Phylogenomic evidence for ancient hybridization in the genomes of living cats (Felidae)

Gang Li,1 Brian W. Davis,1,2,4 Eduardo Eizirik,3 and William J. Murphy1,2

1Department of Veterinary Integrative Biosciences, Texas A&M University, College Station, Texas 77843, USA; 2Interdisciplinary Program in Genetics, Texas A&M University, College Station, Texas 77843, USA; 3Faculdade de Biociências, PUCRS, Porto Alegre, RS 90619-900, Brazil

Inter-species hybridization has been recently recognized as potentially common in wild animals, but the extent to which it shapes modern genomes is still poorly understood. Distinguishing historical hybridization events from other processes leading to phylogenetic discordance among different markers requires a well-resolved species tree that considers all modes of inheritance and overcomes systematic problems due to rapid lineage diversification by sampling large genomic character sets. Here, we assessed genome-wide phylogenetic variation across a diverse mammalian family, Felidae (cats). We combined genotypes from a genome-wide SNP array with additional autosomal, X- and Y-linked variants to sample ~150 kb of nuclear sequence, in addition to complete mitochondrial genomes generated using light-coverage Illumina sequencing. We present the first robust felid time tree that accounts for unique maternal, paternal, and biparental evolutionary histories. Signatures of phylogenetic discordance were abundant in the genomes of modern cats, in many cases indicating hybridization as the most likely cause. Comparison of big cat whole-genome sequences revealed a substantial reduction of X-linked divergence times across several large recombination cold spots, which were highly enriched for signatures of selection-driven post-divergence hybridization between the ancestors of the snow leopard and lion lineages. These results highlight the mosaic origin of modern felid genomes and the influence of sex chromosomes and sex-biased dispersal in post-speciation gene flow. A complete resolution of the tree of life will require comprehensive genomic sampling of biparental and sex-limited genetic variation to identify and control for phylogenetic conflict caused by ancient admixture and sex-biased differences in genomic transmission.

[Supplemental material is available for this article.]

There is an emerging consensus that gene flow frequently occurs following speciation despite the establishment of reproductive barriers that otherwise maintain species-level distinctiveness (Roca et al. 2005; Good et al. 2008; Ellegren et al. 2012; Garrigan et al. 2012; Toews and Brelsford 2012; Cahill et al. 2013, 2014; Cui et al. 2013; Martin et al. 2013; Kutschea et al. 2014; Sullivan et al. 2014). However, incomplete lineage sorting (ILS) is assumed by default to underpin most cases of phylogenetic discordance. Few studies in the literature account for hybridization by analyzing all inheritance patterns (uniparental, sex-biased, biparental) with high-resolution data, or instead have focused on only a few species (Roca et al. 2005; Cahill et al. 2013, 2014; Trigo et al. 2013; Khan et al. 2014). The cat family Felidae contains 38 recognized species within eight lineages (designated henceforth by a capitalized name, e.g., Puma lineage) that vary in the breadth of their geographic occurrence (Buckley-Beason et al. 2006; Johnson et al. 2006). While the relationships within many felid clades are robust to variation in subgenomic sampling, several inter-generic and inter-specific relationships remain unresolved and have not been assessed genome-wide to determine the specific drivers of discordance observed in previous studies (Johnson et al. 2006; Davis et al. 2010).

Although hybrid zones between related cat species have been reported (Schwartz et al. 2004; Homyack et al. 2008; Trigo et al. 2008, 2013), the extent to which ancient and contemporary introgression has occurred is poorly understood on broad geographic and genomic scales. Recent genetic evidence suggests complex patterns of admixture in felids of the Neotropical genus *Leopardus*, including the presence of cryptic species (Trigo et al. 2008, 2013). These observations are matched by the prevalence of felid hybridization in captivity (Gray 1972), which has generated numerous hybrids of both large cats and medium to small cats. These include the gigantic liger, a hybrid between a male lion and female tiger, as well as domestic cat inter-specific hybrid breeds, including the Bengal and Savannah, which are common household pets worldwide. This proclivity for hybridization is facilitated by the strong colinearity among felid genomes coupled with recent genetic divergence (Wurster-Hill and Centerwall 1982; Davis et al. 2009; Cho et al. 2013). The genomes of modern felids thus present a unique resource to study the dynamics of introgression and the genetic basis of reproductive isolation in both controlled crosses and natural populations (Davis et al. 2015).

Previous studies have demonstrated that robust phylogenetic signal can be obtained by querying domestic animal SNP arrays with DNA from related species of the same genus, family, or order, despite having diverged tens of millions of years from the array reference genome (Decker et al. 2009; McCue et al. 2012). Here, we generated genome-wide SNP data from 38 cat species and analyzed
these separately and together with Y-linked variation and whole mitogenomes, allowing us to disentangle different maternal, paternal, and biparental histories within a diverse family of mammals. We assessed genome-wide patterns of intra-lineage phylogenetic discordance and identified signals of ancient hybridization throughout the genomes of many cat species. Many of these nuclear signatures were accompanied by patterns of mitonuclear discordance. Our results allow further insight into the evolutionary processes leading to the diversification of extant cats of the world and provide a roadmap for future in-depth population genomics in this group as a model system for better understanding the speciation process.

