

Evolutionary dynamics of conserved non-exonic elements and distinctive X-linked regulatory evolution in Felidae

Ke-Xin Li^{1,2,#}, Xin Du^{1,#}, Ye-Hao Zhang^{1,2}, Yi-Bo Hu^{1,2,*}

¹ State Key Laboratory of Animal Biodiversity Conservation and Integrated Pest Management, Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, China

² University of Chinese Academy of Sciences, Beijing 101408, China

Authors contributed equally to this work

* Corresponding author, E-mail: ybhu@ioz.ac.cn

Funds: This research was supported by the National Natural Science Foundation of China (32325010 and U24A20361), the National Key Program of Research and Development of Ministry of Science and Technology (2023YFF1304800), and the Science and Technology Project of Xizang Autonomous Region, China (XZ202402ZD0005).

ABSTRACT

Conserved non-exonic elements (CNEEs) are widely recognized as key cis-regulatory elements contributing to developmental regulation and phenotypic evolution in vertebrates. However, their chromosomal organization and evolutionary dynamics remain poorly understood, particularly in non-model animals. The family Felidae, characterized by rapid evolutionary radiation and a highly conserved karyotype, provides an ideal system for investigating these questions. Here, we leveraged long-read genome assemblies from 17 felids and two carnivoran outgroups to identify CNEEs and assess their chromosomal organization, lineage-specific loss, and accelerated evolution. We identified 420 225 feline CNEEs and further defined 490 high-density CNEE clusters, which are associated with developmental regulatory landscapes. The X chromosome exhibited distinct regulatory evolutionary features, with an enrichment of high-density CNEE clusters and higher putative CNEE loss rates relative to autosomes. In contrast, CNEE acceleration was not elevated on the X chromosome relative to autosomes. Additionally, we characterized lineage-specific accelerated CNEEs associated with genes involved in locomotion, sensory ecology, hypoxia adaptation, and musculoskeletal traits, including a Pantherinae-shared 5-bp deletion in a candidate regulatory element associated with *LAMA2*. Our study provides a genome-wide view of CNEE organization and evolution in felids, highlights distinct regulatory dynamics on the X chromosome, and offers new insights into the non-coding genomic mechanisms underlying felid diversification and adaptation.

Keywords: Conserved non-exonic elements; Felidae; Pantherinae; Regulatory evolution; X chromosome

INTRODUCTION

Mammalian genomes harbor numerous conserved non-coding sequences, many of which function as cis-regulatory elements located within introns or in the flanking regions of their target genes (Pennacchio et al., 2006). Among these conserved non-coding sequences, conserved non-exonic elements (CNEEs) are genomic regions that evolve at rates significantly lower than the neutral substitution rate and do not overlap with any annotated exons or coding sequences. CNEEs serve as valuable phylogenomic markers and provide an important framework for investigating the regulatory basis of phenotypic evolution (Lowe et al., 2011; Edwards et al., 2017). While abundant across the genome, these elements are heterogeneously distributed (Leypold & Speicher, 2021). They are particularly enriched in large "gene deserts" and frequently cluster around key developmental regulators within gene regulatory networks (Woolfe et al., 2004; Woolfe & Elgar, 2008).

Lineage-specific changes in CNEEs, including sequence acceleration, deletions, or loss of conserved elements, can modify regulatory constraints, alter the spatiotemporal expression patterns of target genes, and thereby contribute to phenotypic innovation and adaptive evolution. For example, a lineage-specific deletion in a conserved non-coding element upstream of *FLT1* has been linked to sensory adaptation in tapirs, where disruption of a transcription factor binding motif may have contributed to reduced visual sensitivity and enhanced auditory sensitivity in nocturnal environments (Zhou et al., 2026). Similarly, in paleognaths, accelerated CNEEs near limb-development genes have been implicated in convergent flight loss, and functional validation of an accelerated element within an intron of *TEAD1* demonstrated loss of forelimb enhancer activity (Sackton et al., 2019). These studies highlight CNEEs as powerful markers for tracing the regulatory mechanisms underlying adaptation and

phenotypic diversification.

A parallel question is whether sex chromosomes exhibit distinctive patterns of conserved regulatory evolution. The X chromosome (ChrX) differs from autosomes in inheritance mode, effective population size, recombination environment, and dosage regulation, all of which may influence the organization and evolutionary turnover of regulatory elements (Charlesworth et al., 1987; Ellegren, 2011; Disteche, 2012). Previous studies have reported enrichment of conserved non-coding elements and enhancer-associated features on the X chromosome in several taxa, suggesting that X-linked genomic regions may harbor distinctive cis-regulatory architectures (Vavouri et al., 2007; Naville et al., 2015). However, whether such chromosome-specific regulatory patterns are broadly present in non-model mammalian lineages remains unclear.

The family Felidae represents an ideal model system for investigating X-linked regulatory evolution and the contribution of CNEEs to adaptive diversification. Felids are distributed across all continents except Australia and Antarctica, inhabiting a wide range of terrestrial ecosystems (Sunquist & Sunquist, 2002). They experienced a rapid evolutionary radiation during the late Miocene, with major lineages diverging within a narrow time window of approximately 10–11 million years ago (Ma) (Johnson et al., 2006), ultimately giving rise to around 41 extant species (Kitchener et al., 2017). During this diversification, felids separated into two principal lineages: the subfamily Pantherinae, consisting primarily of large-bodied apex predators (7 species), and the subfamily Felinae, which includes mostly small- to medium-sized species (34 species). These two groups exhibit pronounced differences in body size, ecological strategies, and physiological characteristics (Werdelin et al., 2010), making felids a compelling system for studying the regulatory basis of phenotypic divergence.

Despite their ecological and morphological diversity, felids show exceptional



genomic macrostructural stability. Most species retain a highly conserved diploid karyotype ($2n=38$), with extensive preservation of gene content and chromosomal synteny across taxa (Nash & O'Brien, 1982; Wurster-Hill & Centerwall, 1982). This architectural stability provides a favorable comparative framework for identifying and tracking CNEEs across species. Intriguingly, the X chromosome contrasts with this overall stability, presenting as a hotspot for structural variation (Bredemeyer et al., 2023). This dichotomy raises the possibility that X-linked regulatory elements follow distinct evolutionary trajectories relative to their autosomal counterparts. Recently, the proliferation of long-read felid reference genomes (Bredemeyer et al., 2023; Winter et al., 2023; Yuan et al., 2024; Supple et al., 2025; Yang et al., 2025) has provided an unprecedented foundation for comparative genomics. However, comparative studies in felids have predominantly focused on protein-coding sequences. For instance, coding variations in *TAQPEP* govern coat-pattern diversity in domestic cats (*Felis catus*) and cheetahs (*Acinonyx jubatus*) (Kaelin et al., 2012, 2021), while parallel amino acid substitutions in *MYSM1* and *GOLGB1* are associated with arboreal adaptation in the clouded leopard (*Neofelis nebulosa*) and marbled cat (*Pardofelis marmorata*) (Yuan et al., 2024). Consequently, the evolutionary significance of CNEEs in felid diversification remains largely unexplored.

Here, we leveraged long-read genome assemblies from 17 felids and two carnivoran outgroups to systematically identify CNEEs across Felidae and examine X-linked regulatory evolution in this non-model mammalian family. We compared X-linked and autosomal CNEEs in terms of chromosomal distribution, local clustering, lineage-specific loss, and substitution-based acceleration to evaluate chromosome-specific patterns of conserved regulatory element organization and evolution. We further characterized lineage- and species-specific accelerated CNEEs to identify



candidate regulatory signatures associated with ecological and phenotypic diversification, including locomotion, sensory ecology, hypoxia adaptation, and musculoskeletal traits. This study evaluates whether X-linked CNEs exhibit evolutionary patterns distinct from autosomal CNEs and provides a genome-wide perspective on regulatory evolution in felids.

MATERIALS AND METHODS

Genome collection and whole-genome alignment

To identify CNEs, we retrieved genome assemblies of 17 felid species, alongside the gray wolf (*Canis lupus*) and brown bear (*Ursus arctos*) as outgroups, from NCBI. Genome assembly completeness was evaluated using Benchmarking Universal Single-Copy Orthologs (BUSCO) (Simão et al., 2015) against the mammalia_odb10 lineage dataset, revealing high contiguity and completeness across all species, with contig N50 values exceeding 5 Mb and BUSCO completeness scores $\geq 96\%$ (Supplementary Table S1).

Pairwise whole-genome alignments against the domestic cat reference genome were generated using LASTZ (Harris, 2007) with the following parameters: --hspthresh 2400, --gappedthresh 3000, --ydrop 9400, --inner 2000, --linearGap medium, and --score HOXD70. Raw alignments were chained using axtChain. To construct a high-quality multiple sequence alignment (MSA), we sequentially processed the chains using RepeatFiller (Osipova et al., 2019) to recover repeat-masked alignments, and chainCleaner (Suarez et al., 2017) to enhance alignment specificity. The refined chains were netted, filtered via NetFilterNonNested.perl, converted to MAF format, and finally merged into a genome-wide MSA using MULTIZ (Blanchette et al., 2004). To address potential chromosome-specific alignment bias, we additionally summarized



MAF-based coverage separately for ChrX and autosomes in felid species. Autosomal coverage was calculated as a length-weighted average across autosomes (Supplementary Table S2).

Phylogenetic reconstruction and divergence time estimation

To reconstruct the phylogeny and estimate divergence times, we extracted fourfold degenerate sites (4dSs) from the whole-genome multiple sequence alignment based on the domestic cat genome annotation. Sites with >20% missing data across felid species were removed. The remaining alignment was concatenated and filtered using Gblocks (Talavera & Castresana, 2007), yielding a final supermatrix of 5 006 939 bp. Maximum likelihood (ML) phylogenetic inference was performed using RAxML (Stamatakis, 2014) under the GTRGAMMA substitution model with 1 000 bootstrap replicates, using the gray wolf and brown bear as outgroups.

Species divergence times were estimated using MCMCtree in PAML (Yang, 2007). To optimize computational efficiency, we employed the approximate likelihood method (usedata=2), deriving the gradient vector and Hessian matrix of branch lengths via BASEML under the REV model. The Feliformia-Caniformia divergence, estimated at approximately 55 Ma in TimeTree (Kumar et al., 2022), was used to calibrate the root (RootAge='<0.6') and estimate the genome-wide average substitution rate, guiding the overall substitution rate prior (rgene_gamma=2 12.72). Three additional calibration constraints derived from TimeTree were incorporated as node age constraints (Supplementary Table S3). Two independent MCMC runs (burn-in: 2 000 000; sample frequency: 3 000; samples: 20 000) were executed. Convergence was verified using Tracer (Rambaut et al., 2018), ensuring effective sample sizes (ESS)>200 for all parameters.

Identification and filtering of CNEEs

To identify CNEEs across the 19 evaluated species, we utilized the PHAST suite (Hubisz et al., 2011). Initially, a neutral model of sequence evolution was estimated using phyloFit (--sym-freqs --subst-mod REV), based on our phylogenetic topology and 5 Mb of randomly selected repetitive sequences from the domestic cat genome. This model was subsequently GC-adjusted via modFreqs, and ancestral nodes were labeled using tree_doctor (--name-ancestors). To estimate the parameter rho (the expected substitution rate of conserved elements relative to neutrality), phastCons (Siepel et al., 2005) (--estimate-rho --no-post-probs) was run on 3-Mb non-overlapping alignment blocks. The resulting outputs were averaged into a single conserved and a single non-conserved model using phyloBoot.

Relying on these averaged models, we re-ran phastCons (--expected-length 45 --target-coverage 0.3 --most-conserved) across 40-Mb non-overlapping alignments to predict conserved elements, merging adjacent elements separated by <5 bp. To isolate non-coding regions, we used BEDTools (Quinlan & Hall, 2010) to exclude any conserved element that overlapped (≥ 1 bp) with annotated exon features in the domestic cat genome. This filtering step removed conserved elements overlapping protein-coding exons, untranslated regions (UTRs), non-coding RNA exons, exons of alternative transcripts, and pseudogene exons. After this filtering, only CNEEs with a length of at least 50 bp were retained as the felid CNEE dataset.

Genomic partitioning and delineation of high-density CNEE clusters

To characterize the genomic distribution of CNEEs and identify large-scale CNEE-enriched clusters, overlapping CNEEs were merged into continuous genomic intervals using BEDTools. The domestic cat reference genome was then partitioned into 300-kb sliding windows with a 50-kb step size. For each window, CNEE enrichment was quantified as fold-change (FC) relative to the genome-wide background CNEE density.



191 The background density was calculated as the total length of merged CNEEs divided
192 by the effective genome size, whereas the local density of each window was calculated
193 as the total CNEE-covered length within the window divided by the window size.

194 A geometric inflection-point approach was used to define the FC threshold for
195 identifying candidate high-density windows. Briefly, candidate windows were ranked
196 by increasing FC values, and a linear baseline was defined between the two endpoints
197 of the rank-ordered distribution. The orthogonal distance between each window and
198 this baseline was calculated, and the point showing the maximum deviation was defined
199 as the empirical inflection point. This procedure identified an FC threshold of 1.98 for
200 the main 300-kb window analysis (Supplementary Figure S1). Windows with $FC < 0.5$
201 were classified as CNEE deserts, those with $FC \geq 1.98$ were considered candidate high-
202 density windows, and the remaining windows were classified as background regions.

203 Adjacent candidate high-density windows were merged into continuous intervals,
204 allowing a maximum gap of 50 kb. The number of unique CNEEs within each merged
205 candidate cluster was then quantified using BEDTools. To exclude small regions
206 potentially representing isolated high-density signals rather than broad CNEE-enriched
207 domains, we applied an empirical count-based quality-control threshold based on the
208 distribution of CNEE numbers per candidate cluster. In the primary 300-kb window
209 analysis, candidate clusters contained a median of 191.5 CNEEs, with the lower 1st
210 percentile corresponding to 84 CNEEs (Supplementary Figure S2). We therefore used
211 80 CNEEs as the final cutoff, retaining only merged intervals containing at least 80
212 CNEEs as final high-density clusters.

213 To evaluate the robustness of high-density CNEE cluster identification and
214 associated chromosomal enrichment patterns to parameter choices, we performed one-
215 at-a-time sensitivity analyses. First, we repeated the sliding-window analysis using

alternative window sizes of 200 kb and 500 kb, in addition to the 300-kb window used in the main analysis. Because FC values depend on local CNEE density calculated within each window, the FC threshold was independently re-estimated for each window size using the same geometric inflection-point approach. Second, under the main 300-kb window setting, we tested alternative maximum merging distances of 30 kb and 100 kb compared with the 50-kb distance used in the primary analysis. For each parameter combination, high-density clusters were redefined, the distribution of CNEE counts per candidate cluster was recalculated, and the minimum CNEE-count threshold was determined using the same empirical lower-tail distribution-based criterion. The resulting cluster sets were used to reassess chromosome-level enrichment patterns, including ChrX enrichment (Supplementary Table S4).

Genomic characterization and functional enrichment of high-density CNEE clusters

We performed chromosome-level and region-level analyses to characterize the genomic features of high-density CNEE clusters. Chromosome sizes and final cluster coordinates were used to calculate chromosome-level cluster density (number of clusters per Mb) and cluster coverage (proportion of chromosome length occupied by clusters). Differences in cluster occurrence rates between ChrX and autosomes were evaluated using two-sided Poisson rate tests, with chromosome length considered as the exposure term. The relationship between chromosome size and cluster abundance was assessed using Pearson correlation analysis.

To compare sequence features among genomic partitions, GC content and repeat coverage were calculated for CNEE deserts, background regions, and high-density CNEE clusters. GC content was calculated from the domestic cat reference genome for each interval. Repeat coverage was calculated as the proportion of each interval

overlapping annotated repetitive elements. Differences among the three genomic partitions were assessed using Kruskal–Wallis tests, followed by pairwise Wilcoxon rank-sum tests with Benjamini–Hochberg correction.

Coding density was calculated as the length of merged non-redundant coding sequence (CDS) regions per megabase (kb/Mb), with overlapping CDS annotations merged to avoid redundancy from transcript isoforms. CDS density was compared between high-density CNEE clusters and non-cluster genomic regions. Chromosomal differences in CDS density between ChrX and autosomes were assessed using non-overlapping 1-Mb genomic windows and two-sided Wilcoxon rank-sum tests. Within high-density CNEE clusters, Spearman rank correlation analyses were performed to examine the relationship between CDS density and CNEE coverage density, defined as the CNEE-covered sequence length per megabase (kb/Mb). CNEE element density, defined as the number of CNEE elements per megabase, was also analyzed as a complementary count-based metric.

To investigate potential biological functions associated with high-density CNEE clusters, genes located within 500 kb upstream or downstream of each cluster were extracted and deduplicated. GO biological process enrichment analysis was performed using clusterProfiler (Yu et al., 2012), with all deduplicated genes associated with testable CNEEs used as the background gene set. Multiple testing correction was performed using the Benjamini–Hochberg method, and GO terms with adjusted *P* values <0.05 were considered significantly enriched.

