequencing (RNA-seq) on spontaneously immortalized fibroblasts from wild-type (WT) ($n=4$) and pzTR-depleted ($n=3$) plateau zokors. Additionally, to examine whether centromere-related genes contribute to hypoxia adaptation in plateau zokor, we performed RNA-seq on fibroblasts from both WT ($n=3$) and hypoxic ($n=3$) groups of plateau zokors and rats. The hypoxic condition was induced by exposing cells to 0.3% oxygen for 48 h.

Raw sequencing reads were pre-processed to remove adapter sequences, reads containing more than 10% N bases, and those with over 50% low-quality bases, using fastp v0.23.4 with the following parameters: -q 5 -u 50 -n 15[47]. Clean reads were then aligned to the reference genomes of plateau zokor and rat using hisat2 v2.2.1 [90]. The reference genome for plateau zokor was generated in this study, while the rat genome was obtained from the Ensembl database (http://www.ensembl.org/index.html; Assembly: mRatBN7.2). Gene expression levels were quantified using StringTie v2.2.1 [91]. Differentially expressed genes (DEGs) were identified using DESeq2 v1.42.1 (https://bioconductor.org/packages/release/bioc/html/DESeq2.html) [92].

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13059-026-04159-y.

Additional file 1: Includes Tables S1, S2, S3, S4, S5, S7, S8, S9, S10, S11, S12, S13, S14, S15.

Additional file 2: Includes Figs. S1, S2, S3, S4, S5, S6, S7, S8, S9, S10.

Additional file 3: Includes Tables S6, S16, S17, S18, S19, S20.

Acknowledgements

We would like to thank the Core Technology Facility of Kunming Institute of Zoology(KIZ), Chinese Academy of Sciences(CAS) for providing us with supercomputing resources and laser confocal microscopy.

Peer review information

Andrew Cosgrove and Wenjing She were the primary editors of this article and managed its editorial process and peer review in collaboration with the rest of the editorial team. The peer-review history is available in the online version of this article.

Authors' contributions

T.Z., L.S., Z.W., and P.S. conceived and supervised the study. L.S., Q.W., T.Z., M.L., and H.W. performed genome assembly and refinement for the four zokor species. T.Z., L.S., and H.W. identified and characterized pzTRs in the plateau zokor genome. X.L., T.Z., B.Z., and X.Y. established and propagated the zokor cell lines. T.Z. and L.F. designed FISH probes. X.L. conducted FISH experiments and developed pzTR or rat satellite I depletion cell lines. X.Y. identified chromosome stability-related phenotypes. L.S., T.Z., and M.L. analyzed RNA-seq data. L.S., X.Y., H.G., G.C., and L.W., collected the genomic and transcriptomic sequencing samples of zokors. L.S. T.Z., and P.S. wrote the manuscript with contributions from all authors. All authors read and approved the final manuscript.

Funding
This work was funded by the National Natural Science Foundation of China (grant no. U2004152, 32388102, 32570493, 32470501, and 32200351), the Yunnan Revitalization Talent Support Program Science & Technology Champion Project (grant no. 202305AB350002), the Yunnan Revitalization Talent Support Program Top team (grant no. 202405AS350022 and 202505AT350003), the Youth Innovation Promotion Association, Chinese Academy of Sciences (funding recipient: T.Z.), and the Basic Research Project of Yunnan Province (grant no. 202301AT070301, 202301AT070289).

Data availability
All sequencing data and genome assembly generated for this study are available under NCBI BioProject accession PRJNA1179672 [93]. Microscopy image data have been deposited in Figshare [94]. Previously published sequencing datasets reanalyzed in this study include: plateau zokor genome v2.1 [95] and v3.0 [96], resequencing reads of several zokor species [97] and laboratory rat [98].

Declarations

Ethics approval and consent to participate
Ethical approval for these experiments was obtained from the Life Science Ethics Review Committee of Zhengzhou University (ZZUIRB-2022–44) and the Ethics Committee of the Kunming Institute of Zoology, Chinese Academy of Sciences (IACUC-RE-2022–08-014). China's Ministry of Science and Technology developed the 'Guidelines on the Ethical Treatment of Experimental Animals', which governed all sample operations.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests except that Qiang-Hui Wang is currently employed by Novogene Co., Ltd.