Results

We generated Illumina whole-genome genotyping data (~63,000 SNPs) for 100 felid DNA samples covering virtually all recognized species and validated the consistency of genotyping quality and utility for phylogenomic inference (see Supplemental Table S1; Supplemental Fig. S1; Methods). SNP call rates ranged from 92% to 99% across felids and were generally correlated with previous estimates of phylogenetic divergence (Johnson et al. 2006). Within each felid lineage, SNP call rates for each species were very similar (Supplemental Table S1), indicating that comparisons between members of the same felid lineage (i.e., admixture tests) should not be compromised by array ascertainment bias.

After excluding low-quality SNPs and heterozygous sites, we generated maximum likelihood (ML) phylogenies for the combined SNP supermatrix (Fig. 1) and for each chromosome (Supplemental Fig. S2). No individual chromosome was strongly discordant with the biparental SNP-based phylogeny, except for the Y Chromosome data set from Johnson et al. (2006), which strongly supported a closer association of the Bay cat lineage with the Panthera lineage (also observed by Luo et al. 2014). To confirm that SNP-based characters were robust for phylogenetic inference, we compared our results to trees derived from an independent 50-kb supermatrix of complete feld vomeronasal receptor (V1R) gene repertoires for 27 felids (Montague et al. 2014). The V1R ML topology showed strong congruence with the SNP-based phylogeny and confirmed the nested position of the Bay lineage.
cat lineage and more basal positions for the Caracal and Ocelot lineages (Supplemental Fig. S3). We then generated a robust felid phylogeny (Fig. 1A) based on biparentally inherited data (~130-kb from SNPs + V1R genes + published genes) that removed the confounding effects of paternal evolutionary history (Fig. 1B).

Mitonuclear discordance

Next, we contrasted the maternal and biparental evolutionary history of felids by assembling complete mitochondrial genomes of nearly all recognized species (except *Neofelis diardi* and *Leopardus guttulus*) using light coverage whole-genome sequencing (see Methods; Supplemental Table S1). Remapping the raw reads to each assembly revealed extremely high and even depth of coverage for each mitochondrial genome (avg. 135-fold) compared to the expected nuclear genome coverage (avg. 0.3-fold coverage). This process supported the true cytoplasmic origin of each mitogenome assembly. When we compared the mitogenome-based phylogeny to our nuclear genome species tree, we identified nine topological conflicts, including strongly supported differences within *Panthera* (i.e., the position of the snow leopard) as well as the relative position of the Puma and Caracal lineages within Felidae (Fig. 1B). The mitogenomic phylogeny supported a sister-group relationship between the Bay cat and Lynx lineages that was also observed in the biparental phylogeny but not in the Y Chromosome phylogeny (Fig. 1B), indicating the Bay cat lineage Y Chromosome tracks a unique history relative to the other genomic partitions.

One potential and underappreciated source of mitonuclear discordance is the inadvertent amplification and sequencing of nuclear mitochondrial pseudogenes (numts) (Antunes et al. 2007). When we compared our reference mitogenomes to GenBank entries, we found that half of the published big cat (i.e., genus *Panthera* and *Puma concolor*) mitogenomes contained long stretches of high sequence divergence that were consistent with numt contamination (Supplemental Fig. S4; Supplemental Table S2). These results underscore the confounding impact that numts may have on phylogenetic relationships (Antunes et al. 2007; Davis et al. 2010) and caution against indiscriminate use of published mitochondrial DNA (mtDNA) sequences.

Felid diversification and evidence for hybridization

We then applied Bayesian divergence time and ancestral area reconstruction approaches to explore spatial and temporal patterns of diversification throughout modern felid evolution (Supplemental Figs. S5, S6; Supplemental Tables S3, S4). Our time tree indicated that the ancestors of each felid lineage originated within and dispersed out of Asia in the late Miocene, obviating the need to invoke hypotheses requiring multiple dispersal events back to Asia from North America (Fig. 1; Johnson et al. 2006). Rather, the ancestors of the Puma and Lynx lineages probably dispersed simultaneously to the Americas via the reopening of the Bering Land Bridge ~5.9 million years ago (MYA) (Koufos et al. 2005) as documented by the first occurrence of fossil Felinae in North America shortly thereafter (Qiu 2003). Older divergence times (~10 MYA) for the progenitors of the Caracal and Ocelot lineages leave open the possibility that the ancestors of these two lineages dispersed out of Asia into Africa and the Americas, respectively, via land bridges established earlier rather than later in the Miocene (Haq et al. 1987; Koufos et al. 2005).

Given our observation of substantial phylogenetic discordance across the felid phylogeny, along with biogeographic reconstructions that indicate ample spatial overlap in the history of cat lineages, we scanned the genome-wide SNP matrix for phylogenetic patterns suggestive of inter-specific hybridization. Our first approach used sliding window-based likelihood ratio tests to identify broad (>1 Mb) chromosomal regions that harbored significantly different phylogenetic signal from the species tree (Fig. 2A,B), which would not be expected when serial species divergences were separated by millions of years (Hoboth et al. 2007, 2011). In order to distinguish discordant regions that are a result of hybridization from ILS, we estimated the D-statistic from the ABBA/BABA test (Table 1), which assesses phylogenetic asymmetry of nonspecies trees and the proportion of the genome that is shared between two taxa due to admixture (Durand et al. 2011). Taken together, these approaches provide genome-wide evidence for historical gene flow within the majority of the eight felid lineages (Table 1; Fig. 2).

