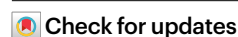


Genomic structural variation is associated with hypoxia adaptation in high-altitude zokors

Received: 21 March 2023

Accepted: 20 November 2023

Published online: 9 January 2024



Xuan An^{1,6}, Leyan Mao^{1,6}, Yinjia Wang^{1,6}, Qinqin Xu^{2,6}, Xi Liu¹, Shangzhe Zhang¹, Zhenglei Qiao³, Bowen Li¹, Fang Li³, Zhuoran Kuang¹, Na Wan¹, Xiaolong Liang¹, Qijiao Duan¹, Zhilong Feng¹, Xiaojie Yang¹, Sanyuan Liu¹, Eviatar Nevo⁴, Jianquan Liu¹✉, Jay F. Storz⁵✉ & Kexin Li¹✉

Zokors, an Asiatic group of subterranean rodents, originated in lowlands and colonized high-elevation zones following the uplift of the Qinghai–Tibet plateau about 3.6 million years ago. Zokors live at high elevation in subterranean burrows and experience hypobaric hypoxia, including both hypoxia (low oxygen concentration) and hypercapnia (elevated partial pressure of CO₂). Here we report a genomic analysis of six zokor species (genus *Eospalax*) with different elevational ranges to identify structural variants (deletions and inversions) that may have contributed to high-elevation adaptation. Based on an assembly of a chromosome-level genome of the high-elevation species, *Eospalax baileyi*, we identified 18 large inversions that distinguished this species from congeners native to lower elevations. Small-scale structural variants in the introns of *EGLN1*, *HIF1A*, *HSF1* and *SFTPD* of *E. baileyi* were associated with the upregulated expression of those genes. A rearrangement on chromosome 1 was associated with altered chromatin accessibility, leading to modified gene expression profiles of key genes involved in the physiological response to hypoxia. Multigene families that underwent copy-number expansions in *E. baileyi* were enriched for autophagy, HIF1 signalling and immune response. *E. baileyi* show a significantly larger lung mass than those of other *Eospalax* species. These findings highlight the key role of structural variants underlying hypoxia adaptation of high-elevation species in *Eospalax*.

Animals native to high elevations are forced to cope with major environmental stressors, including low temperature, strong ultraviolet exposure and reduced oxygen availability¹. Genomic studies of highland humans and non-human mammals have suggested candidate genes for hypoxia adaptation^{1–13}, and experimental physiological studies have

identified specific mechanisms of adaptation^{14–16}. However, much less is known about the extent to which structural variation in the genome has contributed to high-altitude adaptation.

Structural variants (SVs) can affect the orientation and linkage arrangements of genes, thereby potentially altering patterns of gene

¹State Key Laboratory of Herbage Improvement and Grassland Agro-ecosystems, College of Ecology, Lanzhou University, Lanzhou, China. ²Department of Medical Oncology, Qinghai Provincial People's Hospital, Xining, China. ³College of Life Sciences and Technology, Mudanjiang Normal University, Mudanjiang, China. ⁴Institute of Evolution, University of Haifa, Haifa, Israel. ⁵School of Biological Sciences, University of Nebraska, Lincoln, NE, USA.

⁶These authors contributed equally: Xuan An, Leyan Mao, Yinjia Wang, Qinqin Xu. ✉e-mail: liujq@nwipb.cas.cn; jstorz2@unl.edu; likexin@lzu.edu.cn

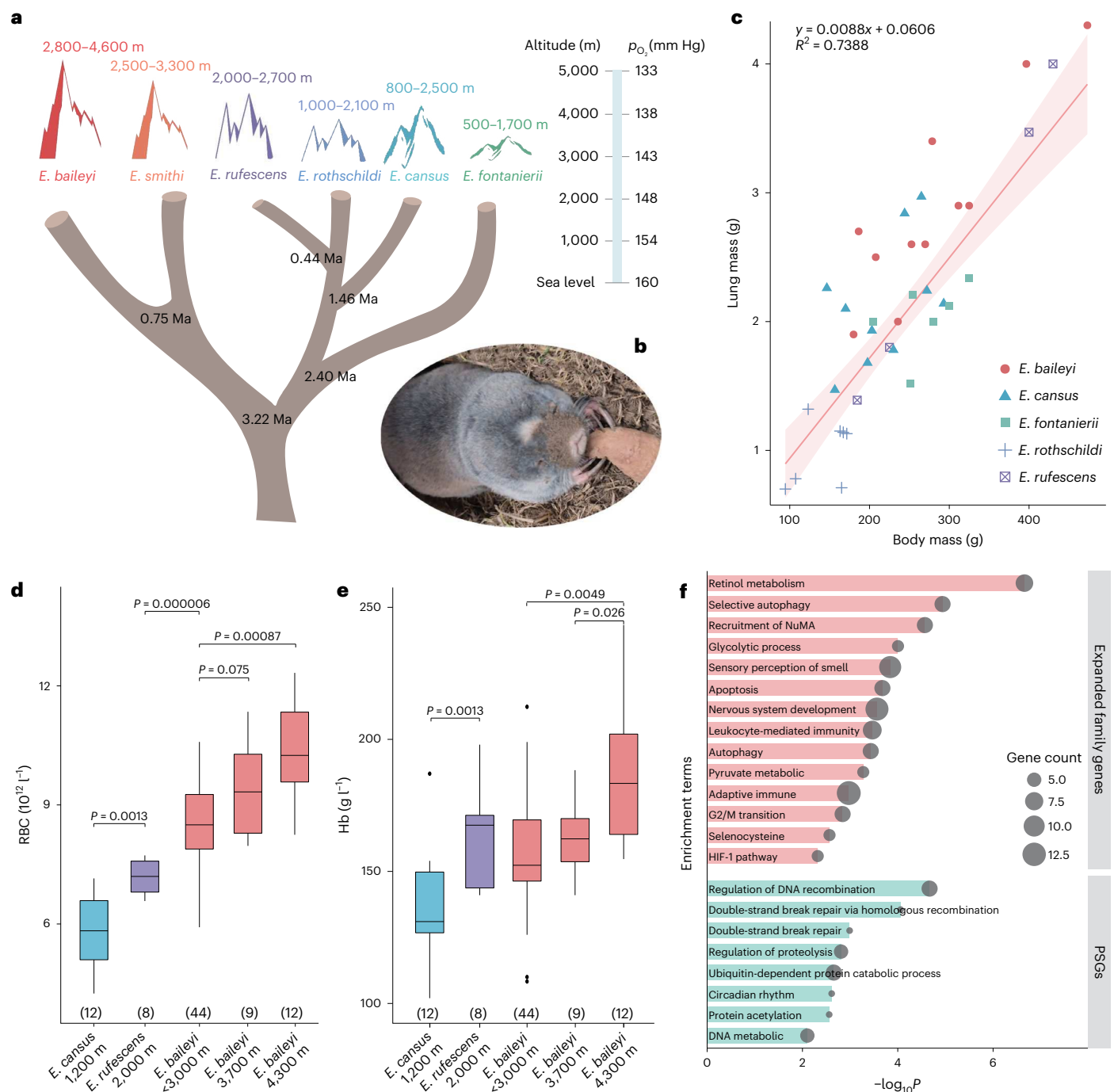


Fig. 1 | Phylogenetic relationships, elevational distributions and patterns of phenotypic variation among *Eospalax*. **a**, Phylogenetic trees show the relationship and elevation ranges of *Eospalax*. The bar shows the partial pressures of oxygen (p_{O_2}) at each elevation. **b**, Primary subject of this study, *E. baileyi* (plateau zokor). Credit: Xuan An, Lanzhou University. **c**, Comparison of the relationship between lung mass and body mass among *Eospalax*. The linear regression line is shown as a solid red line and the 95% confidence intervals are shown as pink shadows. **d, e**, Difference in RBC counts ($10^{12} l^{-1}$) (**d**) and blood Hb ($g l^{-1}$) (**e**) among *E. baileyi*, *E. rufescens* and *E. cansus*. There are also differences in *E. baileyi* populations at different elevations. The numbers in parentheses are

sample sizes. In the box plots, the lower and higher hinges correspond to the first and third quartiles, respectively. The median is represented by the line inside the box. The whisker extends from the hinge to at most $1.5 \times$ interquartile range (IQR). The other points are outliers. The difference in RBC and Hb supporting different populations was tested using a two-sided t -test. **f**, Functional enrichment of expanded-family genes (pink) and PSGs (green) in high-altitude *E. baileyi*. The size of the round shape represents the number of genes in this enrichment term. The PSGs (likelihood ratio test, $P < 0.05$, FDR < 0.05) and expanded family genes (likelihood ratio test, $P < 0.05$, FDR < 0.05) were identified with multiple corrections.

expression¹⁷. Structural changes can also affect gene expression by altering the distance between *cis*-acting regulators and their targets and by modifying chromatin accessibility to nuclear macromolecules^{18,19}. Moreover, SVs can affect gene expression and regulation by

altering three-dimensional (3D) genomic structures, including active (A) or inactive (B) compartments, topologically associating domain (TAD) boundaries and chromatin loops. Large chromosome inversions, mediated mainly through non-homologous end joining and non-allelic

homologous recombination (NAHR), are known to contribute to adaptive phenotypic change^{20–22}, and they can suppress recombination, thereby helping to maintain co-adapted combinations of alleles at linked loci (supergenes)^{23–25}. Recent studies have documented that SVs contribute to variation in physiological responses to hypoxia, thereby affecting susceptibility to altitude-related pathologies such as pulmonary oedema and cerebral hypoxia¹⁹.

Zokors (genus *Eospalax*, Spalacidae: Rodentia) live most of their lives underground, where they have to cope with hypoxia and hypercapnia in closed burrow systems. O₂ and CO₂ levels in sealed tunnels have been recorded at 7.2% and 6.1%, respectively, during the rainy season when wet soil further impedes gas exchange with the outside air²⁶. *Eospalax* comprises six East Asian species that have well-documented phylogenetic relationships²⁷ (Fig. 1a,b). With the uplift of the Qinghai–Tibet Plateau (~3.6 million years ago (Ma)), *Eospalax* split into the high-altitude (including *E. baileyi*, 2,800–4,600 m, and *E. smithi*, 2,500–3,300 m) and low-altitude clades at ~3.22 Ma. *E. baileyi* and *E. smithi* split from their ancestor at ~1.18 Ma and 0.75 Ma, respectively, due to the generation of the Yellow River and a glaciation. For the low-altitude lineage, four *Eospalax* species have elevational ranges under 2,700 m. *E. fontanierii* speciated first, *E. cansus* diverged northwards from the ancestor at ~1.43 Ma, whereas *E. rothschildi* speciated southwards because of a later glaciation (Fig. 1a). Comparisons between highland and lowland species provide an opportunity to explore genetic changes that may have contributed to high-elevation adaptation.

Here we investigated the role of SVs (deletion and inversion) in hypoxia adaptation of high-elevation species in *Eospalax*. We used a long-read sequencing approach to obtain a chromosome-level genome assembly of *E. baileyi*, and we assessed species-level variation using a long-read resequencing approach. The effects of SVs on hypoxia-related genes were examined using RNA sequencing (RNA-seq) and reporter assays. We also investigated TAD rearrangements and A/B compartment transitions to assess how evolved changes in 3D genomic architecture may have contributed to hypoxia adaptation.

Results

Phenotypic variation among zokors native to different elevations

We measured several phenotypes of relevance to high-altitude acclimatization and/or adaptation: the relationship between lung mass and body mass, red blood cell (RBC) counts, and blood haemoglobin concentration ([Hb]). The high-elevation *E. baileyi* showed a larger lung mass compared with other *Eospalax* species (Fig. 1c), which can be expected to enhance pulmonary diffusing capacity for gas exchange¹. RBC counts and [Hb] were also higher in *E. baileyi* relative to *E. cansus* and *E. rufescens*, suggesting a higher O₂-transport capacity per unit of cardiac output. In *E. baileyi*, RBC counts and [Hb] increased as a positive function of native elevation (Fig. 1d,e), consistent with expected acclimatization responses to chronic hypoxia²⁸.

Genome assembly, scaffolding and annotation

We generated the first chromosome-level genome assembly of *E. baileyi*. The initial assembly consisted of ~209.0 Gb of long reads and was subsequently corrected using ~255.1 Gb of Illumina paired-end reads (Extended Data Table 1). About 99.8% of corrected contigs were anchored onto 31 pseudochromosomes (2n = 62) through Hi-C data (Supplementary Table 1). This is consistent with the fibroblast karyotype analysis (Extended Data Fig. 1a,b,e). The finally assembled genome size was 2.47 Gb. Compared with the previously published scaffold-level assembly of the plateau zokor (v.2.1)³, the contig N50 increased from 323 kb to 38.0 Mb, the scaffold N50 increased from 6.7 to 82.3 Mb, and gaps decreased from 1.64% to 0.01% (Table 1). The completeness of the assembly, tested using Benchmarking Universal Single-Copy Orthologs (BUSCO), was 96.8%, indicating high precision,

Table 1 | Comparison of assembly statistics between Plateau zokor v.2.1 and Plateau zokor v.3.0

Plateau zokor	Plateau zokor v.2.1	Plateau zokor v.3.0
Sequencing platform	Illumina, Pacbio (10x)	ONT, Hi-C
Assembly size (bp)	2,569,336,452	2,471,808,144
Total length of chromosome (bp)	–	2,385,578,628
Number of contigs	42,044	758
Contig N50 (bp)	322,513	37,952,775
Number of scaffolds	14,401	465
Scaffold N50 (bp)	6,701,949	82,317,615
Number of gaps	102,445	293
Total gap length (bp)	42,024,904	124,000
Ratio of gap (%)	1.64	0.01
GC content (%)	41.20	41.26
BUSCO (%)	93.30	96.80

continuity and integrity. The proportion of genomic sequence consisting of repeats was 44.4%, where short interspersed nuclear elements (SINEs) and long interspersed nuclear elements (LINEs) accounted for 18.1% and 22.2%, respectively (Supplementary Table 2). We annotated 21,642 protein-coding genes, consistent with the annotation results for the genome of the closely related species *E. fontanierii*²⁷. The BUSCO evaluation of the annotation completeness showed that 95.6% of genes were predicted in this *E. baileyi* assembly (Extended Data Fig. 1c,d).

Genome evolution and possible mechanisms of hypoxia adaptation

We estimated a phylogenetic tree and species divergence time of 9 rodent species based on 9,258 single-copy orthologues (Extended Data Fig. 1f). Compared with the other 8 rodents, we identified 469 gene families in the *E. baileyi* lineage that underwent significant expansions relative to the expectations of a null model of birth-and-death evolution (likelihood ratio test, $P < 0.05$). Functional enrichment analysis of the expanded gene families revealed an over-representation of the following terms: hypoxia-inducible factor (*HIF*) signalling pathway, DNA damage repair, autophagy, cell cycle, selenocysteine synthesis, purine and retinol metabolism, recruitment of nuclear mitotic apparatus (NuMA) to mitotic centrosomes, immune responses and nutrition functions (Fig. 1f and Supplementary Table 3). We also identified a total of 1,609 gene families that underwent significant contractions in the *E. baileyi* lineage. These genes were related to cell cycle, signalling by WNT, metabolism and meiotic recombination. Using the adaptive branch-site random effects likelihood (aBSREL) model, we tested for accelerated rates of non-synonymous substitution across all genes in the branch leading to *E. baileyi* and identified 71 lineage-specific positively selected genes (PSGs) ($\omega > 1$, likelihood ratio test, $P < 0.05$, false discovery rate (FDR) < 0.05 ; Fig. 1f and Supplementary Tables 3 and 4). These PSGs were enriched for the following terms: double-strand break repair, circadian rhythm and ubiquitin-dependent protein catabolic functions.

