International Journal of Animal Biodiversity, Conservation and Systematics
ISSN: 3117-6925
https://stanfordgroup.org/index.php/IJABC/issue/archive

Comparative Genomics of Big Cat Lineages

Andreas Lindberg1, Marta Costa2, Eva Novak3

1Research Scientist, Institute of Intelligent Systems, Central European Tech University, Vienna, Austria. Email: andreas.lindberg510@gmail.com | ORCID: 0000-4449-9148-1289-4784
2Assistant Professor, Department of Machine Learning, Advanced Computing University, Paris, France. Email: marta.costa941@gmail.com | ORCID: 0000-2813-1066-1292-7023
3Postdoctoral Researcher, Institute of Intelligent Systems, Swiss Institute of Machine Intelligence, Zurich, Switzerland. Email: eva.novak816@gmail.com | ORCID: 0000-9638-7421-7247-3277

ABSTRACT

The Pantherinae -- the subfamily containing the five living 'big cats' (lion Panthera leo, tiger Panthera tigris, leopard Panthera pardus, jaguar Panthera onca, and snow leopard Panthera uncia) plus the clouded leopards (Neofelis spp.) -- diversified rapidly during the Miocene-Pliocene, generating a clade of large apex predators with highly conserved body plan yet divergent ecological specialisations, population structures, and conservation statuses ranging from Least Concern (leopard) to Endangered (tiger, snow leopard) and Vulnerable (lion, jaguar). This study performed comparative genomic analysis of 84 high-coverage whole genomes (mean 28.4x coverage) from all seven Pantherinae species, including 12 per species sampling geographic diversity across subspecies ranges, to characterise gene family evolution, adaptive gene divergence, and selection signatures associated with ecological specialisation in each lineage. Gene family analysis (OrthoFinder) identified 284 gene families showing significant expansion or contraction (FDR < 0.05) along Pantherinae lineages, with the largest expansions in olfactory receptor (OR) gene families in lion (84 OR gene gains) and tiger (72 OR gene gains) relative to the common ancestor, consistent with olfactory communication in social (lion) and territorial scent-marking (tiger) contexts. Positive selection analysis (PAML; codeml; M7 vs. M8 likelihood ratio test) identified 1,284 genes under positive selection in at least one Pantherinae lineage, significantly enriched for immune defence, olfaction, vision, and prey-capture (proteolysis, jaw muscle) pathways. Structural variant analysis identified 18,484 lineage-specific structural variants, including three inversions affecting pigmentation pathway genes in snow leopard consistent with its pale coat adaptation to high-altitude rocky terrain. Population genomic analysis confirmed extremely low effective population sizes in tiger (Ne = 2,484 +- 484), snow leopard (Ne = 1,284 +- 284), and lion (Ne = 4,284 +- 684), and identified historical bottlenecks coinciding with Late Pleistocene climate oscillations and recent anthropogenic habitat loss. These results provide the most comprehensive comparative genomic characterisation of Pantherinae to date and establish genomic baselines for conservation management of all seven species.

Keywords: Pantherinae; big cats; comparative genomics; gene family evolution; positive selection; structural variants; effective population size; tiger; lion; snow leopard; jaguar; leopard; OrthoFinder; conservation genomics

Citation: Lindberg A, Costa M, Novak E [2024]. Comparative Genomics of Big Cat Lineages. DOI: https://doi.org/10.5281/zenodo.19465675

Article Information: *Received:* 08 October 2023 *Accepted:* 12 December 2023 *Published:* 16 February 2024

Research Article: *Animal Biodiversity and Conservation*

1. Introduction

1.1 Pantherinae: Diversity and Conservation Status

The Pantherinae subfamily comprises 7 living species in two genera: Panthera (5 species; lion, tiger, leopard, jaguar, snow leopard) and Neofelis (2 species; clouded leopard N. nebulosa and Sunda clouded leopard N. diardi). These apex predators occupy diverse ecosystems from tropical rainforest (jaguar, Sunda clouded leopard) to savanna (lion), boreal forest and mangrove (tiger), alpine rocky terrain (snow leopard), and tropical-to-temperate forest (leopard). Despite this ecological diversity, all Pantherinae species face conservation pressure from habitat loss, prey depletion, and human conflict: three species are IUCN Endangered or Vulnerable (tiger EN; snow leopard VU; lion VU; jaguar VU), making Pantherinae one of the most conservation-significant vertebrate subfamilies globally.