Identification and filtering of putative CNEE losses

Initial gap-based putative CNEE losses were identified from multi-species genome alignments using the domestic cat genome as the reference genome. For each non-reference species, the fraction of each reference CNEE interval lacking aligned



sequence was calculated. CNEEs for which >40% of the reference interval lacked aligned sequence were classified as raw putative CNEE losses, including both partially and completely missing elements. Because the domestic cat genome was used as the reference genome, no putative losses were assigned to this species by definition; it was included only for visualization and reference comparison.

To reduce false-positive loss calls from the initial gap-based screening, raw putative CNEE losses were further evaluated using a synteny-based validation strategy. For each raw putative CNEE loss, upstream and downstream syntenic anchors were searched within a 200-kb flanking region of the domestic cat genome. The nearest valid upstream and downstream aligned blocks were selected as the corresponding syntenic anchors, and the distance between each anchor and the CNEE was recorded for quality control.

A raw putative CNEE loss was retained as a high-confidence putative CNEE loss only when all of the following criteria were satisfied in the target species: (i) both flanking anchors were successfully identified; (ii) both anchors were located on the same chromosome or scaffold with consistent strand orientation; and (iii) the distance between the two anchors in the target genome (target gap) was ≤ 500 bp or showed only a minor overlap between anchors ($\leq 1\,000$ bp). Raw putative CNEE losses failing any of these criteria were excluded from downstream analyses. Unless otherwise stated, the term “putative CNEE losses” refers to synteny-validated high-confidence loss calls throughout the manuscript.

To evaluate the robustness of chromosome-level loss-rate estimates and ChrX enrichment analyses, we performed sensitivity analyses using two additional missing-sequence thresholds (60% and 80%) in addition to the 40% threshold used in the main analysis. Raw putative CNEE losses identified under each threshold were subjected to

the same synteny-based validation procedure described above.

Overlap between putative CNEE loss and high-density CNEE clusters

To assess whether putative CNEE losses were depleted from high-density CNEE clusters, species-specific high-confidence putative losses were mapped to the genome-wide CNEE annotation and intersected with previously defined high-density CNEE cluster regions. For each species, filtered-out putative loss CNEEs were removed from all annotated CNEEs to define a species-specific analyzable background set. Fisher's exact tests were performed using a 2×2 contingency table comparing putatively lost CNEEs and remaining analyzable CNEEs inside and outside high-density clusters. *P* values were adjusted for multiple testing using the Benjamini–Hochberg procedure. Odds ratios below 1 indicate depletion of high-confidence putative losses within high-density CNEE clusters.

Chromosome-level putative CNEE loss rate and ChrX enrichment analysis

Chromosome-level putative CNEE loss rates were calculated for each species and chromosome based on synteny-filtered loss calls. For each chromosome, the number of effective CNEEs was defined as the total number of annotated CNEEs minus species-specific candidates excluded during synteny validation. The loss rate was calculated as the number of retained high-confidence putative CNEE losses divided by the corresponding number of effective CNEEs.

ChrX enrichment of CNEE losses was evaluated for each species by calculating the ratio between the ChrX loss rate and the mean autosomal loss rate. The enrichment ratio was log₂-transformed, with positive values indicating higher loss rates on ChrX relative to autosomes. To account for phylogenetic non-independence among species, phylogenetic generalized least squares (PGLS) models were fitted using the felid phylogeny under a Brownian-motion correlation structure. ChrX enrichment was

assessed by testing whether the estimated mean $\log_2$ -transformed ChrX-to-autosome loss-rate ratio was greater than zero. One-sided P values were calculated.

To evaluate the robustness of ChrX enrichment signals, the analyses were repeated using CNEE loss calls identified under both 60% and 80% missing-sequence thresholds. Consistent analytical procedures, including loss-rate calculation and PGLS testing, were applied under each threshold. For visualization of chromosome-level loss rates on a log scale, a pseudo-count of 0.001 was added to loss-rate values prior to log-scale transformation.

Putative CNEE-loss block analysis

To examine whether putative CNEE losses were spatially dispersed or locally aggregated, species-specific putative CNEE losses were merged into CNEE-loss blocks. Only synteny-validated high-confidence losses were included, whereas filtered-out candidates were excluded. Putative losses located within 50 kb of each other on the same chromosome were assigned to the same block. Block size was defined as the number of high-confidence putative losses contained within each block. Blocks were classified as singletons, small blocks containing 2–5 losses, medium blocks containing 6–20 losses, or large blocks containing more than 20 losses. For each species, the proportion of blocks in each size category was calculated.

Estimation of effective population size (N_e)

To test whether variation in putative CNEE loss rate among felid species could be explained by genetic drift associated with differences in N_e , we estimated historical N_e using PSMC (Li & Durbin, 2011). For each felid species, excluding the domestic cat reference individual, whole-genome resequencing reads were obtained (Supplementary Table S5). Reads were quality-filtered using fastp (Chen et al., 2018) and mapped to the corresponding genome assembly with BWA-MEM (Li & Durbin, 2009). After



sorting and duplicate removal using SAMtools (Li et al., 2009) and Picard (Scholak et al., 2021), diploid consensus sequences were generated using bcftools (Danecek et al., 2021), retaining only sites with read depth between 10 and 100. The resulting consensus FASTQ files were used for PSMC analysis. For each species, we summarized historical N_e over the period from 10 thousand to 1 Ma by calculating the time-weighted harmonic mean of N_e across PSMC-inferred time intervals (Wilder et al., 2023). The duration of each interval was used as the weight.

We tested the association between species-level CNEE loss rate and N_e using a PGLS model. CNEE loss rate was defined as high-confidence putative losses divided by effective CNEEs, and effective CNEEs were calculated as total annotated CNEEs minus filtered-out CNEEs. N_e was $\log_{10}$ -transformed, and loss rate was logit-transformed.

Identification of Pantherinae-shared putative CNEE losses

To identify subfamily-associated putative CNEE loss patterns, synteny-validated high-confidence putative CNEE losses were screened across all sampled Pantherinae species. CNEEs classified as putative losses in all sampled Pantherinae species were defined as Pantherinae-shared putative CNEE losses. Filtered-out loss candidates were excluded from this analysis. Loss states for each CNEE were summarized across felid species and the outgroup and visualized as a heatmap.

CNEE-associated gene assignment

CNEE-associated genes were assigned using a modified GREAT-like basal-plus-extension strategy (McLean et al., 2010). Candidate genes for each CNEE were obtained from the CNEE annotation file and matched to gene models in the reference GFF annotation. For each candidate gene, the transcription start site (TSS) was determined according to strand orientation. The basal regulatory domain was defined



as 5 kb upstream and 1 kb downstream of the TSS for genes on the positive strand, and 1 kb upstream and 5 kb downstream of the TSS for genes on the negative strand. Each basal domain was extended up to 1 Mb in both directions, with extension boundaries truncated by the basal domains of neighboring genes. A CNEE was assigned to a gene if its interval overlapped either the basal domain or the extended regulatory domain. CNEEs overlapping multiple regulatory domains were assigned to all corresponding genes, whereas CNEEs without overlapping regulatory domains were treated as unassigned.

Functional enrichment analysis of Pantherinae-shared putative CNEE losses

Functional enrichment analysis was performed for genes associated with Pantherinae-shared high-confidence putative CNEE losses. Only putative losses retained after synteny-based validation were included in this analysis. For enrichment analysis, genes associated with all analyzable CNEEs, excluding CNEEs corresponding to filtered-out loss candidates, were used as the background gene set. Enrichment analysis was performed using Metascape (Zhou et al., 2019).

Identification of accelerated CNEEs and GC-biased gene conversion (gBGC)

filtering

To minimize potential artifacts caused by incomplete alignments or lineage-specific missing sequences, CNEEs with >40% missing sequence in any analyzed species were excluded before testing for accelerated evolution. We identified accelerated CNEEs across the feline phylogeny using phyloP (Pollard et al., 2010), defining significant acceleration by negative scores and a Benjamini–Hochberg correction with false discovery rate (FDR) <0.05. Tests were conducted at two phylogenetic scales: (i) species-specific acceleration, evaluating each terminal branch independently (--branch), and (ii) Pantherinae-specific acceleration, targeting both the entire Pantherinae clade



(from lion (*Panthera leo*) to clouded leopard; --subtree) and its ancestral root (--branch).

To reduce potential gBGC-related false-positive acceleration signals, accelerated elements were subsequently evaluated using phastBias. Accelerated regions with a gBGC posterior probability ≥ 0.5 were excluded from downstream analyses to conservatively remove potential gBGC-associated signals. Genes associated with accelerated CNEEs were assigned using the same modified GREAT-like basal-plus-extension strategy described above. Functional enrichment analysis of accelerated CNEE-associated genes was conducted using Metascape. Genes associated with all analyzable CNEEs, after excluding CNEEs with more than 40% missing sequence, were used as the background gene set. To assess chromosome-level enrichment of accelerated CNEEs, we calculated the acceleration rate for each chromosome as the number of retained accelerated CNEEs divided by the total number of analyzable CNEEs remaining after filtering for missing sequence. ChrX enrichment was evaluated using the same chromosome-level comparison framework described above for putative loss-rate-based analyses.

Prediction of transcription factor binding sites (TFBSs)

To determine whether sequence variations within CNEEs alter potential transcription factor binding sites, we scanned the orthologous sequences using the FIMO tool (Grant et al., 2011) from the MEME Suite. Putative TFBSs were identified by searching against the JASPAR2026_CORE Vertebrates Non-redundant motif database. The domestic cat sequence was used as the baseline reference, and significant motif matches were called using a significance threshold of $P < 1 \times 10^{-4}$.

Validation and regulatory annotation of the *LAMA2*-associated accelerated CNEE

To validate the predicted Pantherinae-specific deletion within the *LAMA2*-associated

416 accelerated CNEE (mCE205254), we reanalyzed publicly available whole-genome
417 resequencing data from Pantherinae species. Raw sequencing reads were mapped to the
418 domestic cat reference genome using minimap2 (Li, 2018) with the short-read
419 alignment mode (-ax sr). The resulting alignments were sorted and indexed with
420 SAMtools, and read coverage across the candidate CNEE and its flanking regions was
421 manually inspected using Integrative Genomics Viewer (IGV) (Thorvaldsdottir et al.,
422 2013).

423 To assess whether the acceleration signal at the *LAMA2*-associated CNEE was
424 influenced by local phylogenetic discordance, we reconstructed a local phylogeny using
425 the aligned sequences of mCE205254 from the analyzed species. The acceleration
426 analysis was then repeated using this CNEE-derived topology, with the neutral
427 substitution model re-estimated following the same repeat-based procedure described
428 above for CNEE identification and acceleration testing. The significance of
429 acceleration for mCE205254 was evaluated under this alternative phylogenetic
430 framework.

431 The regulatory context of mCE205254 was further investigated using publicly
432 available domestic cat epigenomic datasets. Brain ATAC-seq (GEO: GSE303046) and
433 heart H3K27ac ChIP-seq datasets (GEO: GSE182952) were analyzed to examine
434 whether mCE205254 overlapped genomic regions associated with regulatory activity.
435 Paired-end reads were quality-checked and mapped to the domestic cat reference
436 genome using Bowtie2 (Langmead & Salzberg, 2012). The resulting alignments were
437 sorted with SAMtools, and duplicate reads were removed using Picard. Peaks were
438 called for each biological replicate using MACS3 (Zhang et al., 2008), and reproducible
439 peaks between replicates were assessed using Irreproducible Discovery Rate (IDR) (Li
440 et al., 2011). The coordinates of mCE205254 were intersected with the resulting peak



sets to examine whether this CNEE overlapped accessible chromatin or H3K27ac-marked active regulatory regions in the available datasets.

RESULTS

Reconstruction of the feline phylogenetic tree

To provide an evolutionary context for comparative genomic analyses, we integrated 17 long-read-based felid genomes, including 11 Felinae species and six Pantherinae species, with the brown bear and gray wolf as outgroups (Supplementary Table S1). Using approximately 5 million 4dSs extracted from whole-genome alignments, we reconstructed the felid phylogeny and estimated divergence times (Figure 1A; Supplementary Table S6). The resulting topology is highly congruent with previously published studies (Li et al., 2016; Figueiró et al., 2017; Yang et al., 2026). The divergence between Pantherinae and Felinae is estimated at 12.59 Ma, a posterior estimate consistent with the adjusted TimeTree prior interval of 11.6–14.7 Ma. This timescale reveals a late Miocene diversification of major Felinae lineages and a rapid early Pliocene radiation of modern *Panthera* species, dating the *Panthera* crown group to 4.93 Ma. These results provide a robust phylogenetic framework and evolutionary timescale for subsequent comparative analyses of CNEEs.

Genome-wide identification and distribution of feline CNEEs

We applied the phastCons software to the whole-genome alignment of 19 species and identified 420 225 candidate feline CNEEs, spanning approximately 168.88 Mb of the genome. These elements had a median length of 323 bp, with lengths ranging from 53 bp to 6 999 bp (Figure 1B). Genomic distribution analysis showed that 56% of CNEEs ($n=235\,251$) map to intergenic regions, while the remaining 44% ($n=184\,974$) are located in introns (Figure 1C).

To examine the chromosomal distribution of CNEEs, we calculated CNEE density across all chromosomes. Overall, CNEEs showed a relatively even distribution at the chromosome scale, with no single chromosome showing exceptionally high clustering in total CNEE density (Figure 1D). Notably, the CNEE density on the X chromosome was comparable to, or lower than, those of several autosomes, such as ChrC1 and ChrA3. This suggests that the X chromosome does not show an overall excess of CNEE accumulation at the individual element level when considered at the whole-chromosome scale.

Because CNEEs often act as cis-regulatory elements and can occur in spatial clusters near genes involved in developmental regulation, we further examined their local density patterns across the genome, identifying 490 high-density CNEE clusters across the feline genome. The genome-wide CNEE density landscape revealed pronounced local heterogeneity in CNEE distribution, with distinct peaks of CNEE accumulation observed within multiple chromosomes (Figure 1E).

High-density CNEE clusters show an X-chromosome bias

We next characterized the genomic distribution and functional properties of the 490 high-density CNEE clusters identified across the feline genome. Although the density of individual CNEEs is broadly comparable across chromosomes, high-density CNEE clusters exhibit a clear chromosome-specific pattern. The X chromosome exhibited the highest cluster density among all chromosomes (Figure 2A; Supplementary Table S7), representing a significant enrichment relative to the autosomal background (Poisson rate test, $P=1.60\times 10^{-3}$; Supplementary Figure S3A). We further examined the relationship between the total physical length of high-density CNEE clusters and chromosome length. Although total high-density CNEE cluster length was strongly correlated with chromosome length (Pearson's $r=0.86$, $P=1.79\times 10^{-6}$), ChrX and ChrC1

showed notable positive deviations from the genome-wide trend (Figure 2B). ChrC1 had the greatest total length of high-density CNEE clusters, whereas ChrX showed the highest proportional coverage relative to chromosome length (Figure 2B; Supplementary Table S7).

The ChrX enrichment remained consistent across alternative cluster-identification settings, with X/autosome density ratios ranging from 1.53 to 1.72 and remaining statistically significant across all tested parameter combinations (Poisson rate test, P values ranged from 0.0016 to 0.0427; Supplementary Table S4). To evaluate whether the ChrX enrichment of high-density CNEE clusters could be explained by heterogeneous coding sequence distribution, we compared CDS density between ChrX and autosomes. ChrX showed significantly lower CDS density than autosomes (Wilcoxon rank-sum test, $P=2.94\times 10^{-2}$; Supplementary Figure S3B), suggesting that the observed enrichment of high-density CNEE clusters is unlikely to result from elevated coding sequence density on ChrX.

To investigate the genomic features associated with these clusters, we compared their sequence composition and local genomic context with CNEE deserts and genomic background regions. GC content differed significantly among the three genomic contexts (Kruskal–Wallis test, $P<2.2\times 10^{-16}$). High-density CNEE clusters exhibited significantly lower GC content than both CNEE deserts and background regions (BH-adjusted Wilcoxon tests, $P<2\times 10^{-16}$), whereas CNEE deserts and background regions showed no significant difference ($P=0.33$; Figure 2C; Supplementary Table S8). Repeat coverage also differed significantly among genomic partitions (Kruskal–Wallis test, $P=1.98\times 10^{-11}$). High-density CNEE clusters showed significantly reduced repeat coverage relative to both CNEE deserts and background regions, with BH-adjusted Wilcoxon P values of 7.9×10^{-11} and 4.3×10^{-4} , respectively (Figure 2D; Supplementary

Table S8). Together, these results indicate that high-density CNEE clusters preferentially occur within genomic regions characterized by reduced repeat coverage and lower GC content.

We next examined whether high-density CNEE clusters were associated with local coding density. CDS density within high-density CNEE clusters was markedly lower than that in non-cluster regions (6.57 kb/Mb vs. 16.13 kb/Mb; Figure 2E), indicating that these regions are preferentially located within relatively coding-poor genomic contexts rather than being simply driven by local enrichment of coding sequences. Within high-density CNEE clusters, CDS density showed only a weak positive correlation with CNEE coverage density, a length-based measure of CNEE abundance expressed as covered sequence length per megabase (Spearman's $\rho=0.227$, $P=3.71\times10^{-7}$; Supplementary Figure S4A), while no significant association was detected for the complementary count-based CNEE element density metric (Spearman's $\rho=-0.073$, $P=0.105$; Supplementary Figure S4B).