Received: 6 February 2025 Accepted: 9 June 2026
Published online: 15 June 2026

References

1. Guo YT, Zhang J, Xu DM, Tang LZ, Liu Z. Phylogenomic relationships and molecular convergences to subterranean life in rodent family Spalacidae. Zool Res. 2021;42:671–4. https://doi.org/10.24272/j.issn.2095-8137.2021.240.
2. Lin GH, Wang K, Deng XG, Nevo E, Zhao F, Su JP, et al. Transcriptome sequencing and phylogenomic resolution within Spalacidae (Rodentia). BMC Genomics. 2014;15:32. https://doi.org/10.1186/1471-2164-15-32.
3. Norris RW, Zhou K, Zhou C, Yang G, William Kilpatrick C, Honeycutt RL. The phylogenetic position of the zokors (Myospalacinae) and comments on the families of muroids (Rodentia). Mol Phylogenet Evol. 2004;31:972–8. https://doi.org/10.1016/j.ympev.2003.10.020.
4. 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. https://doi.org/10.24272/j.issn.2095-8137.2022.045.
5. Tang LZ, Wang LY, Cai ZY, Zhang TZ, Ci HX, Lin GH, et al. Allopatric divergence and phylogeographic structure of the plateau zokor (*Eospalax baileyi*), a fossorial rodent endemic to the Qinghai–Tibetan Plateau. J Biogeogr. 2010;37:657–68. https://doi.org/10.1111/j.1365-2699.2009.02232.x.
6. Wei Y, Zhang T, Li Z, Hua Q, Yin L, Lei M, et al. Evolutionary divergence on the Qinghai-Xizang Plateau: how life-history traits shape the diversity of plateau zokor and pika populations. J Genet Genomics. 2025;52:1096–108. https://doi.org/10.1016/j.jgg.2025.04.019.
7. An X, Mao L, Wang Y, Xu Q, Liu X, Zhang S, et al. Genomic structural variation is associated with hypoxia adaptation in high-altitude zokors. Nat Ecol Evol. 2024;8:339–51. https://doi.org/10.1038/s41559-023-02275-7.
8. 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. https://doi.org/10.1073/pnas.2121819119.
9. Wei DB, Wei L, Zhang JM, Yu HY. Blood-gas properties of plateau zokor (*Myospalax baileyi*). Comp Biochem Physiol A Mol Integr Physiol. 2006;145:372–5. https://doi.org/10.1016/j.cbpa.2006.07.011.
10. 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. https://doi.org/10.1111/mec.16174.
11. Cai Z, Wang L, Song X, Tagore S, Li X, Wang H, et al. Adaptive transcriptome profiling of subterranean zokor, *Myospalax baileyi*, to high- altitude stresses in Tibet. Sci Rep. 2018;8:4671. https://doi.org/10.1038/s41598-018-22483-7.
12. Xu D, Yang C, Shen Q, Pan S, Liu Z, Zhang T, et al. A single mutation underlying phenotypic convergence for hypoxia adaptation on the Qinghai-Tibetan Plateau. Cell Res. 2021;31:1032–5. https://doi.org/10.1038/s41422-021-00517-6.
13. Biscotti MA, Olmo E, Heslop-Harrison JS. Repetitive DNA in eukaryotic genomes. Chromosome Res. 2015;23:415–20. https://doi.org/10.1007/s10577-015-9499-z.
14. Liao X, Zhu W, Zhou J, Li H, Xu X, Zhang B, et al. Repetitive DNA sequence detection and its role in the human genome. Commun Biol. 2023;6:954. https://doi.org/10.1038/s42003-023-05322-y.
15. Thakur J, Packiaraj J, Henikoff S. Sequence, chromatin and evolution of satellite DNA. Int J Mol Sci. 2021;22. https://doi.org/10.3390/ijms22094309.

