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

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

Genome Res. published online October 30, 2015
Access the most recent version at doi:10.1101/gr.186668.114

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

Gang Li[1], Brian W. Davis[1,2#], Eduardo Eizirik[3], and William J. Murphy[1,2]

[1]Department of Veterinary Integrative Biosciences, Texas A&M University, College Station, TX 77843, USA.

[2]Interdisciplinary Program in Genetics, Texas A&M University, College Station, TX 77843, USA.

[3]Faculdade de Biociências, PUCRS, Porto Alegre, RS 90619-900, Brazil.

[#]Current address: National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892, USA.

Corresponding author:
William Murphy
Department of Veterinary Integrative Biosciences
4458 TAMU
Texas A&M University
College Station, TX 77843-4458
Tel: (979)458-0906
Fax: (979)845-9972
wmurphy@cvm.tamu.edu

Running Title: Cat speciation and hybridization
Keywords: cats, hybridization, speciation, introgression

Abstract

Interspecies 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 kilobases of nuclear sequence, in addition to complete mitochondrial genomes generated using light-coverage Illumina sequencing. We present the first robust felid timetree 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 coldspots, 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.

Introduction

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; Cahill et al. 2013, 2014; Cui et al. 2013; Martin et al. 2013; Kutchera et al. 2014; Sullivan et al. 2014; Toews and Brelsford 2012). 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 thirty-eight 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 sub-genomic sampling, several intergeneric and interspecific 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 interspecific 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 discoredance. 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 Table S1, Fig. S1 and 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 (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 (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-kilobase supermatrix of complete felid 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 cat lineage and more basal positions for the Caracal and Ocelot lineages (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).

Mito-nuclear 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

and 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 that the Bay Cat lineage Y chromosome tracks a unique history relative to the other genomic partitions.

One potential and underappreciated source of mito-nuclear 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 (Fig. S4, 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 (Figs. S5 and S6, Table S3-S4). Our timetree 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 (Johnson et al. 2006) (Fig. 1). 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 ($\geq$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 landbridges 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 phylogenomic patterns suggestive of interspecific 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&2B), which would not be expected when serial species divergences were separated by millions of years (Hobolth 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 non-species 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 (Fig. 2, Table 1).

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 (Fig. 2, Table 1). 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 (Fig. 2, Table 1). 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; Homyack 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 (Fig. S8), but also contained nuclear signatures of ancient hybridization with the Geoffroy's cat (Fig. 3A, Table 1), suggesting an extended history of admixture between the three species lineages. A Central American tigrina (Fig. 1, 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 (MY), which was further supported by deep mtDNA divergence (Fig. 3B). Interestingly, mitogenome phylogenies support a closer relationship between the fishing cat and Indochinese leopard cat populations than between the latter and supposed conspecifics, a pattern consistent with ancient hybridization when coupled with evidence of inter-species gene flow from SNP-based admixture tests (Table 1). This unique pattern of mito-nuclear discordance suggests capture of an unsampled or extinct *Prionailurus sp.* mitochondrion in the ancestral lineage of the fishing cat. As with the Ocelot lineage, further phylogeographic sampling of Asian leopard cats throughout

their broad distribution, along with their closely related congeners will further aid in delimiting new species boundaries more conclusively.

Within the Domestic Cat lineage, species/subspecies of the closely related *Felis silvestris* complex are known to hybridize in nature (e.g. Nussberger et al. 2013, Le Roux et al. 2015) and hybrids of the domestic cat and Jungle cat are the progenitors of an exotic cat breed known as the Chausie. Similarly, we find evidence for two episodes of ancient admixture within the Domestic Cat lineage derived from *D*-statistics inferred from the SNP matrix (Table 1). To demonstrate that our SNP-based *D*-statistics were not artifacts due to ascertainment bias driven by factors such as genotyping error or recurrent mutation, we generated ~20-30× Illumina whole genome sequence coverage for three of the non-*silvestris* members of the Domestic Cat lineage: Black-footed cat, Sand cat, and Jungle cat, and a close outgroup species, the Asian leopard cat. We then aligned these reads to the domestic cat v6.2 assembly, generated reference assemblies, and calculated *D*-statistics after stringent filtering of the SNV data set (mean=26.6 million SNVs per species, see Methods). Our results (Fig. 3C, Table S5) provide highly significant *D* and *Z*-scores between the same pairs of taxa observed with the SNP array-based statistics (Table 1).

As an additional validation of the SNP-based *D*-statistics, we remapped the domestic cat SNP probes to the new *Felis* genome sequences and recalculated the statistics from a reduced set of highest confidence orthologous SNPs, for which we required 99% pairwise identity between the domestic cat probe sequence and 100% identity to the SNP call in the reference assemblies of the other felids. The results (Table S6) demonstrated that even with a smaller, high confidence set of orthologous SNPs that match the sequence probe (and the SNP call) precisely, 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 jacknifed the outgroup taxon for the significant comparisons shown in Table 1 of the manuscript, and in nearly all cases the *Z*-scores remained significant (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 mito-nuclear discordance in the phylogenetic position of the snow leopard (Fig. 1, Fig. 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 (Figure 3, Table S8). 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 (Fig. 4A, Figure S9), similar to the mitogenome 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, Fig. S5B).