Finally, we investigated the potential biological functions associated with genes located near high-density CNEE clusters. GO analysis identified significant enrichment of developmental processes, including pattern specification, regionalization, anterior/posterior pattern specification, limb development, appendage development, and positive regulation of cartilage development (Figure 2F; Supplementary Table S9). As a representative example, the ChrA2 region containing the *HOXA* gene cluster showed a broad high-density CNEE cluster with multiple local density peaks (Figure 2G). This region includes a dense accumulation of CNEEs adjacent to the developmentally important *HOXA* gene cluster, providing an example of the association between high-density CNEE clusters and conserved developmental regulatory loci.

Putative CNEE losses show reduced occurrence in high-density CNEE clusters but are enriched on ChrX

To investigate the genomic distribution of putative CNEE losses across felids, we used the domestic cat genome as the reference for CNEE coordinates and loss calling. In each non-reference species, candidate putative CNEE losses were first identified as CNEEs with more than 40% missing sequence and then filtered by synteny validation to obtain high-confidence putative losses. This filtering removed 3 313 of 8 292 raw loss candidates across species, accounting for approximately 39.95% of all initial candidates. The overall high-confidence CNEE retention rate on ChrX was 34.29%, compared with 66.06% on autosomes (Supplementary Table S10).

We next tested whether these high-confidence putative CNEE losses were preferentially located in high-density CNEE clusters. Most felid species showed odds ratios below 1, indicating that putative CNEE losses are generally underrepresented in these cluster regions (Figure 3A). After FDR correction, this pattern was significant in six species, including Canada lynx (*Lynx canadensis*), puma (*Puma concolor*), lion, fishing cat (*Prionailurus viverrinus*), cheetah, and black-footed cat (*F. nigripes*), with odds ratios ranging from 0.43 to 0.64. Leopard (*P. pardus*) was the only species with an odds ratio above 1, indicating a weak non-significant tendency toward enrichment. Overall, these results suggest a non-random genomic distribution of putative CNEE losses, with a tendency toward depletion in CNEE-rich cluster regions in several felid lineages.

To determine whether putative CNEE loss rates varied across chromosomes, we used PGLS to account for phylogenetic non-independence among species. The estimated mean $\log_2$ (ChrX-to-autosome loss-rate ratio) was significantly greater than zero across felids (estimate=1.35, SE=0.53, $t=2.55$, $df=15$, one-sided $P=0.011$; Figure

3B; Supplementary Figure S5). To evaluate the robustness of this pattern, we repeated the analysis using both 60% and 80% missing-sequence thresholds, both of which consistently supported elevated putative CNEE loss rates on ChrX (Supplementary Table S11). Species-specific ChrX-to-autosome loss-rate ratios were greater than 1 in all examined felids, but varied in magnitude from 1.07 in puma to 6.87 in jungle cat (*F. chaus*) (Supplementary Figure S6). Overall, these results suggest that putative CNEE losses tend to occur at higher rates on ChrX than on autosomes across felids.

We further examined whether putative CNEE losses were spatially isolated or aggregated into local blocks. Putative CNEE losses located within 50 kb of each other on the same chromosome were merged into CNEE-loss blocks, and block size was measured as the number of putative losses contained in each merged interval. Across species, most blocks were singletons, indicating that putative CNEE losses were generally spatially dispersed rather than broadly aggregated (Figure 3C).

Phylogenetically corrected regression revealed a significant negative association between N_e and species-level CNEE loss rate (PGLS, $\beta = -2.102$, $P = 0.0155$) (Figure 3D). Species with smaller N_e tended to exhibit higher CNEE loss rates, consistent with a stronger role of genetic drift in the CNEE losses. Notably, Canada lynx showed the highest loss rate among the analyzed species.

At the subfamily level, we identified a set of high-confidence putative CNEE losses shared across all sampled Pantherinae species. As shown in the heatmap, the majority of these elements exhibited near-complete absence of the corresponding CNEE sequences across Pantherinae lineages (Figure 3E). In contrast, no comparable set of putative losses shared across all sampled Felinae species was detected. To explore the potential functional context of these Pantherinae-shared putative losses, enrichment analysis of the associated genes revealed terms related to signaling, neurodevelopment,

and other developmental processes, including receptor-type tyrosine-protein phosphatases, DAG and IP3 signaling, positive regulation of neuron projection development, forebrain development, RHOA GTPase cycle, and signaling by the TGF- β receptor complex (Figure 3F). These enrichments suggest that genes associated with Pantherinae-shared CNEE losses are involved in developmental and signaling-related biological processes, including pathways related to neural development and cellular regulation.

Lineage-specific acceleration of CNEEs reveals candidate regulatory signatures associated with felid phenotypic diversification

Accelerated evolution of CNEEs may represent lineage-specific changes in conserved regulatory regions associated with phenotypic diversification. We identified species-specific accelerated CNEEs across the feline phylogeny and found substantial interspecific variation in their abundance. The cheetah showed the largest number of accelerated CNEEs ($n=314$), followed by the Canada lynx ($n=250$) and the snow leopard (*P. uncia*) ($n=132$) (Figure 4A). To determine whether this variation could be explained by lineage-specific differences in genome-wide neutral evolutionary rates, we compared accelerated CNEE counts with 4dSs-based neutral branch lengths. No significant correlation was detected between these variables ($R^2=0$, $P=0.97$; Figure 4B), suggesting that the lineage-biased distribution of accelerated CNEEs is unlikely to simply reflect differences in neutral substitution accumulation. We further tested for chromosomal and regional distribution biases of accelerated CNEEs. No evidence was found for an elevated acceleration rate on the X chromosome relative to autosomes across species (PGLS, $P=1.00$; Supplementary Figure S7). Similarly, accelerated CNEEs did not show consistent enrichment within high-density CNEE clusters across species (Supplementary Figure S8). These results indicate that accelerated CNEEs are

broadly distributed across the genome rather than being preferentially concentrated on specific chromosome types or CNEE-rich regions.

To investigate the potential biological relevance of accelerated CNEEs, we classified CNEE-associated genes predicted using a modified GREAT-like basal-plus-extension strategy into shared and strictly species-specific sets. CNEE-associated genes shared across species were defined as genes associated with accelerated CNEEs in more than one species, whereas strictly species-specific CNEE-associated genes were detected only in a single lineage.

Among shared associated genes, we identified candidate regulatory associations involving genes potentially related to environmental specialization. For example, *MYOCD* was identified as a shared associated gene of accelerated CNEEs in the Canada lynx and snow leopard, two phylogenetically distinct felids inhabiting cold environments. *MYOCD* is involved in cardiac muscle regulation and calcium-related signaling, and previous studies have linked its expression to the maintenance of cardiac contraction during deep hypothermia in hibernating mammals (Yang et al., 2021). Thus, accelerated CNEEs associated with *MYOCD* may therefore be relevant to cardiovascular-related processes under cold environmental conditions. Similarly, *GAS1* was identified as a shared associated gene of accelerated CNEEs in the snow leopard and Pallas's cat (*Otocolobus manul*), two felids living in cold, hypoxic montane environments. *GAS1* participates in Hedgehog signaling and developmental regulation and has been implicated in endothelial cell survival and vascular integrity (Spagnuolo et al., 2004), suggesting that *GAS1*-associated accelerated CNEEs may be relevant to vascular-related processes under environmental stress.

Among strictly species-specific accelerated CNEEs, the cheetah exhibited the largest number of lineage-specific candidates, with 231 predicted associated genes.

Functional enrichment analysis revealed significant associations with pathways related to neural signaling, cardiac function, and metabolic homeostasis (Figure 4C). These categories included genes involved in neural signaling, such as *GRIN2B*; muscle development and contraction, including *MEF2A* and *NEBL*; and glucose homeostasis, including *IGF1R* and *SESN3*. These functional categories are broadly consistent with physiological processes potentially relevant to the cheetah's extreme sprinting performance. Previous anatomical studies have shown that cheetah locomotor muscles contain a high proportion of fast-twitch fibers, up to 83%, together with elevated glycolytic enzyme capacities, which are critical for the mechanical power and anaerobic metabolic bursts required for explosive acceleration (Williams et al., 1997). These results suggest that species-specific accelerated CNEEs in the cheetah may be associated with genes involved in neuromuscular coordination, muscle function, and metabolic processes relevant to high-speed pursuit.

Beyond pathway-level enrichment in the cheetah, we identified additional lineage-specific accelerated CNEEs associated with genes involved in specialized biological processes (Figure 4D, E). In the clouded leopard, accelerated CNEEs were associated with *ATXN1* and *AH11*, genes involved in cerebellar development, motor coordination, and spatial orientation (Ferland et al., 2004; Rousseaux et al., 2018). These associations may be relevant to biological processes underlying specialized arboreal locomotion and three-dimensional movement. In the Pallas's cat, accelerated CNEEs were associated with *TWIST1*, a conserved basic helix-loop-helix transcription factor involved in neural crest development and craniofacial morphogenesis (Bildsoe et al., 2009), suggesting a candidate link to craniofacial morphological variation in this lineage. Dental-related associations were detected in the bobcat (*L. rufus*), where species-specific accelerated CNEEs were associated with *ODAPH*, a gene required for tooth enamel maturation

(Parry et al., 2012). Dental integrity is critical for bobcat predation and prey processing, and a survey of California bobcat skulls reported high frequencies of attrition/abrasion (85.2%) and tooth fractures (50.9%), indicating substantial functional demands on the dentition (Aghashani et al., 2016). *ODAPH*-associated accelerated CNEEs may therefore warrant further investigation in relation to enamel maturation and dental durability.

Extreme environments provide powerful natural systems for studying adaptive evolution, as strong ecological pressures can leave detectable genomic signatures associated with physiological and morphological specialization (Qiu et al., 2012; Li et al., 2025). Several lineage-specific candidates were associated with genes involved in physiological responses to extreme environments or specialized ecological niches. In the snow leopard, accelerated CNEEs were linked to *SQOR* and *ITGA6*, two genes related to hypoxia-associated cellular protection and vascular regulation. *SQOR* participates in hydrogen sulfide metabolism and has been implicated in protection against hypoxia-induced mitochondrial dysfunction and tissue injury (Marutani et al., 2021). *ITGA6* encodes an integrin involved in cell adhesion, angiogenesis, and vascular integrity, and has been reported to be regulated by hypoxia-inducible factors (Zhao et al., 2017). Notably, *ITGA6* has also been identified as a candidate gene under polygenic adaptation in Tibetan and Sherpa populations, where variants are associated with hemoglobin modulation and protection against high-altitude polycythemia (Zhao et al., 2017; Gneccchi-Ruscone et al., 2018). These findings suggest that *ITGA6*-associated accelerated CNEEs may be relevant to vascular homeostasis under high-elevation hypoxic stress. Sensory-related candidates were detected in the black-footed cat, including *KCNK2*, a thermosensitive potassium channel involved in thermal nociception (Alloui et al., 2006), and *ADGRV1*, a gene critical for auditory function



(McGee et al., 2006). These associations may be relevant to sensory demands in arid environments characterized by large diurnal temperature fluctuations and nocturnal hunting. A distinct physiological signal was observed in the fishing cat, where accelerated CNEEs were linked to the urea transporter genes *SLC14A1* and *SLC14A2* (Stewart, 2011), suggesting potential relevance to renal urea handling, osmotic balance, and nitrogen metabolism in the context of a high-protein piscivorous diet in wetland ecosystems.

Collectively, these lineage-specific and shared associations highlight accelerated CNEEs as candidate regulatory regions potentially linked to diverse morphological, physiological, sensory, and ecological processes across Felidae. Because these associations were inferred computationally through predefined regulatory domains rather than experimentally measured enhancer–promoter contacts, they should be interpreted as putative CNEE–gene links for functional interpretation, not as direct evidence of regulatory effects or phenotypic causation.

Accelerated CNEEs in the subfamily Pantherinae

The Pantherinae subfamily represents the major radiation of large felids and is characterized by pronounced differences in body size, musculoskeletal robustness, and predatory behavior relative to Felinae. To investigate whether CNEEs contributed to this evolutionary divergence, we used phyloP to identify CNEEs showing accelerated evolution across the Pantherinae subtree and along the ancestral Pantherinae branch.

Across the Pantherinae subtree, we identified 541 accelerated CNEEs associated with 690 putative target genes (Figure 5A; Supplementary Table S12). Enrichment analysis revealed that genes associated with Pantherinae-accelerated CNEEs were primarily involved in developmental and tissue organization processes (Figure 5B). The enriched functional modules were mainly related to cell morphogenesis, cell–cell



interactions, tissue morphogenesis, and nervous system development, with additional associations with organ developmental processes. These enrichment patterns suggest that Pantherinae-accelerated CNEEs are preferentially linked to regulatory networks controlling cellular organization, developmental patterning, and neural processes, which may be relevant to lineage-specific morphological diversification in large cats.

We next focused on the ancestral Pantherinae branch to pinpoint candidate regulatory changes emerging near the origin of this clade. This analysis identified 57 accelerated CNEEs and 87 putative target genes, among which 3 CNEEs and 19 genes overlapped with the results from the Pantherinae subtree analysis (Figure 5A; Supplementary Table S13). A striking candidate among these was mCE205254, which was assigned to *LAMA2*—a gene encoding laminin $\alpha 2$, an essential extracellular matrix component governing skeletal muscle integrity and neuromuscular function (Xu et al., 1994; Helbling-Leclerc et al., 1995; Gawlik & Durbeej, 2011). To evaluate whether local phylogenetic discordance affected the inference of accelerated evolution, we re-examined the highlighted *LAMA2*-associated CNEE. Phylogenetic reconstruction based on the aligned CNEE sequence produced a topology that differed from the genome-wide species tree. We therefore re-estimated the neutral model using this alternative topology and repeated the phyloP analysis. The *LAMA2*-associated CNEE remained significantly accelerated (acceleration test, $P=2.6\times 10^{-4}$), indicating that its accelerated signal is robust to the assumed phylogenetic topology. BLAT alignment mapped mCE205254 to an intronic region of the human *LAMA2* locus (Chr6:128 915 171–128 915 846 in hg38), directly intersecting with the ENCODE candidate cis-regulatory element (cCRE) EH38E4416809, an active chromatin region (Figure 5C).

Notably, local multiple sequence alignment revealed a Pantherinae-fixed 5-bp deletion. While this 5-bp segment is strictly conserved in non-Pantherinae felids and

carnivoran outgroups, it is completely deleted across all examined Pantherinae species (Figure 5D). To validate that this deletion was not caused by genome assembly artifacts, we mapped independent whole-genome resequencing reads from six Pantherinae species to this CNEE region. Visual inspection of the alignments in IGV consistently confirmed the presence of the 5-bp deletion in all six species (Supplementary Figure S9), providing independent support for the authenticity of this lineage-specific variant. To explore the potential regulatory implications of this deletion, we performed FIMO motif scanning to identify TFBSs that were disrupted or lost within the deleted sequence. The analysis predicted that the domestic cat sequence contains ZNF-like binding motifs overlapping the deleted segment. Relative to the domestic cat sequence, the Pantherinae 5-bp deletion is predicted to disrupt these ZNF-like binding motifs while creating a putative Pgr binding motif (JASPAR MA2323.1, FIMO $P=3.98\times 10^{-5}$) spanning the deletion junction (Figure 5E; Supplementary Table S14).

We further examined publicly available domestic cat epigenomic datasets to assess whether mCE205254 is associated with active regulatory chromatin. In domestic cat heart H3K27ac ChIP-seq data, a peak was detected in one biological replicate (rep1) across the mCE205254 locus, overlapping both the conserved element and the Pantherinae-fixed 5-bp deletion site (Supplementary Figure S10). In contrast, domestic cat brain ATAC-seq data did not show an accessibility peak at this locus, suggesting that the regulatory activity of mCE205254 may be tissue-specific rather than broadly active across tissues.

Together, these observations nominate mCE205254 as a candidate regulatory element associated with *LAMA2* that experienced a Pantherinae-specific micro-deletion. Although the conservation pattern, resequencing validation, cCRE overlap, H3K27ac signal, and predicted TFBS changes support a potential regulatory role, future

validation in relevant tissues (such as musculoskeletal epigenomic profiling and experimental assays like reporter assays or genome editing) will be necessary to determine whether this CNEE possesses true enhancer activity and whether the Pantherinae-specific 5-bp deletion affects gene regulation.

DISCUSSION

In this study, we leveraged long-read felid genome assemblies to establish a comprehensive genome-wide catalog of CNEEs and investigate their evolutionary dynamics across Felidae. CNEEs were unevenly distributed across felid genomes, forming discrete high-density clusters. The X chromosome displayed a distinctive CNEE-related pattern, with enrichment of high-density CNEE clusters and higher rates of putative CNEE loss than autosomes. Furthermore, across multiple phylogenetic scales—from individual species to the Pantherinae subfamily—we identified accelerated CNEEs and predicted CNEE–gene associations that suggest links to sensory, physiological, and morphological traits. Collectively, our findings illuminate the non-coding architecture of felid genome evolution, demonstrating that CNEEs show distinct evolutionary patterns across genomic regions and may provide candidate regulatory substrates associated with chromosome-specific evolutionary patterns and the broader phenotypic diversification of the felid radiation.