SV identification, validation and distribution

Distribution of <1 Mb SVs in *Eospalax* genomes. Using long-read sequencing, we identified 398,087 uniquely located SVs in a sample of 2 or 3 individuals from each *Eospalax* species, including 218,607 deletions (DEL), 168,333 insertions (INS), 5,183 duplications (DUP), 5,062 inversions (INV) and 902 translocations (TRA) (Fig. 2a and Supplementary Table 5). The frequencies of these SVs were estimated from 146 individuals, including 55 *E. baileyi*, 25 *E. cansus*, 21 *E. smithi*, 18 *E. rothschildi*, 14 *E. rufescens* and 13 *E. fontanierii* (Extended Data

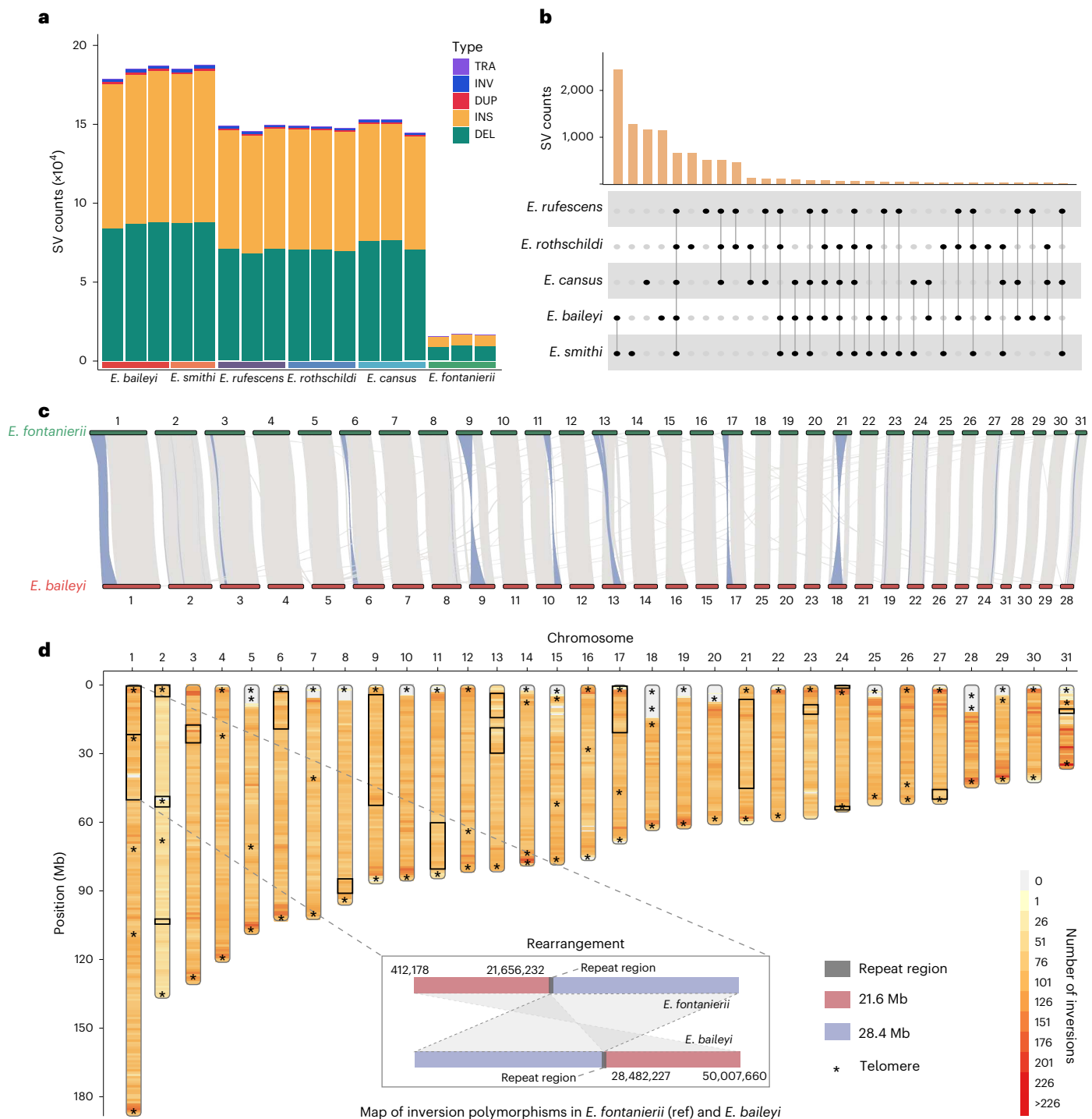


Fig. 2 | Identification and distribution of large inversions (>1 Mb) and SVs (<1 Mb, INS, DEL, INV, TRA and DUP). **a**, Stacked bar graph showing uniquely located SV (<1 Mb) types and numbers for 17 individuals from six *Eospalax* species. **b**, Species-specific and shared SVs of six species in *Eospalax*. The high-altitude branch of *E. baileyi* and *E. smithi* showed the highest number of shared SVs. **c**, Chromosomal synteny between *E. fontanierii* and *E. baileyi* at chromosome-level genome. Syntenic blocks are linked by shaded bands. The

blue lines show the 18 inversion positions. **d**, Distribution of all inversions across *E. baileyi*, annotated with predicted telomeres (black asterisks). The color gradient represents the number of inversions (<1 Mb) in a 1 Mb window based on the long-read-SV dataset. The black box indicated 18 inversions (>1 Mb) positions between *E. fontanierii* (reference genome, ref) and *E. baileyi* genomes. The chromosomal rearrangement (21.6 and 28.4 Mb) located in chromosome 1 is highlighted.

Fig. 2a). We investigated the phylogenetic distribution of SVs across *Eospalax* and identified those that were species specific (Fig. 2b). Most of the INDELs were less than 300 bp, and there was another small peak at ~400 bp, which was mainly influenced by SINEs (Extended Data Fig. 2b,c). The majority of SVs (53.63%, 296,873) overlapped with

simple repetitive elements and transposable elements. For transposable elements, SVs were enriched in regions with SINEs (35%) and long terminal repeats (LTRs, 22.8%). Most SVs (69.16%, 275,307) were distributed in intergenic regions. Meanwhile, 122,780 (30.84%) SVs were found within 68.15% (14,758 out of 21,656) of the genic regions (that is, in

exons, introns, and 5' and 3' flanking regions within 2 kb of start and stop codons, respectively). Intronic SVs accounted for 28.25% of all SVs, and only 5.11% of SVs affected coding sequences (Extended Data Fig. 2d,e).

Large inversions (>1 Mb) in the *E. baileyi* genome. We identified 18 inversions longer than 1 Mb in *E. baileyi* (Fig. 2c,d and Extended Data Fig. 3) based on the chromosome-level genome assembly. These large inversions were validated by aligning the breakpoint-spanning contigs of *E. baileyi* to the genome of *E. fontanierii* (Extended Data Fig. 4). Accurate breakpoints (Supplementary Table 6) were identified in three ways (Methods). We identified 81 telomere-enriched regions, with 60 appearing at expected chromosome ends. We found 11 inversions located at or near the chromosome termini (Fig. 2c), with most of the breakpoints near the telomere-enriched regions. Seven of these inversions contained telomeric regions, suggesting that the inversions probably altered the position of the telomeres.

Possible functional effects of SVs

Differentially expressed genes associated with SVs (<1 Mb). To evaluate whether fixed SVs could potentially affect the expression of associated genes, we estimated the expression of all 3,748 fixed SV-associated genes using RNA-seq from animals after common garden experiments (Methods). Compared with *E. fontanierii*, we found 253, 187 and 122 differentially expressed genes (DEGs) in the lung, liver and heart of the high-altitude *E. baileyi*, respectively, accounting for 6.75%, 5.00% and 3.26% of all fixed SV-associated genes (Supplementary Tables 12–14). The DEGs in the lung were associated with immunity, inflammatory response, cell adhesion and cell developmental cycle (*ILDR2*, *IL20*, *CDON*, *LAMB3*, *IL2RB*, *OPCM*, *RAB3C* and *NEK10*). The DEGs of *TBX20* and *KCND2* in the lung contained intronic SVs and were involved in cardiac development, conduction and diseases. In the liver, differentially expressed SV-associated genes were involved in haematopoiesis, vascular development, platelet activation and vascular endothelial growth factor (*GLRX5*, *CD109*, *NRP1* and *ADTRP*). Notably, immunity-related genes of *SFTPD* and *NRP1*, both of which contained intronic SVs, were significantly upregulated in the livers of *E. baileyi*. We found a number of differentially expressed SV-associated genes in the heart significantly enriched for the HIF1 pathway function (*NOS1*, *NQO1*, *HIPK2* and *BMP7*), and genes related to cell adhesion and leucocyte proliferation (*CD1D*, *NCAM2* and *CDC26*) were also upregulated in the high-altitude *E. baileyi* (Extended Data Fig. 5a–d).

Functional effects of SVs (<1 Mb). SVs specific to the two species with the highest range limits, *E. smithi* and *E. baileyi*, were located in 345 genes (that is, in exons, introns, or 5' and 3' flanking regions within 2 kb of start and stop codons), which were enriched in cellular response to hypoxia, DNA damage and repair, autophagy, heart morphogenesis and contraction, lung epithelium development, response to viruses,

and regulation of vascular-associated smooth muscle cell migration. These same pathways have been implicated in hypoxia adaptation in other highland mammals^{29,30} (Extended Data Fig. 5e and Supplementary Table 7).

We detected an intronic SV in *EGLN1* with both heterozygous inversion (NI) and homologous inversion (II) existing only in *E. baileyi* (1,442 bp, LG01:1,889,355–1,890,797). Frequencies of inversion were 100% present at a frequency of 100% in animals above 4,300 m ($n = 6$) and 84.7% in those from <4,300 m ($n = 49$) (Fig. 3a). DEGs analyses showed that the *E. baileyi* *EGLN1* allele with intronic inversion had a significantly higher expression compared with low-altitude species in *Eospalax* (Fig. 3b), and the dual-luciferase reporter assay (DLRA) confirmed that the inversion significantly increased luciferase activity (Fig. 3c and Extended Data Table 2).

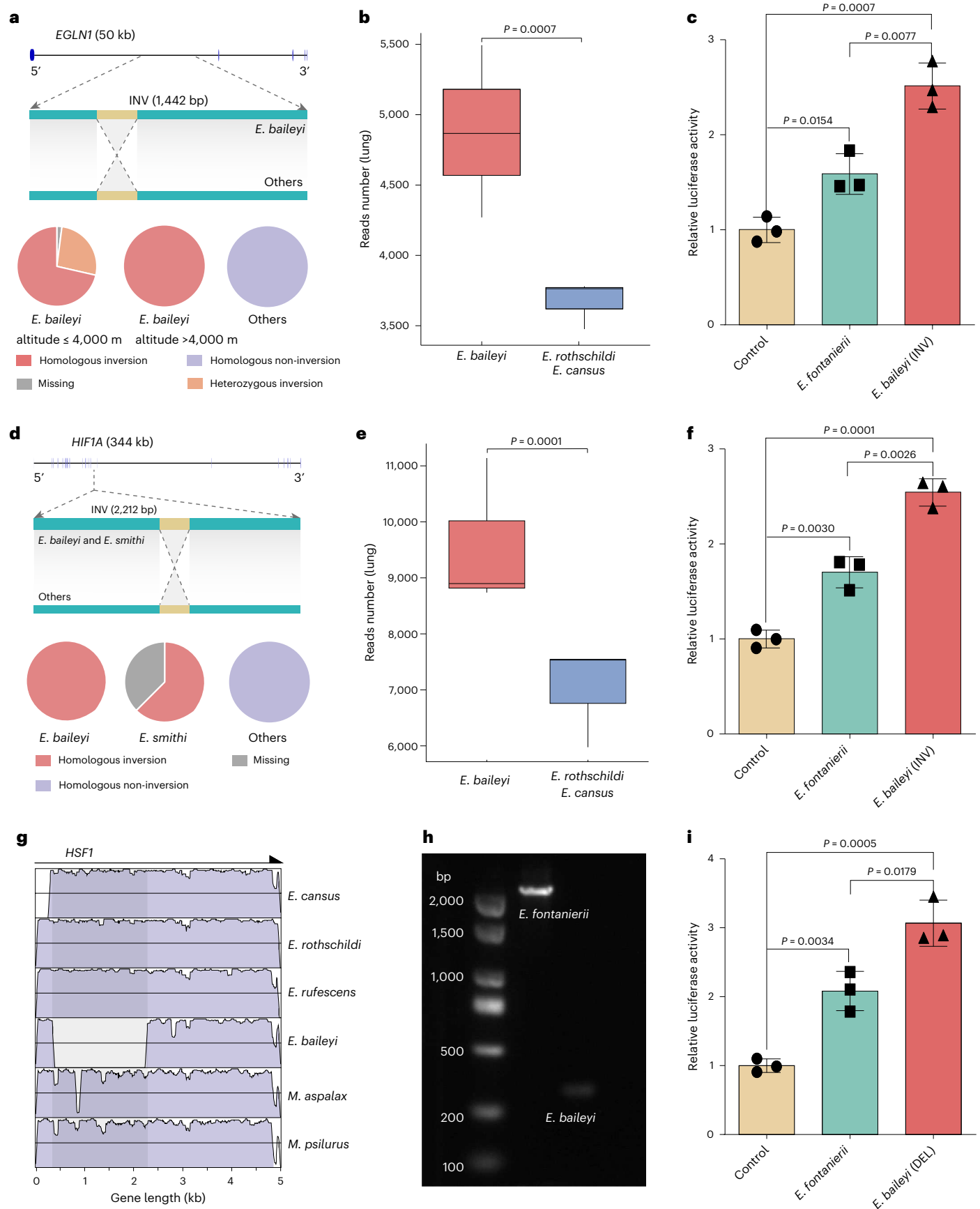
A homozygous intronic inversion (2,212 bp, LG13: 44,667,409–44,669,626) inside *HIF1A* was fixed in *E. smithi* and *E. baileyi* (Fig. 3d and Extended Data Fig. 6a). The inversion allele in *E. baileyi* showed upregulated gene expression (Fig. 3e). As with *EGLN1*, the inversion within the *HIF1A* intron also resulted in significantly higher expression using DLRA (Fig. 3f and Extended Data Table 2). An intronic deletion in the *HSF1* was identified (1,849 bp, LG09: 81,333,952–81,335,801) (Fig. 3g) and confirmed by polymerase chain reaction (PCR; Fig. 3h). DLRA suggested that the presence of this deletion significantly increased the *HSF1* expression (Fig. 3i).

SVs were also detected in three genes related to DNA repair and immunity. A homozygous 12,732 bp deletion that included an intron and the second exon of *SFTPD* in *E. baileyi* resulted in a deletion of 55 amino acids from the N terminus of the protein. Protein structure predictions of *E. fontanierii* and *E. baileyi* suggested that the deletion resulted in missing α -helix and β -fold in *E. baileyi* (Extended Data Fig. 6b). We also detected a 9,280 bp intronic deletion in *XRCC4* in *E. baileyi* and a 1,984-bp intronic deletion in *UVRAG* of *E. smithi* and *E. baileyi*, which were validated through read coverage and long-read sequence alignment (Extended Data Fig. 6c,d).

Limited deleterious effects of large inversions. We found that 7 of the 36 large inversion breakpoints occurred in genic regions, with possible consequences for gene function and gene expression. For example, the breakpoints of LG01.1 disrupted the *SPSI* gene, which has been reported to be involved in selenocysteine synthesis. Breakpoints for LG02.1, LG06.1, LG11.1 and LG17.1 disrupted the *P2R2B*, *CB016*, *CFA91* and *DLG5* genes, respectively. Breakpoints may also affect gene expression by disrupting regulatory elements (Fig. 4a). Inversions were not disproportionately associated with loss-of-function single nucleotide polymorphism (LOF SNP) (Fig. 4b and Supplementary Table 8). Some large inversions potentially rearranged TAD boundaries, which could alter interactions between genes/regulators. We found that most inversion breakpoints occurred near TAD boundaries (Fig. 4c).