On one end of this continuum, the Puma, Caracal, and Bay cat lineages showed fewer discordant windows and possessed D-statistics and Z-scores lower than most other clades (Table 1; Fig. 2). These lineages each contain three ecologically distinct species that diverged over long evolutionary time frames relative to other felid groups, thus ILS-based discordance is less expected. In contrast, the remaining five felid lineages possessed numerous signatures of phylogenetic discordance, some of which may be attributable to historical hybridization (Table 1; Fig. 2). One of the best documented felid hybrid zones is between the bobcat and Canada lynx, which share a broad trans-continental range overlap in North America that has likely persisted to varying degrees as climate fluctuated across time. Genetic studies have identified several populations where hybridization is common along the US/Canada border (Schwartz et al. 2004; Homayack et al. 2008; Koen et al. 2014). ABBA/BABA tests indicate ancient signatures of bobcat-Canada lynx gene flow (Table 1), consistent with an extended period of gene flow that continues to the present day. Our results also indicate introgression between the Canada lynx and Eurasian lynx (Table 1), which likely occurred during recurrent emergence of the Bering Land Bridge during the Pleistocene.

The greatest amount of phylogenetic discordance was observed within the Ocelot lineage, which diversified within the past 2–3 million years (MY) in the Neotropics, where many species coexist in sympatry or parapatry. Recent genetic analyses demonstrated complex speciation/hybridization affecting at least four species within this group: two tigrina species, one from northeastern Brazil (*Leopardus tigrinus*) that possesses pampas cat (*L. colocolo*) mitochondria within a tigrina-like nuclear background, and a recently proposed sister species from southeastern Brazil (*L. guttulus*, not sampled here) which hybridizes with the Geoffroy’s cat (*L. geoffroyi*) in a narrow hybrid zone in southern Brazil (Trigo et al. 2008, 2013, 2014). As predicted, the northeastern tigrina we genotyped possessed pampas cat mtDNA within a tigrina nuclear DNA background (Supplemental Fig. S8) but also contained nuclear signatures of ancient hybridization with the Geoffroy’s cat (Table 1; Fig. 3A), suggesting an extended history of admixture between the three species lineages. A Central American tigrina (Fig. 1; Supplemental Fig. S8) showed large mitochondrial and nuclear divergence (11.0%–15.3% and 0.5%–0.6%, respectively) from Brazilian tigrinas, as well as from the Geoffroy’s cat and kodkod (*L. guigna*). Tigrinas have not been extensively sampled for genetic variation across the northern part of their range; however, our data support previous observations based solely on mtDNA (Johnson et al. 1999; Trigo et al. 2008) of the potential existence of an additional, presently unrecognized Central American cat species.
The Asian Leopard cat lineage (genus *Prionailurus*) includes a similar radiation of small-bodied cats distributed throughout Southeast Asia. The leopard cat is broadly sympatric with other members of the genus, and recent molecular evidence indicates that its Indochinese and Sundaic populations display species-level measures of Y Chromosome and mitochondrial divergence (Luo et al. 2014). Whole-genome SNP data from 13 Asian leopard cats sampled across these regions confirmed a species-level separation of more than one million years, which was further supported by deep mtDNA divergence (Fig. 3B). Interestingly, mitogenome phylogenies support a closer relationship between the fishing cat and members of the Domestic cat lineage: Black-footed cat, Sand cat, and Jungle cat, and a close outgroup species, the Asian leopard cat. We aligned these reads to the domestic cat v6.2 assembly, generated reference assemblies, and calculated pairwise identity between the domestic cat probe sequence and sequences and recalculated the statistics from a reduced set of high-confidence orthologous SNPs, for which we required 99% sequence and recalculated the statistics from a reduced set of high-confidence orthologous SNPs, for which we required 99% identity to the SNP call in the reference assemblies of the species. As an additional validation of the SNP-based *D*-statistics, we remapped the domestic cat SNP probes to the new *Felis* genome sequence and recalculated the statistics from a reduced set of high-confidence orthologous SNPs, for which we required 99% identity to the SNP call in the reference assemblies of the other felids. The results (Supplemental Table S6) demonstrated that even with a smaller, high confidence set of orthologous SNPs that match the sequence probe (and the SNP call) precisely,

---

**Figure 2.** Signatures of genome-wide phylogenetic discordance. (A) Results of sliding window-based approximately unbiased (AU) tests for each of the eight feline lineages (see Methods for details). Horizontal panels indicate genomic regions (vertical bars) that produced significantly better topologies than the species tree topology. Felid chromosomes are displayed in bars below the scale spanning the length of the feline genome. Abbreviations for lineages are as follows: (ALC) Asian leopard cat, (DOM) Domestic cat, (LYN) Lynx, (BAY) Bay cat, (PUM) Puma, (CAR) Caracal, (OCE) Ocelot, (PAN) *Panthera*. (B) Summary statistics for results in Figure 1A. The white bars represent the percentage of sliding windows supporting the species tree topology (black bars) with bootstrap support ≥ 70% (Hillis and Bull 1993). The gray bars indicate the proportion of nonspecies tree topologies identified by the AU test that were significantly better than the species tree (Fig. 1A).
Table 1. Measures of phylogenetic discordance and admixture within felid lineages