16. Courret C, Hemmer LW, Wei X, Patel PD, Chabot BJ, Fuda NJ, et al. Turnover of retroelements and satellite DNA drives centromere reorganization over short evolutionary timescales in *Drosophila*. PLoS Biol. 2024;22:e3002911. https://doi.org/10.1371/journal.pbio.3002911.
17. Mora P, Rico-Porras JM, Palomeque T, Montiel EE, Pita S, Cabral-de-Mello DC, et al. Satellitome analysis of *Adalia bipunctata* (Coleoptera): Revealing centromeric turnover and potential chromosome rearrangements in a comparative interspecific study. Int J Mol Sci. 2024;25. https://doi.org/10.3390/ijms25179214.
18. Betancourt AJ, Wei KH, Huang Y, Lee YCG. Causes and consequences of varying transposable element activity: An evolutionary perspective. Annu Rev Genomics Hum Genet. 2024;25:1–25. https://doi.org/10.1146/annurev-genom-120822-105708.
19. Torresen OK, Star B, Mier P, Andrade-Navarro MA, Bateman A, Jarnot P, et al. Tandem repeats lead to sequence assembly errors and impose multi-level challenges for genome and protein databases. Nucleic Acids Res. 2019;47:10994–1006. https://doi.org/10.1093/nar/gkz841.
20. Casper J, Speir ML, Raney BJ, Perez G, Nassar LR, Lee CM, et al. The UCSC Genome Browser database: 2026 update. Nucleic Acids Res. 2026;54:D1331–5. https://doi.org/10.1093/nar/gkaf1250.
21. Audano PA, Sulovari A, Graves-Lindsay TA, Cantsilieris S, Sorensen M, Welch AE, et al. Characterizing the major structural variant alleles of the human genome. Cell. 2019;176:663-75e19. https://doi.org/10.1016/j.cell.2018.12.019.
22. Melters DP, Bradnam KR, Young HA, Telis N, May MR, Ruby JG, et al. Comparative analysis of tandem repeats from hundreds of species reveals unique insights into centromere evolution. Genome Biol. 2013;14:R10. https://doi.org/10.1186/gb-2013-14-1-r10.
23. Wan N, Duan Q, Cai Z, Zhu Z, Wang J, Tian Y, et al. *Aplf/Dna2* variants drive chromosomal fission and accelerate speciation in zokors. Sci Adv. 2025;11:eadt2282. https://doi.org/10.1126/sciadv.adt2282.
24. Kuang Z, Yang X, Wan N, Chen J, Duan Q, Li B, et al. Genomic insights into chromosomal fusion and its evolutionary implications for zokors. Mol Biol Evol. 2026;43. https://doi.org/10.1093/molbev/msag032.
25. Luo LY, Wu H, Zhao LM, Zhang YH, Huang JH, Liu QY, et al. Telomere-to-telomere sheep genome assembly identifies variants associated with wool fineness. Nat Genet. 2025;57:218–30. https://doi.org/10.1038/s41588-024-02037-6.
26. Luo YB, Huang N, Zha CW, Yang LX, Xue PX, Xu Q, et al. Telomere-to-telomere genome assembly of a male pig provides insight into population structure and selection for body stature. Nat Genet. 2026;58:195–205. https://doi.org/10.1038/s41588-025-02433-6.
27. Logsdon GA, Vollger MR, Hsieh P, Mao Y, Liskovykh MA, Koren S, et al. The structure, function and evolution of a complete human chromosome 8. Nature. 2021;593:101–7. https://doi.org/10.1038/s41586-021-03420-7.
28. Fachinetti D, Han JS, McMahon MA, Ly P, Abdullah A, Wong AJ, et al. DNA sequence-specific binding of CENP-B enhances the fidelity of human centromere function. Dev Cell. 2015;33:314–27. https://doi.org/10.1016/j.devcel.2015.03.020.
29. McKinley KL, Cheeseman IM. The molecular basis for centromere identity and function. Nat Rev Mol Cell Biol. 2016;17:16–29. https://doi.org/10.1038/nrm.2015.5.