Fig. 3 | Intronic SVs upregulated gene expression and facilitated hypoxia adaptation. a–c, A 1,442-bp intronic inversion was detected in *EGLN1*, as well as its effect on RNA expression in *E. baileyi*. a, Schematic inversion located in an intron of *EGLN1* of the high-altitude *E. baileyi*. The inversion frequency was estimated from 146 individuals of all *Eospalax* species, and this inversion tended to be fixed in *E. baileyi* but absent in others. b, The number of mRNA reads of *EGLN1* was significantly higher in the lung of *E. baileyi* ($n = 3$) than those of *E. rothschildi* and *E. cansus* ($n = 3$). The analysis utilizing a two-sided t -test. In the box plots, the centre line represents the median, whereas box boundaries represent the upper and lower quartiles. The whisker reaches no farther than $1.5 \times \text{IQR}$. c, The DLRA showed that the fluorescence content was significantly higher in the HEK293T cells containing the inversion that occurred in *E. baileyi*. Values are shown as means \pm s.d. from three biological replicates ($n = 3$); Exact P values are shown (two-sided t -test). d–f, A 2,212-bp intronic inversion in *HIF1A* is shared by the high-altitude *E. baileyi* and *E. smithi*. d, Schematic inversion located

in an intron of *HIF1A* of the high-altitude *E. baileyi* and *E. smithi*. e, Expression differences of the *HIF1A* gene (two-sided t -test) in the lung ($n = 3$). In the box plots, the centre line represents the median, whereas the box boundaries represent the upper and lower quartiles. The whisker reaches no farther than $1.5 \times \text{IQR}$. f, Luciferase signals and expression of two SV-haplotypes of *HIF1A*. Values are shown as means \pm s.d. from three biological replicates ($n = 3$); Exact P values are shown (two-sided t -test). g–i, Schematic of an 1849 bp intronic deletion located at *HSF1* in *E. baileyi* compared with other species in *Eospalax*. g, Sequence comparison with other species in Myospalacinae showed a deletion specific to *E. baileyi*. The horizontal coordinate shows the length of the gene (kb). h, Agarose gel electrophoresis of PCR amplification supports the deletion in *HSF1*. i, Results of reporter assays suggest that the deletion in the intronic regions of the *HSF1* gene leads to the upregulation of gene expression. Values are shown as means \pm s.d. from three biological replicates ($n = 3$). Exact P values are shown (two-sided t -test). Source data are provided in Extended Data Table 2.



Rearrangement on chromosome 1 in *Eospalax*. Compared with the *E. fontanierii* genome, we noticed a 50 Mb chromosomal rearrangement involving two parts with lengths of 21.6 Mb and 28.4 Mb. The chromosomal rearrangement was observed in each of the four possible homozygous or heterozygous genotypic combinations (Fig. 2d). The two lowest-altitude species of *E. fontanierii* and *E. cansus* showed the same genotype I. Although very close to *E. rufescens*, the humid-region species of *E. rothschildi* showed a specific genotype II. The third highest-altitude species of *E. rufescens* showed a unique genotype III, and the two highest-elevation species showed the same genotype IV (Fig. 4d). Notably, repeated elements were present around their breakpoints. Remarkably, the hypoxia-relevant *EGLN1* gene located in this region shifted from the chromosome terminal to the middle of the chromosome arm through inversions in all three variant genotypes (genotypes II, III and IV) in species from the top four highest elevations (Fig. 4d). To further verify the frequency of the rearrangement, we identified the breakpoint in three unrelated individuals per species on chromosome 1 (Supplementary Fig. 2). Interestingly, the species corresponding to genotypes II–IV lived in more hypoxic environments caused by higher elevation or greater soil humidity. Remarkably, most genes located in large inversions, such as *EGLN1*, have been reported to be involved in hypoxia adaptation, DNA damage repair, inflammation, immunity, apoptosis, platelets and vascular development (Fig. 4f). The number of genes located in the remaining 17 inversion regions ranged from 35 to 384.

To further clarify the potential impact of SV on the 3D chromatin architecture at compartment A/B, TADs and significant interaction levels (loop), we constructed a chromosome interaction matrix at different resolutions with homologous bins across two genomes. The GC content and gene density in compartment A were significantly higher than in compartment B (Extended Data Fig. 7a). Although the compartment types (A or B) were consistent between the high-altitude *E. baileyi* and low-altitude *E. fontanierii* across most of the genome, those within the rearrangements in chromosome 1 were less consistent (Fig. 4e). Within the chromosome 1 rearrangement, many homologous bins were identified as compartment B in *E. fontanierii* but as compartment A in *E. baileyi*, indicating altered gene expression. In *E. baileyi*, the breakpoints of this SV were located at TAD boundaries with significantly lower insulation scores than those of other nearby TAD boundaries, indicating that interactions were weakest at the inversion breakpoints (Extended Data Fig. 7b). There were more species-specific TADs in the rearrangement region than found in other regions of chromosome 1 (Extended Data Fig. 7c). We examined the genes in the *E. baileyi*-specific TADs and found that six genes (*AKR1C1*, *AKR1C18*, *AKR1C13*, *AKR1C5*, *AKR1C6* and *AKR1C15*) of the *AKR1* family were in the same TAD. These genes were reported to be related to apoptosis and autophagy³¹. Five genes (*CYP3A25*, *CYP3A1*, *CYP3A11*, *CYP3A6* and *CYP3A9*) in the *CYP3A* subfamily were also located in the

same TAD, and these reportedly catalyse cholesterol and lipid synthesis reactions^{32,33} (Extended Data Fig. 7d). Three consecutive TADs within the rearrangement of *E. fontanierii* fused into a single TAD in *E. baileyi*, and multiple genes within this TAD, including *EGLN1*, *GNPAT*, *EXOC8*, *SPRTN*, *DISC1* and *TSNAX*, are known to play a role in the cellular response to hypoxia^{12,34–36}. The *DISC1*, *TSNAX* and *GSK3B* complex was reportedly associated with DNA repair. After TAD fusion, *DISC1* and *TSNAX* entered the same TAD (Fig. 4f).

We also compared the significant chromatin interactions of *E. baileyi* and *E. fontanierii* in the rearrangement region. The results showed that there were few significant interaction regions shared by two species (Extended Data Fig. 7e). Consistent with the TAD fusion event in the rearrangement region, some TADs in the region of *E. baileyi* were expanded and interaction regions extended significantly (Fig. 4g). In summary, the chromosome rearrangement in *E. fontanierii* and *E. baileyi* showed differences in compartment type, TAD boundary and significant interaction.

Mechanisms of inversion formation

Mechanism of short inversion (<1Mb) formation by LINE-1. LINE-1 insertions can result in genomic inversions generated during retrotransposition via ‘twin priming’³⁷. Of all transposable elements identified in short inversions, 31%, 25%, 25% and 19% were SINE, LINE, LTR and DNA elements, respectively. Almost all LINEs belonged to the LINE-1 superfamily (Extended Data Fig. 8a,b). More than 200 inversions were found to consist almost entirely of LINE-1 retrotransposon sequences, suggesting that the retrotransposition of LINE-1 elements is an important mechanism of chromosomal evolution in *Eospalax*.

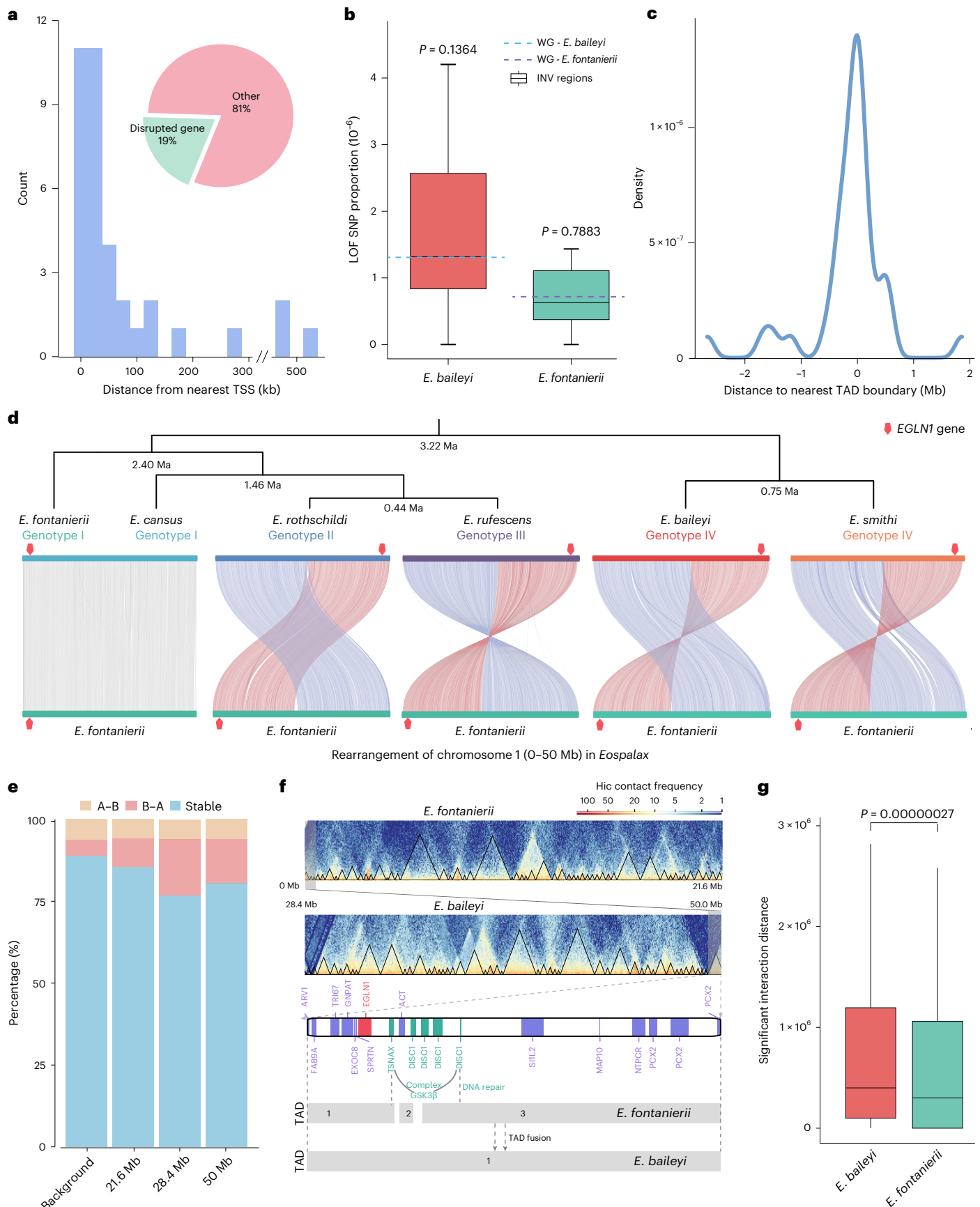
Mechanism of large inversion (>1Mb) formation by NAHR. To determine the mechanism of inversion formation in *E. baileyi*, we identified inverted repeats near each breakpoint. At least one repeat ranging from 200 bp to 70 kb in length was identified near both breakpoints in 17 of the 18 inversions (Extended Data Fig. 8c,d), suggesting that most inversions were mediated by NAHR, consistent with patterns documented in deer mice²⁵. To explore whether breakpoints were in repetitive regions, we identified segmental duplications across the genome, and we found significant enrichment of segmental duplications (75%) (Extended Data Fig. 8e) near the inversion breakpoints (1 Mb) compared with the whole genome (Kolmogorov–Smirnov test: $P < 0.0001$). In conclusion, segmental duplication-enriched regions may be particularly prone to structural change via NAHR in *Eospalax*.

Discussion

Genome-wide scan of nucleotide variation has been used extensively to identify candidate genes for hypoxia adaptation in high-altitude humans and other animals^{12,13,35,38,39}, but the role of structural genomic variation has not been assessed systematically. Here we capitalized on

Fig. 4 | Possible functional effects of large inversions (>1 Mb) and rearrangement. **a**, Distance from inversion breakpoints to the transcriptional start sites of the nearest genes. Seven (19%) of these breakpoints disrupted genes. **b**, Proportion of LOF SNP in inversion/non-inversion regions compared with that in the respective genome-wide regions. A one-sample two-sided *t*-test was used ($n = 18$ inversions). In the box plots, the centre line represents the median, whereas box boundaries represent the upper and lower quartiles. The whisker reaches no farther than $1.5 \times \text{IQR}$. **c**, Distance from the nearest TAD boundary to each inversion breakpoint in the *E. fontanierii* genome. **d**, Phylogenetic trees are constructed based on the long-read-SV dataset. The previously reported divergence time of species was plotted on phylogenetic trees. The synteny comparison showed that the rearrangement types on chromosome 1 (0–50 Mb) of *E. cansus*, *E. rothschildi*, *E. rufescens*, *E. baileyi* and *E. smithi* aligned with that of *E. fontanierii*, respectively. The terminal region with the *EGLN1* gene (red arrow) was inverted to the middle of the chromosome arm in the three high-altitude species and the humid-region species. **e**, Comparison of A/B compartments

between *E. fontanierii* and *E. baileyi* on chromosome 1. The background represents the regions of chromosome 1, excluding the area of rearrangement. ‘A–B’ means the homologous bins were compartment A in *E. fontanierii* but compartment B in *E. baileyi*. ‘B–A’ was the opposite. ‘Stable’ means the homologous bins in two genomes had the same compartment type. **f**, TAD distribution in the 21.6 Mb region of chromosome 1. Distribution of genes associated with adaptation to hypoxia in the inversion region of chromosome 1 in the *E. fontanierii* genome. The shaded area contains three consecutive TADs in the *E. fontanierii* genome, and the homologous bin appears as a single TAD in the *E. baileyi* genome. After TAD fusion, *TSNAX* and *DISC1* were located in the same TAD. **g**, Comparison of the significant cis interaction distance between *E. fontanierii* ($n = 1,002$) and *E. baileyi* ($n = 1,771$) in the rearrangement region. A two-sided *t*-test was used. In the box plots, the centre line represents the median, whereas box boundaries represent the upper and lower quartiles. The whisker reaches no farther than $1.5 \times \text{IQR}$.



technological advances in long-read sequencing to investigate how SVs may contribute to hypoxia adaptation in *Eospalax*.

The high-quality, chromosome-level genome assembly of *E. baileyi* enabled us to identify SVs, gene family expansions and PSGs. The genes we identified (Fig. 1f and Supplementary Tables 3 and 4) have been reported to play important roles in hypoxia adaptation^{30,40}. In *E. baileyi*, significantly expanded lungs were suggested to maximize the pulmonary diffusing capacity for respiratory gases with proliferated alveolar and increased surface area¹, suggesting more efficient oxygen absorption^{1,41–43}.

EGLN1 and *HIF1A* are known to play important roles in cellular responses to hypoxia, and mutations within these genes (even in introns) are reported to have functional significance^{44,45}. Intronic inversions in *EGLN1* and *HIF1A* in high-altitude *Eospalax* were associated with significant upregulation in the lung. Increased *EGLN1* gene expression could result in increased *PHD2* expression, which could blunt the expression of *HIF1A* in conditions of chronic hypoxia⁴⁶. *HSF1* was a key transactivator of stress-inducible genes, regulating cellular and oxidative stress, immune, autophagy and other such responses to hypoxia^{47,48}. Moreover, *E. baileyi*-specific intronic deletion of the *SFTPD* genes was possibly involved in response to hypoxia. *XRCC4* and *UVRAG* genes were associated with DNA damage. Thus, these results suggested that SVs could affect gene expression and take part in hypoxia response. Additionally, *E. baileyi* has been divided into two lineages. *E. baileyi2* split from *E. baileyi1* at ~1.18 Ma earlier than the split of *E. smithi* at ~0.75 Ma. We speculate that the shared SVs may account for the same phenotype in *E. baileyi* and *E. smithi*. We think the adaptations evolved in the ancestry of the two highland species.

We speculate that the rearrangement in chromosome 1 occurred to avoid silencing by telomeres^{49,50}, thereby maintaining the required gene expression under hypoxia. Heterochromatin has been reported to be dynamic at telomeres⁵¹, which could alter adjacent nucleosomal chromatin to silence the expression of some nearby genes. The 3D chromosome model of chromosome 1 suggested that rearrangement was more tightly structured in *E. fontanierii* than *E. baileyi* (Supplementary Fig. 3). More compartment transformation from B to A in *E. baileyi* in the rearrangement suggested that genes within these regions might be more highly expressed than that in *E. fontanierii*³⁹. This rearrangement contributed to a fusion of TADs and expansions of interacting regions in *E. baileyi* (Fig. 4f,g), suggesting changes could alter patterns of some gene expression. Although we speculate that large rearrangements harbouring important genes such as *EGLN1* should have functions facilitating hypoxia adaptation, functional tests of even whether they were fixed require further investigation.