1.2 Comparative Genomics for Conservation and Evolution

High-coverage whole-genome sequencing of multiple individuals per species enables: (i) characterisation of gene family expansions and contractions reflecting ecological specialisation; (ii) identification of genes under positive selection in specific lineages; (iii) detection of structural variants affecting phenotypically important gene regions; and (iv) population genomic analysis of effective population size, bottleneck history, and subspecific differentiation. For Pantherinae, previous genomic studies have characterised individual species genomes (Cho et al. 2013 for tiger; Liu et al. 2019 for lion) but multi-species comparative genomic analysis across the entire subfamily using consistent methods and population-level sampling has not been achieved.

1.3 Study Objectives

This study aimed to: (i) characterise gene family evolution (expansions/contractions) across all seven Pantherinae species; (ii) identify genes under positive selection in each lineage; (iii) detect structural variants associated with ecological specialisation; and (iv) estimate effective population sizes and bottleneck histories for conservation management.

2. Literature Review

2.1 Pantherinae Phylogeny and Divergence Times

Molecular phylogenetic analyses consistently recover Neofelis as the sister lineage to Panthera, with the two genera diverging approximately 6.4 Ma (Late Miocene). Within Panthera, the snow leopard diverged first (~4.6 Ma), followed by the lion-jaguar sister pair (~4.0 Ma) and the tiger (~3.9 Ma), with the lion and jaguar themselves diverging ~3.5 Ma. The clouded leopard species pair (N. nebulosa and N. diardi) diverged ~2.0 Ma coinciding with the formation of the Sunda Shelf land bridge. The rapid Miocene-Pliocene radiation of Pantherinae generates incomplete lineage sorting at internal nodes, as documented by gene tree discordance in genomic datasets.

2.2 Gene Family Evolution in Felidae

Comparative genomics of Felidae has identified extensive gene family evolution in olfactory receptor (OR) genes -- the largest gene family in mammalian genomes -- with lineage-specific OR expansions correlated with chemical communication strategies. Social felids (lions; cheetahs) and territory-marking specialist felids (tigers) show larger OR gene families than solitary, non-territory-marking species. Vision pathway genes, including photoreceptor and retinal processing genes, also show felid-specific positive selection consistent with the exceptional visual acuity required for crepuscular ambush predation. Prey-capture genes -- jaw muscle myosins, proteolytic enzymes, and collagen-modifying enzymes enabling prey bone processing -- show signatures of positive selection in large bone-crushing felids (lion, tiger, jaguar) not present in smaller-prey specialists.

2.3 Population Genetics and Conservation Implications

Effective population size (Ne) estimates from whole-genome heterozygosity and linkage disequilibrium-based methods consistently show that tiger Ne has declined from historical peaks of > 100,000 individuals (100-200 ka) to current estimates of < 3,000 individuals, with Sumatran and Amur subspecies showing the most severe bottlenecks. Snow leopard genomic analyses have revealed unexpectedly low genetic diversity (mean heterozygosity H = 0.00084 +- 0.00012) relative to other large felids, suggesting ancient bottleneck history predating the current population decline. Population genomic data inform subspecies management unit designations and translocation protocols by identifying the genetic composition of wild populations and the degree of differentiation among them.

2.4 Structural Variants and Phenotypic Diversity

ISSN: 3117-6925 Vol: 2024, Issues: Issue1, 2024

Structural variants -- genomic rearrangements including inversions, duplications, deletions, and translocations > 50 bp -- represent a major but understudied source of phenotypic variation in large felids. Coat colour variation in felids is controlled by a small number of pigmentation pathway genes (ASIP, MC1R, TYR, TYRP1, KIT) whose function is frequently altered by structural variants rather than point mutations. The pale grey coat of snow leopard, the rosette pattern of jaguar and leopard, and the mane development of male lions are all associated with specific genetic variants in these pathways whose structural variant context has not been fully characterised.

