

Genome-wide SNPs clarify a complex radiation and support recognition of an additional cat species

Fernanda J. Trindade¹, Maíra R. Rodrigues^{1,2}, Henrique V. Figueiró^{1,3}, Gang Li^{4,5}, William J. Murphy⁴, Eduardo Eizirik^{1,6}

¹ PUCRS, Escola de Ciências da Saúde e da Vida, Laboratório de Biologia Genômica e Molecular, Porto Alegre, RS, Brazil

² USP, Department of Genetics and Evolutionary Biology, Institute of Biosciences, São Paulo, Brazil.

³ Center for Species Survival, Smithsonian Conservation Biology Institute, National Zoological Park, Front Royal, VA 22630, USA

⁴ Department of Veterinary Integrative Biosciences, Texas A&M University, College Station, TX 77843-4458, USA

⁵ College of Life Sciences, Shaanxi Normal University, Xi'an, Shaanxi, China

⁶ Instituto Pró-Carnívoros, Atibaia, SP, Brazil.

Corresponding author:

Eduardo Eizirik

PUCRS, Escola de Ciências da Saúde e da Vida, Laboratório de Biologia Genômica e Molecular, Porto Alegre, RS, Brazil. eduardo.eizirik@pucrs.br

Abstract

Phylogenetic reconstruction and species delimitation are often challenging in the case of recent evolutionary radiations, especially when post-speciation gene flow is present. *Leopardus* is a Neotropical cat genus that has a long history of recalcitrant taxonomic problems, along with both ancient and current episodes of interspecies admixture. Here we employ genome-wide SNP data from all presently recognized *Leopardus* species, including several individuals from the tigrina complex (representing *L. guttulus* and two distinct populations of *L. tigrinus*), to investigate the evolutionary history of this genus. Our results reveal that the tigrina complex is paraphyletic, containing at least three distinct species. While one can be assigned to *L. guttulus*, the other two remain uncertain regarding their taxonomic assignment. Our findings indicate that the ‘tigrina’ morphology may be plesiomorphic within this group, which has led to a longstanding taxonomic trend of lumping these poorly known felids into a single species.

Main text

Introduction

Genome-wide data hold great potential to address complex evolutionary problems, such as resolving the phylogenetic relationships and dissecting introgression histories among closely related species (e.g. Li et al. 2016, 2019; Edelman et al., 2019; Pulido-Santacruz et al., 2020). Within the mammalian order Carnivora, several genera have undergone recent radiations, leading to complex networks that challenge phylogenetic resolution using traditional approaches (e.g. Figueiró et al., 2017). In the Neotropics (encompassing South and Central America, Mexico and Southern USA), at least two genera (*Leopardus* in the Felidae and *Lycalopex* in the Canidae) have diversified recently, each of them following a single episode of colonization from North America during the Pliocene or Pleistocene (Eizirik 2012). Accurately resolving the phylogenetic structure of these clades is critical to stabilize their taxonomy, enable adequate conservation assessment and actions on behalf of these threatened organisms, and allow a better understanding of the intricate evolutionary and biogeographic history of Neotropical biotas.

Leopardus is a Neotropical-endemic clade of small to medium-sized wild cats that diverged from other felid lineages ca. 10 million years ago (MYA), and underwent a radiation starting ca. 3-4 MYA (Li et al. 2016). It comprises at least seven extant species

that have been traditionally recognized by taxonomists since the mid-20th century: ocelot (*L. pardalis*), margay (*L. wiedii*), Andean mountain cat (*L. jacobita*), pampas cat (*L. colocola*), Geoffroy's cat (*L. geoffroyi*), huiña (*L. guigna*), and tigrina (*L. tigrinus*). The latter has recently been found to represent at least two distinct species, the northern tigrina (*L. tigrinus*) and southern tigrina (*L. guttulus*), based on the analysis of multiple molecular markers (Trigo et al. 2013). Subsequent morphological analyses supported the distinctiveness of *L. guttulus*, and further proposed the separation of *L. emiliae* (occurring in northeastern Brazil) from *L. tigrinus*, which would be restricted to northern and western South America, as well as Central America (Nascimento & Feijó 2017) (see species distribution in supplementary fig. S1). This arrangement has so far not been tested with molecular data, for the reasons outlined below.