30. Frankel N, Wang S, Stern DL. Conserved regulatory architecture underlies parallel genetic changes and convergent phenotypic evolution. Proc Natl Acad Sci U S A. 2012;109:20975–9. https://doi.org/10.1073/pnas.1207715109.
31. Lamichhaney S, Card DC, Grayson P, Tonini JFR, Bravo GA, Napflin K, et al. Integrating natural history collections and comparative genomics to study the genetic architecture of convergent evolution. Philos Trans R Soc Lond B Biol Sci. 2019;374:20180248. https://doi.org/10.1098/rstb.2018.0248.
32. Sackton TB, Grayson P, Cloutier A, Hu Z, Liu JS, Wheeler NE, et al. Convergent regulatory evolution and loss of flight in paleognathous birds. Science. 2019;364:74–8. https://doi.org/10.1126/science.aat7244.
33. Treangen TJ, Salzberg SL. Repetitive DNA and next-generation sequencing: computational challenges and solutions. Nat Rev Genet. 2011;13:36–46. https://doi.org/10.1038/nrg3117.
34. Amarasinghe SL, Su S, Dong X, Zappia L, Ritchie ME, Gouil Q. Opportunities and challenges in long-read sequencing data analysis. Genome Biol. 2020;21:30. https://doi.org/10.1186/s13059-020-1935-5.
35. Nurk S, Koren S, Rhie A, Rautiainen M, Bzikadze AV, Mikheenko A, et al. The complete sequence of a human genome. Science. 2022;376:44–53. https://doi.org/10.1126/science.abj6987.
36. Henikoff S, Ahmad K, Malik HS. The centromere paradox: Stable inheritance with rapidly evolving DNA. Science. 2001;293:1098–102. https://doi.org/10.1126/science.1062939.
37. Talbert P, Henikoff S. Centromere drive: Chromatin conflict in meiosis. Curr Opin Genet Dev. 2022;77:102005. https://doi.org/10.1016/j.gde.2022.102005.
38. Gambogi CW, Pandey N, Dawicki-McKenna JM, Arora UP, Liskovykh MA, Ma J, et al. Centromere innovations within a mouse species. Sci Adv. 2023;9:eadi5764. https://doi.org/10.1126/sciadv.adi5764.
39. Iwata-Otsubo A, Dawicki-McKenna JM, Akera T, Falk SJ, Chmatal L, Yang K, et al. Expanded satellite repeats amplify a discrete CENP-A nucleosome assembly site on chromosomes that drive in female meiosis. Curr Biol. 2017;27:2365-73e8. https://doi.org/10.1016/j.cub.2017.06.069.
40. Henikoff S. Malik HS. Centromeres: Selfish drivers. Nature. 2002;417:227. https://doi.org/10.1038/417227a.
41. Dumont M, Gamba R, Gestraud P, Klaasen S, Worrall JT, De Vries SG, et al. Human chromosome-specific aneuploidy is influenced by DNA-dependent centromeric features. EMBO J. 2020;39:e102924. https://doi.org/10.15252/embj.2019102924.
42. Pardo-Manuel de Villena F, Sapienza C. Female meiosis drives karyotypic evolution in mammals. Genetics. 2001;159:1179–89. https://doi.org/10.1093/genetics/159.3.1179.
43. Garagna S, Page J, Fernandez-Donoso R, Zuccotti M, Searle JB. The Robertsonian phenomenon in the house mouse: Mutation, meiosis and speciation. Chromosoma. 2014;123:529–44. https://doi.org/10.1007/s00412-014-0477-6.
44. Ichikawa K, Tomioka S, Suzuki Y, Nakamura R, Doi K, Yoshimura J, et al. Centromere evolution and CpG methylation during vertebrate speciation. Nat Commun. 2017;8:1833. https://doi.org/10.1038/s41467-017-01982-7.
45. Kalitsis P, Griffiths B, Choo KH. Mouse telocentric sequences reveal a high rate of homogenization and possible role in Robertsonian translocation. Proc Natl Acad Sci U S A. 2006;103:8786–91. https://doi.org/10.1073/pnas.0600250103.
46. Wickham H. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag NewYork; 2016.