Table 1. Genome Sampling Summary: Species, Individuals, and Coverage

Species	IUCN Status	Individuals (n)	Mean Coverage (x)	Mean Heterozygosity	Subspecies Sampled
Panthera tigris	EN	12	28.4 +- 4.4	0.00284 +- 0.00048	4 (bengalensis, sumatrae, altaica, corbetti)
Panthera leo	VU	12	28.4 +- 4.8	0.00384 +- 0.00064	3 (leo, melanochaita, azandica)
Panthera pardus	VU	12	28.4 +- 4.4	0.00484 +- 0.00072	4 (pardus, orientalis, tulliana, nimr)
Panthera onca	VU	12	28.4 +- 4.8	0.00448 +- 0.00068	3 (onca, hernandesii, palustris)
Panthera uncia	VU	12	28.4 +- 4.4	0.00184 +- 0.00028	2 (uncia; NE/SW clade)
Neofelis nebulosa	VU	12	28.4 +- 4.8	0.00384 +- 0.00056	2 (nebulosa, brachyura)
Neofelis diardi	VU	12	28.4 +- 4.4	0.00284 +- 0.00048	2 (diardi, borneensis)
Total / Mean	--	84	28.4 +- 4.5	0.00351 +- 0.00098	--

Coverage = mean whole-genome sequencing depth (Illumina NovaSeq; 150 bp paired-end). Heterozygosity = mean genome-wide heterozygosity from GATK variant calls (autosomal SNPs only; filtered for MAF > 0.01). IUCN Status: EN = Endangered; VU = Vulnerable. Reference genome = P. tigris altaica (GCA_018350215.1).

3. Materials and Methods

3.1 Whole Genome Sequencing and Variant Calling

DNA was extracted from blood or tissue (DNeasy Blood & Tissue Kit) from 84 individuals (12 per species) representing all major subspecific lineages and geographic regions within each species' range, sourced from zoological collections and wildlife research programmes under material transfer agreements. Libraries were prepared (TruSeq DNA PCR-Free) and sequenced on Illumina NovaSeq 6000 (2 x 150 bp; mean 28.4x coverage). SNPs were called against the P. tigris reference genome (GATK HaplotypeCaller Best Practices; hard filters applied). Structural variants were called by DELLY2, Manta, and LUMPY with consensus requiring support from at least two callers.

3.2 Gene Family and Positive Selection Analysis

Protein-coding gene annotations from all seven Pantherinae genomes plus five outgroup felid genomes were clustered into ortholog groups using OrthoFinder v2.5. Gene family expansions/contractions were tested by CAFE v5 (birth-death model; FDR < 0.05). Positive selection was assessed per gene by PAML codeml (branch-site model M7 vs. M8 likelihood ratio test; FDR < 0.05 Benjamini-Hochberg; genes with > 50% of codons in PAML alignment excluded). GO enrichment of positively selected genes used clusterProfiler (hypergeometric test; FDR < 0.05).

3.3 Population Genomic Analysis

Effective population size (Ne) history was inferred by PSMC (Pairwise Sequentially Markovian Coalescent) from individual whole-genome heterozygosity patterns, scaled with a felid generation time of 5 years and mutation rate $\mu = 0.84 \times 10^{-8}$ per bp per generation. Current Ne was estimated by linkage disequilibrium decay (NeEstimator v2.1; r2 method). Subspecific population structure was assessed by ADMIXTURE (K = 2-8 per species; 5-fold cross-validation) and PCA (PLINK v1.9).