Several molecular studies focusing on this genus have revealed that it has had a complex evolutionary history, including different episodes of interspecies hybridization. Previous work has shown that the southern tigrina (*L. guttulus*) is currently hybridizing with Geoffroy's cat in southern Brazil (Trigo et al. 2008, 2013, 2014). In contrast, tigrina populations from northeastern Brazil (NE tigrina), identified as *L. tigrinus* or *L. emiliae* depending on the assumed classification, bear molecular signatures of ancient hybridization with pampas cats (Trigo et al. 2013). This ancient interspecies admixture has resulted in remarkable cyto-nuclear discordance in the NE tigrina, with complete replacement of its mitochondrial genome with introgressed mtDNA from the pampas cat (Trigo et al. 2013; Santos et al. 2018). The latter taxon (*L. colocola*) has recently been proposed to actually comprise five distinct species (Nascimento et al. 2021); under this scheme, the hybridization of NE tigrina would have occurred with the central/northeastern Brazilian pampas cat (*L. braccatus*), based on mtDNA phylogeographic analyses (Santos et al. 2018). Importantly, this mitochondrial replacement has precluded any mtDNA-based phylogenetic analysis comparing NE tigrinas with other members of the tigrina complex. At the same time, the nuclear markers analyzed so far (Trigo et al. 2013) did not contain enough phylogenetic signal to reliably resolve their relationships, suggesting that NE tigrinas and *L. guttulus* could be sister-species (based on Y-chromosome markers) or that the former was in fact more closely related to Geoffroy's cat (based on X-chromosome markers).

Furthermore, these earlier studies did not include representatives of additional, geographically distant, tigrina populations. This is especially relevant since early mtDNA data (Johnson et al. 1999; Trigo et al. 2008) had indicated that Central American tigrinas

(presently recognized as *L. tigrinus oncilla* [Kitchener et al. 2017]) were very divergent from southern South American populations (now recognized as *L. guttulus*). Genome-wide SNP data supported this deep divergence (Li et al. 2016), but that study included only the Central American lineage and the NE tigrina, with no representation of *L. guttulus*. Therefore, no phylogenetic assessment thus far has included all three tigrina units.

Here we expand on the genome-wide SNP dataset reported by Li et al. (2016) by including several *L. guttulus* individuals as well as additional Geoffroy's cat specimens. We also genotyped previously identified hybrids between these two species, as well as a captive-bred hybrid between *L. guttulus* and the pampas cat, aiming to assess the effects of including admixed individuals in genome-wide assessments of species-level monophyly and phylogenetic relationships. Our results indicate that this impact can be quite relevant in phylogenetic analyses of recent radiations, and robustly demonstrate that the tigrina complex comprises at least three different species.

Results and Discussion

We performed multiple sets of analyses to investigate the impacts of varying taxon sampling, filtering schemes for missing data, treatment of heterozygous sites and inclusion of hybrid individuals (see Materials and Methods and Supplementary Information for details). Principal Component Analyses (PCAs) revealed a clear separation among the recognized species, and indicated that the three sampled tigrina units (southern (S) tigrina [*L. guttulus*], NE tigrina and Central American tigrina) were very distinct from each other (fig. 1A, supplementary fig. S2). This finding was corroborated by the Admixture analyses, regardless of the taxon sampling scheme (supplementary fig. S3-S5). In the PCA plots, S tigrina and NE tigrina were at least as distinct from each other as Geoffroy's cats vs. huiñas. In the Admixture plots, S tigrina and NE tigrina exhibited completely different ancestry assignments from K=3 (taxon subgroups 2 and 3) or K=4 (subgroup 1) upwards, much below the optimal K inferred for each dataset. In addition to the separation between the two South American tigrina units, the PCA and Admixture results also supported the distinctiveness of the Central American tigrina (e.g. it was the most distinct unit in PC1 for the focal taxonomic group).