High-density CNEE clusters highlight candidate conserved developmental regulatory landscapes in felids

At the broad chromosomal scale, felid CNEEs exhibit a relatively uniform distribution, with no obvious chromosome-wide enrichment. However, this overall pattern masks pronounced local heterogeneity. The identification of high-density CNEE clusters indicates that these conserved non-coding elements are not randomly dispersed, but

instead are concentrated in discrete genomic neighborhoods, particularly around genes involved in developmental regulation. This clustered architecture is consistent with previous observations that CNEEs often accumulate in gene deserts flanking key developmental regulators, where they may function as long-range cis-regulatory elements within extended regulatory domains (Bejerano et al., 2004; Ovcharenko et al., 2005; Elgar & Vavouri, 2008).

The genomic features of these high-density clusters further support their interpretation as candidate conserved regulatory architectures. These clusters were generally located in regions with significantly reduced CDS density, relatively low GC content, and slightly reduced repeat coverage compared with background genomic regions. Because gene density and GC content are often positively correlated in vertebrate genomes, the concordance between low CDS density and low GC content supports the interpretation that these clusters occur in relatively gene-poor genomic contexts. Such gene-poor contexts are characteristic of extended developmental regulatory landscapes, as observed in other vertebrate genomes (Lander et al., 2001; Bejerano et al., 2004; Ovcharenko et al., 2005).

We further observed that putative lineage-specific CNEE losses tended to be less frequent within high-density clusters. Although this pattern was not consistently significant across species after multiple-testing correction, its direction is consistent with stronger evolutionary constraint in CNEE-rich regulatory regions. Together with their developmental associations and gene-poor genomic context, this reduced tendency toward CNEE loss supports the interpretation of high-density CNEE clusters as candidate conserved regulatory landscapes, although their degree of constraint may vary across lineages.

Distinctive evolutionary patterns of X-linked CNEEs



816 Against this conserved regulatory background, the felid X chromosome shows
817 distinctive CNEE evolutionary patterns. ChrX is disproportionately enriched for high-
818 density CNEE clusters and generally elevated putative CNEE loss rates across felid
819 species. These patterns highlight distinctive evolutionary dynamics of X-linked
820 conserved regulatory elements.

821 The enrichment of high-density CNEE clusters on ChrX may partly reflect the
822 distinctive regulatory architecture of X-linked genomic regions. Because X-linked
823 genes are hemizygous in males and subject to dosage compensation, their associated
824 regulatory elements may experience functional constraints that differ from those acting
825 on autosomal elements. CNEEs have been reported to be enriched on the X
826 chromosome in *Caenorhabditis elegans*, where this enrichment was proposed to reflect
827 a greater requirement for complex cis-regulatory architectures among X-linked genes
828 (Vavouri et al., 2007). Similarly, functional analyses of CNEEs on the human X
829 chromosome have shown strong associations with enhancer-related chromatin features
830 and experimentally supported regulatory activity (Naville et al., 2015). While these
831 observations provide a plausible functional context for the concentration of CNEEs in
832 X-linked regulatory landscapes, the precise contribution of functional constraint to this
833 elevated cluster density in felids remains to be tested experimentally.

834 The elevated putative CNEE loss rate on ChrX reflects chromosome-specific
835 features. Under idealized demographic assumptions, the smaller N_e of ChrX can
836 increase the relative influence of drift, thereby reducing the efficacy of purifying
837 selection against weakly deleterious losses. This mechanism is consistent with our
838 species-level association between smaller N_e and higher overall CNEE loss rates. The
839 net effect is nevertheless difficult to predict, because X-linked hemizyosity also
840 exposes deleterious alleles directly to purifying selection in males. In addition, this



effect may be modulated by species-specific evolutionary histories. Although smaller N_e was generally associated with higher overall putative CNEE loss rates at the species level, the Canada lynx and puma deviated from this trend, and their ChrX loss rates were not among the highest. These observations suggest that species-specific demographic histories and selective pressures may jointly shape the magnitude and distribution of putative CNEE loss. Thus, the elevated loss rate on ChrX likely reflects the combined effects of drift, selection, and chromosome-specific inheritance rather than a single dominant mechanism. Beyond population genetic processes, some putative CNEE losses may also contribute to regulatory remodeling. Because cis-regulatory elements are modular, the loss of individual enhancers can provide a substrate for evolutionary change. This is particularly relevant for CNEE losses shared across Pantherinae, which are nonrandomly distributed and are associated with genes enriched for developmental and signaling-related processes. Under the “less is more” model of evolutionary adaptation, loss of ancestral regulatory modules can contribute to phenotypic divergence (Kvon et al. 2016). Thus, Pantherinae-shared CNEE losses may represent candidate non-coding changes associated with lineage-specific regulatory evolution. However, their regulatory consequences and phenotypic relevance remain to be tested experimentally.

In contrast to elevated putative CNEE loss, ChrX showed no corresponding enrichment of accelerated CNEEs. This phenomenon may be linked to the lower neutral substitution rate characteristic of the X chromosome; as male-biased mutation in cats causes the X chromosome to spend less evolutionary time in the male germline than autosomes, it exhibits a slightly slower baseline substitution rate (Sayres et al., 2011). In addition, X-linked genes are subject to dosage compensation, and differences in X-inactivation status, together with hemizygous exposure and other features of X-



chromosome biology, mean that their associated regulatory elements may experience distinct functional constraints. This situation suggests that the currently identified accelerated CNEEs may have been subject to selective pressure to some extent, warranting further functional investigation into their potential roles in adaptive phenotypic diversification in Felidae.

Accelerated CNEEs as candidate regulatory substrates for felid ecological and phenotypic diversification

Understanding the molecular basis of adaptive evolution in wild animals is an important theme in evolutionary genomics, especially as genome resources expand for non-model species (Hu et al., 2023). Previous genomic studies of felid adaptation have predominantly focused on protein-coding variation, leaving the non-coding regulatory landscape largely underexplored. However, evidence from diverse taxa has shown that conserved non-coding elements often function as developmental regulatory elements, and that their lineage-specific divergence or acceleration can contribute to phenotypic evolution and adaptation (Pollard et al., 2006; Prabhakar et al., 2008; Kvon et al., 2016). More recent comparative genomic studies have further linked CNEE evolution to major morphological and ecological transitions across vertebrates (Peng et al., 2020; Wang et al., 2023; Chai et al., 2024; Du et al., 2026). By systematically identifying lineage-specific accelerated CNEEs at both the species and subfamily levels, our study addresses this knowledge gap in Felidae, a rapidly radiating mammalian clade.

Notably, the number of accelerated CNEEs detected across felid lineages showed no significant correlation with neutral branch lengths, indicating that their lineage-specific distribution cannot be explained solely by differences in genome-wide neutral substitution accumulation. Thus, these elements represent lineage-specific sequence divergence within otherwise conserved non-coding regions, making them plausible



891 candidates for investigating regulatory evolution during felid diversification.

892 Consistent with this interpretation, we identified accelerated CNEEs at multiple
893 evolutionary scales, with predicted target genes involved in biological processes
894 relevant to felid ecological specialization. At the species level, GREAT-based analyses
895 identified accelerated CNEEs associated with genes involved in physiological
896 regulation, sensory function, neural processes, and metabolic pathways. At the
897 subfamily level, Pantherinae-accelerated CNEEs were significantly enriched in
898 pathways related to tissue morphogenesis and skeletal system development. Among
899 these candidates, we identified a Pantherinae-specific 5-bp deletion and motif changes
900 within a putative intronic enhancer associated with *LAMA2*, a gene involved in skeletal
901 muscle basement membrane integrity and muscle fiber stability. Resequencing data
902 from six Pantherinae species supported the presence of this lineage-specific deletion,
903 and both comparative conservation analyses and phyloP results indicated that this
904 region has undergone accelerated evolution. Together, these observations nominate
905 mCE205254 as a candidate Pantherinae-shared regulatory variant associated with
906 *LAMA2* and potentially relevant to muscle biology; however, they do not establish
907 enhancer activity, target-gene regulation, or phenotypic effects.

908 These candidate regulatory changes can be viewed within the broader evolutionary
909 history of Felidae. Our phylogenomic estimates place the divergence between
910 Pantherinae and Felinae at approximately 12.59 Ma, consistent with a relatively recent
911 and rapid radiation of modern felids. During this interval, felids expanded globally and
912 colonized highly diverse environments, including hypoxic alpine regions, arid deserts,
913 and dense tropical forests. This ecological expansion was accompanied by substantial
914 variation in body size, locomotor strategies, and physiological specialization among
915 major lineages. The lineage-specific acceleration of conserved non-coding elements



may therefore represent an additional genomic layer associated with this diversification, complementing previously described changes in protein-coding regions.

An important limitation of this study is that the functional interpretation of accelerated CNEEs remains based on candidate associations rather than direct evidence of regulatory activity or adaptive causality. Gene associations inferred from the GREAT-like basal-plus-extension strategy, together with functional enrichment analyses, provide useful hypotheses about potential biological relevance, but they cannot determine whether individual accelerated CNEEs directly regulate associated genes, alter gene expression, or contribute to phenotypic traits. Likewise, resequencing validation of the Pantherinae-shared *LAMA2*-associated deletion confirms the presence of the sequence variant, but does not establish enhancer function. In addition, our genome-wide acceleration analyses rely on a species tree, whereas extensive incomplete lineage sorting and historical introgression in Felidae may result in local phylogenetic discordance that could influence lineage-specific acceleration inference. For the highlighted *LAMA2*-associated CNEE, local sequence alignment and phylogenetic reconstruction did not indicate that the acceleration signal was driven by local phylogenetic discordance, but such effects cannot be fully excluded across the genome. Future studies integrating locus-specific phylogenetic analyses, reporter assays, and genome editing will be necessary to further evaluate the evolutionary and regulatory significance of individual accelerated CNEEs and their potential molecular and phenotypic consequences.

COMPETING INTERESTS

The authors declare that they have no competing interests.

AUTHORS' CONTRIBUTIONS

K.X.L. and Y.B.H. conceived the ideas and designed the methodology; K.X.L. wrote the scripts, conducted the formal analyses, generated the visualizations, and led the writing of the manuscript; X.D. contributed to script development and data analyses, and revised the manuscript; Y.H.Z. contributed to data analyses and manuscript revision; Y.B.H. supervised the study, provided validation, acquired funding, and revised the manuscript. All authors read and approved the final version of the manuscript.

ACKNOWLEDGEMENTS

We thank Zheng Yan and Yu-Bo Hao from the Institute of Zoology, Chinese Academy of Sciences, for their valuable assistance with data analysis.

955 REFERENCES

- 956 Aghashani A, Kim AS, Kass PH, et al. 2016. Dental Pathology of the California Bobcat (*Lynx*
957 *rufus californicus*). *Journal of Comparative Pathology*, **154**(4): 329–340.
- 958 Alloui A, Zimmermann K, Mamet J, et al. 2006. TREK-1, a K⁺ channel involved in polymodal
959 pain perception. *The EMBO Journal*, **25**(11): 2368–2376.
- 960 Bejerano G, Pheasant M, Makunin I, et al. 2004. Ultraconserved Elements in the Human
961 Genome. *Science*, **304**(5675): 1321–1325.
- 962 Bildsoe H, Loebel DAF, Jones VJ, et al. 2009. Requirement for *Twist1* in frontonasal and skull
963 vault development in the mouse embryo. *Developmental Biology*, **331**(2): 176–188.
- 964 Blanchette M, Kent WJ, Riemer C, et al. 2004. Aligning multiple genomic sequences with the
965 threaded blockset aligner. *Genome Research*, **14**(4): 708–715.
- 966 Bredemeyer KR, Hillier L, Harris AJ, et al. 2023. Single-haplotype comparative genomics
967 provides insights into lineage-specific structural variation during cat evolution. *Nature*
968 *Genetics*, **55**(11): 1953–1963.
- 969 Chai S, Chong Y, Yin D, et al. 2024. Genomic insights into adaptation to bipedal saltation and
970 desert-like habitats of jerboas. *Science China Life Sciences*, **67**(9): 2003–2015.
- 971 Charlesworth B, Coyne JA, Barton NH. 1987. The Relative Rates of Evolution of Sex
972 Chromosomes and Autosomes. *The American Naturalist*, **130**(1): 113–146.
- 973 Chen S, Zhou Y, Chen Y, et al. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor.
974 *Bioinformatics*, **34**(17): i884–i890.
- 975 Danecek P, Bonfield JK, Liddle J, et al. 2021. Twelve years of SAMtools and BCFtools.
976 *GigaScience*, **10**(2): giab008.
- 977 Distech CM. 2012. Dosage compensation of the sex chromosomes. *Annual review of genetics*,
978 **46**(1): 537–560.
- 979 Du X, Fan H, Hu Y, et al. 2026. Functional convergence of regulatory regions provides vital
980 insights into mammalian gliding adaptation. *Science China Life Sciences*,
981 <https://doi.org/10.1007/s11427-026-3340-3>.



- 982 Edwards SV, Cloutier A, Baker AJ. 2017. Conserved nonexonic elements: A novel class of
983 marker for phylogenomics. *Systematic Biology*, **66**(6): 1028–1044.
- 984 Elgar G, Vavouri T. 2008. Tuning in to the signals: noncoding sequence conservation in
985 vertebrate genomes. *Trends in Genetics*, **24**(7): 344–352.
- 986 Ellegren H. 2011. Sex-chromosome evolution: recent progress and the influence of male and
987 female heterogamety. *Nature Reviews Genetics*, **12**(3): 157–166.
- 988 Ferland RJ, Eyaid W, Collura RV, et al. 2004. Abnormal cerebellar development and axonal
989 decussation due to mutations in *AH1* in Joubert syndrome. *Nature Genetics*, **36**(9):
990 1008–1013.
- 991 Figueiró HV, Li G, Trindade FJ, et al. 2017. Genome-wide signatures of complex introgression
992 and adaptive evolution in the big cats. *Science Advances*, **3**(7): e1700299.
- 993 Gawlik KI, Durbeej M. 2011. Skeletal muscle laminin and MDC1A: pathogenesis and
994 treatment strategies. *Skeletal Muscle*, **1**(1): 9.
- 995 Gneccchi-Ruscione GA, Abondio P, De Fanti S, et al. 2018. Evidence of polygenic adaptation to
996 high altitude from Tibetan and Sherpa genomes. *Genome Biology and Evolution*,
997 <https://doi.org/10.1093/gbe/evy233>.
- 998 Grant CE, Bailey TL, Noble WS. 2011. FIMO: scanning for occurrences of a given motif.
999 *Bioinformatics*, **27**(7): 1017–1018.
- 1000 Harris RS. 2007. Improved pairwise alignment of genomic DNA. *The Pennsylvania State*
1001 *University ProQuest Dissertations & Theses*.
- 1002 Helbling-Leclerc A, Zhang X, Topaloglu H, et al. 1995. Mutations in the laminin alpha 2-chain
1003 gene (*LAMA2*) cause merosin-deficient congenital muscular dystrophy. *Nature*
1004 *genetics*, **11**(2): 216–218.
- 1005 Hu Y, Wang X, Xu Y, et al. 2023. Molecular mechanisms of adaptive evolution in wild animals
1006 and plants. *Science China Life Sciences*, **66**(3): 453–495.
- 1007 Hubisz MJ, Pollard KS, Siepel A. 2011. PHAST and RPHAST: phylogenetic analysis with
1008 space/time models. *Briefings in Bioinformatics*, **12**(1): 41–51.
- 1009 Johnson WE, Eizirik E, Pecon-Slaterry J, et al. 2006. The Late Miocene radiation of modern



1010 Felidae: A genetic assessment. *Science*, **311**(5757): 73–77.

1011 Kaelin CB, McGowan KA, Barsh GS. 2021. Developmental genetics of color pattern
1012 establishment in cats. *Nature Communications*, **12**(1): 5127.

1013 Kaelin CB, Xu X, Hong LZ, et al. 2012. Specifying and sustaining pigmentation patterns in
1014 domestic and wild Cats. *Science*, **337**(6101): 1536–1541.

1015 Kitchener AC, Breitenmoser-Würsten C, Eizirik E, et al. 2017. A revised taxonomy of the
1016 Felidae: The final report of the Cat Classification Task Force of the IUCN Cat
1017 Specialist Group. *Cat News*, Special Issue 11: 1–80.

1018 Kumar S, Suleski M, Craig JM, et al. 2022. TimeTree 5: An expanded resource for species
1019 divergence times. *Molecular Biology and Evolution*, **39**(8): msac174.

1020 Kvon EZ, Kamneva OK, Melo US, et al. 2016. Progressive loss of function in a limb enhancer
1021 during snake evolution. *Cell*, **167**(3): 633–642.

1022 Lander ES, Linton LM, Birren B, et al. 2001. Initial sequencing and analysis of the human
1023 genome. *Nature*, **409**(6822): 860–921.

1024 Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods*,
1025 **9**(4): 357–359.

1026 Leybold NA, Speicher MR. 2021. Evolutionary conservation in noncoding genomic regions.
1027 *Trends in Genetics*, **37**(10): 903–918.