<table>
<thead>
<tr>
<th>Lineage</th>
<th>% of windows matching species tree</th>
<th>% of nonspecies-tree topologies supported by likelihood ratio tests</th>
<th>Species tree (P1,P2) P3</th>
<th>Outgroup</th>
<th>ABBA/BABA results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ocelot</td>
<td>1.5</td>
<td>26.0</td>
<td>(Kodkod, Tigrina, S.Am.), Pampas cat *</td>
<td>Serval</td>
<td>$0.102 \pm 0.057$ $1.782$ $0.0748$ $173$ $141$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(Kodkod, Geoffroy's cat), Tigrina (S.Am.)</td>
<td>Serval</td>
<td>$0.146 \pm 0.048$ $3.064$ $0.0022$ $232$ $173$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(Ocelot, Margay), Pampas cat</td>
<td>Serval</td>
<td>$0.063 \pm 0.063$ $1.466$ $0.1427$ $271$ $239$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(Ocelot, Margay), Tigrina (S.Am.)</td>
<td>Serval</td>
<td>$0.029 \pm 0.045$ $0.079$ $0.9368$ $293$ $295$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(Leopard, Lion), Jaguar</td>
<td>Sunda Clouded leopard</td>
<td>$0.013 \pm 0.043$ $0.310$ $0.7565$ $264$ $257$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(Lion, Leopard), Tiger</td>
<td>Sunda Clouded leopard</td>
<td>$0.031 \pm 0.048$ $0.652$ $0.5144$ $215$ $202$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(Snow leopard, Tiger), Leopard</td>
<td>Sunda Clouded leopard</td>
<td>$0.056 \pm 0.042$ $1.345$ $0.1788$ $291$ $260$</td>
</tr>
<tr>
<td>Panthera</td>
<td>14</td>
<td>14.3</td>
<td>(Tiger, Snow leopard), Lion *</td>
<td>Domestic cat</td>
<td>$0.041 \pm 0.001$ $33.39$ $0$ $357,875$ $329,510$</td>
</tr>
<tr>
<td>Lynx</td>
<td>37.3</td>
<td>8.7</td>
<td>(Iberian lynx, Canada lynx), Bobcat *</td>
<td>Asian Leopard cat</td>
<td>$0.176 \pm 0.056$ $3.182$ $0.0015$ $204$ $143$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(Iberian lynx, Eurasian lynx), Canada lynx</td>
<td>Asian Leopard cat</td>
<td>$0.199 \pm 0.047$ $6.112$ $1.9 \times 10^{-6}$ $274$ $183$</td>
</tr>
<tr>
<td>Leopard cat</td>
<td>26.6</td>
<td>7.6</td>
<td>(Sunda ALC, Indochinese ALC), Fishing cat *</td>
<td>Pallas cat</td>
<td>$0.179 \pm 0.047$ $3.847$ $0.0001$ $277$ $193$</td>
</tr>
<tr>
<td>Domestic cat</td>
<td>33.1</td>
<td>4.4</td>
<td>(Sand cat, Domestic cat), Black-footed cat</td>
<td>Jungle cat</td>
<td>$0.152 \pm 0.023$ $6.630$ $3.4 \times 10^{-11}$ $1033$ $761$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(Black-footed cat, Sand cat), Jungle cat</td>
<td>Bobcat</td>
<td>$0.090 \pm 0.028$ $3.223$ $0.0013$ $690$ $577$</td>
</tr>
<tr>
<td>Puma</td>
<td>65.7</td>
<td>1.5</td>
<td>(Domestic cat, Sand cat), Jungle cat</td>
<td>Bobcat</td>
<td>$0.186 \pm 0.024$ $7.626$ $2.4 \times 10^{-14}$ $926$ $636$</td>
</tr>
<tr>
<td>Bay cat</td>
<td>80.4</td>
<td>1.0</td>
<td>(Bay cat, Asian golden cat), Marbled cat</td>
<td>Tiger</td>
<td>$0.139 \pm 0.042$ $2.976$ $0.0029$ $312$ $243$</td>
</tr>
<tr>
<td>Caracal</td>
<td>98.0</td>
<td>0.0</td>
<td>(African golden cat, Caracal), Serval</td>
<td>Asian golden cat</td>
<td>$0.058 \pm 0.069$ $0.842$ $0.3997$ $109$ $97$</td>
</tr>
</tbody>
</table>

Boldface indicates significant $D$-statistics ($Z \geq 3.0$) with SNP and/or SNV data.
*Species comparisons with documented mitochondrial introgression or hybridization.
**Whole-genome SNV alignment results.
the D-statistics remain significant. These results confirm that phylogenetic noise or genotyping errors were not biasing the results and further validated the use of the array-based D-statistics. As a third measure to assess the potential impact of phylogenetic noise in the SNP array data generated for the non-Felis lineages, we jackknifed the outgroup taxon for the significant comparisons shown in Table 1 of the manuscript, and in nearly all cases the Z-scores remained significant (Supplemental Table S7).

**Big cat hybridization and remarkable patterns of X Chromosome divergence**

Big cats of the genus *Panthera* readily reproduce in captivity (Gray 1972), yielding many possible hybrids among the parent species, including the liger (male lion × female tiger). The recent sequencing of several big cat genomes (Cho et al. 2013) provided an opportunity to perform a high-resolution test of the hypothesis that ancient introgression may have led to the mitonuclear discordance in the phylogenetic position of the snow leopard (Figs. 1, 3D). Genome-wide analysis of tiger, lion, and snow leopard single nucleotide variants (mean = 20.8 million SNVs/species), using the domestic cat as the outgroup, identified significant signatures of admixture between the lion and snow leopard genomes (Supplemental Table S8; Fig. 3). The most striking signal was observed for the X Chromosome, where sliding window-based divergence time estimates were significantly younger than autosomes and notably enriched for topologies supporting a sister relationship between snow leopard and lion (Supplemental Fig. S9; Fig. 4A), similar to the mitogenomic phylogeny where the snow leopard is sister to lion and leopard. Furthermore, our estimates for lion/snow leopard X Chromosome divergence were very similar to the mean mitogenomic divergence (∼2.1 MYA) (Supplemental Fig. S5B).