In addition to assessing the distinctiveness among the three tigrina units, we also investigated inter-species hybridization. Our data supported the inference of admixed ancestry in one field-collected individual (bLti135) that had been previously reported to

be a hybrid between *S. tigrina* and Geoffroy's cat using traditional molecular markers (Trigo et al. 2013). We also found that another individual (LCO.2), previously suspected to derive from admixture in captivity between *S. tigrina* and pampas cat (Trigo et al. 2008), was indeed a hybrid (likely F1) between these species. Importantly, we did not find any evidence of nuclear introgression from pampas cat into NE *tigrina* (supplementary fig. S3), in striking contrast to the complete substitution of the latter's mtDNA with that of the former (Trigo et al. 2013; Santos et al. 2018). This indicates that signatures from the ancient hybridization episode between these species may have been erased from the nuclear genome by cumulative backcrossing, and highlights the remarkable cyto-nuclear discordance present in NE tigrinas.

We performed extensive phylogenetic analyses with our SNP dataset using several combinations of individuals, filters for missing data and treatments of heterozygous sites (see Supplementary Information). We also employed different phylogenetic approaches: Maximum Likelihood (ML) on a supermatrix comprising the concatenation of all sites (including both variable and invariant positions), and SNP-based phylogenetic reconstructions. All analyses converged on the conclusion that the *tigrina* complex is paraphyletic (fig. 1B; supplementary fig. S6-S10). ML trees reconstructed the NE *tigrina* and *S. tigrina* as sister-groups, but strongly supported the placement of the Central American *tigrina* at a more external position, outside of the clade that also included two other, well-recognized *Leopardus* species (Geoffroy's cat and huiña). Coalescent-based trees also strongly supported this inference, and further indicated paraphyly of NE tigrinas and *S. tigrinas* relative to Geoffroy's cat (supplementary figs. S9, S10C).

In addition to dissecting *tigrina* relationships, our analyses also helped understand the effects of including hybrid individuals in SNP-based phylogenetic inference (supplementary fig. S6). In the case of the pampas cat vs. *S. tigrina* hybrid (LCO.2), its inclusion rendered the pampas cat paraphyletic, as it was drawn with high support towards the focal clade comprising the *tigrina* complex, Geoffroy's cat and huiña. The other individual detected as a hybrid with our SNP data (bLti135) also led to distortions in the topology, as it was drawn to the Geoffroy's cat clade, disrupting *S. tigrina* monophyly, altering the position of the huiña, and lowering support for the affected nodes. Inclusion of another putative hybrid (bLge094) between Geoffroy's cat and *S. tigrina* (previously inferred with traditional markers, but not with this SNP dataset) also led to a distortion in the huiña's position. In this case, the distortion may also have been induced by the extensive amount of missing data for bLge094. The presence of missing data also seems

to affect the position of the Andean mountain cat, as previously noted by Li et al. (2016) when using these same SNP data (relative to a larger dataset employed in that study). Regardless of its local instability, the Andean mountain cat was strongly supported as being more closely related to the ocelot and margay than to our focal clade comprising the tigrina complex, Geoffroy's cat and huiña (supplementary fig. S6-S9), so that its local instability does not affect our conclusions.

Our final phylogenetic analyses (excluding putative hybrids and individuals with extensive missing data) provided robust support for all nodes pertaining to the focal clade (fig. 1B; see Supplementary Information for additional analyses). Molecular dating analyses indicated that genus *Leopardus* began its diversification *ca.* 4.6 MYA, and that the pampas cat diverged from the focal clade >4 MYA (see fig. 1B). The Central American tigrina diverged from the inner group *ca.* 2.4 MYA, and clearly represents a distinct, species-level lineage. Within the inner group, the divergence between NE tigrina and *L. guttulus* was estimated at *ca.* 1.5 MYA, similar to the depth between the huiña and Geoffroy's cat (*ca.* 1.8 MYA). This result adds weight to the recognition of these two tigrina units as distinct species, corroborating previous genetic data indicating lack of gene flow between them (Trigo et al. 2013) and morphological analyses that support this taxonomic separation (Nascimento and Feijó 2017). Therefore, we conclude that the tigrina complex comprises at least three different species, one of which (S tigrina) has already been formally recognized as *L. guttulus*. The taxonomic assignment of NE tigrinas and Central American tigrinas will depend on additional geographic sampling of the complex, especially in the Guiana shield, which includes the type locality for *L. tigrinus* (Kitchener et al. 2017; Nascimento and Feijó 2017). This region remains unsampled for molecular data, and holds the key for resolving the taxonomy of this complex. Moreover, tigrinas from other regions in northern South America (e.g. Colombia, Peru) must also be analyzed to assess their affinities with the groups identified here. Interestingly, our results suggest that the 'tigrina' morphology may be plesiomorphic in this felid clade, leading to the existence of cryptic species that have remained undetected for decades. More broadly, our results illustrate how genomic data can be used to dissect complex histories of speciation and hybridization, uncover cryptic diversity, and inform the design of phylogenetic analyses in the face of potentially challenging confounding factors.