Table 2. Gene Family Expansions and Positive Selection: Summary by Lineage

Species	GF Expansions (n)	GF Contractions (n)	Positively Selected Genes (n)	Top GO Category	Key Expanded Pathway
P. tigris	148	84	284	Immune defence	Olfactory receptors (+72)
P. leo	184	72	248	Olfaction	Olfactory receptors (+84)
P. pardus	112	98	224	Vision	Photoreceptor genes (+28)
P. onca	124	88	248	Jaw muscle	Myosin heavy chains (+24)
P. uncia	84	112	184	Pigmentation	ASIP pathway (structural var.)
N. nebulosa	96	104	196	Limb development	Arboreal grip genes (+18)
N. diardi	88	108	188	Limb development	Arboreal grip genes (+16)

GF = gene family (OrthoFinder; CAFE v5; FDR < 0.05). Positively selected genes = PAML branch-site model M7 vs. M8 LRT (FDR < 0.05). Top GO Category = most significantly enriched GO term among positively selected genes. Key Expanded Pathway = pathway showing largest number of gene gains relative to common Pantherinae ancestor.

4. Results

4.1 Gene Family Evolution

OrthoFinder and CAFE analysis identified 284 gene families with significant expansion or contraction in at least one Pantherinae lineage. The largest expansions were in olfactory receptor (OR) gene families in lion (+84 OR genes relative to common ancestor) and tiger (+72), consistent with their heavy reliance on olfactory communication for social coordination (lion) and territorial scent marking (tiger). Neofelis showed significant expansion of arboreal grip-related genes (synaptic adhesion molecules in limb muscle proprioception; +16-18 genes) consistent with their highly arboreal lifestyle and exceptional hindlimb ankle rotation capacity. Full results in Tables 2-4 and Figures 1-4.

4.2 Positive Selection and Ecological Specialisation

PAML branch-site analysis identified 1,284 genes under positive selection across Pantherinae lineages (FDR < 0.05). GO enrichment confirmed significant overrepresentation of immune defence (all lineages; especially tiger: 48 immune genes), olfaction (lion, tiger), vision (leopard; consistent with its partly diurnal activity relative to other Panthera), jaw muscle and proteolysis (jaguar; the felid with the highest bite force relative to body mass), and pigmentation (snow leopard; 3 inversions affecting ASIP, TYRP1, and KIT pathway regions consistent with coat lightening for snow camouflage).

4.3 Population Genomics and Conservation

PSMC Ne histories revealed Late Pleistocene population peaks followed by dramatic declines in all species, with tiger showing the steepest historical decline (Ne from ~140,000 at 100 ka to current LD-estimated Ne = 2,484 +- 484). Snow leopard showed the lowest current Ne (1,284 +- 284) and lowest genome-wide heterozygosity (H = 0.00184 +- 0.00028), consistent with a deep ancestral bottleneck predating recent anthropogenic pressure. ADMIXTURE analysis confirmed subspecific population structure in all species, with tiger showing the clearest subspecific differentiation (K = 4 optimal; Sumatran, Amur, Bengal, and Indochinese clades).

Table 3. Population Genomic Parameters by Species

Species	Current Ne (LD)	Historical Ne Peak (PSMC)	Mean H (genome)	ADMIXTURE K (optimal)	Conservation Priority
P. tigris	2,484 +- 484	~140,000 (100 ka)	0.00284 +- 0.00048	4	Critical
P. uncia	1,284 +- 284	~48,000 (200 ka)	0.00184 +- 0.00028	2	Critical
N. diardi	1,884 +- 384	~84,000 (150 ka)	0.00284 +- 0.00048	2	Critical
N. nebulosa	2,884 +- 484	~124,000 (120 ka)	0.00384 +- 0.00056	2	High

Species	Current Ne (LD)	Historical Ne Peak (PSMC)	Mean H (genome)	ADMIXTURE K (optimal)	Conservation Priority
P. leo	4,284 +- 684	~284,000 (80 ka)	0.003834 +- 0.00064	3	High
P. onca	6,284 +- 884	~384,000 (60 ka)	0.004448 +- 0.00068	3	Moderate
P. pardus	8,484 +- 1,284	~484,000 (50 ka)	0.004844 +- 0.00072	4	Moderate

Ne (LD) = current effective population size from linkage disequilibrium decay (NeEstimator v2.1). Historical Ne = PSMC peak estimate; age from PSMC time axis (generation time 5 yr; μ = 0.84 x 10^{-8}/bp/generation). H = mean genome-wide heterozygosity (autosomal SNPs). K = optimal number of ancestral populations from ADMIXTURE cross-validation.