The distribution of the alternative topologies on *Panthera* X Chromosomes was nonrandom, with more than half occupied by massive blocks of reduced divergence time. These blocks correspond to one ∼45-Mb recombination cold spot and at least two smaller S- to 10-Mb cold spots on the domestic cat X Chromosome (Schmidt-Kuntzel et al. 2009), a genomic feature virtually absent on autosomes (Fig. 4B). Intriguingly, a large recombination cold spot, also flanked by high recombination rate regions, is found on the domestic pig X Chromosome in nearly the same syntenic region (Ma et al. 2010; Ai et al. 2015). Both species share colinearity with the human chromosome over the vast majority of the

**Figure 3.** Patterns of hybridization within felid lineages. (A) Ocelot lineage, showing the discordant position of the northern tigrina (green) between nuclear (red) and mitogenome (black) phylogenies. Within the species tree is shown nuclear (red) and mitogenome (black) genealogies, with dashed red lines indicating genomic evidence for ancient hybridization (not scaled to time). (B) Asian leopard cat lineage, showing the discordant position of the fishing cat (green) relative to the Asian leopard cat populations in nuclear and mitogenome trees. (C) Domestic cat lineage, showing the results of admixture tests based on whole-genome SNV alignments which reveal two strong introgression signals (a and b) across the whole genome of *Felis* species. (D) *Panthera* lineage, showing the discordant position of the snow leopard (green) between nuclear and mitogenome phylogenies. Admixture tests from whole-genome SNV alignments support gene flow from the lion/leopard lineage to the snow leopard. Note the significantly skewed D-statistic for the X Chromosome (right).
Figure 4. Evidence for admixture within Panthera genome sequences. (A) Species tree topology of the three Panthera species defined by black outline. Blue lines represent coalescent patterns of tiger + snow leopard shared alleles, green indicates lion + tiger shared alleles, and red represents lion + snow leopard shared alleles. (B) Distribution of trees supporting each topology from the genome-wide sliding window analysis (y-axis) plotted against divergence time (x-axis), shown for autosomes and the X Chromosome (see Supplemental Fig. S9 for individual chromosome plots). (C) Mean and standard error for divergence times for each category of sister-species relationship within the three taxa. (D) Phylogeny and relative divergence time (y-axis, in millions of years ago) for each window plotted along the largest autosome (Chromosome A1, above) and the X Chromosome (below). Dots are color-coded for each of the three sister-species relationships (see legend). The gray dots indicate the age of the base of Panthera. Note the drop in divergence times for both internal and basal nodes in regions corresponding to extremely low recombination rates (bottom). Recombination data are from the domestic cat linkage map (Schmidt-Kuntzel et al. 2009).

Like most mammals, felid males disperse farther than females (Sunquist and Sunquist 2002), hence the maternally transmitted mitochondrial genome is not predicted to reflect species boundaries as accurately as Y Chromosome and autosomal markers (Currat et al. 2008; Petit and Excoffier 2009). This prediction is supported similar to that invoked to explain recent X Chromosome divergence between chimpanzee and human (Patterson et al. 2006). If fertile F1 female hybrids, derived from snow leopard-ancestor males and lion/leopard-ancestor females, backcrossed to males from the snow leopard ancestor, a strong selective sweep favoring a newly beneficial X Chromosome allele(s), coupled with severely limited recombination across a large portion of the X Chromosome (Fig. 4), could account for the presence of this large introgressed region within the genome of modern snow leopards. The persistence of phylogenetically discordant X Chromosome haplotypes is expected as they maintain a mitochondrion-like signature longer than other nuclear markers, being carried twice as often by females than males. This trend would be exacerbated in X Chromosome regions with extremely reduced recombination.

Discussion

Like most mammals, felid males disperse farther than females (Sunquist and Sunquist 2002), hence the maternally transmitted mitochondrial genome is not predicted to reflect species boundaries as accurately as Y Chromosome and autosomal markers (Currat et al. 2008; Petit and Excoffier 2009). This prediction is supported...
by the greater discordance between the felid mitochondrial phylogeny and different nuclear partitions, the latter of which are more similar (Fig. 1). In addition, felid inter-species hybrids follow Haldane’s Rule (Haldane 1922) specifically as it relates to hybrid sterility: Females are generally fertile, while males are overwhelmingly sterile (Gray 1972; Davis et al. 2015). Unisexual sterility coupled with male-biased dispersal promotes situations where the mitochondrial of one species or lineage may persist within the genomic background of a different species following introgression into one of the parental populations (e.g., Roca et al. 2005). This pattern is exemplified by the northeastern tigrina and the fishing cat genomes, which possess mitochondrial DNA more similar to that of a different species.