1028 Li G, Davis BW, Eizirik E, et al. 2016. Phylogenomic evidence for ancient hybridization in the
1029 genomes of living cats (Felidae). *Genome Research*, **26**(1): 1–11.

1030 Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*, **34**(18):
1031 3094–3100.

1032 Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform.
1033 *Bioinformatics*, **25**(14): 1754–1760.

1034 Li H, Durbin R. 2011. Inference of human population history from individual whole-genome
1035 sequences. *Nature*, **475**(7357): 493–496.

1036 Li H, Handsaker B, Wysoker A, et al. 2009. The Sequence Alignment/Map format and
1037 SAMtools. *Bioinformatics*, **25**(16): 2078–2079.



1038 Li K, Liu Y, Zhang Y, et al. 2025. Host genomic convergence, rather than gut microbiome
 1039 convergence, underlies the convergent evolution of subterranean adaptation in
 1040 mammals. *Cell Reports*, **44**(8): 116110.

1041 Li Q, Brown JB, Huang H, et al. 2011. Measuring reproducibility of high-throughput
 1042 experiments. *The Annals of Applied Statistics*, **5**(3): 1752–1779.

1043 Lowe CB, Kellis M, Siepel A, et al. 2011. Three periods of regulatory innovation during
 1044 vertebrate evolution. *Science*, **333**(6045): 1019–1024.

1045 Marutani E, Morita M, Hirai S, et al. 2021. Sulfide catabolism ameliorates hypoxic brain injury.
 1046 *Nature Communications*, **12**(1): 3108.

1047 McGee J, Goodyear RJ, McMillan DR, et al. 2006. The very large G-protein-coupled receptor
 1048 VLGR1: A component of the ankle link complex required for the normal development
 1049 of auditory hair bundles. *Journal of Neuroscience*, **26**(24): 6543–6553.

1050 McLean CY, Bristor D, Hiller M, et al. 2010. GREAT improves functional interpretation of
 1051 cis-regulatory regions. *Nature Biotechnology*, **28**(5): 495–501.

1052 Nash WG, O'Brien SJ. 1982. Conserved regions of homologous G-banded chromosomes
 1053 between orders in mammalian evolution: carnivores and primates. *Proceedings of the*
 1054 *National Academy of Sciences*, **79**(21): 6631–6635.

1055 Naville M, Ishibashi M, Ferg M, et al. 2015. Long-range evolutionary constraints reveal cis-
 1056 regulatory interactions on the human X chromosome. *Nature Communications*, **6**(1):
 1057 6904.

1058 Osipova E, Hecker N, Hiller M. 2019. RepeatFiller newly identifies megabases of aligning
 1059 repetitive sequences and improves annotations of conserved non-exonic elements.
 1060 *GigaScience*, **8**(11): giz132.

1061 Ovcharenko I, Loots GG, Nobrega MA, et al. 2005. Evolution and functional classification of
 1062 vertebrate gene deserts. *Genome Research*, **15**(1): 137–145.

1063 Parry DA, Brookes SJ, Logan CV, et al. 2012. Mutations in *C4orf26*, encoding a peptide with
 1064 in vitro hydroxyapatite crystal nucleation and growth activity, cause amelogenesis
 1065 imperfecta. *The American Journal of Human Genetics*, **91**(3): 565–571.



- 1066 Peng C, Ren J-L, Deng C, et al. 2020. The genome of Shaw's sea snake (*Hydrophis curtus*)
1067 reveals secondary adaptation to its marine environment. *Molecular biology and*
1068 *evolution*, **37**(6): 1744–1760.
- 1069 Pennacchio LA, Ahituv N, Moses AM, et al. 2006. In vivo enhancer analysis of human
1070 conserved non-coding sequences. *Nature*, **444**(7118): 499–502.
- 1071 Pollard KS, Hubisz MJ, Rosenbloom KR, et al. 2010. Detection of nonneutral substitution rates
1072 on mammalian phylogenies. *Genome Research*, **20**(1): 110–121.
- 1073 Pollard KS, Salama SR, Lambert N, et al. 2006. An RNA gene expressed during cortical
1074 development evolved rapidly in humans. *Nature*, **443**(7108): 167–172.
- 1075 Prabhakar S, Visel A, Akiyama JA, et al. 2008. Human-specific gain of function in a
1076 developmental enhancer. *Science*, **321**(5894): 1346–1350.
- 1077 Qiu Q, Zhang G, Ma T, et al. 2012. The yak genome and adaptation to life at high altitude.
1078 *Nature Genetics*, **44**(8): 946–949.
- 1079 Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic
1080 features. *Bioinformatics*, **26**(6): 841–842.
- 1081 Rambaut A, Drummond AJ, Xie D, et al. 2018. Posterior summarization in bayesian
1082 phylogenetics using tracer 1.7. *Systematic Biology*, **67**(5): 901–904.
- 1083 Rousseaux MWC, Tschumperlin T, Lu H-C, et al. 2018. ATXN1-CIC complex is the primary
1084 driver of cerebellar pathology in spinocerebellar ataxia type 1 through a gain-of-
1085 function mechanism. *Neuron*, **97**(6): 1235–1243.
- 1086 Sackton TB, Grayson P, Cloutier A, et al. 2019. Convergent regulatory evolution and loss of
1087 flight in paleognathous birds. *Science*, **364**(6435): 74–78.
- 1088 Sayres MAW, Venditti C, Pagel M, et al. 2011. Do variations in substitution rates and male
1089 mutation bias correlate with life-history traits? A study of 32 mammalian genomes.
1090 *Evolution*, **65**(10): 2800–2815.
- 1091 Scholak T, Schucher N, Bahdanau D. 2021. PICARD: parsing incrementally for constrained
1092 auto-regressive decoding from language models. *In Proceedings of the 2021*
1093 *Conference on Empirical Methods in Natural Language Processing*, 9895–9901.



- 1094 Siepel A, Bejerano G, Pedersen JS, et al. 2005. Evolutionarily conserved elements in vertebrate,
1095 insect, worm, and yeast genomes. *Genome Research*, **15**(8): 1034–1050.
- 1096 Simão FA, Waterhouse RM, Ioannidis P, et al. 2015. BUSCO: assessing genome assembly and
1097 annotation completeness with single-copy orthologs. *Bioinformatics*, **31**(19): 3210–
1098 3212.
- 1099 Spagnuolo R, Corada M, Orsenigo F. 2004. Gas1 is induced by VE-cadherin and vascular
1100 endothelial growth factor and inhibits endothelial cell apoptosis. *Blood*, **103**(8): 3005–
1101 3012.
- 1102 Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of
1103 large phylogenies. *Bioinformatics*, **30**(9): 1312–1313.
- 1104 Stewart G. 2011. The emerging physiological roles of the *SLC14A* family of urea transporters.
1105 *British Journal of Pharmacology*, **164**(7): 1780–1792.
- 1106 Suarez HG, Langer BE, Ladde P, et al. 2017. chainCleaner improves genome alignment
1107 specificity and sensitivity. *Bioinformatics*, **33**(11): 1596–1603.
- 1108 Sunquist M, Sunquist F. 2002. *Wild cats of the world*. University of Chicago press.
- 1109 Supple MA, Escalona M, Alexandre N, et al. 2025. A chromosome-level genome assembly of
1110 the mountain lion, *Puma concolor*. *Journal of Heredity*, **116**(4): 479–487.
- 1111 Talavera G, Castresana J. 2007. Improvement of phylogenies after removing divergent and
1112 ambiguously aligned blocks from protein sequence alignments. *Systematic Biology*,
1113 **56**(4): 564–577.
- 1114 Thorvaldsdottir H, Robinson JT, Mesirov JP. 2013. Integrative Genomics Viewer (IGV): high-
1115 performance genomics data visualization and exploration. *Briefings in Bioinformatics*,
1116 **14**(2): 178–192.
- 1117 Vavouri T, Walter K, Gilks WR, et al. 2007. Parallel evolution of conserved non-coding
1118 elements that target a common set of developmental regulatory genes from worms to
1119 humans. *Genome Biology*, **8**(2): R15.
- 1120 Wang Z, Peng C, Wu W, et al. 2023. Developmental regulation of conserved non-coding
1121 element evolution provides insights into limb loss in squamates. *Science China Life*



- 1122 *Sciences*, **66**(10): 2399–2414.
- 1123 Werdelin L, Yamaguchi N, Johnson WE, et al. 2010. Phylogeny and evolution of cats (Felidae).
- 1124 In: *Biology and conservation of wild felids*, 59–82.
- 1125 Wilder AP, Supple MA, Subramanian A, et al. 2023. The contribution of historical processes
- 1126 to contemporary extinction risk in placental mammals. *Science*, **380**(6643): eabn5856.
- 1127 Williams TM, Dobson GP, Mathieu-Costello O, et al. 1997. Skeletal muscle histology and
- 1128 biochemistry of an elite sprinter, the African cheetah. *Journal of Comparative*
- 1129 *Physiology B: Biochemical, Systemic, and Environmental Physiology*, **167**(8): 527–535.
- 1130 Winter S, Meißner R, Greve C, et al. 2023. A chromosome-scale high-contiguity genome
- 1131 assembly of the cheetah (*Acinonyx jubatus*). *Journal of Heredity*, **114**(3): 271–278.
- 1132 Woolfe A, Elgar G. 2008. Chapter 12 Organization of conserved elements near key
- 1133 developmental regulators in vertebrate genomes. In: *Advances in Genetics*. Elsevier,
- 1134 307–338.
- 1135 Woolfe A, Goodson M, Goode DK, et al. 2004. Highly conserved non-coding sequences are
- 1136 associated with vertebrate development. *PLoS Biology*, **3**(1): e7.
- 1137 Wurster-Hill D, Centerwall W. 1982. The interrelationships of chromosome banding patterns
- 1138 in canids, mustelids, hyena, and felids. *Cytogenetic and Genome Research*, **34**(1–2):
- 1139 178–192.
- 1140 Xu H, Christmas P, Wu XR, et al. 1994. Defective muscle basement membrane and lack of M-
- 1141 laminin in the dystrophic *dy/dy* mouse. *Proceedings of the National Academy of*
- 1142 *Sciences*, **91**(12): 5572–5576.
- 1143 Yang L, Jin H, Yang Q, et al. 2025. Genomic evidence for low genetic diversity but purging of
- 1144 strong deleterious variants in snow leopards. *Genome Biology*, **26**(1): 94.
- 1145 Yang L, Li R-C, Xiang B, et al. 2021. Transcriptional regulation of intermolecular Ca²⁺
- 1146 signaling in hibernating ground squirrel cardiomyocytes: The myocardin–junctophilin
- 1147 axis. *Proceedings of the National Academy of Sciences*, **118**(14): e2025333118.
- 1148 Yang Y, Pang X-X, Ding Y-M, et al. 2026. Synergizing bayesian and heuristic approaches: D-
- 1149 BPP uncovers ghost introgression in *Panthera* and *Thuja*. *Systematic Biology*, **0**(0): 1–



- 1150 19.
- 1151 Yang Z. 2007. PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Molecular Biology*
1152 *and Evolution*, **24**(8): 1586–1591.
- 1153 Yu G, Wang L-G, Han Y, et al. 2012. clusterProfiler: an R package for comparing biological
1154 themes among gene clusters. *OMICS: A Journal of Integrative Biology*, **16**(5): 284–
1155 287.
- 1156 Yuan J, Kitchener AC, Lackey LB, et al. 2024. The genome of the black-footed cat: Revealing
1157 a rich natural history and urgent conservation priorities for small felids. *Proceedings*
1158 *of the National Academy of Sciences*, **121**(2): e2310763120.
- 1159 Zhang Y, Liu T, Meyer CA, et al. 2008. Model-based Analysis of ChIP-Seq (MACS). *Genome*
1160 *Biology*, **9**(9): R137.
- 1161 Zhao Y, Zhang Z, Liu L, et al. 2017. Associations of high altitude polycythemia with
1162 polymorphisms in *EPAS1*, *ITGA6* and *ERBB4* in Chinese Han and Tibetan populations.
1163 *Oncotarget*, **8**(49): 86736–86746.
- 1164 Zhou X, Pan D, Zhou J, et al. 2026. Cis-regulatory evolution reveals sensory trade-offs as a
1165 genetic basis for temporal niche evolution in tapirs. *Science Advances*, **12**: eadz4758.
- 1166 Zhou Y, Zhou B, Pache L, et al. 2019. Metascape provides a biologist-oriented resource for the
1167 analysis of systems-level datasets. *Nature Communications*, **10**(1): 1523.

1168

1169

1170

1171

1172

1173

1174

1175

1176



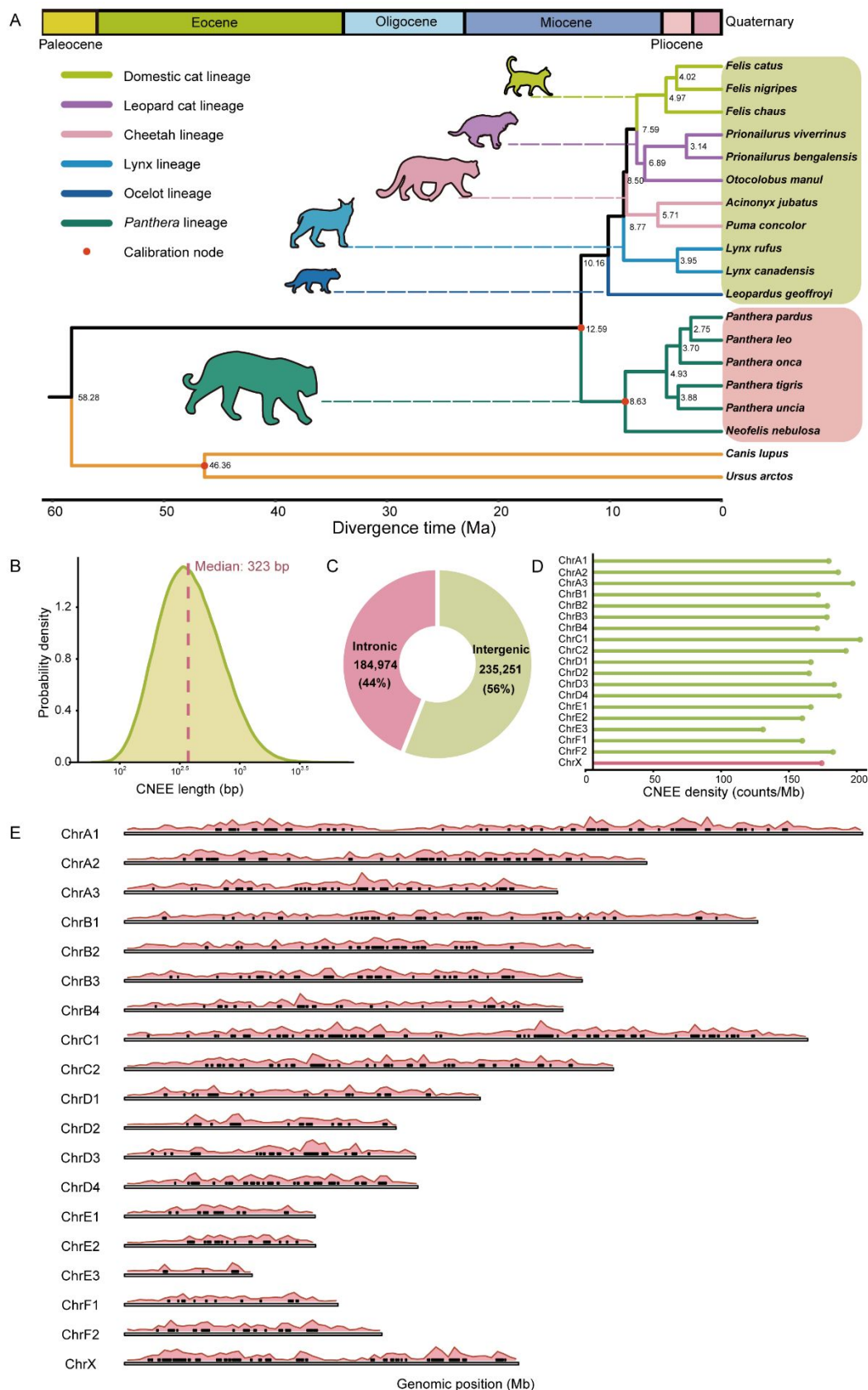


Figure 1 Phylogenetic reconstruction and genome-wide characterization of feline CNEEs

1180 A: Time-calibrated phylogeny of 17 felid species, with brown bear and gray wolf as
1181 outgroups. Node values indicate divergence times in Ma, and red circles mark
1182 calibration nodes. Calibration constraints and 95% HPD intervals are listed in
1183 Supplementary Tables S3 and S6, respectively. B: Length distribution of identified
1184 feline CNEEs. The dashed line indicates the median length. C: Proportions of CNEEs
1185 located in intergenic and intronic regions. D: CNEE density across chromosomes. E:
1186 Genome-wide local density landscape of CNEEs within chromosomes. Black blocks
1187 indicate high-density CNEE clusters. CNEEs: Conserved non-exonic elements; HPD:
1188 Highest posterior density; Ma: Million years ago.