Long-read sequencing enabled us to identify SVs and breakpoint regions, which were mainly distributed in repeated regions. A large proportion of short SVs and 56% (10 out of 18) of large SVs occurred at chromosomal ends (Fig. 2c), possibly because telomeric concentrations of repetitive elements facilitated NAHR^{25,52}. Alternatively, inversions with breakpoints near telomeres could be less likely to be removed by purifying selection because these regions are relatively sparse in genes^{25,52}. Large inversions mainly originated via inverted repeats through NAHR in *Eospalax*, similar to other mammals^{25,53}.

In this study, the patterns of genomic structural variation in *E. baileyi* and the documented effects on the expression of hypoxia-relevant genes suggest numerous hypotheses about possible mechanisms of hypoxia adaptation. Such hypotheses can now be tested using genome-editing tools in conjunction with experimental studies of organismal physiology.

Methods

Sample information

We collected 17 specimens of *Eospalax* in northern China: 3 *E. baileyi*, 3 *E. fontanierii*, 3 *E. cansus*, 3 *E. rufescens*, 3 *E. rothschildi* and 2 *E. smithi* (Supplementary Table 9). Live zokors were caught and euthanized

using chloral hydrate. Tissues were collected and preserved in liquid nitrogen.

Phenotypic data of *Eospalax*

We measured the mass of the body and tissues for all specimens. We did a linear regression analysis of the body and lung mass for all samples. Blood was collected with EDTA as anticoagulant. RBC and Hb were measured using Mindray BC-2800Vet. Three replicates were measured for each individual. The data for *E. baileyi* came from ref. 3.

Fibroblast culture and karyotype analysis

Skin fibroblast culture. The skin of zokor was sterilized with 75% alcohol, washed with PBS, clipped by adding a DMEM/F12 medium and digested with collagenase at 37 °C for 1 h. After centrifugation at 3,700g for 10 min, the supernatant was poured off and the precipitate was mixed with DMEM/F12 medium using 10% FBS, and it was spread out flat in a Petri dish.

Cellular chromosome karyotype analysis. Vigorous growing cells with 70–80% confluence were selected and treated with 0.1 µg ml⁻¹ of colchicine for 1 h and 0.075 mol l⁻¹ of KCl solution at low osmolality. Trypsin was used to digest the cells and then added with a newly configured fixative (methanol:glacial acetic acid = 3:1), and the fixation was repeated three times to make the slides, which were placed in the oven for ageing. The slides were stained with Giemsa stain for 15 min, and the intermediate schizogony phases with high sorting index and well-dispersed chromosomes were selected for karyotyping through microscopic examination.

DNA extraction and library construction

Genomic DNA was extracted from muscle using SDS-based methods. The purity, concentrations and integrity of genomic DNA were measured using a Nanodrop spectrophotometer and Qubit 3.0 fluorometer. After the samples passed quality control, a BluePippin automated nucleic acid recovery instrument (Sage Sciences) was used to collect the target fragments (10–50 kb). AMPure XP bead purification, DNA damage repair, DNA end-repair and adaptor ligation (SQK-LSL109 Ligation Sequencing Kit) were performed to construct sequencing libraries. The Qubit fluorometer was used to quantify the sequencing-ready DNA libraries.

Data generation

Long-read sequencing of *E. baileyi* using Oxford Nanopore Technologies (ONT, 209 Gb) was performed to assemble contigs; a 51× coverage of short paired-end sequencing was performed on the Illumina platform for subsequent read correction and polishing. The assembled contigs were anchored to chromosomes using Hi-C data (315 Gb, Extended Data Table 1).

Standard operating procedures were followed to load the long-read sequencing libraries onto PromethION R9.4.1 flow cells for the long-read–SV dataset (Supplementary Table 10). The number of reads and N50 length for raw long-read data were calculated using NanoPlot v.1.38.0.

Short reads from 146 *Eospalax* individuals were utilized for SV frequency estimation (Supplementary Table 11), and 143 of the 146 were downloaded from public databases^{3,27,54}, including 55 *E. baileyi*. Of these, 40 belonged to *E. baileyi1* and 15 belonged to *E. baileyi2*.

De novo genome assembly and annotation

Genome assembly. The NextCorrect module of NextDenovo v.2.3.0 was used to obtain the consistency sequence using clean long reads from the ONT platform. NextGraph was used for the initial assembly of the genome. NextPolish was then used to subject the assembly to two rounds of corrections using Illumina reads and three rounds of self-polish corrections using ONT data. Mitochondrial sequences

identified using blastn (blast v.2.2.29) were removed. BUSCO v.5.2.2 was used to assess the genome integrity.

The paired-end Hi-C reads were mapped to the assembled contigs using BWA-MEM v.0.7.17-r1188. Chromosomes were organized, ordered and oriented using automated 3D-DNA v.180922 and Juicer v.1.6 processes. Juicebox v.1.5.1 was used to correct the chromosomal-level genome.

Repetitive DNA annotation. LTR_FINDER, Tandem Repeat Finder (TRF) v.4.07b, RepeatModeler v.4.07b, RepeatMasker v.1.331 and Repeat-ProteinMask v.4.1.0 were used to identify repetitive sequences. The overlapping transposable elements belonging to the same repeat class within the genome's repetitive sequences were sorted and merged.

Gene annotation. Combining ab initio predictions and homology and transcript evidence, we generated models for zokors' protein-coding genes. AUGUSTUS v.3.3.1 and Genscan gene models based on repeat-masked genomes were annotated for ab initio prediction. The protein sequences of *Mesocricetus auratus*, *Rattus norvegicus*, *Microtus ochrogaster*, *E. fontanierii* and *E. baileyi* were used as templates in GeMoMa v.1.6.1. Exonerate v.2.4.0 (<http://www.ebi.ac.uk/~guy/exonerate/>) was utilized to align homologous protein sequences from the UniProt database to the *E. baileyi* genome. RNA-seq reads were mapped to genomes using HISAT2 v.2.1.0 and re-assembled using StringTie v.2.2.1. Program Assemble Spliced Alignment v.2.3.3 was subsequently used to generate transcript predictions. TransDecoder v.2.0 (<http://transdecoder.github.io>) was then used to predict the genes. Evidence Modeler v.1.1.1 was used to integrate all predictions. Using BUSCO, the completeness of gene annotations was assessed.

Analysis of gene family expansions and contractions

We downloaded the genome and annotation files of *E. fontanierii*, *Rhizomys pruinosus*, *Spalax galili*, *Rattus norvegicus*, *Mus musculus*, *Peromyscus maniculatus*, *Mesocricetus auratus* and *Jaculus jaculus* from the database of NCBI. Together with the *E. baileyi* assembly, nine species were used to examine lineage-specific changes in gene family size. Orthofinder v.2.5.4 was used to identify orthogroups among nine species based on the longest transcripts extracted with a custom Perl script. We generated a tree with divergence times based on the species tree and single-copy gene sequences using MCMCTree in PAML v.4.10.3. The species tree was calibrated using TIMETREE (<http://www.timetree.org/>) fossil times. The contracted and expanded gene families were identified using CAFÉ v.4.2.1 with orthogroup counts and a tree containing divergence time as input files. The likelihood approach modelled the evolution of gene family size using a stochastic birth and death process in CAFÉ v.4.2.1. We performed a multiple correction of *P* value (FDR < 0.05; Extended Data Table 1).

Identification of genes subject to selection

We identified PSGs in *E. baileyi* based on homologous single-copy genes from nine species. The codon sequences of each homologous single-copy gene were extracted for subsequent analysis using a Perl script (Epal2nal.pl, <https://github.com/wonaya/ParaAT>) based on the amino acid and CDS sequences. The selected genes were identified using the 'aBSREL' parameter of HyPhy (<https://www.hyphy.org/>), with an omega value greater than 1 indicating PSGs and an omega value less than 1 indicating negatively selected genes. *P* values were multiples corrected by FDR values (<0.05).

SV (<1 Mb) calling, description and verification

Illumina SV calling. Quality control was conducted on the sequencing data using FastQC. The clean short reads were mapped against the *E. fontanierii* genome using BWA-MEM v.0.7.17-r1188. PCR duplicates were removed using Picard (<https://broadinstitute.github.io/picard>) with default parameters. Three SV-detecting tools (DELLY v.0.7.6, LUMPY

v.0.2.13 and Manta v.1.6.0) with respective default parameters were used to call SVs in each individual. Only SVs with the 'PASS' and 'PRECISE' flags using all three approaches were kept and merged in svimmer (<https://github.com/DecodeGenetics/svimmer>). GraphTyper2 was used to refine the genotype results based on the alignment information in BAM files. Sites with a high ratio of missing data across all samples (--max missing 0.90) were filtered out using VCFtools v.0.1.16.

Nanopore SV calling. We mapped raw long reads to the *E. fontanierii* genome using NGM-LR v.0.2.7 with default parameters. SAMtools v.1.7 was used to sort and convert alignments into BAM format and calculate depth and coverage. SVs longer than 50 bp were first called in each sample using Sniffles v.1.0.12 with the option '-l50'. After merging using surpyvor v.0.10.0, genotypes at each SV across all samples were identified again using Sniffles. The sequences of insertions and deletions were added and polished using Iris v.1.0.5. We merged SVs using surpyvor again, which was based on the comparison algorithm of SURVIVOR v.1.0.3. Complex SVs (that is, SVs with multiple types, overlapping SVs and SVs longer than 1 Mb) were considered as false negatives and removed. SVs with a missing rate greater than 0.15 across all samples were also removed for carrying insufficient genotype information.

Genomic region and repeat annotation for SVs. The VCF format file with SV information was converted to bed format to get the genomic coordinates of each SV. The intersection between repetitive elements and SVs was then calculated using the intersect method in BEDtools v.2.29.2. The VCF format file of the long-read-SV dataset contained 17 individuals. The VCF format file of the SR-SV dataset had 146 individuals (55 *E. baileyi*, 25 *E. cansus*, 21 *E. smithi*, 18 *E. rothschildi*, 14 *E. rufescens* and 13 *E. fontanierii*).

Gene function of associated SVs. Using a custom Python script, the frequencies of each SV were calculated among all species. SVs fixed in one species and not identified in others were defined as species-specific SVs. Notably, *E. fontanierii* was used as the reference genome; thus, *E. fontanierii*-specific SVs were identified as those SVs that were fixed in all other samples except those from *E. fontanierii*. We focused on species-specific and species-shared SV-associated genes in this study. Genes associated with these SVs were then extracted from the annotation files in each species. The functional enrichments of genes associated with species-specific SVs were performed using Metascape.

PCR validation of SVs. Genomic DNA was extracted from the muscles of *E. fontanierii* and *E. baileyi*. PCR reactions were performed using a 2× Taq PCR Master mix (TIANGEN). The *HSF1* gene deletion was confirmed via PCR genotyping using the following primer pairs: (F: CCTGCCCTCTCACTGAGTTG, R: TCCGGGTCTTCTGTTCTGTC).

Large inversion (>1 Mb)

Identification and verification of large SVs. The chromosome-level genome of *E. baileyi* was compared with that of *E. fontanierii* using the nucmer model of MUMmer v.4.0.0 with default parameters. Alignments shorter than 100 bp were filtered using the delta-filter, and their coordinates were obtained using the show-coords module. SyRI v.1.5.6 was used to identify large inversions (>1 Mb).

We aligned the contig-level assemblies of *E. baileyi* to the *E. fontanierii* genome using nucmer with default parameters. Inversion breakpoints were identifiable if (1) a contig contains the entire inversion, with the inversion sequence aligned in reverse to the reference genome, and (2) a contig includes only a partial region of the inversion, with a sequence from one end of the contig to the inversion breakpoint aligned in the same direction as the reference genome and a sequence from the other end of the contig to the inversion breakpoint aligned in reverse to the reference genome. The breakpoints of each inversion were identified based on the alignments at the contig level.

Genomes from two additional individuals of *E. baileyi* were de novo assembled, and one public scaffold-level genome³ was used to verify the SV on chromosome 1. First, we used NextDenovo to de novo assemble the ONT sequencing data to obtain contig-level genomes. Then, we used the 'correct' module of Ragtag v.2.0.1 to scaffold the contigs using the chromosome-level *E. baileyi* genome as reference. Then, the same method was used to identify the presence of SV on chromosome 1 in both genomes.

Predicting telomere locations. We used tolkit v.0.2.2 (<https://github.com/tolkit/telomeric-identifier>) to identify telomeres in the *E. fontanierii* genome.

Inverted repeat and segmental duplications at inversion breakpoints. We identified segmental duplications near inversion breakpoints using SEDEF with the soft-masked repeat sequence as input. Segmental duplications were then screened for length ≥ 1 kb and soft-masked repeat sequences $< 70\%$. We calculated the segmental duplication density around 500 kb of each inversion breakpoint. To compare the density at the breakpoints with the density in random regions of the whole genome, we randomly selected 1,000 sites from the genome and calculated the segmental duplication density in the region 500 kb upstream and downstream from each site. Inverted repeats were identified in three ways: (1) segmental duplications (> 1 kb) located near two breakpoint regions of an SV that were inverted and highly similar were regarded as inverted repeats; (2) IRF v.3.05 was used to identify inverted repeats longer than 500 bp; (3) short (> 100 bp) inverted repeats were obtained by aligning each chromosome to itself.

Potential effect of large inversions. We checked whether any inversion breakpoints were located within the genes. We calculated the distance between each inversion breakpoint and the nearest transcriptional start site. LOF SNPs were utilized to represent genetic load. SnpEff v.5.1 was used to identify LOF SNPs (stop gained, start lost, splice acceptor variant and splice donor variant). We calculated the number of LOF SNPs in each inverted region of *E. fontanierii* and *E. baileyi*, in addition to the overall number of LOF SNPs found in their genomes.

Construction of homologous bin pairs. To identify homologous bins between *E. fontanierii* and *E. baileyi*, we used LAST v.885 to align the query *E. baileyi* sequence to the target *E. fontanierii* sequence. According to published pipelines⁵⁵, the MAF file obtained from LAST was converted into a bin pairs file. Using the University of California, Santa Cruz genome browser, the MAF file was converted to PSL format using mafToPsl and then converted to a chain file using axtChain (-linearGap = medium). The liftOver tool (-minMatch = 0.85) was used to construct homologous bin pairs of varying sizes.

Comparison of 3D chromatin architecture. We compared the 3D chromatin architecture in different regions (within the chromosome rearrangement region and in other regions) of *E. baileyi* and *E. fontanierii*, including the A/B compartments, TADs and significant interaction bins.

Principal component analysis was performed using the reciprocal matrix at 150 kb resolution. The A and B compartments were separated along PC1, with positive and negative values for A and B compartments, respectively. The gene density was higher in the A compartment than in the B compartment. The homologous bins of the two species were aligned, and the counterpart A/B compartments in each genome were identified. The homologous bins were identified as switched regions if the two bins showed different compartments, whereas the remaining bins were regarded as stable regions. The ratio of the two region types was counted (A to B: A in *E. fontanierii* but B in *E. baileyi*; B to A: B in *E. fontanierii* but A in *E. baileyi*; stable: bins in both genomes contained the same compartment type).

We used cworld-dekker (<https://github.com/dekkerlab/cworld-dekker/releases>) to identify TAD boundaries based on the 40 kb resolution matrix and obtained an insulation score. For a given TAD in one genome, we identified all bins that showed overlap with this TAD. Counterpart TADs between the two species with more than 70% overlap were defined as shared TADs, and all other TADs were classified as species-specific TADs.

Finally, the significance of genome-wide interactions between each pair of loci (at 20 kb resolution) was determined using FitHic v.1.1.3. Interactions with $P \leq 0.01$, $q \leq 0.01$, and > 9 reads supporting the interactions were significant interactions. We limited our analysis to significant interaction regions within the same chromosomes. The bins of significant interaction regions were determined according to their coordinates. We then tallied the number of shared and species-specific bins between the two genomes and calculated the distance between the interaction bins.