Such sex-biased asymmetries in fertility and dispersal may also explain the observed discordances at deeper branches of the felid phylogeny when interpreted jointly with Bayesian ancestral geographic reconstructions (Supplemental Fig. S6). In one example, the mitogenomic phylogeny strongly supports the Puma lineage as sister to the Lynx+Bay cat clade, while the nuclear genome places the Puma lineage as sister to the Asian leopard cat+Domestic cat clade. Although ILS could explain this pattern, we interpret the strongly supported alternate topologies and the reduced relative divergence times between genomic partitions (Supplemental Fig. SSA,B) as evidence for historical admixture and mitochondrial capture between ancestors of the Puma lineage and Lynx+Bay cat clade within Asia, preceding migration to North America in the Late Miocene (Fig. 1). Similarly, the Caracal clade origin is ∼2 MY and multiple inter-nodes younger in the mitogenome than in the biparental tree (Fig. 1A,B; Supplemental Fig. SSA,B), a pattern that is also consistent with a scenario involving mitochondrial capture.

Sex-biased asymmetry can also produce rare instances of Y Chromosome discordance, the nuclear marker that is least expected to show discordance in species with male-biased dispersal (Petit and Excoffier 2009). Such an example exists within the felid phylogeny, where the 9.5 MYA coalescence of the Bay cat and Panthera lineage Y Chromosomes (Luo et al. 2014) post-dates the estimated biparental lineage divergence by >2 MY, spanning several internal nodes that are discordant between the two topologies. We interpret this pattern as indication of an episode of ancient admixture within Eurasia where these two lineages originated and diversified (Supplemental Fig. S6). Introggression of a Panthera-like Y Chromosome into the Bay cat lineage ancestor early during post-speciation divergence could have been derived from the generation of fertile male backcross hybrids, and their transient increased fitness over nonhybrid males. This could effectively result in a selective sweep by a Panthera-like Y Chromosome within the nuclear genome background of a Bay cat lineage ancestor. In this particular scenario, it is worth noting that the ancestors of each of these cat lineages were much less divergent at the putative time of hybridization than many living felids that currently produce viable and fertile inter-specific hybrids (Gray 1972; Schwartz et al. 2004).

Although phylogenetic discordance is often attributed to ILS, the low effective population size of mitochondrial and Y Chromosome DNA, and that of felid species in general, coupled with the influence of male-mediated dispersal and Haldane’s Rule on introgression, suggests that the conflicting phylogenetic signals between genomic partitions are best explained by hybridization (Currat et al. 2008; Petit and Excoffier 2009; Cahill et al. 2013). Our observations of historical admixture between many pairs of cat species are consistent with evidence for contemporary inter-specific hybridization in nature and captivity spanning large phylogenetic distances. Ongoing whole-genome sequencing efforts in different felid species should provide much greater resolution and insight into the genomic landscape of admixture and provide further tests of these SNP array-based inferences.

Hybridization is a natural component of the evolutionary process, yet anthropogenic influences may further promote inter-specific hybridization in low-density carnivore populations (Allendorf et al. 2001). This is especially critical for the majority of cat species worldwide, where habitat encroachment, poaching, deforestation, and climate change are drastically reducing felid numbers. We have highlighted groups of species where further population genomic sampling will better define the architecture of admixture and genes underlying adaptive introgression and divergence. Differentiating between natural and human-mediated hybridization will be critical to develop effective conservation efforts on behalf of these threatened carnivores.

Methods

SNP genotyping and analysis

We used 150–300 ng of DNA from each felid species to genotype ∼63,000 SNPs on the Illumina feline array. The 63K array was designed using the domestic genome sequence as a reference with 62,897 SNPs ascertained from six domestic cats and an African wild cat (Felis lybica) (Mullikin et al. 2010), distributed across all 18 autosomes and the X Chromosome. The vast majority of the SNPs on the array were ascertained in a pool of domestic cat breeds from low-coverage whole-genome sequencing (Mullikin et al. 2010). Illumina array genotypes for all individuals were filtered to include only SNP call rates >95%. We excluded all SNPs included in the array design that assay felid lineage-specific apomorphies and phenotype-associated mutations (Eizirik et al. 2003; Johnson et al. 2006). Heterozygous sites (predominantly found in the domestic cat and its closest relatives, the European wild cat, African wild cat, and Chinese desert cat) (Supplemental Table S1) were scored as ambiguities in phylogenetic analyses only.

Since the creation of the 63K chip, revisions to the feline genome assembly required reassessment of the physical marker location of the SNP probe sequences. We compared each probe sequence to the Felis catus v6.2 (felCat5) assembly using BLAST (Altschul et al. 1997). After removal of SNPs with negative and/or highly duplicated probe mapping results, 59,628 SNPs remained in our analyses. The highest nondomestic cat SNP calling rate was observed between the three wildcat subspecies/species of the Felis silvestris complex (i.e., European wild cat, African wild cat, and Chinese desert cat; average call rate >99.4%). The lowest SNP call rate (86%) was observed for the Andean mountain cat DNA sample, extracted from a museum hide, which may have contributed to its poor phylogenetic placement within the Ocelot lineage.

To demonstrate that a potential bias of nucleotide identification caused by the two-dye Illumina genotyping system on nondomestic cat species did not influence phylogenetic accuracy, we also performed a second phylogenetic analysis in which the nuclear SNP matrix was translated into a binary matrix (“0” representing G or C; “1” representing A or T), following Decker et al. (2009).