1189



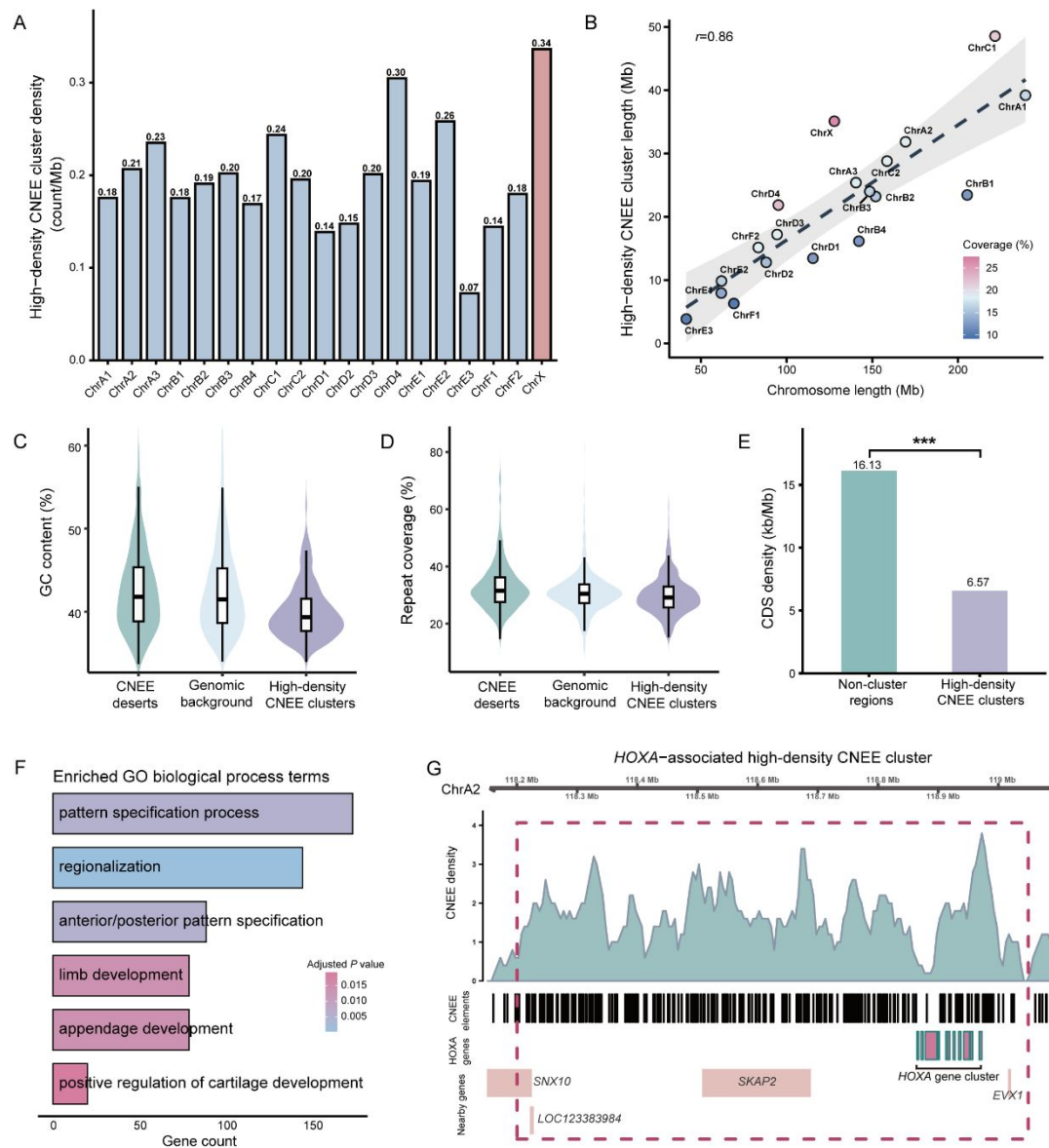


Figure 2 Characterization of high-density CNEE clusters in the feline genome

A: Density of high-density CNEE clusters across chromosomes. B: Relationship between chromosome length and total length of high-density CNEE clusters. Point color indicates the percentage of chromosome length covered by high-density CNEE clusters. C: GC content comparison among CNEE deserts, genomic background regions, and high-density CNEE clusters. D: Repeat coverage comparison among CNEE deserts, genomic background regions, and high-density CNEE clusters. E: Comparison of CDS density between non-cluster genomic regions and high-density CNEE clusters. Statistical significance was determined using a two-sided Wilcoxon rank-sum test. ***:

$P < 0.001$. F: Enriched Gene Ontology biological process terms for genes associated with high-density CNEE clusters. G: *HOXA*-associated high-density CNEE cluster on ChrA2. The dashed box indicates the high-density CNEE cluster region. CNEE: Conserved non-exonic element; CDS: Coding sequence; GO: Gene Ontology.

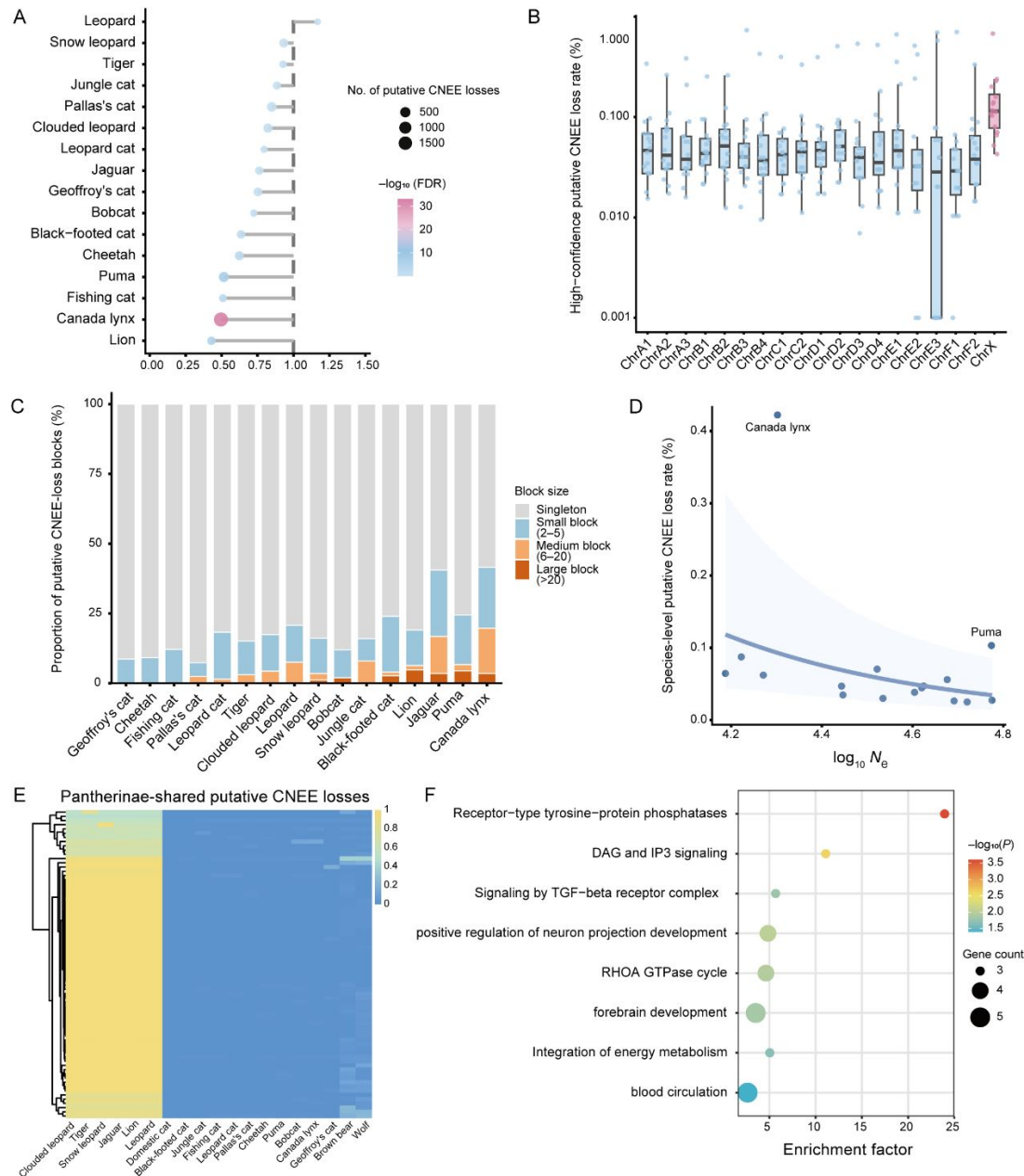


Figure 3 Landscape of putative CNEE losses across felid genomes

A: Fisher's exact tests assessing the depletion of putative CNEE losses from high-density CNEE clusters. The dashed line indicates an odds ratio of 1. Dot size represents

the number of high-confidence putative CNEE losses, and color indicates $-\log_{10}(\text{FDR})$.
B: Chromosome-level putative CNEE loss rates across felid species. ChrX is highlighted in pink. The y-axis is shown on a $\log_{10}$ scale, with a pseudo-count of 0.001 added for visualization. C: Proportions of putative CNEE-loss blocks by block size. Blocks were generated by merging adjacent putative lost CNEEs located within 50 kb of each other. D: PGLS relationship between N_e and putative CNEE loss rate ($\beta=-2.102$, $P=0.0155$). Points represent species, the line shows the PGLS fit, and shading indicates the 95% confidence interval. E: Heatmap showing loss rates of Pantherinae-shared putative CNEE losses across felid species and two outgroups. F: Functional enrichment of genes associated with Pantherinae-shared putative CNEE losses. Dot size represents gene count, and color represents $-\log_{10}(P)$. CNEE: Conserved non-exonic element; FDR: False discovery rate; PGLS: Phylogenetic generalized least squares; N_e : Effective population size.



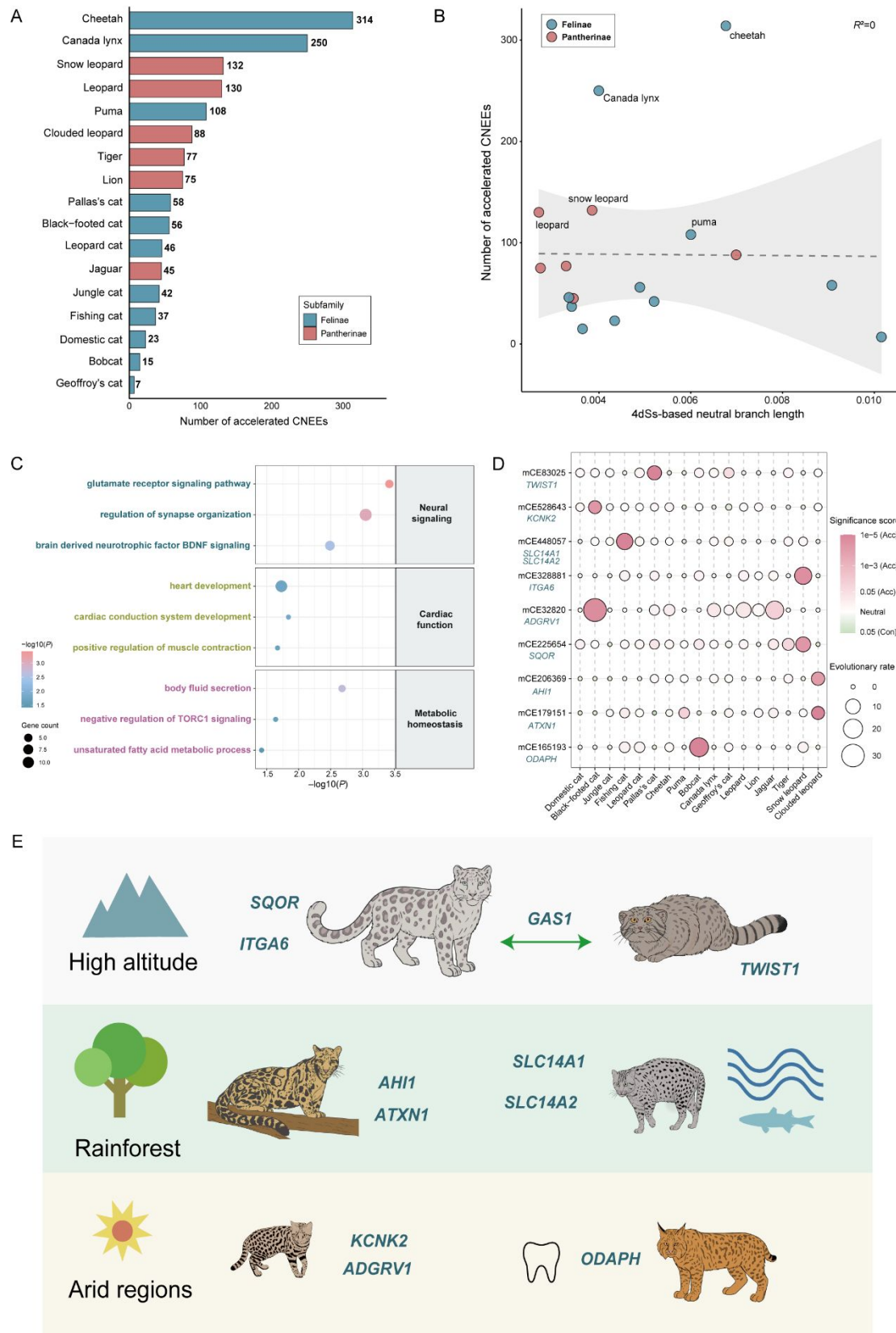


Figure 4 Lineage-specific acceleration of CNEs and putative target genes across Felidae

1226 A: Numbers of species-specific accelerated CNEEs across felid lineages. Bars are
1227 colored by subfamily. B: Relationship between species-specific accelerated CNEE
1228 counts and 4dSs-based neutral branch length. The dashed line represents the fitted
1229 regression line, and the shaded area indicates the confidence interval. C: Functional
1230 enrichment of strictly species-specific putative target genes associated with accelerated
1231 CNEEs in the cheetah. Point size represents gene count, and color represents $-\log_{10}(P)$.
1232 D: Species-specific acceleration patterns of selected CNEEs linked to candidate genes
1233 across felids. Point color indicates the significance category, with red denoting
1234 accelerated evolution and green denoting conserved evolution; point size indicates
1235 evolutionary rate. E: Schematic overview of representative putative target genes
1236 associated with accelerated CNEEs in selected lineage-specific ecological or
1237 phenotypic contexts. CNEE: Conserved non-exonic element; 4dSs: Fourfold
1238 degenerate synonymous sites; Acc: Accelerated evolution; Con: Conserved evolution.
1239
1240

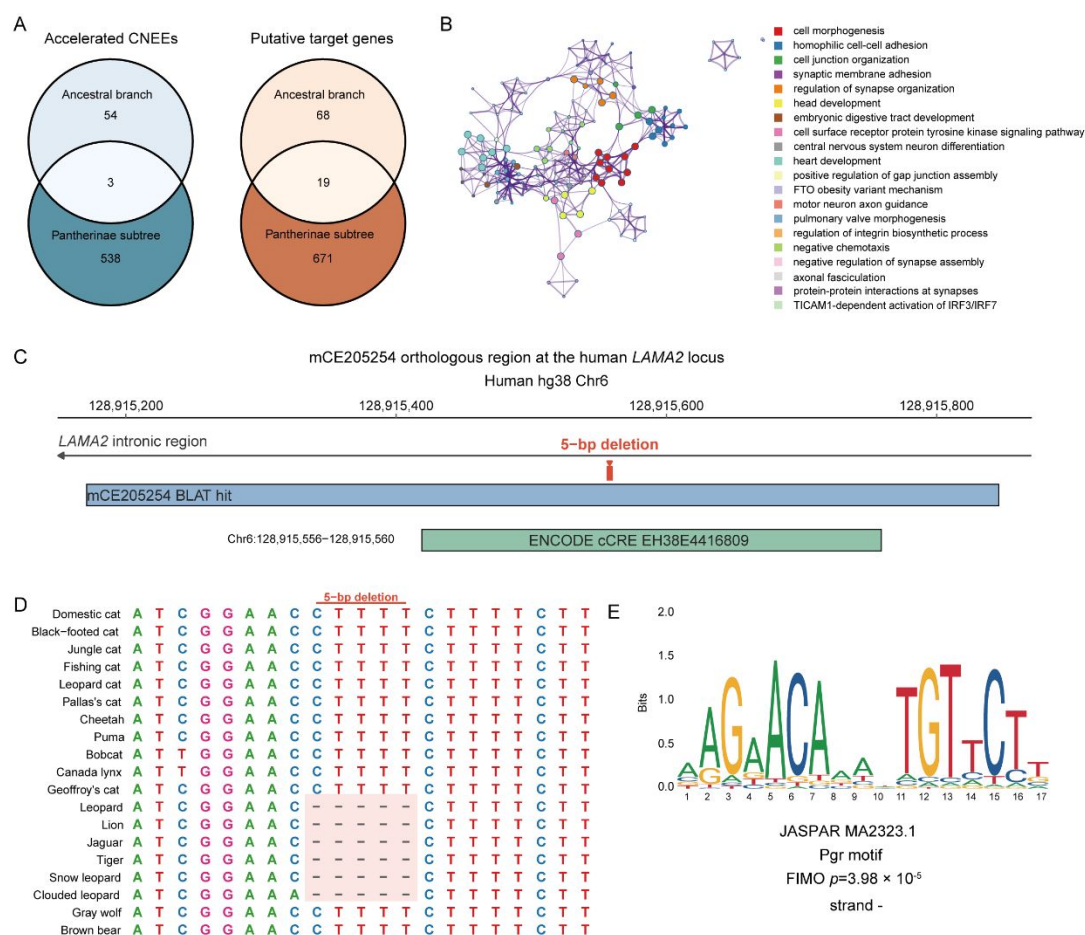


Figure 5 Pantherinae-accelerated CNEEs and a candidate TFBS turnover event identified within a *LAMA2*-associated CNEE

A: Overlap of accelerated CNEEs and putative target genes identified on the ancestral Pantherinae branch and across the Pantherinae subtree. B: Enrichment network of biological terms enriched among accelerated CNEE-associated genes in the Pantherinae subtree. C: Human orthologous region of mCE205254 within the *LAMA2* locus. The BLAT hit overlaps ENCODE cCRE EH38E4416809, and the Pantherinae-shared 5-bp deletion maps within this cCRE. D: Multiple sequence alignment of mCE205254 showing the Pantherinae-shared 5-bp deletion. E: JASPAR MA2323.1 Pgr motif predicted by FIMO at the deletion junction. CNEE: Conserved non-exonic element; ENCODE: Encyclopedia of DNA Elements; cCRE: Candidate cis-regulatory element; FIMO: Find Individual Motif Occurrences; Pgr: Progesterone receptor.