RNA-seq analysis

The transcriptomes of the hearts, livers and lungs of *E. baileyi*, *E. cansus* and *E. rothschildi* from common garden experiments were sequenced. We collected the zokors on the way from our university to our field station located at Guoluo, Qinghai Province, with an altitude of ~ 3800 m. All the male adults were in the same environment there for at least 1 week. We anaesthetized the zokor with chloral hydrate, collected tissue and froze it in liquid nitrogen and stored it in a -80°C refrigerator. We extracted RNA using RNAPrep pure Tissue Kit (TIANGEN). NanoPhotometer (IMPLEN) was used to detect RNA purity. The RNA Nano 6000 Assay Kit of the Bioanalyzer 2100 system (Agilent Technologies) was used to assess RNA integrity and concentration. Then $1-2\ \mu\text{g}$ of RNA per sample was used to construct the cDNA library. Qubit3.0 was used for preliminary quantification. Agilent Bioanalyzer 2100 was used to detect the library insert size. StepOnePlus Real-Time PCR System was used to accurately quantitate the insert size (library valid concentration > 10 nM). *E. baileyi* was used as the high-altitude species, whereas *E. cansus* and *E. rothschildi* were the low-altitude species. The raw reads were filtered using Trim Galore to remove low-quality reads, and the filtered data were assessed using FastQC. HISAT2 was used to align clean reads to the genome. We calculated the number of transcripts of each gene mapped against the genome using Stringtie v.2.0.4.

DESeq v.2 was used to identify DEGs (adjusted $P \leq 0.05$). We used the result of multiple corrections (Benjamini–Hochberg model). We corrected RNA-seq data using phylogenetic comparative methods, such as phylogenetic generalized least squares. We increased statistical power and controlled the FDR using mashr (<https://github.com/stephenslab/mashr>; Extended Data Table 2). Metascape was used for functional enrichments.

Dual-luciferase assay

HEK293T cells were cultured using an RPMI1640 medium supplemented with penicillin, streptomycin and horse serum. The synthesized sequences of *HIF1A* (Ensembl: Chr13: 44667409–44669626), *EGLN1* (Ensembl: Chr1: 1889355–1890797) and *HSF1* (Ensembl: Chr9: 81333952–81335801) were inserted into pGL13-Promoter (Promega). We prepared and transformed competent cells and then extracted plasmid DNA. Both the plasmid with the target gene (pGL13) and the control plasmid (pRL-TK) were transfected into HEK293T cells using Lipofectamine 2000 (Thermo Fisher Scientific) and OPTI-MEM medium (GIBCO, Invitrogen). Then 24 h after transfection, cellular lysates were collected. The ratio of FLUC and RLUC values (that is, luciferase activity) was calculated per well. All data were presented as means with standard deviations, and statistical differences between groups were examined using one-way analysis of variance and Tukey's test, with a P value less than 0.05 indicating significance, and we used a two-sided t -test.

Prediction of *SFTPD* protein structure

We used RoseTTAFold with default parameters to predict the protein structure of *SFTPD* of *E. fontanierii* and *E. baileyi*. PyMOL (<https://pymol.org>) was used to visualize the predicted results.

3D model reconstruction of genomes

The 'LorDG' module of Genomeflow (<https://github.com/jianlin-cheng/GenomeFlow>) was used to build 3D chromosomes and genome models. Then we viewed the 3D structure using PyMOL.

Ethics statement

The Ethics Committee of the College of Ecology, Lanzhou University (ethical approval form no. EAF2022007) authorized all experimental methods and sample collection protocols. China's Ministry of Science and Technology developed the 'Guidelines on the Ethical Treatment of Experimental Animals', which governed all sample operations.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

DNA sequence data are available from NCBI SRA under BioProject accession [PRJNA947478](https://ncbi.nlm.nih.gov/bioproject/PRJNA947478). The RNA sequence data reported in this paper are deposited in the Genome Sequence Archive in National Genomics Data Center, China National Center for Bioinformation/Beijing Institute of Genomics, Chinese Academy of Sciences ([PRJCA019941: CRA012698](https://ngdc.cncb.ac.cn/gsa)), which are publicly accessible at <https://ngdc.cncb.ac.cn/gsa>. The chromosomal level genome assembly and annotation reported in this paper are deposited in the China National GeneBank DataBase (<https://db.cngb.org>) with BioProject ID CNP0005002 and uploaded in figshare (<https://doi.org/10.6084/m9.figshare.24085119>). Source data are provided with this paper.

Code availability

The code used for the analyses is available from GitHub (<https://github.com/anxuan-web/Structural-variants-in-Eospalax>).

References

- Chiou, K. L. et al. Genomic signatures of high-altitude adaptation and chromosomal polymorphism in geladas. *Nat. Ecol. Evol.* **6**, 630–643 (2022).
- Cai, Z. et al. Adaptive transcriptome profiling of subterranean Zokor, *Myospalax baileyi*, to high-altitude stresses in Tibet. *Sci. Rep.* **8**, 4671 (2018).
- Zhang, T. et al. Phenotypic and genomic adaptations to the extremely high elevation in plateau zokor (*Myospalax baileyi*). *Mol. Ecol.* **30**, 5765–5779 (2021).
- Simonson, T. S. et al. Genetic evidence for high-altitude adaptation in Tibet. *Science* **329**, 72–75 (2010).
- Yi, X. et al. Sequencing of 50 human exomes reveals adaptation to high altitude. *Science* **329**, 75–78 (2010).
- Peng, Y. et al. Genetic variations in Tibetan populations and high-altitude adaptation at the Himalayas. *Mol. Biol. Evol.* **28**, 1075–1081 (2011).
- Peng, Y. et al. Down-regulation of *EPAS1* transcription and genetic adaptation of Tibetans to high-altitude hypoxia. *Mol. Biol. Evol.* **34**, 818–830 (2017).
- Huerta-Sánchez, E. et al. Genetic signatures reveal high-altitude adaptation in a set of Ethiopian populations. *Mol. Biol. Evol.* **30**, 1877–1888 (2013).
- Schweizer, R. M. et al. Physiological and genomic evidence that selection on the transcription factor *Epas1* has altered cardiovascular function in high-altitude deer mice. *PLoS Genet.* **15**, e1008420 (2019).
- Schweizer, R. M. et al. Broad concordance in the spatial distribution of adaptive and neutral genetic variation across an elevational gradient in deer mice. *Mol. Biol. Evol.* **38**, 4286–4300 (2021).
- Scott, G. R. et al. Adaptive modifications of muscle phenotype in high-altitude deer mice are associated with evolved changes in gene regulation. *Mol. Biol. Evol.* **32**, 1962–1976 (2015).
- Storz, J. F. High-altitude adaptation: mechanistic insights from integrated genomics and physiology. *Mol. Biol. Evol.* **38**, 2677–2691 (2021).
- Storz, J. F. & Cheviron, Z. A. Physiological genomics of adaptation to high-altitude hypoxia. *Annu. Rev. Anim. Biosci.* **9**, 149–171 (2021).
- McClelland, G. B. & Scott, G. R. Evolved mechanisms of aerobic performance and hypoxia resistance in high-altitude natives. *Annu. Rev. Physiol.* **81**, 561–583 (2019).
- Storz, J. F. & Scott, G. R. Life ascending: mechanism and process in physiological adaptation to high-altitude hypoxia. *Annu. Rev. Ecol. Syst.* **50**, 503 (2019).
- Storz, J. F. et al. Evolution of physiological performance capacities and environmental adaptation: insights from high-elevation deer mice (*Peromyscus maniculatus*). *J. Mammal.* **100**, 910–922 (2019).
- Stewart, N. B. & Rogers, R. L. Chromosomal rearrangements as a source of new gene formation in *Drosophila yakuba*. *PLoS Genet.* **15**, e1008314 (2019).
- Shi, J. et al. Structural variants involved in high-altitude adaptation detected using single-molecule long-read sequencing. *Nat. Commun.* **14**, 8282 (2023).
- Quan, C. et al. Characterization of structural variation in Tibetans reveals new evidence of high-altitude adaptation and introgression. *Genome Biol.* **22**, 159 (2021).
- Almarri, M. A. et al. Population structure, stratification, and introgression of human structural variation. *Cell* **182**, 189–199. e115 (2020).
- Mérot, C. et al. A roadmap for understanding the evolutionary significance of structural genomic variation. *Trends Ecol. Evol.* **35**, 561–572 (2020).
- Villoutreix, R. et al. Large-scale mutation in the evolution of a gene complex for cryptic coloration. *Science* **369**, 460–466 (2020).
- Yan, Z. et al. Evolution of a supergene that regulates a trans-species social polymorphism. *Nat. Ecol. Evol.* **4**, 240–249 (2020).
- Hager, E. R. et al. A chromosomal inversion contributes to divergence in multiple traits between deer mouse ecotypes. *Science* **377**, 399–405 (2022).
- Harrington, O. S. & Hoekstra, H. E. Chromosomal inversion polymorphisms shape the genomic landscape of deer mice. *Nat. Ecol. Evol.* **6**, 1–15 (2022).
- Shams, I. et al. Oxygen and carbon dioxide fluctuations in burrows of subterranean blind mole rats indicate tolerance to hypoxic-hypercapnic stresses. *Comp. Biochem. Physiol. A* **142**, 376–382 (2005).
- Liu, X. et al. Genomic insights into zokors' phylogeny and speciation in China. *Proc. Natl Acad. Sci. USA* **119**, e2121819119 (2022).
- Storz, J. F. & Bautista, N. M. Altitude acclimatization, hemoglobin-oxygen affinity, and circulatory oxygen transport in hypoxia. *Mol. Asp. Med.* **84**, 101052 (2022).
- Partha, R. et al. Subterranean mammals show convergent regression in ocular genes and enhancers, along with adaptation to tunneling. *eLife* **6**, e25884 (2017).
- Davies, K. T. J. et al. Limited evidence for parallel molecular adaptations associated with the subterranean niche in mammals: a comparative study of three superorders. *Mol. Biol. Evol.* **35**, 2544–2559 (2018).
- Chang, L. et al. Akr1c1 connects autophagy and oxidative stress by interacting with sqstm1 in a catalytic-independent manner. *Acta Pharmacol. Sin.* **43**, 703–711 (2022).

32. Bochud, M. et al. Association of CYP3A5 genotypes with blood pressure and renal function in African families. *J. Hypertens.* **24**, 923–929 (2006).
33. Martinez, C. et al. Expression of paclitaxel-inactivating CYP3A activity in human colorectal cancer: implications for drug therapy. *Br. J. Cancer* **87**, 681–686 (2002).
34. Jain, I. H. et al. Genetic screen for cell fitness in high or low oxygen highlights mitochondrial and lipid metabolism. *Cell* **181**, 716–727. e711 (2020).
35. Brutsaert, T. D. et al. Association of *EGLN1* gene with high aerobic capacity of Peruvian Quechua at high altitude. *Proc. Natl Acad. Sci. USA* **116**, 24006–24011 (2019).
36. Chien, T. et al. GSK3 β negatively regulates TRAX, a scaffold protein implicated in mental disorders, for NHEJ-mediated DNA repair in neurons. *Mol. Psychiatry* **23**, 2375–2390 (2018).
37. Ostertag, E. M. & Kazazian, H. H. Twin priming: a proposed mechanism for the creation of inversions in L1 retrotransposition. *Genome Res.* **11**, 2059–2065 (2001).
38. Yang, J. et al. Genetic signatures of high-altitude adaptation in Tibetans. *Proc. Natl Acad. Sci. USA* **114**, 4189–4194 (2017).
39. Hu, L. et al. Arctic introgression and chromatin regulation facilitated rapid Qinghai-Tibet Plateau colonization by an avian predator. *Nat. Commun.* **13**, 6413 (2022).
40. Davies, K. T. et al. Family wide molecular adaptations to underground life in African mole-rats revealed by phylogenomic analysis. *Mol. Biol. Evol.* **32**, 3089–3107 (2015).
41. Hsia, C. C. et al. Enhanced alveolar growth and remodeling in guinea pigs raised at high altitude. *Respir. Physiol. Neurobiol.* **147**, 105–115 (2005).
42. Llapur, C. J. et al. Increased lung volume in infants and toddlers at high compared to low altitude. *Pediatr. Pulmonol.* **48**, 1224–1230 (2013).
43. Li, J. et al. A new homotetramer hemoglobin in the pulmonary surfactant of plateau zokors (*Myospalax Baileyi*). *Front. Genet.* **13**, 824049 (2022).
44. Cimmino, F. et al. HIF-1 transcription activity: HIF1A driven response in normoxia and in hypoxia. *BMC Med. Genet.* **20**, 37 (2019).
45. Aggarwal, S. et al. *EGLN1* involvement in high-altitude adaptation revealed through genetic analysis of extreme constitution types defined in Ayurveda. *Proc. Natl Acad. Sci. USA* **107**, 18961–18966 (2010).
46. Parathath, S. et al. Hypoxia is present in murine atherosclerotic plaques and has multiple adverse effects on macrophage lipid metabolism. *Circ. Res.* **109**, 1141–1152 (2011).
47. Mendillo, M. L. et al. HSF1 drives a transcriptional program distinct from heat shock to support highly malignant human cancers. *Cell* **150**, 549–562 (2012).
48. Santagata, S. et al. High levels of nuclear heat-shock factor 1 (HSF1) are associated with poor prognosis in breast cancer. *Proc. Natl Acad. Sci. USA* **108**, 18378–18383 (2011).
49. Gottschling, D. E. et al. Position effect at *S. cerevisiae* telomeres: reversible repression of Pol II transcription. *Cell* **63**, 751–762 (1990).
50. Nimmo, E. R. et al. Telomere-associated chromosome breakage in fission yeast results in variegated expression of adjacent genes. *EMBO J.* **13**, 3801–3811 (1994).
51. Wang, J. et al. A heterochromatin domain forms gradually at a new telomere and is dynamic at stable telomeres. *Mol. Cell. Biol.* **38**, e00393–17 (2018).
52. Samonte, R. V. & Eichler, E. E. Segmental duplications and the evolution of the primate genome. *Nat. Rev. Genet.* **3**, 65–72 (2002).
53. Porubsky, D. et al. Recurrent inversion polymorphisms in humans associate with genetic instability and genomic disorders. *Cell* **185**, 1986–2005. e1926 (2022).
54. Zhang, T. et al. 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.* **43**, 331–342 (2022).
55. Kielbasa, S. M. et al. Adaptive seeds tame genomic sequence comparison. *Genome Res.* **21**, 487–493 (2011).

Acknowledgements

We thank members of the Hoekstra laboratory for commenting on the paper. We thank X. Luo from the Kunming Institute of Zoology. This project was supported by the National Natural Science Foundation of China (grants 32271691 and 32071487 to K.L.), the National Key Research and Development Programs (grant 2021YFD1200901 to K.L.), the Fundamental Research Funds for Central Universities, LZU (grants lzujbky-2021-ey17 to K.L. and lzujbky-2022-it01 to Y.W.), the Science Fund for Creative Research Groups of Gansu Province (grant 21JR7RA533 to K.L.), Lanzhou University's 'Double First-Class' Guided Project-Team Building Funding-Research Startup Fee for K. Li, a grant from State Key Laboratory of Herbage Improvement and Grassland Agro-ecosystems (Lanzhou University) (grant SKLGAE-202001, -202009 and -202010 to K.L.) and the Key Basic Research Project of Qinghai Provincial Department of Science and Technology (grant 2022-ZJ-733 to Q.X.). 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 contributions

X.A., X. Liu., Z.Q. and N.W. collected the samples. X.A., L.M., Y.W., Q.X., X. Liu., S.Z., Z.Q., B.L., F.L., Z.K., N.W., X. Liang., Q.D., Z.F., X.Y. and S.L. performed the data analyses. X.A., E.N., J.L., J.F.S. and K.L. wrote the paper. K.L. designed this study. All authors read, revised and approved the final paper.

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41559-023-02275-7>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41559-023-02275-7>.

Correspondence and requests for materials should be addressed to Jianquan Liu, Jay F. Storz or Kexin Li.