Mitochondrial genome sequencing and assembly

To obtain high-quality complete mitochondrial genome sequences and to avoid sequencing nuclear mitochondrial pseudogenes, we isolated DNA from mitochondria-enriched preparations of...
fibroblast cells for 35 felid species where cell lines or tissues were available. Mitochondrial enrichment was performed with a two-step sucrose gradient procedure (Jones et al. 1988) that yields both mitochondrial and nuclear-enriched pellets. Each pellet was extracted using the Qiagen DNeasy kit. Standard Illumina fragment libraries (~300-bp average insert size) were prepared for each mitochondria-enriched DNA isolate and were sequenced to ~0.3× genome-wide depth of coverage on the Illumina HiSeq 2000 platform. We generated de novo mitochondrial genome assemblies with SOAPdenovo2 (Luo et al. 2012) by evaluating a series of k-mer sizes. Following assembly, all raw Illumina reads were mapped to assembled contigs using default settings in BWA-MEM to assess coverage depth (Li and Durbin 2009). Mapping results were analyzed using SAMtools, including removal of PCR duplicates (Li et al. 2009). Mitochondrial contigs were identified with BLAST (Altschul et al. 1997) comparisons to the domestic cat mitochondrial genome and confirmed by read-depth statistics.

Sequence alignment, phylogenetic reconstruction, and tree topology comparison

The SNP matrix was combined with several published genes and gene supermatrices, including a 5.6-kb CES7 gene matrix (Li et al. 2011), the 23-kb multigene supermatrix of Johnson et al. (2006), and a 49.7-kb vomeronasal (V1R) gene supermatrix (Montague et al. 2014). Sequence alignments were performed with MAFFT (Katoh and Toh 2010). We applied MODESTTEST (Posada and Crandall 1998) to select the best-fitting substitution model for each data set. RAxML 7.2.8 (Stamatakis 2006) was used for maximum likelihood tree searching and bootstrapping. We also analyzed a more conservative subset of the SNP matrix (40,225 SNPs) that excluded all SNP markers whose flanking probe sequence failed to produce a single unique BLAST hit with 100% nucleotide identity to the felCat5 (v6.2) genome reference sequence. This precaution was taken because SNPs on the array were identified from previous low coverage (1.9×) assemblies of the domestic cat genome, i.e., versions felCat3/4 (Pontius et al. 2007; Mullikin et al. 2010).

Divergence time estimation and ancestral area reconstruction

We used the MCMCtree v4.8a software in the PAML-4 package (Yang 2007) to estimate divergence times using previously published fossil constraints for Felidae (Johnson et al. 2006). Analyses were run for 100,000 generations with a burn-in of 10,000 generations. Analyses were run twice to check for convergence. The final time trees in Supplemental Figure S5, A and B are based on the mean of analyses that varied fossil constraint (hard- versus soft-bounded), rate model (independent versus autocorrelated), and character sampling (only for the nuclear gene supermatrix, both with and without the SNP matrix) (Supplemental Tables S3, S4). Ancestral area reconstructions were performed with the Bayesian Binary MCMC (BBM) approach implemented in the software RASP (Ali et al. 2012).

Detection of historical gene flow

We used Consel (Shimodaira and Hasegawa 2001) to perform approximately unbiased (AU) tests to compare different tree topologies across the whole-genome SNP matrix. We divided the whole-genome matrix into 10-Mb sliding windows (step = 100-kb), chosen to contain adequate sequence variation within each felid lineage (Supplemental Table S9) to allow for examination of variation in phylogenetic signal across each chromosome. Maximum likelihood trees for each window were built using RAxML (GTR + Γ substitution model) and compared to the null hypothesis (i.e., the species tree topology).

To detect admixture over longer evolutionary time scales, we applied the ABBA/BABA approach of Green et al. (2010) to the SNP matrix as well as whole-genome SNV data sets (see below). This method identifies imbalances in alternative topology frequencies under a four-taxon scenario (three ingroup taxa and one outgroup). For analyses based on the Illumina SNP matrix, we tested alternative phylogenetic scenarios only within members of the eight felid lineages, to eliminate any ascertainment bias that would result from the SNPs being discovered in the domestic cat (Felis silvestris catus). Although each SNP character state that is observed within any of the non-Felis cat lineages may be due to recurrent mutation, relative to the domestic cat from which they were discovered, these mutations are unique within each of the felid lineages and will track lineage-specific changes within that clade and in its immediate ancestors. Whether the mutation occurred within the immediate ancestor of the domestic cat and the ancestors of another felid lineage is irrelevant because we did not include members of the Domestic cat lineage as outgroup taxa when calculating SNP array-based D-statistics for the other seven felid lineages. All D-statistics calculated within the Domestic cat lineage were validated using whole-genome sequencing SNV data, to rule out ascertainment bias effects within the members of the genus Felis.

D-statistics and Z-scores calculations were performed with the R package evobitR (http://www.uta.edu/karyodb/evobit/index.html), which measures signatures of alternative phylogenetic asymmetry and the proportion of the genome that is shared between two taxa due to admixture, respectively (Durand et al. 2011). ABBA/BABA tests performed on the SNP data set included heterozygous sites by generating pseudohaploid sequences, following Green et al. (2010) and Cahill et al. (2013, 2014). Statistical significance of the Z-score was assessed for each replicate by converting the Z-score into a two-tailed P-value (Eaton and Ree 2013). One thousand bootstrap iterations were used to measure the standard deviation of the D-statistic derived from the SNP-array data, because the ABBA/BABA SNP sites are largely unlinked (~1 SNP per 4 Mb) (Eaton and Ree 2013, Streicher et al. 2014). We defined significant D-statistics as those having Z-scores ≥ 3 (following Green et al. 2010; Cahill et al. 2013).