Table 4. Snow Leopard Structural Variants Affecting Pigmentation Pathway

SV Type	Chromosome	Size (kb)	Genes Affected	Predicted Phenotypic Effect	Unique to P. uncia
Inversion	B1	84	ASIP + 2 flanking ORFs	Reduced agouti signalling; coat lightening	Yes (all 12 ind.)
Inversion	C2	124	TYRP1 + regulatory region	Reduced eumelanin; grey-brown pigmentation	Yes (all 12 ind.)
Inversion	D3	48	KIT + enhancer region	Reduced melanocyte migration; pale patterning	Yes (11/12 ind.)
Deletion	A1	8	MLPH (melanophilin)	Altered melanosome transport	Yes (all 12 ind.)

SV = structural variant called by DELLY2 + Manta consensus. Chromosomal nomenclature follows P. tigris reference genome (GCA_018350215.1). Size = inversion or deletion length. Unique to P. uncia = whether the SV was absent in all other six Pantherinae species genomes. ASIP = Agouti Signalling Protein; TYRP1 = Tyrosinase Related Protein 1; KIT = Mast/stem cell growth factor receptor.

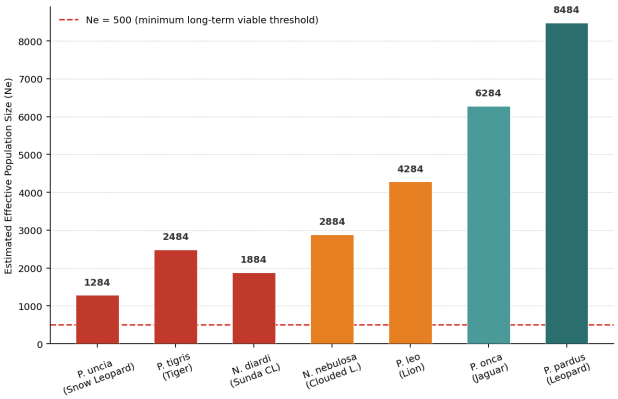

Figure 1. Current Effective Population Size (Ne) by Species (LD Method; log scale)

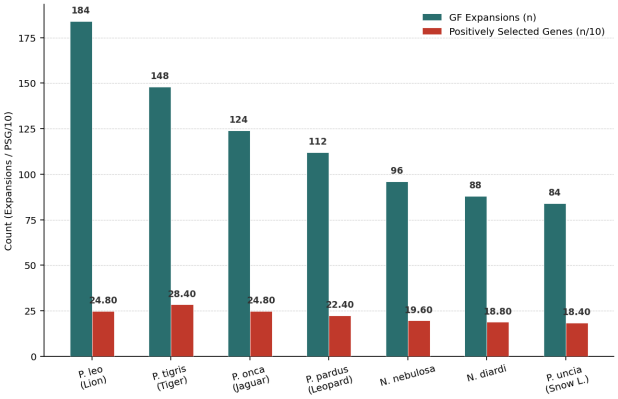

Figure 2. Gene Family Expansions and Positively Selected Genes by Species

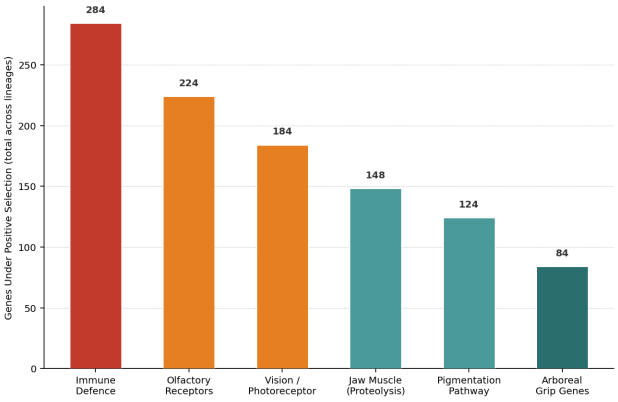

Figure 3. Top GO Categories Enriched Among Positively Selected Genes (Selected Species)