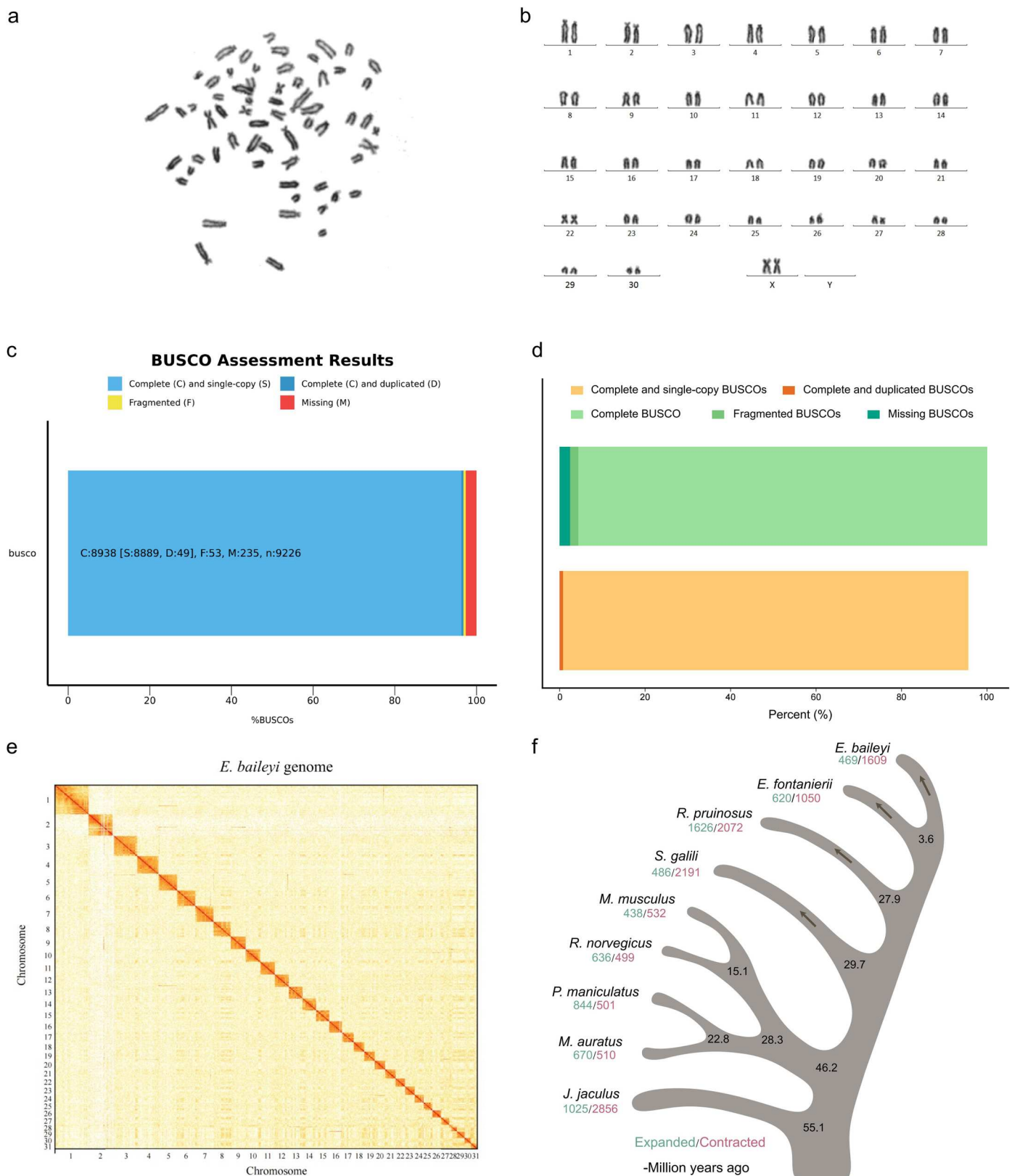
Peer review information *Nature Ecology & Evolution* thanks Erica Heinrich and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permissions information is available at www.nature.com/reprints.

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

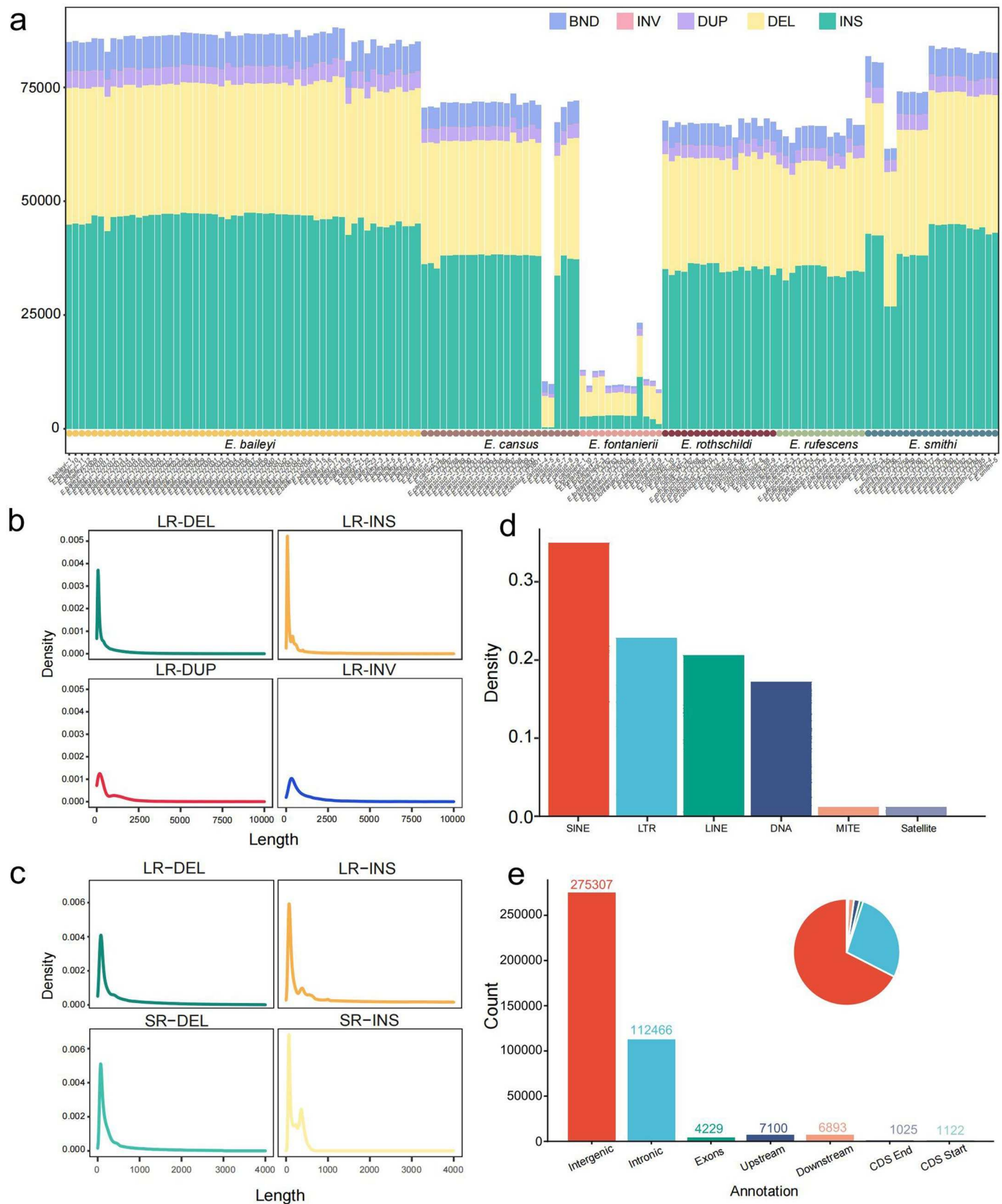
Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© The Author(s), under exclusive licence to Springer Nature Limited 2024



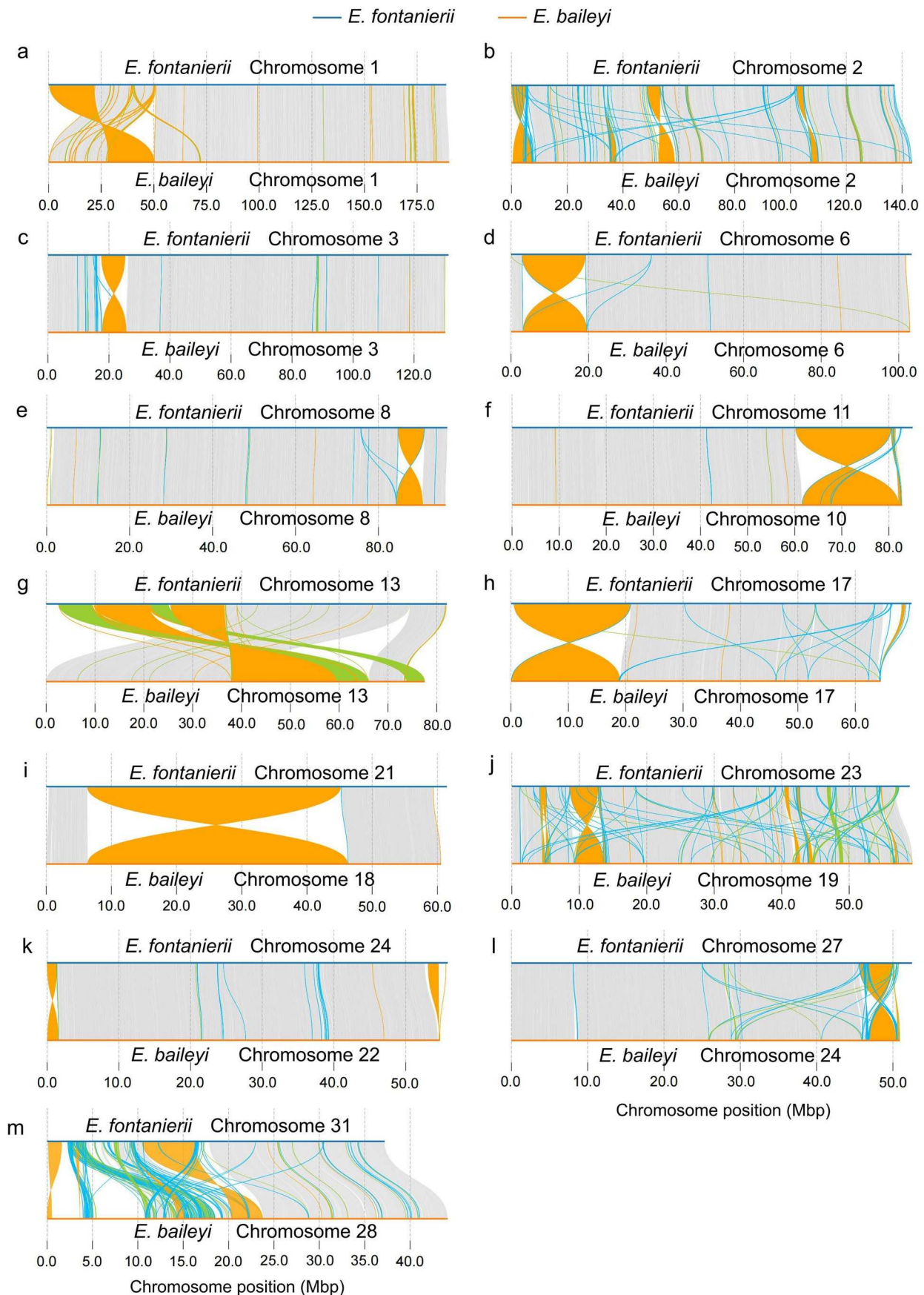
Extended Data Fig. 1 | Genome assembly and annotation of *E. baileyi* (plateau zokor v3.0). (a) Full karyotype of a female *E. baileyi*. (b) Example G-banded chromosome spread with 62 counted chromosomes. (c–d) Genome assembly and annotated genes completeness were evaluated by BUSCO. (e) Heatmap

of the *E. baileyi* chromosome-level genome interaction matrix. (f) MCMCTree was constructed with nine rodent species. Arrows indicate species living underground.

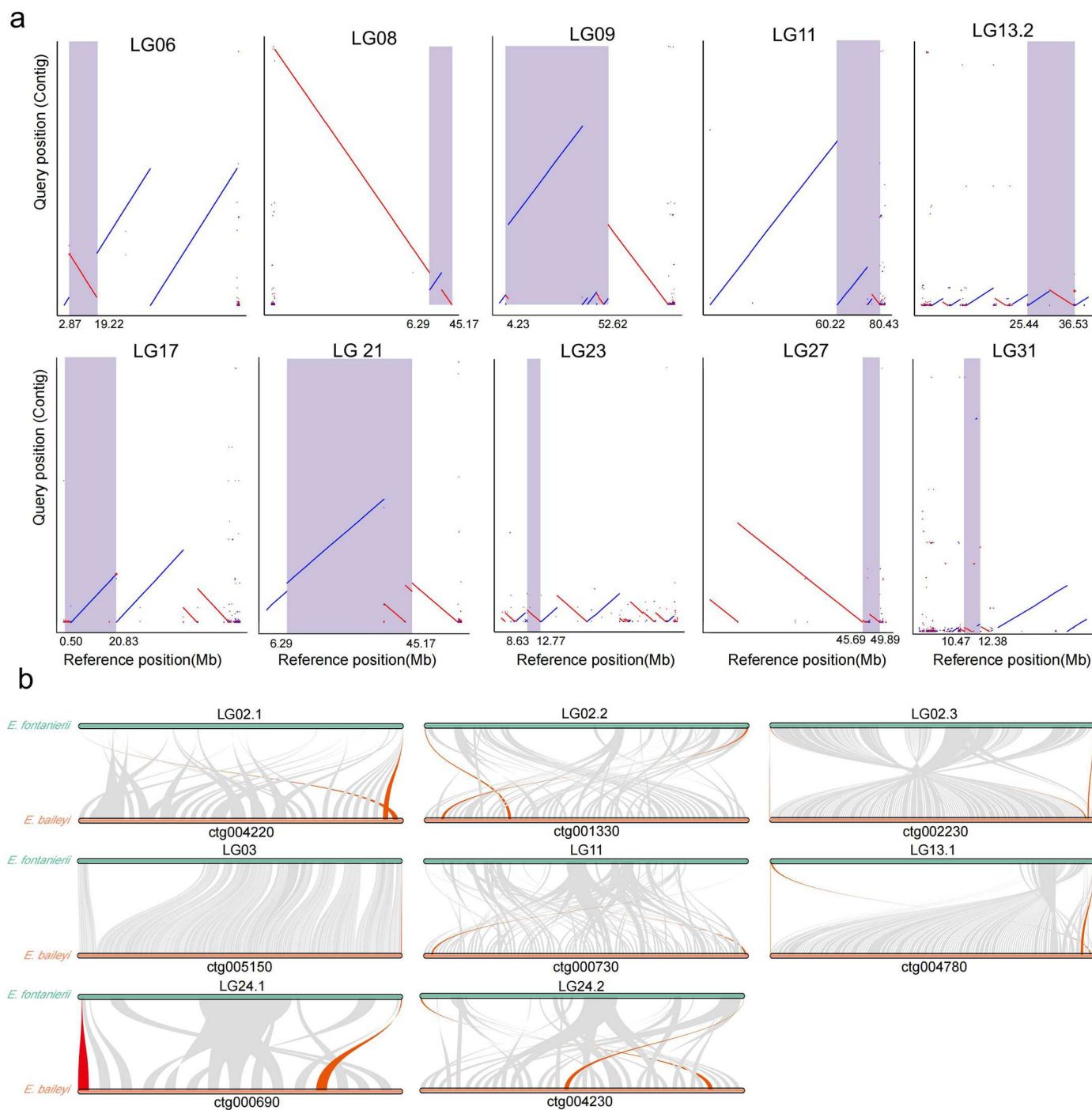


Extended Data Fig. 2 | SV (<1 Mb) dataset summary (LR-SV and SR-SV). (a) The quantity and kinds of SVs were displayed as stacked bar graphs based on short-read sequencing. Bars and dots of various hues denote the type of variation and each species, respectively. (b) Diagram of length distribution for various types of

SVs shorter than 10,000 bp. (c) The length distribution of INDELs generated from different sequencing platforms. (d) Annotation of repeat sequences of the LR-SV. (e) The genomic coordinates of LR-SV.