Material and Methods

Our initial dataset comprised all *Leopardus* individuals sampled by Li et al. (2016), which were genotyped with an Illumina array targeting genome-wide SNPs identified in the domestic cat (Mullikin et al. 2010). We complemented this dataset by genotyping the same markers in five additional *L. geoffroyi* individuals, two of which had suggestive evidence of admixture with *L. tigrinus* from the Brazilian northeast (Trigo et al. 2013) and six additional *L. guttulus* individuals (a species that had not been included in Li et al.'s [2016] study) with known geographic origin. This represents the most complete dataset assembled so far for this genus, including 22 individuals from all currently recognized species (supplementary table S1). Beginning from the 62,771 sites surveyed by this array, we applied filters using PLINK (Purcell et al. 2007), excluding individuals with more than 10% of missing data, and sites with 10% missing genotypes. Although the sites were originally selected in the domestic cat as SNPs, in our *Leopardus* datasets most of them (>90%) were invariant, which fits the goal of randomly surveying genomic sites, while still allowing the recovery of substantial evolutionary information. From the full genotype matrix, we constructed several different datasets that varied in the inclusion of putative hybrids and in the treatment of heterozygous and invariant sites (see Supplementary Information for details).

To characterize genetic structure, we conducted a Principal Component Analysis (PCA) using SmartPCA within the EIGENSOFT package (Patterson et al. 2006). PCA plots (for PCs 1-10) were then generated in R. Unsupervised analyses with Admixture (Alexander et al. 2009) were performed with three taxon subgroups: subgroup 1 included the focal clade (tigrina complex, Geoffroy's cat and huiña) and the pampas cat; subgroup 2 included the tigrina complex and Geoffroy's cat; and subgroup 3 included S tigrina, NE tigrina, Geoffroy's cat and huiña. To reduce bias, we implemented the penalized estimation using the best-fit lambda for each subgroup. All Admixture analyses were performed with a 5-fold cross-validation; the cross-validation error was calculated to determine the best-fitting K value.

For the phylogenetic analyses, we used two different approaches: (i) a concatenation of all sites (both variable and invariant) into a single supermatrix, followed by Maximum likelihood (ML) analyses with both RAxML v.8.2.5 (Stamatakis 2006) and IQ-TREE V2.1.2 (Nguyen et al., 2015); and (ii) two coalescent-based methods that allow SNP sites to evolve independently: the Bayesian approach implemented in SNAPP (Bryant et al. 2012), and the quartet-based inference implemented in SVDquartets v4.0

(Chifman and Kubatko, 2014). Finally, we used *memctree*, included in the PAML 4.9 package (Yang 2007), to date the inferred divergences, using a correlated rates model and a conservative molecular calibration for the root node (base of *Leopardus*), which was derived from the lower and upper boundaries (1.64 MYA and 5.03 MYA, respectively) reported by Li et al. (2016) for the age of this split.

Acknowledgments

The authors would like to thank CNPq/Brazil, INCT-EECBio/Brazil and FAPERGS/Brazil (PRONEX program: FAPERGS/CNPq 12/2014) for financial support. This study was also financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior/Brazil (CAPES)-(Finance Code 001).