The distribution of the alternative topologies on *Panthera* X chromosomes was non-random, with more than half occupied by massive blocks of reduced divergence time. These blocks correspond to one ~45-Mb recombination coldspot and at least two smaller 5-10-Mb coldspots on the domestic cat X chromosome (Schmidt-Kuntzel et al. 2009), a genomic feature

virtually absent on autosomes (Fig. 4B). Intriguingly a large recombination coldspot, 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 chromosome (Murphy et al. 1999; Davis et al. 2009; Ma et al. 2010), which would preclude an inversion-based mechanism to explain this extremely large non-recombining region that is shared across divergent mammalian orders.

Recently, Ai et al. (2015) found similar patterns of introgression between high and low-altitude populations of pigs within the large X chromosome recombination coldspot, and suggested that these events were driven by an adaptive sweep. Positive selection is most effective in low-recombination regions (Smith and Haigh 1974), particularly on the X chromosome where both beneficial and deleterious recessive alleles are exposed in hemizygous males (Charlesworth et al. 1987). However, distinguishing positive from background selection can be difficult as both types of selection will shorten the time to the most recent common ancestor (O'Fallon 2013). We suggest that the marked internode depth and time reduction of (lion + snow leopard) phylogenies within but not outside of low recombination regions is also consistent with a similar selective sweep across this region (Fig. 4B).

Moreover, we propose that the reduced lion-snow leopard X chromosome divergence, together with the retention of a lion/leopard-related mitochondrion within the snow leopard genome (Fig. 1) can be best explained under a scenario of hybridization, 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 day 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 interspecies 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 mitochondrion 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 discordancies at deeper branches of the felid phylogeny when interpreted jointly with Bayesian

ancestral geographic reconstructions (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 (Figs S5A&B) as evidence for historical admixture and mitochondrial capture between ancestors of the Puma lineage and Lynx+Bay cat lineage within Asia, preceding their migration to North America in the Late Miocene (Fig. 1). Similarly, the Caracal clade origin is approximately 2 MY and multiple internodes younger in the mitogenome tree than in the biparental tree (Fig. 1A&B, Figs. S5A&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 more than 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 (Fig. S6). Introgression 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 non-hybrid 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 less divergent at the putative time of

hybridization than many living felids that currently produce viable and fertile interspecific 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 due to hybridization (Currat et al. 2008; Petit & Excoffier 2009; Cahill et at. 2013). Our observations of historical admixture between many pairs of cat species are consistent with evidence for contemporary interspecific 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 interspecific 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

Previous studies demonstrated that robust and reliable phylogenetic signal can be obtained by applying SNP arrays developed for domestic species to related organisms of the same genus, family or suborder, despite having diverged tens of millions of years from the array reference genome (Decker et al. 2009; McCue et al. 2012). 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 6 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 greater than 95%. Heterozygous sites (predominantly found in the domestic cat and its closest relatives the European wild cat, African wild cat, and Chinese desert cat, 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 non-domestic 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 non-domestic 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 (numts), 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 2-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 MODELTEST (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 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). 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 timetrees in Figures S5A and S5B 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) (Tables S3 and 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 (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 likelihood of the species tree topology).

To detect admixture over longer evolutionary timescales, 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 evobiR (http://www.uta.edu/karyodb/evobiR/index.html), which measure 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). 1,000 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, Steicher et al. 2014). We defined significant *D* statistics as those having *Z*-scores $\geq$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 (2x100-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 $\geq$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 (5-Mb block size) with 100 replicates re-sampled 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 4-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 timetree (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

Sequence data generated for this study has been submitted to GenBank (http://www.ncbi.nlm.nih.gov/Genbank) under accession numbers KP202255-202295. The Illumina genotype matrix is deposited in DRYAD (doi:10.5061/dryad.751cv). Whole genome Illumina sequencing reads from *Felis chaus*, *Felis margarita*, *Felis nigripes*, and *Prionailurus bengalensis* are deposited in the NCBI Sequence Read Archive (SRA, http://www.ncbi.nlm.nih.gov.sra) with accession numbers SRX1058146, SRX1058385, SRS1087700 and SRX1058464.

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 WJM. EE 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.D. prepared mitochondria and nuclear-enriched DNAs. E.E. contributed reagents and sequences.