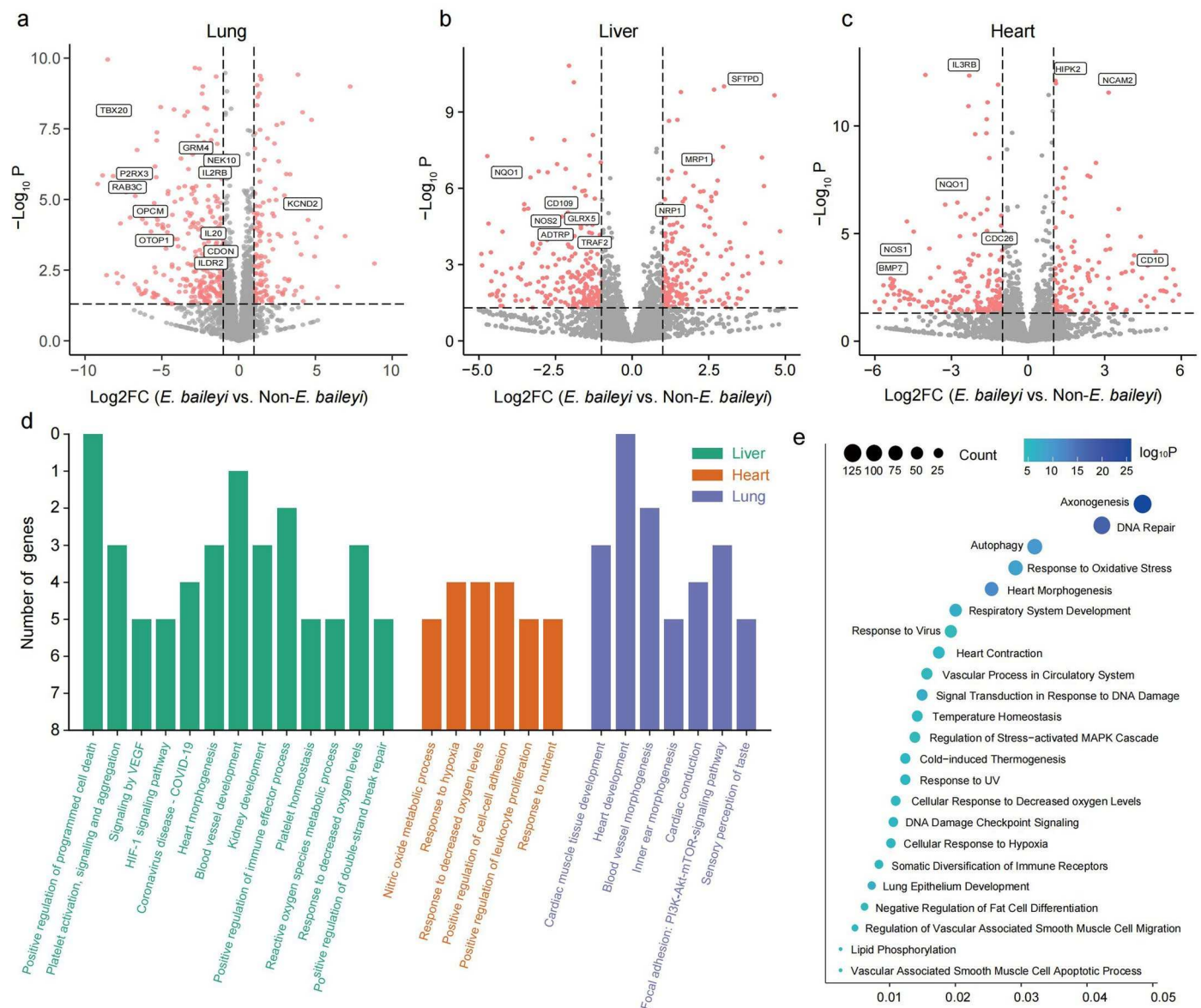


Extended Data Fig. 3 | SV (>1 Mb) dataset summary (LR-SV and SR-SV). The dark blue line at the top of each panel represents the genome of *E. fontanierii*, while the dark orange line represents the genome of *E. baileyi*. The orange block in each panel denotes the inversion position. Each panel (a-m) was one of 32 pseudo-chromosomes.



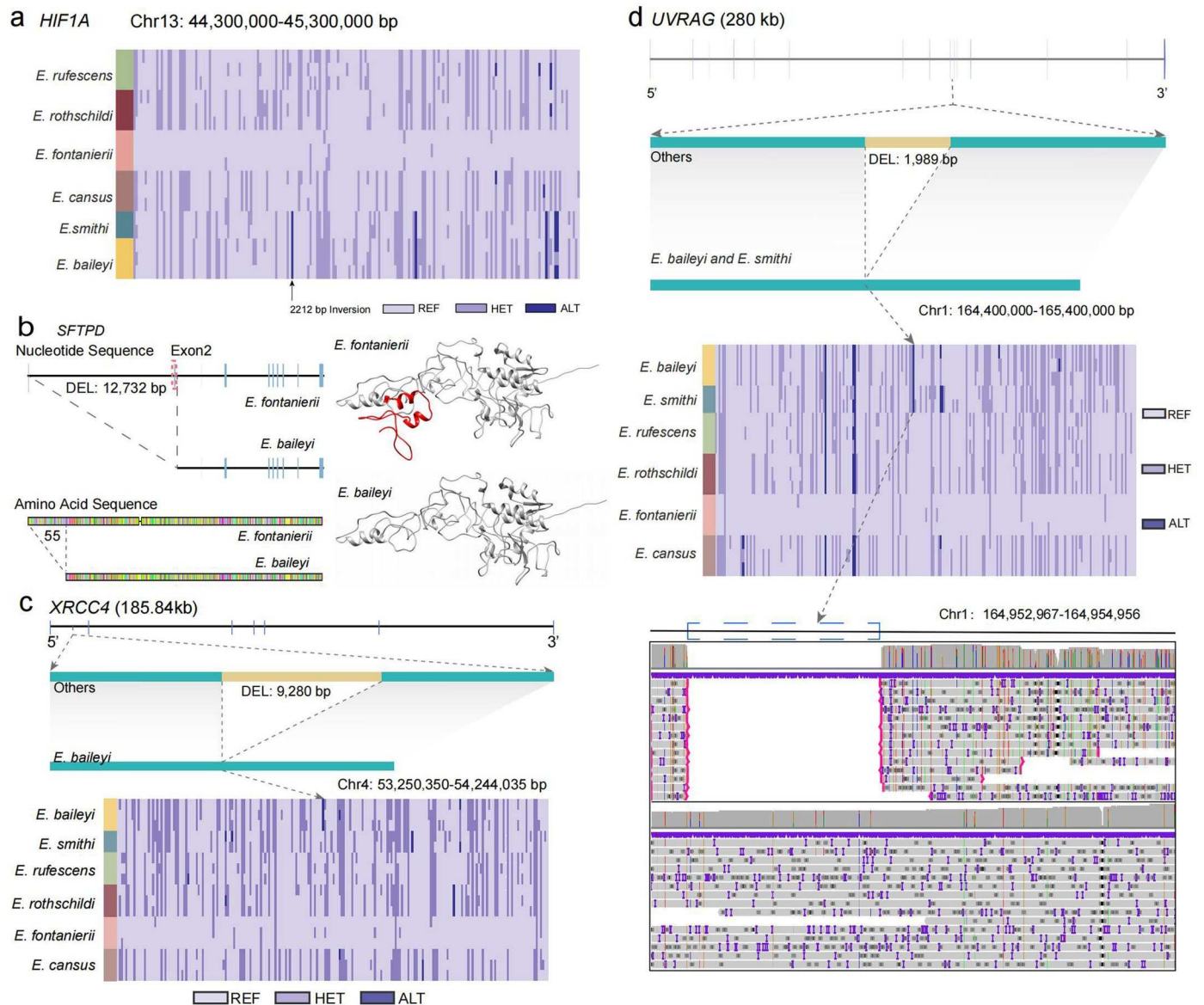
Extended Data Fig. 4 | SV (<1 Mb) dataset summary (LR-SV and SR-SV). (a) Contigs from *de novo* genome assemblies ('query', y-axis) were aligned to the *E. fontanierii* reference genome ('reference', x-axis). Contigs (blue) and those identifying inversion breakpoints (red) are shown. Predicted inversion

boundaries are highlighted (purple box), showing predicted inversion (arrow) above. (b) The Synteny comparison between the genomes of *E. fontanierii* and *E. baileyi* showed that part of the contig of *E. baileyi* was reversed and the other contig was positive.



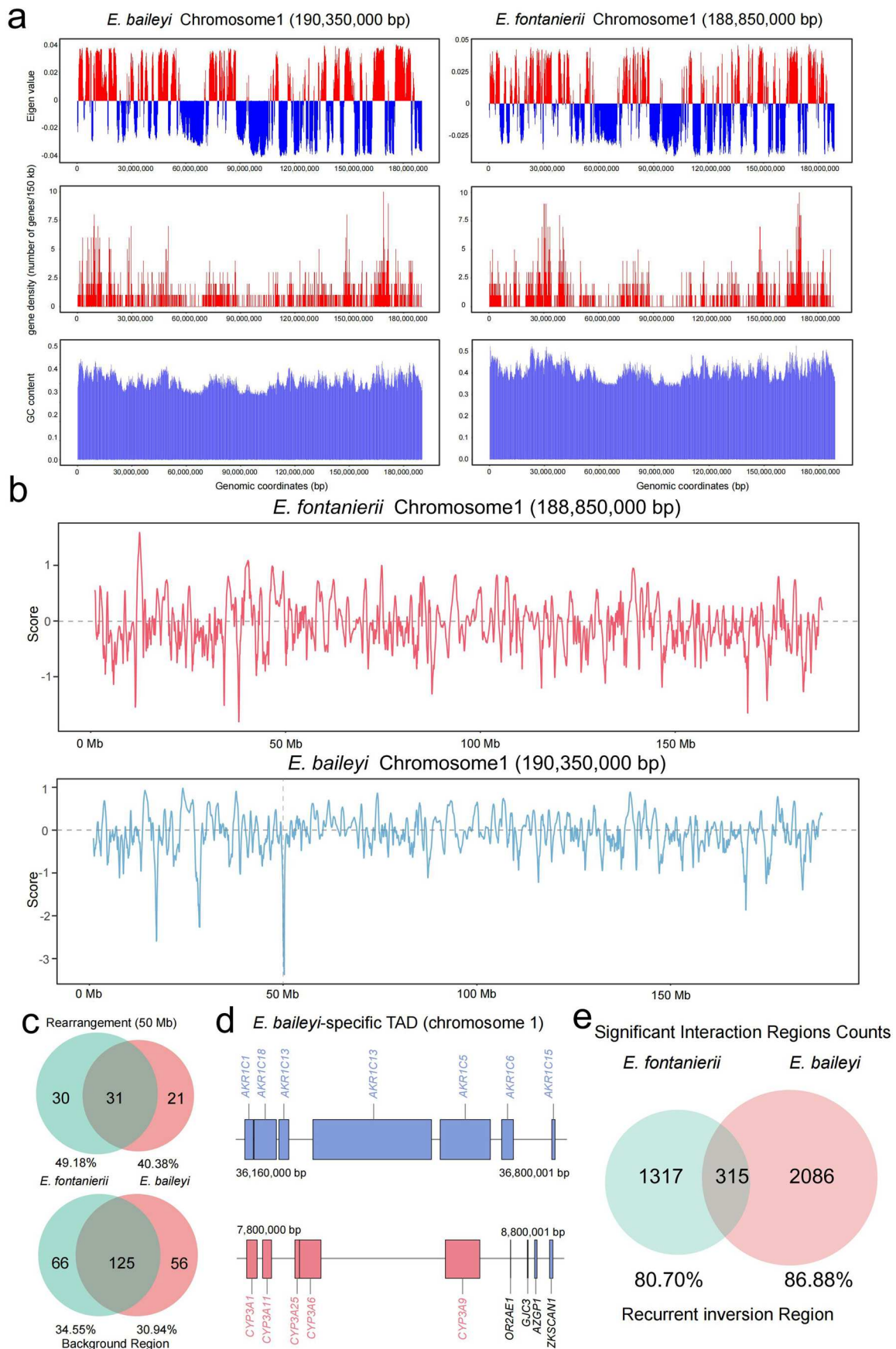
Extended Data Fig. 5 | SV (<1 Mb) dataset summary (LR-SV and SR-SV). The fixed-SV related differentially expressed genes in the lung (a) liver (b), and heart (c) of *E. baileyi*. Negative binomial distribution model and generalized linear model were used to identify DEGs, and we multiple-corrected using Benjamini-Hochberg model ($p_{adj} \leq 0.05$). (d) GO enrichment terms of fixed SV-related genes typically in *E. baileyi*. (e) Enrichment of genes associated with *E. baileyi*-specific SVs, *E. smithi*-specific SVs and the shared SVs between the two species (Hypergeometric test, $p < 0.05$; Benjamini-Hochberg model, $q < 0.05$).

Hochberg model ($p_{adj} \leq 0.05$). (d) GO enrichment terms of fixed SV-related genes typically in *E. baileyi*. (e) Enrichment of genes associated with *E. baileyi*-specific SVs, *E. smithi*-specific SVs and the shared SVs between the two species (Hypergeometric test, $p < 0.05$; Benjamini-Hochberg model, $q < 0.05$).



Extended Data Fig. 6 | SV-related genes were involved in hypoxia adaptation. (a) The frequency of inversion (2,212 bp) in *HIF1A* in *Eospalax*. (b) The nucleotide sequence and amino acid sequence of *SFTPD* in *E. baileyi* reveal a loss of one exon

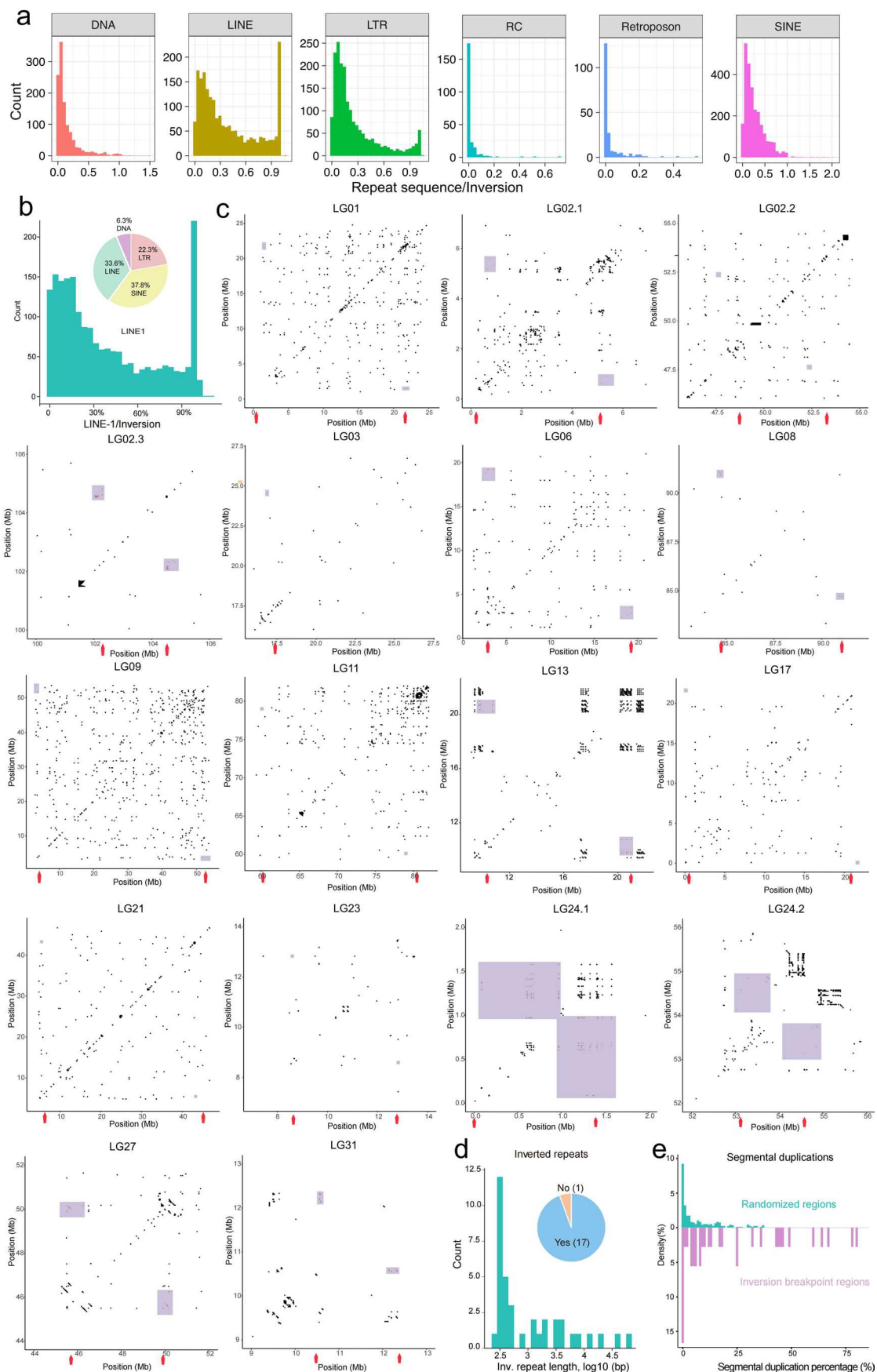
and 55 amino acids. De novo prediction of protein structure for *E. fontanierii* and *E. baileyi*. (c) A 1,989 bp intronic deletion in *UVRAG* was validated by coverage of reads. (d) Map and coverage of reads of deletion in *XRCC4*.