47. Chen S, Zhou Y, Chen Y, Gu J. Fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 2018;34:i884–90. https://doi.org/10.1093/bioinformatics/bty560.
48. Wang H, Liu B, Zhang Y, Jiang F, Ren Y, Yin L, et al. Estimation of genome size using k-mer frequencies from corrected long reads. arXiv. 2020. https://doi.org/10.48550/arXiv.2003.11817.
49. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res. 2017;27:722–36. https://doi.org/10.1101/gr.215087.116.
50. Hu J, Fan J, Sun Z, Liu S. NextPolish: a fast and efficient genome polishing tool for long-read assembly. Bioinformatics. 2020;36:2253–5. https://doi.org/10.1093/bioinformatics/btz891.
51. Zhang X, Zhang S, Zhao Q, Ming R, Tang H. Assembly of allele-aware, chromosomal-scale autopolyploid genomes based on Hi-C data. Nat Plants. 2019;5:833–45. https://doi.org/10.1038/s41477-019-0487-8.
52. Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. 2018;34:3094–100. https://doi.org/10.1093/bioinformatics/bty191.
53. Cheng H, Concepcion GT, Feng X, Zhang H, Li H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. Nat Methods. 2021;18:170–5. https://doi.org/10.1038/s41592-020-01056-5.
54. Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand NC, et al. De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. Science. 2017;356:92–5. https://doi.org/10.1126/science.aal3327.
55. Simao FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015;31:3210–2. https://doi.org/10.1093/bioinformatics/btv351.
56. Parra G, Bradnam K, Korf I. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. Bioinformatics. 2007;23:1061–7. https://doi.org/10.1093/bioinformatics/btm071.
57. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv. 2013. https://doi.org/10.48550/arXiv.1303.3997.
58. 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. https://doi.org/10.1093/bioinformatics/btp352.
59. Xu Z, Wang H. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. Nucleic Acids Res. 2007;35:W265–8. https://doi.org/10.1093/nar/gkm286.
60. Price AL, Jones NC, Pevzner PA. De novo identification of repeat families in large genomes. Bioinformatics. 2005;21(Suppl 1):i351–8. https://doi.org/10.1093/bioinformatics/bti1018.
61. Benson G. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. 1999;27:573–80. https://doi.org/10.1093/nar/27.2.573.
62. Hu K, Ni P, Xu M, Zou Y, Chang J, Gao X, et al. HiTE: a fast and accurate dynamic boundary adjustment approach for full-length transposable element detection and annotation. Nat Commun. 2024;15:5573. https://doi.org/10.1038/s41467-024-49912-8.
63. Nawrocki EP, Eddy SR. Infernal 1.1: 100-fold faster RNA homology searches. Bioinformatics. 2013;29:2933–5. https://doi.org/10.1093/bioinformatics/btt509.
64. Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 1997;25:955–64. https://doi.org/10.1093/nar/25.5.955.
65. McGinnis S, Madden TL. BLAST: at the core of a powerful and diverse set of sequence analysis tools. Nucleic Acids Res. 2004;32:W20–5. https://doi.org/10.1093/nar/gkh435.
66. Yu XJ, Zheng HK, Wang J, Wang W, Su B. Detecting lineage-specific adaptive evolution of brain-expressed genes in human using rhesus macaque as outgroup. Genomics. 2006;88:745–51. https://doi.org/10.1016/j.ygeno.2006.05.008.
67. Birney E, Clamp M, Durbin R. GeneWise and Genomewise. Genome Res. 2004;14:988–95. https://doi.org/10.1101/gr.1865504.
68. Stanke M, Morgenstern B. AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. Nucleic Acids Res. 2005;33:W465–7. https://doi.org/10.1093/nar/gki458.
69. Majoros WH, Pertea M, Salzberg SL. TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. Bioinformatics. 2004;20:2878–9. https://doi.org/10.1093/bioinformatics/bth315.
70. Korf I. Gene finding in novel genomes. BMC Bioinformatics. 2004;5:59. https://doi.org/10.1186/1471-2105-5-59.
71. Guigo R. Assembling genes from predicted exons in linear time with dynamic programming. J Comput Biol. 1998;5:681–702. https://doi.org/10.1089/cmb.1998.5.681.
72. Burge C, Karlin S. Prediction of complete gene structures in human genomic DNA. J Mol Biol. 1997;268:78–94. https://doi.org/10.1006/jmbi.1997.0951.
73. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics. 2009;25:1105–11. https://doi.org/10.1093/bioinformatics/btp120.
74. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol. 2010;28:511–5. https://doi.org/10.1038/nbt.1621.
75. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol. 2011;29:644–52. https://doi.org/10.1038/nbt.1883.
76. Haas BJ, Delcher AL, Mount SM, Wortman JR, Smith RK Jr, Hannick LI, et al. Improving the *Arabidopsis* genome annotation using maximal transcript alignment assemblies. Nucleic Acids Res. 2003;31:5654–66. https://doi.org/10.1093/nar/gkg770.
77. Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, et al. Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. Genome Biol. 2008;9:R7. https://doi.org/10.1186/gb-2008-9-1-r7.
78. Finn RD, Attwood TK, Babbitt PC, Bateman A, Bork P, Bridge AJ, et al. InterPro in 2017-beyond protein family and domain annotations. Nucleic Acids Res. 2017;45:D190–9. https://doi.org/10.1093/nar/gkw1107.