Data Availability Statement

All the novel data reported in this study, along with the complete analyzed data set, will be made available as part of the online supplementary material that accompanies this paper.

Author Contributions

FJT: Helped design data analyses; conducted most data analyses; performed biological interpretations of the results; co-wrote the initial draft.

MRR: Conducted data analyses; performed biological interpretations of the results; edited the original draft.

HVF: Conducted data analyses; performed biological interpretations of the results; edited the original draft.

GL: Performed data collection; performed biological interpretations of the results; edited the original draft.

WJM: Helped design the study; supervised data collection; performed biological interpretations of the results; edited the original draft.

EE: Designed the study; supervised data analyses; performed biological interpretations of the results; co-wrote the initial draft.

References

- Alexander DH, Novembre J, Lange K. 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19(9):1655–1664. doi:10.1101/gr.094052.109.
- Bryant D, Bouckaert R, Felsenstein J, Rosenberg NA, Roychoudhury A. 2012. Inferring species trees directly from biallelic genetic markers: Bypassing gene trees in a full coalescent analysis. *Mol Biol Evol.* 29(8):1917–1932. doi:10.1093/molbev/mss086.
- Chifman J, Kubatko L. 2014. Quartet inference from SNP data under the coalescent model. *Bioinformatics*, 30(23), 3317–3324. doi: 10.1093/bioinformatics/btu530.
- Darriba D, Taboada GL, Doallo R, Posada D. 2012. JModelTest 2: More models, new heuristics and parallel computing. *Nat Methods.* 9(8):772. doi:10.1038/nmeth.2109.
- Edelman NB, Frandsen PB, Miyagi M, Clavijo B, Davey J, Dikow RB, García-Accinelli G, Van Belleghem SM, Patterson N, Neafsey DE, et al. 2019. Genomic architecture and introgression shape a butterfly radiation. *Science*, 366(6465), 594–599.
- Eizirik E. 2012. A Molecular View on the Evolutionary History and Biogeography of Neotropical Carnivores (Mammalia, Carnivora). In: Patterson BD, Costa LP, editors. *Bones, Clones, and Biomes: The History and Geography of Recent Neotropical Mammals*. University of Chicago Press. p. 123.
- Figueiró H V., Li G, Trindade FJ, Assis J, Pais F, Fernandes G, Santos SHD, Hughes GM, Komissarov A, Antunes A, et al. 2017. Genome-wide signatures of complex introgression and adaptive evolution in the big cats. *Sci Adv.* 3(7). doi:10.1126/sciadv.1700299.
- Johnson WE, Eizirik E, Pecon-Slattery J, Murphy WJ, Antunes A, Teeling E, O'Brien SJ. 2006. The Late Miocene Radiation of Modern Felidae: A Genetic Assessment. *Science* (80-). 311(5757):73–77. doi:10.1126/science.1122277.

Johnson WE, Slattery JP, Eizirik E, Kim J-H, Menotti Raymond M, Bonacic C, Cambre R, Crawshaw P, Nunes A, Seuanez HN, et al. 1999. Disparate phylogeographic patterns of molecular genetic variation in four closely related South American small cat species. *Mol Ecol.* 8(s1):S79–S94. doi:10.1046/j.1365-294x.1999.00796.x.

Kitchener AC, Breitenmoser-Würsten C, Eizirik E, Gentry A, Werdelin L, Wilting A, Yamaguchi N, Abramov AV, Christiansen P, Driscoll C, Duckworth JW, et al. 2017. A revised taxonomy of the Felidae: The final report of the Cat Classification Task Force of the IUCN/SSC Cat Specialist Group. *Cat News Special Issue* 11, 80 pp.

Li G, Davis BW, Eizirik E, Murphy WJ. 2016. Phylogenomic evidence for ancient hybridization in the genomes of living cats (Felidae). *Genome Res.* 26(1):1–11. doi:10.1101/gr.186668.114.

Li G, Figueiró HV, Eizirik E, Murphy WJ. 2019. Recombination-aware phylogenomics reveals the structured genomic landscape of hybridizing cat species. *Molecular biology and evolution*, 36(10), 2111-2126.