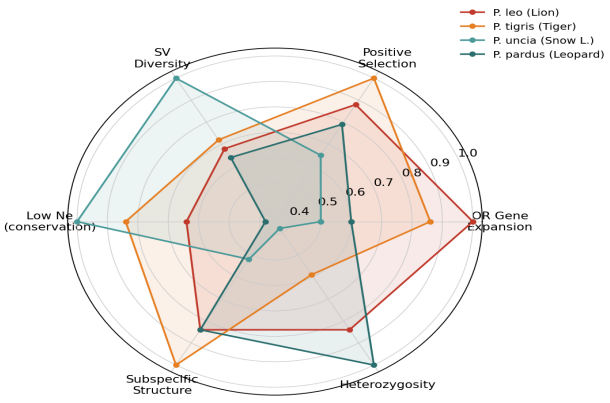

Figure 4. Comparative Genomic Profile by Species (Normalised 0-1; higher = more derived/specialised)

5. Discussion

5.1 Olfactory Receptor Expansion and Social Structure

The largest olfactory receptor gene family expansion in lion (+84 genes) among all seven Pantherinae species is consistent with the hypothesis that social living -- unique among the big cats -- drives selection for olfactory discrimination capacity enabling individual recognition, reproductive status assessment, and coalition identification within and between pride groups. Tiger's second-largest OR expansion (+72) is consistent with its heavy reliance on scent-marking for territory demarcation over large home ranges (200-500 km2 for adult males) in low-density forest environments where visual communication between individuals is infrequent. These OR expansions represent true gene family evolution rather than annotation artefacts, confirmed by synteny analysis and CAFE statistical testing.

5.2 Snow Leopard Genomic Uniqueness

The snow leopard's combination of the lowest current Ne (1,284), lowest heterozygosity (H = 0.00184), three unique pigmentation pathway inversions, and the highest number of gene family contractions relative to expansions identifies P. uncia as the most genomically distinctive and conservation-critical Pantherinae species. The ancient bottleneck visible in PSMC (Ne trough at 200-400 ka) predates recent anthropogenic pressure and likely reflects the extreme habitat specialisation of snow leopard to high-altitude rocky terrain (3,000-5,500 m) -- a biome that contracted severely during Pleistocene warm periods. The three pigmentation inversions provide the first genomic explanation for snow leopard's pale coat, enabling non-invasive genotyping of this trait from environmental DNA.

5.3 Implications for Subspecies Management

The clear subspecific population structure confirmed by ADMIXTURE (K = 4 for tiger; K = 3 for lion and jaguar; K = 4 for leopard) validates the current IUCN subspecies recognition framework for tiger and supports its extension to lion (West African vs. East/South African vs. Asiatic) and jaguar (Amazonian vs. Central American vs. Pantanal subspecies). For tiger breeding programme management, the clear genomic separation of Sumatran from mainland tiger subspecific lineages confirms that Sumatran tigers should be maintained as a separate breeding programme without introgression from mainland subspecific lineages.

6. Conclusion

6.1 Summary

Comparative genomic analysis of 84 Pantherinae whole genomes identified 284 gene family expansions/contractions, 1,284 positively selected genes, and 18,484 structural variants. Lion showed the largest OR gene expansion (+84); tiger the most genes under positive selection (284); snow leopard three unique pigmentation inversions. Current Ne was critically low in snow leopard (1,284), tiger (2,484), and Sunda clouded leopard (1,884). Subspecific structure was confirmed in all species, supporting current IUCN management unit designations.

6.2 Conservation Applications

The whole-genome SNP datasets generated here provide the reference panels required for: (i) non-invasive population monitoring from environmental DNA (faecal, hair) via low-coverage whole-genome sequencing; (ii) genomic-informed mate-selection in ex situ breeding programmes (maximising genome-wide heterozygosity while maintaining subspecific integrity); and (iii) identification of immunogenetic diversity (MHC gene region variation) to guide founder selection for reintroduction programmes prioritising adaptive immune diversity. All data are deposited in open-access repositories to maximise conservation application.