79. Finn RD, Clements J, Arndt W, Miller BL, Wheeler TJ, Schreiber F, et al. HMMER web server: 2015 update. Nucleic Acids Res. 2015;43:W30–8. https://doi.org/10.1093/nar/gkv397.
80. Finn RD, Tate J, Mistry J, Coggill PC, Sammut SJ, Hotz HR, et al. The Pfam protein families database. Nucleic Acids Res. 2008;36:D281–8. https://doi.org/10.1093/nar/gkm960.
81. Zdobnov EM, Apweiler R. InterProScan–an integration platform for the signature-recognition methods in InterPro. Bioinformatics. 2001;17:847–8. https://doi.org/10.1093/bioinformatics/17.9.847.
82. Harris MA, Clark J, Ireland A, Lomax J, Ashburner M, Foulger R, et al. The Gene Ontology (GO) database and informatics resource. Nucleic Acids Res. 2004;32:D258–61. https://doi.org/10.1093/nar/gkh036.
83. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000;28:27–30. https://doi.org/10.1093/nar/28.1.27.
84. Novak P, Avila Robledillo L, Koblizkova A, Vrbova I, Neumann P, Macas J. TAREAN: a computational tool for identification and characterization of satellite DNA from unassembled short reads. Nucleic Acids Res. 2017;45:e111. https://doi.org/10.1093/nar/gkx257.
85. Novak P, Neumann P, Macas J. Global analysis of repetitive DNA from unassembled sequence reads using RepeatExplorer2. Nat Protoc. 2020;15:3745–76. https://doi.org/10.1038/s41596-020-0400-y.
86. Nie W, Wang J, Su W, Wang D, Tanomtong A, Perelman PL, et al. Chromosomal rearrangements and karyotype evolution in carnivores revealed by chromosome painting. Heredity (Edinb). 2012;108:17–27. https://doi.org/10.1038/hdy.2011.107.
87. Levan A, Fredga K, Sandberg AA. Nomenclature for centromeric position on chromosomes. Hereditas. 2009;52:201–20. https://doi.org/10.1111/j.1601-5223.1964.tb01953.x.
88. Petsalaki E, Zachos G. The abscission checkpoint: a guardian of chromosomal stability. Cells. 2021;10. https://doi.org/10.3390/cells10123350.
89. Siri SO, Martino J, Gottifredi V. Structural chromosome instability: Types, origins, consequences, and therapeutic opportunities. Cancers (Basel). 2021;13. https://doi.org/10.3390/cancers13123056.
90. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol. 2019;37:907–15. https://doi.org/10.1038/s41587-019-0201-4.
91. Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. Transcript-level expression analysis of RNA-seq experiments with HISAT. StringTie and Ballgown Nat Protoc. 2016;11:1650–67. https://doi.org/10.1038/nprot.2016.095.
92. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15:550. https://doi.org/10.1186/s13059-014-0550-8.
93. Wang ZL, Shi LY, Liu XY, Yao XQ, Li MY, Wang QH, et al. Four species in Spalacidae: *Eospalax baileyi*, *Eospalax rufescens*, *Eospalax cansus* and *Myospalax psilurus* Genome sequencing and assembly. Datasets. Datasets sequence read archive. 2026. https://www.ncbi.nlm.nih.gov/sra/?term=PRJNA1179672.
94. Wang ZL, Shi LY, Liu XY, Yao XQ, Li MY, Wang QH, et al. Centromeric tandem repeats expansions in plateau zokor enhance chromosome stability for high-altitude adaptation. 2026. Figshare. https://doi.org/10.6084/m9.figshare.32574066.
95. Zhang T. plateau_zokor_v2.1. Datasets. National Genomics Data Center (NGDC). 2020. https://ngdc.cncb.ac.cn/gwh/Assembly/941/show.
96. An X. The genome analysis of *Eospalax baileyi*. Datasets. China National GeneBank DataBase. 2023. https://db.cngb.org/data_resources/project/CNP0005002/.
97. Zhang T. whole-genome resequence of *Eospalax* species. Datasets. National Genomics Data Center (NGDC). 2022. https://ngdc.cncb.ac.cn/gsa/browse/CRA005933.
98. University of Michigan. Whole-genome sequence data for 8 founders of rat heterogeneous stock. Datasets. Datasets sequence read archive. 2018. https://www.ncbi.nlm.nih.gov/sra/?term=SRR7755593.

Publisher's Note