1254 **Supplementary Data:**

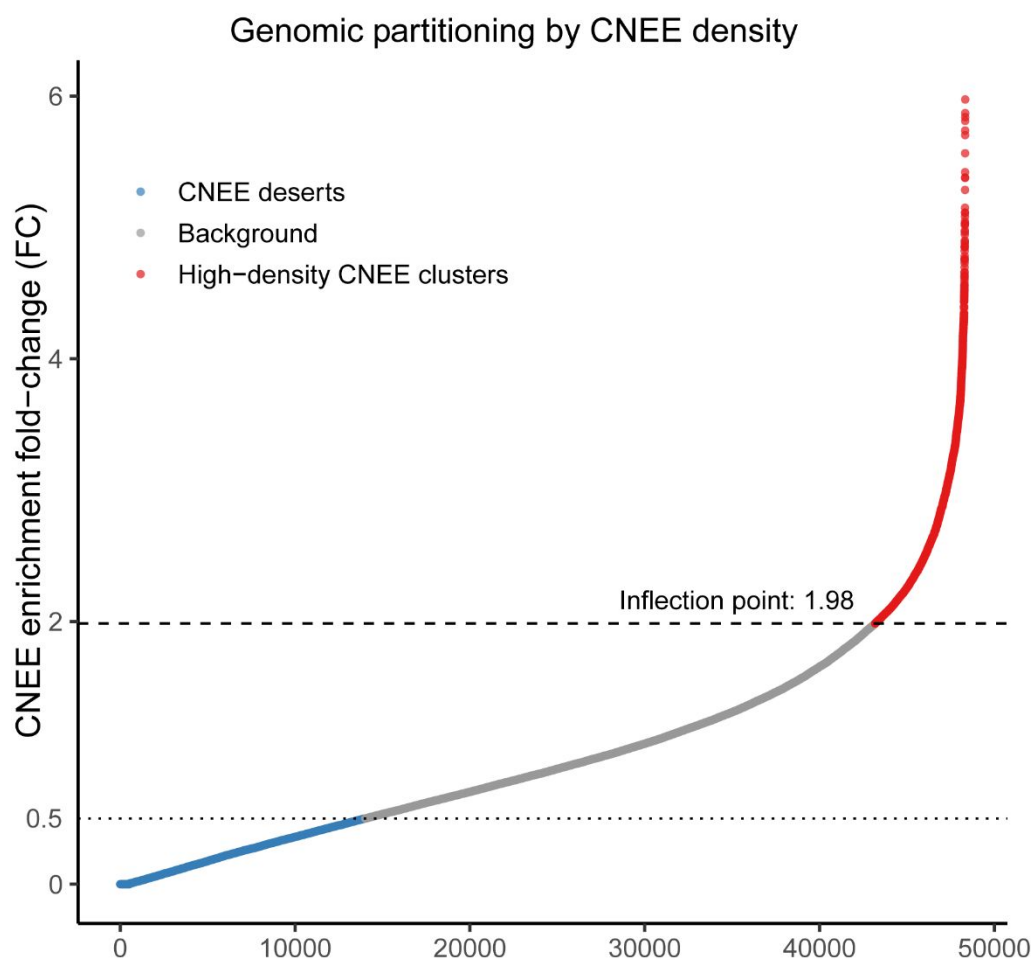
- 1255 1. Supplementary Figure S1-S10
- 1256 2. Supplementary TableS1: Summary of genome assemblies and quality statistics for species
1257 analyzed in this study
- 1258 3. Supplementary TableS2: Comparison of MAF-based coverage between autosomes and
1259 chromosome X across felid species
- 1260 4. Supplementary TableS3: Calibration constraints derived from TimeTree for divergence time
1261 estimation
- 1262 5. Supplementary TableS4: Sensitivity analyses of CNEE high-density cluster calling
1263 parameters and X-chromosome enrichment
- 1264 6. Supplementary TableS5: Whole-genome resequencing data used for effective population size
1265 estimation in felid species
- 1266 7. Supplementary TableS6: Divergence time estimates and 95% HPD intervals of major nodes
1267 in Felidae
- 1268 8. Supplementary TableS7: Chromosome-level statistics of high-density CNEE clusters in the
1269 feline genome
- 1270 9. Supplementary TableS8: Statistical comparisons of GC content and repeat coverage among
1271 CNEE deserts, background regions, and high-density CNEE clusters
- 1272 10. Supplementary TableS9: Gene Ontology enrichment results for genes associated with high-
1273 density CNEE clusters
- 1274 11. Supplementary TableS10: Summary of raw and high-confidence putative CNEE losses after
1275 synteny validation across felid species
- 1276 12. Supplementary TableS11: Comparison of putative loss rates between chromosome X and
1277 autosomes under different thresholds
- 1278 13. Supplementary TableS12: Accelerated CNEEs across the Pantherinae subtree and their
1279 putative target genes
- 1280 14. Supplementary TableS13: Accelerated CNEEs identified along ancestral branches within
1281 Pantherinae and their putative target genes



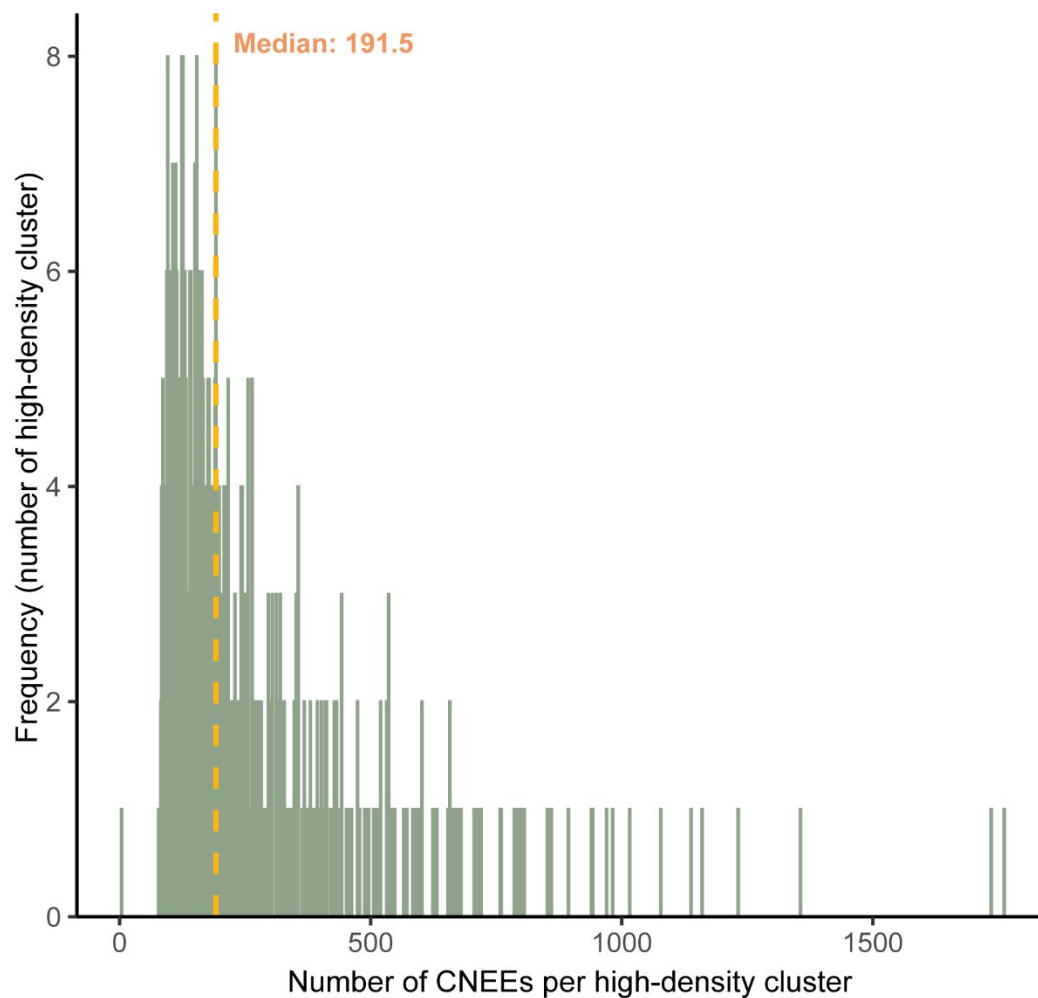
1282 15. Supplementary TableS14: FIMO motif predictions for the *LAMA2*-associated mCE205254
1283 local sequences
1284
1285

uncorrected proof

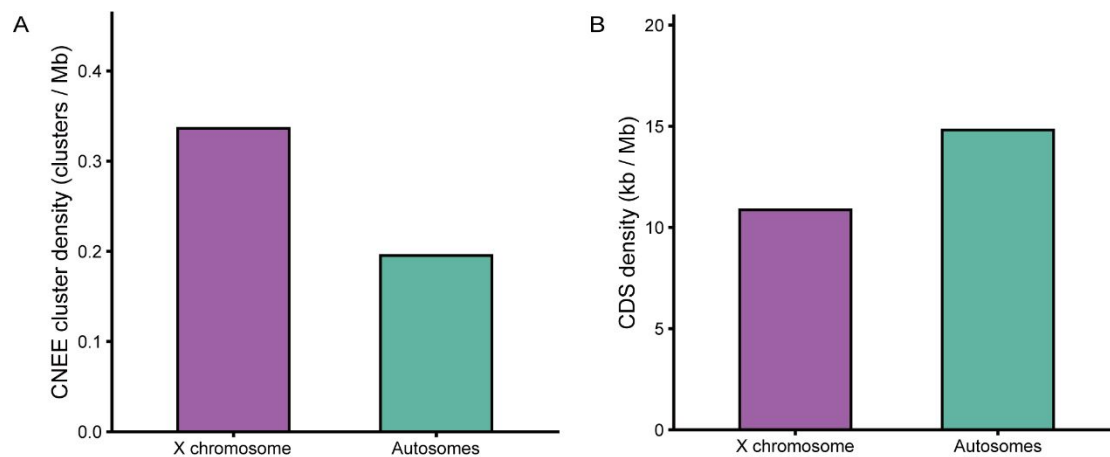




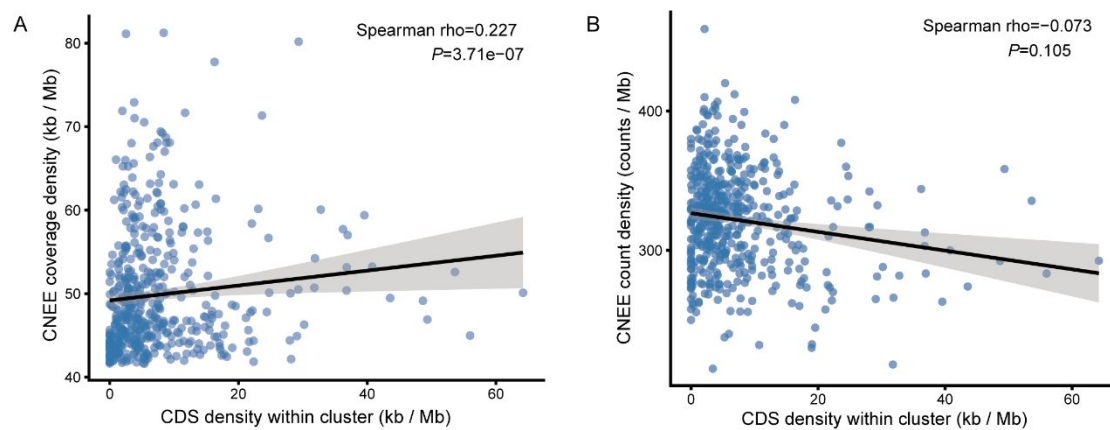
Supplementary Figure S1. Genome-wide 300-kb sliding windows were ranked by CNEE enrichment fold-change (FC). A geometric inflection-point approach identified FC=1.98 as the threshold for candidate high-density CNEE windows. Windows with FC<0.5 were classified as CNEE deserts, and the remaining windows as background regions.



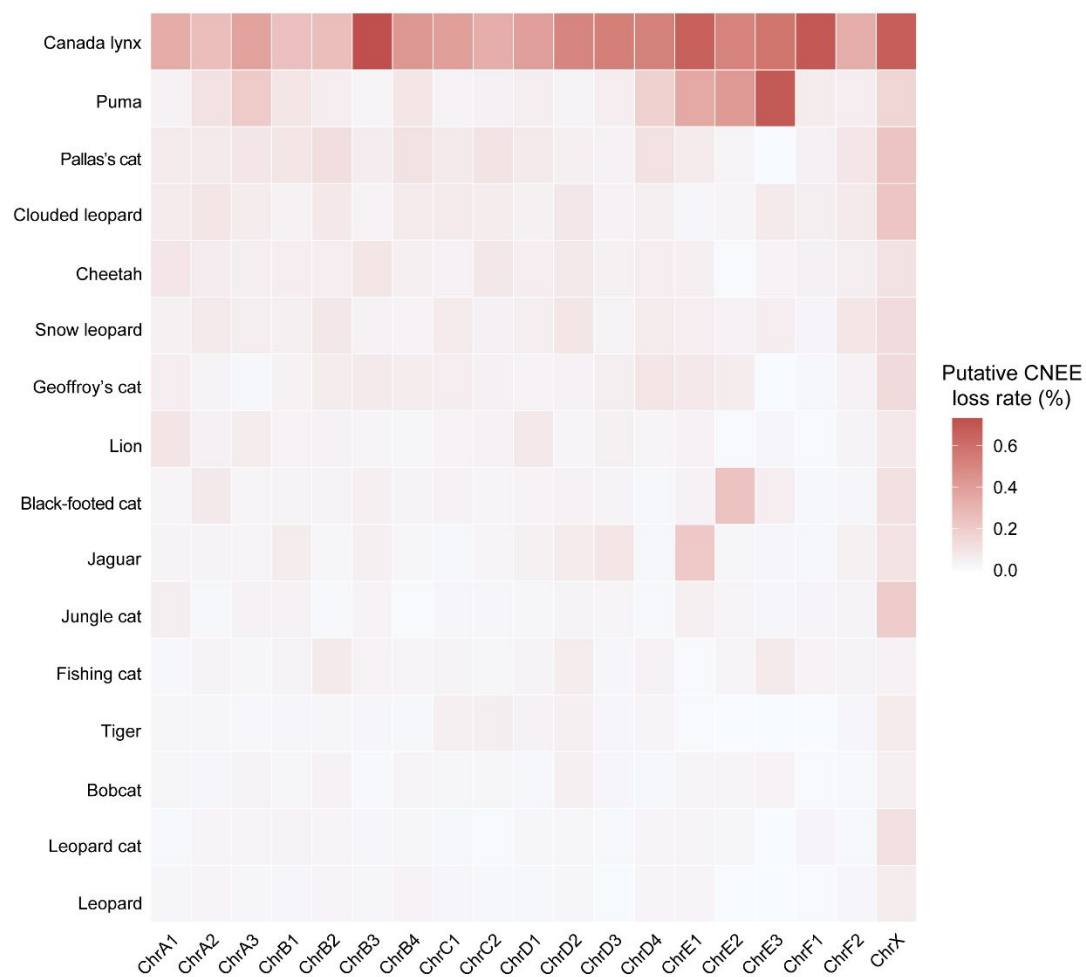
Supplementary Figure S2. Distribution of CNEE counts in candidate high-density clusters. Histogram showing the number of distinct CNEEs per merged candidate high-density cluster. The dashed orange line indicates the median CNEE count per cluster (191.5), which was used to assess the empirical distribution and guide quality control of final high-density CNEE clusters.