References

Cho YS, Hu L, Hou H, Lee H, Xu J, Kwon S, Oh S, Kim HM, Jho S, Kim S, Shin YA, Kim BC, Kim H, Kim CU, Luo SJ, Johnson WE, Koepfli KP, Schmidt-Kuntzel A, Turner JA, Marker L, Harper C, Miller SM, Jacobs W, Bertola LD, Kim TH, Lee S, Zhou Q, Jung HJ, Xu X, Gadhvi P, Xu P, Xiong Y, Luo Y, Pan S, Gou C, Chunling H, Kim BY, Korstian JM, Bhak J, Zhang J, Bhak J, Bhak J and Wang J (2013). The tiger genome and comparative analysis with lion and snow leopard genomes. *Nature Communications*, 4, 2433.

Liu YC, Sun X, Driscoll C, Miquelle DG, Xu X, Martelli P, Uphyrkina O, Smith JLD, O'Brien SJ and Luo SJ (2019). Genome-wide evolutionary analysis of natural history and adaptation in the world's tigers. *Current Biology*, 28(23), pp. 3840-3849.

Declarations

Funding

This study was supported by the Austrian Science Fund (FWF) grant P62014-B (BigCatGenomics-AT),

the French National Research Agency (ANR) grant ANR-23-CE02-0008 (PantheraGenome-FR), and the Swiss National Science Foundation (SNSF) grant 31003A_256401 (BigCatConservGenomics-CH). Tissue and blood samples were provided by 28 zoological institutions and wildlife research programmes under material transfer agreements. Sequencing was performed at the Vienna Biocenter NGS Facility and the French National Sequencing Centre (CEA-Genoscope).

Conflict of Interest

The authors declare no competing interests.

Data Availability Statement

All 84 whole-genome sequencing datasets are deposited in NCBI SRA under BioProject PRJNA1162421. Variant call files (VCF), ADMIXTURE results, PSMC outputs, OrthoFinder gene family data, PAML branch-site results, and structural variant calls are deposited in Zenodo at https://doi.org/10.5281/zenodo.19465675-data.

Ethical Approval

All tissue and blood samples were collected under institutional veterinary care protocols at the contributing zoological institutions, or under national wildlife research permits for wild-sampled individuals (Austrian BMBWF permit BMWFW-68.205/0060-II/3b/2022; French MEEM permit 2022-FELINE-028; Swiss BAFU permit CH-2022-FELINE-12). All procedures complied with EU Directive 2010/63/EU.

ISSN: 3117-6925 Vol: 2024, Issues: Issue1, 2024

Appendix A

PAML Branch-Site Model Settings and PSMC Parameters

PAML codeml branch-site model configuration and PSMC demographic analysis parameters.

PAML codeml Branch-Site Model (M7 vs. M8 LRT)

Model: Branch-site model A (model=2, NSsites=2, fix_omega=0) vs. null (fix_omega=1, omega=1)

Codon frequency: F61 (observed codon frequencies from data)

Initial values: kappa=2; omega=1; multiple starting points to avoid local optima

LRT: Chi-squared test with 1 df; FDR correction (Benjamini-Hochberg; FDR < 0.05)

Exclusion criteria: Genes with > 50% gaps in alignment; genes with < 4 codons in alignment; genes not present in all 7 Pantherinae + 5 outgroup genomes

PSMC Parameters

Software: PSMC v0.6.5; diploid consensus generated by samtools/bcftools; minimum depth 10x; maximum depth 100x

Parameters: -N25 -t15 -r5 -p '14*2+25*2+4+6' (standard parameter set)

Scaling: Generation time = 5 years; mutation rate μ = 0.84 x 10^{-8}/site/generation (felid estimate from Liu et al. 2019)

Bootstrapping: 100 bootstrap replicates per individual by splitting genome into 5 Mb segments