Extended Data Fig. 7 | See next page for caption.

Extended Data Fig. 7 | Three-dimensional genomics and divergence time of the rearrangement on chromosome 1. (a) Compartment A/B, gene density and GC content of chromosome 1 in *E. fontanierii* and *E. baileyi*. A/B compartments are determined by the PC1 value, the red bar indicates compartment A and the blue bar indicates compartment B. (b) Topologically associating domains (TAD)

insulation score of chromosome 1 in *E. fontanierii* and *E. baileyi*. (c) Venn diagram showing species-specific and shared TADs in SV regions and background regions. (d) Two *E. baileyi*-specific TADs associated hypoxia adaptation. (e) The number of shared/species-specific significant interaction regions between *E. fontanierii* and *E. baileyi* in the rearrangement region.



Extended Data Fig. 8 | See next page for caption.

Extended Data Fig. 8 | Three-dimensional genomics and divergence time of the rearrangement on chromosome 1. (a) The overlap of inversions (<1 Mb) with different types of repeats. Different colors represent different repeat sequence types. The X-axis of each figure represents the percentage of the length of the inverted sequence overlapped with different types of repeats as a percentage of the length of the inverted sequence. The Y-axis represents the number of inversions. (b) The type and proportion of repeated sequences in inversion (<1 Mb). The x-coordinate represents the ratio of LINE1 to each inversion sequence. (c) Examples of inversions with inverted repeats at both breakpoints on chromosomes 1, 2, 3, 6, 8, 9, 11, 17, 21, 23, 24, 27 and 31. Dotplot

depicts the inverted repeats near each inversion breakpoint. Locations of breakpoints are denoted by orange arrows, and only alignments longer than 100 bp and within 1 Mb of the breakpoints are displayed. Inverted repeats located within 1 Mb of both breakpoints are depicted (red) and highlighted (purple box). (d) The pie shows the number of inversions (>1 Mb) that have at least one pair of inverted repeats at both inversion breakpoints. The histogram shows the length of inverted repeats near breakpoints. (e) Segmental duplication (SD) percentage of randomly selected 1 Mb regions across the genome and inversion (>1 Mb) regions.

Extended Data Table. 1 | Genome assembly data

Sequencing Platform	Data Size	Sequencing Depth
Oxford Nanopore Technologies, ONT	209.0 Gb	84×
High-through Chromosome Conformation Capture, Hi-C	314.53 Gb	139×
Illumina	255.1 Gb	51×

Extended Data Table. 2 | Dual-luciferase assay of *EGLN1*, *HIF1A*, and *HSF1*

	<i>EGLN1</i>				<i>HIF1A</i>		<i>HSF1</i>		
	pGL3- promoter+Luc	pGL3- promoter+ZH- <i>EGLN1</i> +Luc	pGL3- promoter+GY- <i>EGLN1</i> +Luc	pGL3- promoter+Luc	pGL3- promoter+Z H- <i>HIF1A</i> +Luc	pGL3- promoter+GY - <i>HIF1A</i> +Luc	pGL3- promoter+Luc	pGL3- promoter+ZH - <i>HSF1</i> +Luc	pGL3- promoter+GY - <i>HSF1</i> +Luc
	48,059	81,987	130,256	50,468	99,549	140,249	44,768	112,628	173,763
Luc	46,509	65,681	128,695	58,783	87,829	135,135	51,810	87,089	147,634
	56,059	98,063	125,254	59,134	94,763	149,959	55,337	104,957	150,417
Average	50,209	81,910.33333	128,068.3333	56,128.33333	94,047	141,781	50,638.33333	101,558	157,271.3333
SD	5,125.182924	16,191.13613	2,559.205801	4905.13306	5,892.715164	7,529.808231	5,381.035433	13,104.38976	14,349.82837
	67,107	68,754	64,518	71,279	70,420	68,874	64,869	62,900	66,264
RLuc	57,873	54,668	56,741	75,354	74,283	72,729	68,988	64,160	67,168
	60,176	65,452	66,808	69,098	67,994	72,581	66,022	65,603	69,318
Average	61,718.66667	62,958	62,689	71,910.33333	70,899	71,394.66667	66,626.33333	64,221	67,583.33333
SD	4,806.407633	7,366.741206	5,276.841195	3,175.424434	3,171.744157	2,184.215267	2,124.959843	1,352.53207	1,568.791042

continue Extended Data Table 2

	<i>EGLN1</i>				<i>HIF1A</i>		<i>HSF1</i>		
	pGL3- promoter+Lu c	pGL3- promoter+ZH - <i>EGLN1</i> +Luc	pGL3- promoter+GY - <i>EGLN1</i> +Luc	pGL3- promoter+Luc	pGL3- promoter+Z H- <i>HIF1A</i> +Luc	pGL3- promoter+GY - <i>HIF1A</i> +Luc	pGL3- promoter+Luc	pGL3- promoter+ZH - <i>HSF1</i> +Luc	pGL3- promoter+GY - <i>HSF1</i> +Luc
	0.7162	1.1925	2.0189	0.708	1.4136	2.0363	0.6901	1.7906	2.6223
Luc/RLuc	0.8036	1.2015	2.2681	0.7801	1.1824	1.8581	0.751	1.3574	2.198
	0.9316	1.4982	1.8748	0.8558	1.3937	2.0661	0.8382	1.5999	2.17
Average	0.817133333	1.2974	2.053933333	0.7813	1.3299	1.986833333	0.759766667	1.582633333	2.3301
SD	0.108335836	0.173956115	0.19897669	0.073907307	0.12812568	0.112477613	0.074438185	0.217115553	0.253439598
Relative	0.8765	1.4594	2.4707	0.9062	1.8093	2.6063	0.9083	2.3568	3.4515
luciferase	0.9834	1.4704	2.7757	0.9985	1.5134	2.3782	0.9885	1.7866	2.893
activity	1.1401	1.8335	2.2944	1.0954	1.7838	2.6444	1.1032	2.1058	2.8561
Average	1	1.587766667	2.5136	1.000033333	1.702166667	2.542966667	1	2.083066667	3.066866667
SD	0.132581711	0.21288237	0.243500986	0.094609319	0.163973179	0.14395813	0.097957593	0.285778959	0.333612805

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- n/a

Confirmed
- ☐

☒

The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
- ☐

☒

A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- ☐

☒

The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- ☒

☐

A description of all covariates tested
- ☐

☒

A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- ☐

☒

A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- ☐

☒

For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted
Give P values as exact values whenever suitable.
- ☒

☐

For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- ☒

☐

For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- ☐

☒

Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

No software was used for data collection

Data analysis

In this study, all previous softwares were used, including NextDenovo v2.3.0, blast v2.2.29, BUSCO v5.2.2, BWA-MEM v0.7.17-r1188, 3D-DNA v180419, LTR_FINDER, Tandem Repeat Finder v4.07b, RepeatModeler v4.07b, RepeatMasker v1.331, RepeatProteinMask v4.1.0, AUGUSTUS v3.3.1, Genscan, GeMoMa v1.6.1, Exonerate v2.4.0, HISAT2 v2.1.0, StringTie v2.2.1, PASA v2.3.3, TransDecoder v2.0, EVM v1.1.1, Orthofinder v2.5.4, PAML v4.10.3, CAFÉ v4.2.1, HyPhy, NGM-LR v0.2.7, SAMtools v1.7, Sniffles v1.0.12, surpyvor v0.10.0, Iris v1.0.5, DELLY v0.7.6, LUMPY v0.2.13, Manta v1.6.0, mummer v4.0.0, SyRI v1.5.6, Ragtag v2.0.1, tolkit v0.2.2, SEDEF, IRF v3.05, SnpEff v5.1, LAST v885, UCSC genome browser, Cworld-dekker, FitHic v1.1.3, DESeq v2, mashr, RoseTTAFold, Genomeflow, PyMOL v1.20.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

DNA sequence data are available from NCBI SRA under BioProject accessions: PRJNA947478. The RNA sequence data reported in this paper have been deposited in the Genome Sequence Archive in National Genomics Data Center, China National Center for Bioinformation / Beijing Institute of Genomics, Chinese Academy of Sciences (GSA: CRA012698) that are publicly accessible at <https://ngdc.cncb.ac.cn/gsa>.

Human research participants

Policy information about [studies involving human research participants and Sex and Gender in Research](#).

Reporting on sex and gender

Population characteristics

Recruitment

Ethics oversight

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	To identify structural variants (SVs), we report a genomic analysis of six zokor species (genus <i>Eospalax</i>). Assembly of a chromosome-level genome of the high-elevation species, <i>E. baileyi</i> , enabled us to identify 18 large inversions that distinguished this species from congeners native to lower elevations. Small-scale SVs in the introns of <i>E. baileyi</i> caused upregulated gene expression. A chromosome rearrangement on chromosome 1 altered chromatin accessibility, thereby modifying gene expression profiles of key genes involved in the physiological response to hypoxia. We also found that <i>E. baileyi</i> exhibits a significantly larger lung-mass/body-mass ratio compared to other <i>Eospalax</i> species. These findings highlight the key role of SVs underlying hypoxia adaptation of high-elevation species in <i>Eospalax</i> .
Research sample	We collected six species of <i>Eospalax</i> in China. Seventeen individuals were used for long-read sequencing. Short-read sequencing data from 146 <i>Eospalax</i> individuals were used for SV frequency estimation, including 55 <i>E. baileyi</i> , 25 <i>E. cansus</i> , 21 <i>E. smithi</i> , 18 <i>E. rothschildi</i> , 14 <i>E. rufescens</i> and 13 <i>E. fontanierii</i> . 143 of the 146 were downloaded from public databases.
Sampling strategy	Long-read and short-read sequencing involved six species in <i>Eospalax</i> . On average, three unrelated individuals were selected for each species (there are only two individuals in <i>E. smithi</i>) in the long-read sequence. For the focus subjects (<i>E. baileyi</i>), we estimate variations in large populations (55 individuals) based on short-read data.
Data collection	Genome DNA were extracted from adult zokors. Long-read sequencing for each individual was sequenced to at least 15X depth. The sample used to assemble the genome was sequenced 84X. X.L. collected the raw sequence reads from Sequence platform.
Timing and spatial scale	We collected samples for whole genome sequencing from September to November in 2020. Partial samples for measuring blood routine were collected in April in 2022. The raw data of sequencing and phenotypic data were collected.
Data exclusions	We did not exclude data in the study.
Reproducibility	All attempts to repeat the experiment were successful.
Randomization	All analyses are grouped by different species. In <i>E. baileyi</i> , the group was differentiation by the altitude of the distribution site.

Blinding

Did the study involve field work? ☐ Yes ☒ No

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a Involved in the study

☒ ☐ Antibodies

☐ ☒ Eukaryotic cell lines

☒ ☐ Palaeontology and archaeology

☐ ☒ Animals and other organisms

☒ ☐ Clinical data

☒ ☐ Dual use research of concern

Methods

n/a Involved in the study

☒ ☐ ChIP-seq

☒ ☐ Flow cytometry

☒ ☐ MRI-based neuroimaging

Eukaryotic cell lines

Policy information about [cell lines and Sex and Gender in Research](#)

Cell line source(s)

Authentication

Mycoplasma contamination

Commonly misidentified lines (See [ICLAC](#) register)

Animals and other research organisms

Policy information about [studies involving animals; ARRIVE guidelines](#) recommended for reporting animal research, and [Sex and Gender in Research](#)

Laboratory animals

Wild animals

Reporting on sex

Field-collected samples

Ethics oversight

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Terms and Conditions

Springer Nature journal content, brought to you courtesy of Springer Nature Customer Service Center GmbH (“Springer Nature”).

Springer Nature supports a reasonable amount of sharing of research papers by authors, subscribers and authorised users (“Users”), for small-scale personal, non-commercial use provided that all copyright, trade and service marks and other proprietary notices are maintained. By accessing, sharing, receiving or otherwise using the Springer Nature journal content you agree to these terms of use (“Terms”). For these purposes, Springer Nature considers academic use (by researchers and students) to be non-commercial.

These Terms are supplementary and will apply in addition to any applicable website terms and conditions, a relevant site licence or a personal subscription. These Terms will prevail over any conflict or ambiguity with regards to the relevant terms, a site licence or a personal subscription (to the extent of the conflict or ambiguity only). For Creative Commons-licensed articles, the terms of the Creative Commons license used will apply.

We collect and use personal data to provide access to the Springer Nature journal content. We may also use these personal data internally within ResearchGate and Springer Nature and as agreed share it, in an anonymised way, for purposes of tracking, analysis and reporting. We will not otherwise disclose your personal data outside the ResearchGate or the Springer Nature group of companies unless we have your permission as detailed in the Privacy Policy.

While Users may use the Springer Nature journal content for small scale, personal non-commercial use, it is important to note that Users may not:

1. use such content for the purpose of providing other users with access on a regular or large scale basis or as a means to circumvent access control;
2. use such content where to do so would be considered a criminal or statutory offence in any jurisdiction, or gives rise to civil liability, or is otherwise unlawful;
3. falsely or misleadingly imply or suggest endorsement, approval, sponsorship, or association unless explicitly agreed to by Springer Nature in writing;
4. use bots or other automated methods to access the content or redirect messages
5. override any security feature or exclusionary protocol; or
6. share the content in order to create substitute for Springer Nature products or services or a systematic database of Springer Nature journal content.

In line with the restriction against commercial use, Springer Nature does not permit the creation of a product or service that creates revenue, royalties, rent or income from our content or its inclusion as part of a paid for service or for other commercial gain. Springer Nature journal content cannot be used for inter-library loans and librarians may not upload Springer Nature journal content on a large scale into their, or any other, institutional repository.

These terms of use are reviewed regularly and may be amended at any time. Springer Nature is not obligated to publish any information or content on this website and may remove it or features or functionality at our sole discretion, at any time with or without notice. Springer Nature may revoke this licence to you at any time and remove access to any copies of the Springer Nature journal content which have been saved.

To the fullest extent permitted by law, Springer Nature makes no warranties, representations or guarantees to Users, either express or implied with respect to the Springer nature journal content and all parties disclaim and waive any implied warranties or warranties imposed by law, including merchantability or fitness for any particular purpose.

Please note that these rights do not automatically extend to content, data or other material published by Springer Nature that may be licensed from third parties.

If you would like to use or distribute our Springer Nature journal content to a wider audience or on a regular basis or in any other manner not expressly permitted by these Terms, please contact Springer Nature at

onlineservice@springernature.com