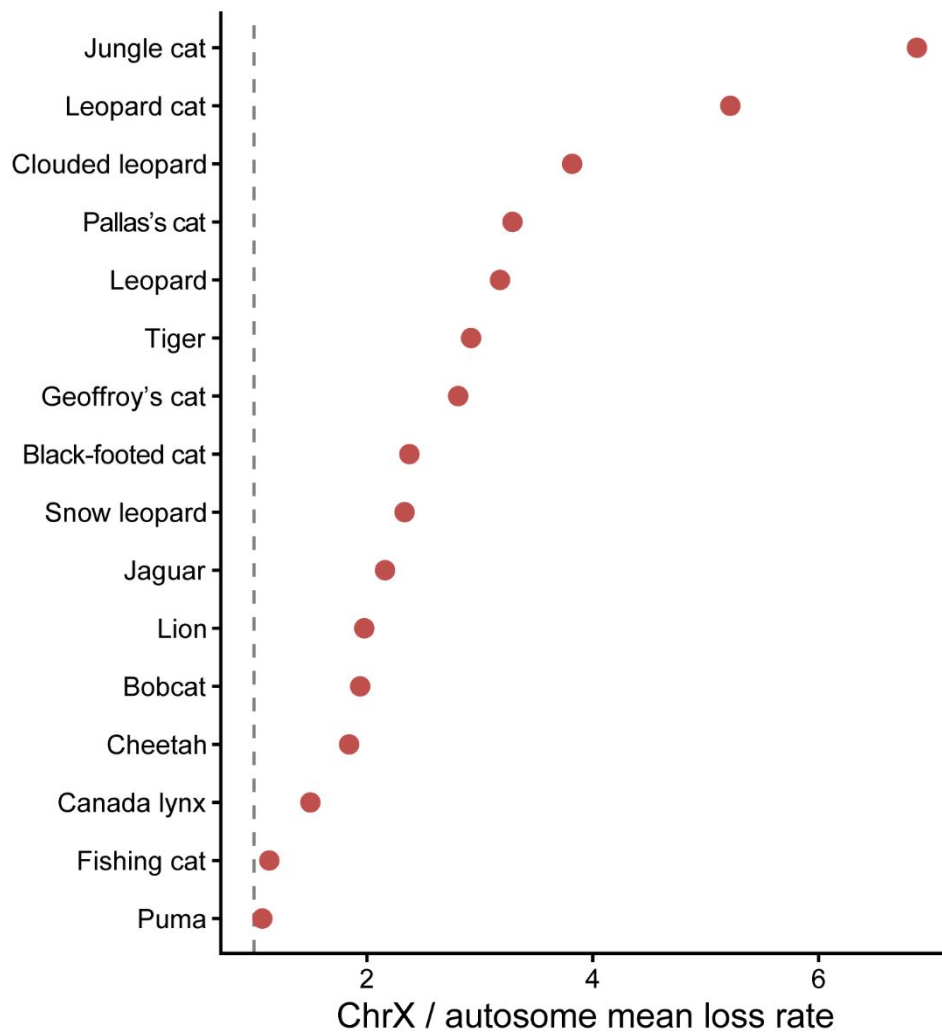
Supplementary Figure S3. Comparison of CNEE cluster density and coding sequence (CDS) density between the X chromosome and autosomes. A: CNEE cluster density (expressed as clusters per Mb) on the X chromosome compared with autosomes. Statistical significance was evaluated using a two-sided Poisson rate test ($P=1.60 \times 10^{-3}$). B: CDS density (expressed as kb per Mb) on the X chromosome compared with autosomes, calculated across non-overlapping 1-Mb genomic windows. Statistical significance was assessed using a two-sided Wilcoxon rank-sum test ($P=2.94 \times 10^{-2}$).



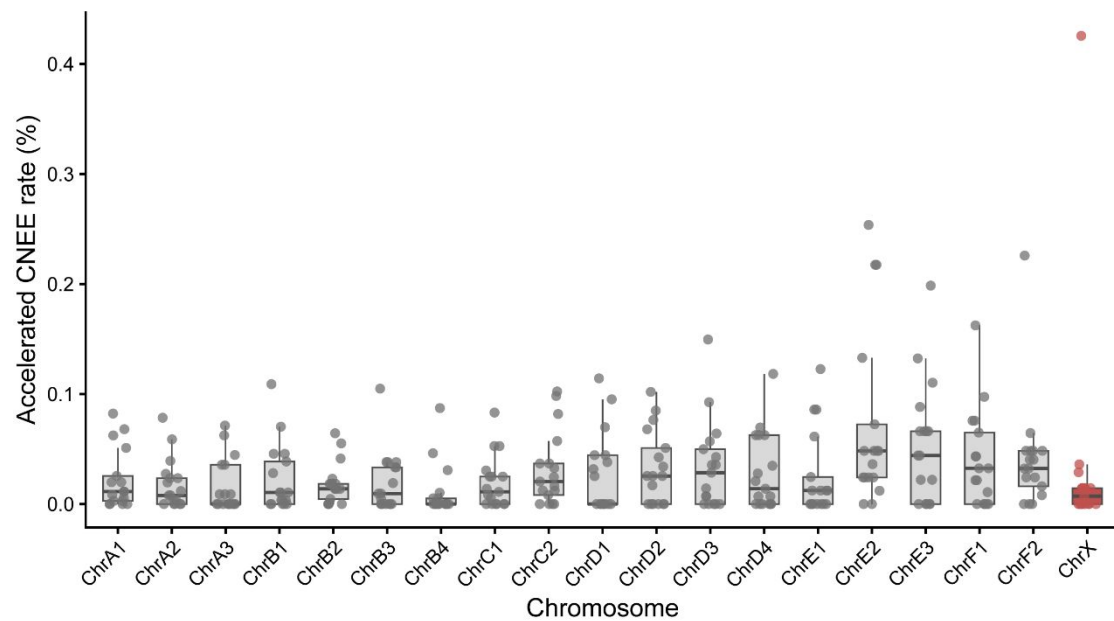
Supplementary Figure S4. Correlation analyses between coding sequence (CDS) density and CNEE densities within high-density CNEE clusters. **A:** Spearman rank correlation between CDS density (kb/Mb) and CNEE coverage density (kb/Mb) within high-density CNEE clusters. **B:** Spearman rank correlation between CDS density (kb/Mb) and CNEE count density (counts/Mb) within high-density CNEE clusters. Solid black lines represent linear regression fits, and gray shaded regions indicate 95% confidence intervals.



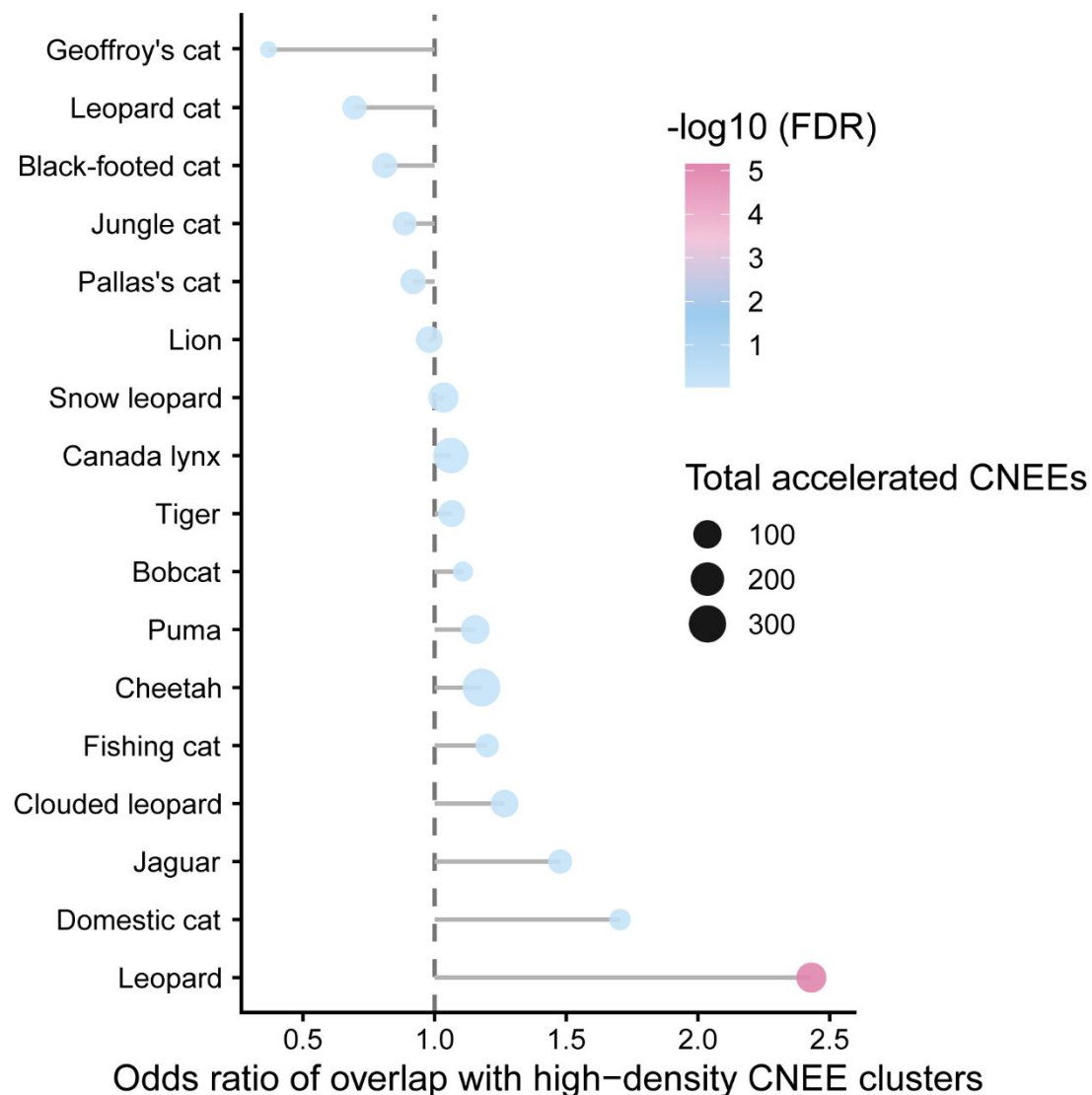
Supplementary Figure S5. Chromosome-level distribution of putative CNEE loss rates across felid species. Heatmap showing putative CNEE loss rates across chromosomes for each felid species. The color scale indicates the percentage of CNEEs putatively lost on each chromosome.



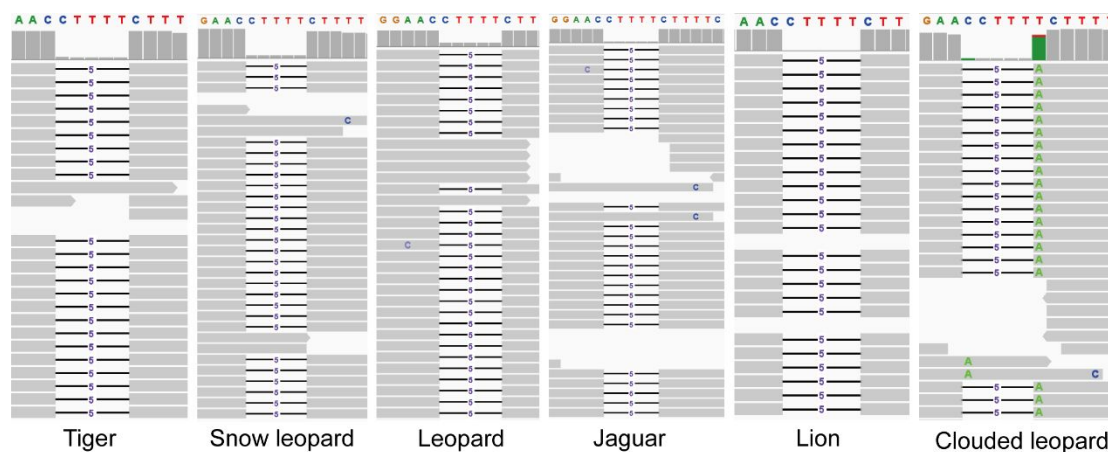
Supplementary Figure S6. ChrX-to-autosome ratios of mean putative CNEE loss rates for each species. The ratio was calculated as ChrX loss rate divided by mean autosomal loss rate. The dashed line indicates a ratio of 1.



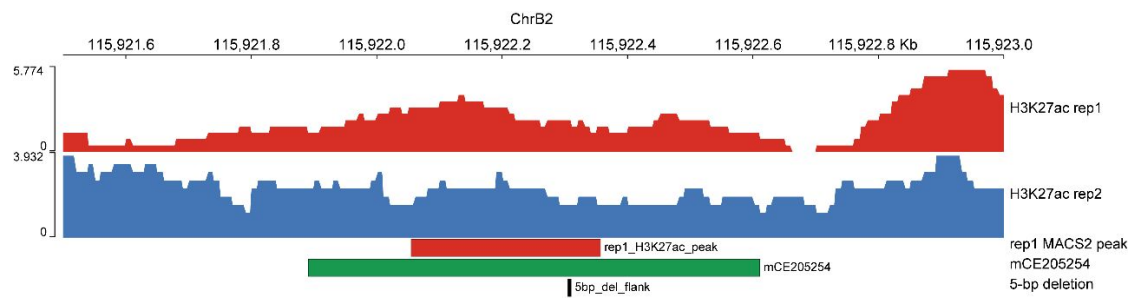
Supplementary Figure S7. Acceleration rates of CNEEs across chromosomes, showing no consistent elevation on the X chromosome relative to autosomes (PGLS one-sided test, $P=1.00$).



Supplementary Figure S8. Odds ratios for the overlap between accelerated CNEEs and high-density CNEE clusters across species. Point size indicates the total number of accelerated CNEEs, and color represents $-\log_{10}(\text{FDR})$. No consistent enrichment pattern was observed across felids.



Supplementary Figure S9. Short-read resequencing validation of a Pantherinae-shared 5-bp deletion in mCE205254. IGV snapshots showing read-level support for a 5-bp deletion at positions 415–419 of the 720-bp felid mCE205254 sequence across Pantherinae species. Species are shown from left to right as tiger (SRR15369229), snow leopard (CRR1429613), leopard (ERR5671315), jaguar (SRR22298149), lion (ERR13719738), and clouded leopard (SRR22827223). Gray reads represent aligned short-read sequencing reads; black gaps denote the 5-bp deletion.



Supplementary Figure S10. Genome browser tracks showing H3K27ac ChIP-seq enrichment in domestic cat heart tissue across the mCE205254 locus on chromosome B2. The red and blue tracks represent two H3K27ac replicates, and the red annotation bar indicates a H3K27ac peak. The green bar marks mCE205254, while the black vertical line denotes the Pantherinae-fixed 5-bp deletion site.

1456 猫科动物保守非外显子元件的演化动态及 X 染色体独特的调控演化

1457 李可欣^{1,2,#}, 杜欣^{1,#}, 张叶浩^{1,2}, 胡义波^{1,2,*}

1458 ¹ 动物多样性保护与有害动物防控全国重点实验室, 中国科学院动物研究所, 北京, 100101

1459 ² 中国科学院大学, 北京, 101408

1460 # 这些作者对本研究具有同等贡献

1461 * 通讯作者: 胡义波 (ybhu@ioz.ac.cn)

1462

1463 摘要

1464 保守非外显子元件 (conserved non-exonic elements, CNEEs) 是一类关键的顺式调控元件,
1465 在脊椎动物的发育调控与表型演化中发挥着重要作用。然而, 目前对其在染色体层面的
1466 分布特征及演化动态仍缺乏系统性认识, 尤其是在非模式动物中。猫科 (Felidae) 动物
1467 经历了快速的辐射演化, 且具有高度保守的核型, 是探讨上述问题的理想模型。本研究
1468 基于 17 种猫科动物及 2 种食肉目外群的长读长 (long-read) 基因组组装数据, 在全基因
1469 组范围内鉴定出 420 225 个 CNEEs, 并系统评估了其染色体分布模式、谱系特异性丢失
1470 以及加速演化特征。我们进一步识别出 490 个与发育调控密切相关的高密度 CNEE 簇。
1471 值得注意的是, X 染色体呈现出独特的调控演化模式: 高密度 CNEE 簇在 X 染色体上
1472 显著富集, 且其 CNEE 丢失率显著高于常染色体; 然而, 该染色体上 CNEE 的加速演化
1473 速率并未超过常染色体。此外, 本研究还鉴定出多个表现出谱系特异性加速演化的
1474 CNEEs, 其相关基因主要参与运动、感知、低氧适应及肌肉骨骼发育等生物学过程。其
1475 中, 豹亚科动物在一个与 *LAMA2* 基因相关的 CNEE 中存在一段共享的 5 bp 缺失。本研
1476 究从全基因组层面揭示了猫科动物 CNEE 的染色体组织规律与演化动态, 阐明了 X 染
1477 色体独特的演化特征, 为理解猫科动物表型多样性与环境适应性的非编码调控机制提供
1478 了全新见解。

1479

1480 关键词: 保守非外显子元件; 猫科; 豹亚科; 调控演化; X 染色体