Mullikin JC, Hansen NF, Shen L, Ebling H, Donahue WF, Tao W, Saranga DJ, Brand A, Rubenfield MJ, Young AC, et al. 2010. Light whole genome sequence for SNP discovery across domestic cat breeds. *BMC Genomics.* 11(1):1–8. doi:10.1186/1471-2164-11-406.

Nascimento FO, Feijó A. 2017. Taxonomic revision of the tigrina *Leopardus tigrinus* (Schreber, 1775) species group (Carnivora, Felidae). *Pap Avulsos Zool.* 57(19):231–264. doi:10.11606/0031-1049.2017.57.19.

Nascimento FO, Cheng J, Feijó A. 2021. Taxonomic revision of the pampas cat *Leopardus colocola* complex (Carnivora: Felidae): an integrative approach. *Zoological Journal of the Linnean Society* 191 (2): 575–611. doi:10.1093/zoolinnean/zlaa043

Nguyen LT, Schmidt HA, Von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution* 32(1), 268-274. doi:10.1093/molbev/msu300.

- Patterson N, Price AL, Reich D. 2006. Population structure and eigenanalysis. PLoS Genet. 2(12):2074–2093. doi:10.1371/journal.pgen.0020190.
- Pulido-Santacruz P, Aleixo A, Weir JT. 2020. Genomic data reveal a protracted window of introgression during the diversification of a Neotropical woodcreeper radiation. Evolution, 74(5), 842-858.
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, De Bakker PIW, Daly MJ, et al. 2007. PLINK: A tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 81(3):559–575. doi:10.1086/519795.
- Santos AS, Trigo TC, de Oliveira TG, Silveira L, Eizirik E. 2018. Phylogeographic analyses of the pampas cat (*Leopardus colocola*; Carnivora, Felidae) reveal a complex demographic history. Genetics and Molecular Biology 41, 273-287.
- Stamatakis A. 2006. RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22(21):2688–2690. doi:10.1093/bioinformatics/btl446.
- Trigo TC, Freitas TRO, Kunzler G, Cardoso L, Silva JCR, Johnson WE, O'Brien SJ, Bonatto SL, Eizirik E. 2008. Inter-species hybridization among Neotropical cats of the genus *Leopardus*, and evidence for an introgressive hybrid zone between *L. geoffroyi* and *L. tigrinus* in southern Brazil. Mol Ecol. 17(19):4317–4333. doi:10.1111/j.1365-294X.2008.03919.x.
- Trigo TC, Schneider A, De Oliveira TG, Lehugeur LM, Silveira L, Freitas TRO, Eizirik E. 2013. Molecular data reveal complex hybridization and a cryptic species of Neotropical wild cat. Curr Biol. 23(24):2528–2533. doi:10.1016/j.cub.2013.10.046.
- Trigo TC, Tirelli FP, De Freitas TRO, Eizirik E. 2014. Comparative assessment of genetic and morphological variation at an extensive hybrid zone between two wild cats in southern Brazil. PLoS One. 9(9). doi:10.1371/journal.pone.0108469.

365 Yang Z. 2007. PAML 4: Phylogenetic analysis by maximum likelihood. *Mol Biol Evol.*
366 24(8):1586–1591. doi:10.1093/molbev/msm088.

367

368

369

Figure Legend

Figure 1. Population genetic and phylogenetic analyses of *Leopardus* based on genome-wide SNP data (see text and Supplementary Information for details). (A) Principal Component Analysis for the focal taxonomic group, comprising the tigrina complex, Geoffroy's cat and huiña; PCs 1-4 are shown, along with their respective variance explanatory power. (B) Maximum likelihood phylogeny of *Leopardus* based on a supermatrix comprising 60,931 SNPs (including 4708 variable sites); hybrid individuals and those with extensive missing data were excluded (see Supplementary Information for details). Bootstrap support values are shown next to nodes (nodes with no values indicate support below 60%). Numbers above branches are indicate divergence times (in Million years ago) for the adjacent node, with credibility intervals shown below the respective branch (see supplementary fig. S11 for true branch heights).



Figure 1