Whole-genome sequencing and alignment of Felis species

We constructed standard 300-bp insert Illumina fragment libraries using high molecular weight DNA extracted from fibroblast cell cultures of sand cat (Felis margarita), jungle cat (Felis chaus), and Asian leopard cat (Prionailurus bengalensis). For each sequencing library, we generated ~25× genome coverage with 2 × 125-bp reads on the Illumina HiSeq 2500 platform. In addition, we analyzed ~713 million raw Illumina sequence reads (2 × 100-bp, Illumina HiSeq 2000) from a black-footed cat, Felis nigripes. Reads were trimmed (Trim Galore, http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) and mapped to the v6.2 domestic cat genome assembly using BWA-MEM (Li and Durbin 2009) with default settings. SAMtools (Li et al. 2009) was used to call raw single nucleotide variants (SNV). We filtered the SNV data as follows: (1) We identified high-quality SNVs (quality > 100); (2) we restricted our analysis to variants with mapping quality ≥ 30 and excluded variants from regions with read-depth variation greater or less than 50% of the genome-wide average; (3) we evaluated the likelihood score for each genotype (i.e., likelihood of double-strand alternative genotype equal to 0); and (4) we merged the filtered SNV data to whole-genome alignments conforming to the structure of the domestic cat v6.2 reference genome assembly. D-statistics were
calculated in a similar fashion as for the SNP data set; however, statistical significance of the Z-score was assessed with a block jackknife test (S-Mb block size) with 100 replicates resampled with replacement.

Whole-genome analysis of Panthera species

To detect signals of phylogenetic discordance and introgression within the whole genomes of tiger, lion, and snow leopard, we generated D-statistics and performed a sliding-window-based divergence time scan across all chromosomes. Illumina short reads from the three Panthera species (Cho et al. 2013) were aligned to the repeat-masked domestic cat genome assembly (serving as the outgroup) using BWA (Li and Durbin 2009) with the following settings: bwa aln -n 0.08 SNV data for three Panthera species was obtained and filtered applying the same strategy used for the Felis whole-genome data. Whole-genome four-way alignment blocks were created and analyzed in 100-kb windows (with 25-kb moving steps) across the whole-genome alignment. D-statistics were calculated in the same manner as the Felis data set (described above). Maximum likelihood trees were constructed for each window with RAxML (Stamatakis 2006) (GTR+Gamma substitution model). We used MCMCtree in PAML to calculate the relative divergence time for each node, assuming independent rates and employing three soft-bounded constraints established from the 95% credibility intervals of our biparental matrix time tree (Supplemental Table S3): a minimum of 8 MY and maximum of 16 MY for the split between the domestic cat and Panthera, and a maximum of 7 MY for the basal Panthera node. These secondary constraints were consistent with the fossil constraints of Johnson et al. (2006).

Data access

Mitochondrial sequence data generated for this study have been submitted to GenBank (http://www.ncbi.nlm.nih.gov/Genbank/) under accession numbers KP202253–KP202295. The Illumina genotype matrix is deposited in DRYAD (doi:10.5061/dryad.751cv). Whole genome Illumina sequencing reads from Felis silvestris, Felis margarita, Felis nigrigiris, and Prionailurus bengalensis are deposited in the NCBI Sequence Read Archive (SRA; http://www.ncbi.nlm.nih.gov/sra/) under accession numbers SRX1058146, SRX1058385, SRS1087700, and SRX1058464, respectively. Scripts used for sliding-window analysis can be found in the Supplemental Material.

Acknowledgments

This work was supported by grants from the Cat Health Network/Morris Animal Foundation (D12FE-502) and the National Science Foundation Assembling the Tree of Life program (EF0629849) awarded to W.J.M. E.E. is supported by CNPq/Brazil. We thank Leslie Lyons and the 99 Lives Project for providing access to the black-footed cat whole-genome sequence data. We thank Melody Roelke and Stephen J. O’Brien for access to felid DNA samples. We thank Mark Springer and Kris Helgen for comments on an earlier version of the manuscript. We also thank Nick Patterson and two anonymous reviewers for their critical and useful comments that improved the manuscript.

Author contributions: G.L. and W.J.M. conceived the study, designed the experiments, and wrote the manuscript, with input from all authors. G.L. and W.J.M. analyzed the data. B.W.D. prepared mitochondria and nuclear-enriched DNAs. E.E. contributed reagents and sequences.

References


Phylogenomic evidence for ancient hybridization in the genomes of living cats (Felidae)

Gang Li, Brian W. Davis, Eduardo Eizirik, et al.

*Genome Res.* 2016 26: 1-11 originally published online October 30, 2015
Access the most recent version at doi:10.1101/gr.186668.114

**Supplemental Material**

http://genome.cshlp.org/content/suppl/2015/10/30/gr.186668.114.DC1

**References**

This article cites 70 articles, 11 of which can be accessed free at:
http://genome.cshlp.org/content/26/1/1.full.html#ref-list-1

**Creative Commons License**

This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first six months after the full-issue publication date (see http://genome.cshlp.org/site/misc/terms.xhtml). After six months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at http://creativecommons.org/licenses/by-nc/4.0/.

**Email Alerting Service**

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or click here.

To subscribe to *Genome Research* go to:
http://genome.cshlp.org/subscriptions

© 2016 Li et al.; Published by Cold Spring Harbor Laboratory Press