Figure legends

Fig. 1. Discordant phylogenetic patterns across maternal, paternal and biparentally inherited subgenomic partitions. A. Comparison of nuclear genome and mitogenome phylogenies. The nuclear phylogeny/timetree excludes the Y chromosome partition. Lines between the two trees indicate alternative placement of particular species (black) or clades (red). Dashed lines represent poorly supported alternative placements. The timetree is based on the average divergence times across 8 individual *mcmctree* analyses (Table S3). Lineages are color-coded based on their current/historical distributions (see inset map). Dashed lines indicate hypothesized dispersal events out of Eurasia. Gray vertical bars indicate periods of extended low sea level (Haq et al. 1987) that may have facilitated dispersal between continents/islands. Asterisks indicate ML bootstrap support values of 100 in all analyses (all SNPs/binary-coded SNPs/SNP+gene supermatrix). **B.** Phylogenies showing relationships and bootstrap support values for eight felid lineages based on the mitogenome, X chr. (5,761-bp), Y chr. (5,352-bp), and autosomal (123,906-bp) SNP+gene supermatrix partitions. Bootstrap support for monophyly of each lineage is 100% (not shown). The Bay Cat clade is displayed in boldface to highlight the different topological positions based on different modes of inheritance.

Fig. 2. Signatures of Genome-Wide Phylogenetic Discordance A. Results of sliding window-based approximately-unbiased (AU) tests for each of the eight felid 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: ALC=Asian leopard cat, DOM=Domestic cat, LYN=Lynx, BAY=Bay cat, PUM=Puma, OCE=Ocelot, CAR=Caracal, 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 non-species tree topologies identified by the AU test that were significantly better than the species tree (Fig. 1A).

Fig. 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).

Fig. 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 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 is from the domestic cat linkage map (Schmidt-Kuntzel et al. 2009).

References

Ali SS, Yu Y, Pfosser M, Wetschnig W. 2012. Inferences of biogeographical histories within subfamily Hyacinthoideae using S-DIVA and Bayesian binary MCMC analysis implemented in RASP (Reconstruct Ancestral State in Phylogenies). *Annals of Botany* **109**: 95-107.

Ai H, Fang X, Yang B, Huang Z, Chen H, Mao L, Zhang F, Zhang L, Cui L, He W et al. 2015. Adaptation and possible ancient interspecies introgression in pigs identified by whole-genome sequencing. *Nat Genet* **47**: 217-225.

Allendorf FW, Leary RF, Spruell P, Wenburg JK. 2001. The problems with hybrids: setting conservation guidelines. *Trends Ecol Evol* **16**: 613-622.

Altschul SF, Madden TL, Schaffer AA, Zhang JH, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**: 3389-3402.

Antunes A, Pontius J, Ramos MJ, O'Brien SJ, Johnson WE. 2007. Mitochondrial introgressions into the nuclear genome of the domestic cat. *Journal of Heredity* **98**: 414-420.

Buckley-Beason VA, Johnson WE, Nash WG, Stanyon R, Menninger JC, Driscoll CA, Howard J, Bush M, Page JE, Roelke ME et al. 2006. Molecular evidence for species-level distinctions in clouded leopards. *Curr Biol* **16**: 2371-2376.

Cahill JA, Green RE, Fulton TL, Stiller M, Jay F, Ovsyanikov N, Salamzade R, St John J, Stirling I, Slatkin M et al. 2013. Genomic evidence for island population conversion resolves conflicting theories of polar bear evolution. *PLOS Genet* **9**: e1003345.

Cahill JA, Stirling I, Kistler L, Salamzade R, Ersmark E, Fulton TL, Stiller M, Green RE, Shapiro B. 2014. Genomic evidence of geographically widespread effect of gene flow from polar bears into brown bears. *Mol Ecol* **24**: 1205-1217.

Charlesworth B, Coyne JA, Barton NH. 1987. The relative rates of evolution of sex-chromosomes and autosomes. *Am Nat* **130**: 113-146.

Cho YS, Hu L, Hou HL, Lee H, Xu JH, Kwon S, Oh S, Kim HM, Jho S, Kim S et al. 2013. The tiger genome and comparative analysis with lion and snow leopard genomes. *Nature Communications* **4**: 2433.

Cui R, Schumer M, Kruesi K, Walter R, Andolfatto P, Rosenthal GG. 2013. Phylogenomics reveals extensive reticulate evolution in *Xiphophorus* fishes. *Evolution* **67**: 2166-2179

Currat M, Ruedi M, Petit RJ, Excoffier L. 2008. The hidden side of invasions: massive introgression by local genes. *Evolution* **62**: 1908-1920.

Davis BW, Raudsepp T, Pearks Wilkerson AJ, Agarwala R, Schäffer AA, Houck M, Chowdhary BP, Murphy WJ. 2009. A high-resolution cat radiation hybrid and integrated FISH mapping resource for phylogenomic studies across Felidae. *Genomics* **93**: 299-304.

Davis BW, Li G, Murphy WJ. 2010. Supermatrix and species tree methods resolve phylogenetic relationships within the big cats, *Panthera* (Carnivora: Felidae). *Mol Phylogen Evol* **56**: 64-76.

Davis BW, Seabury CM, Brashear W, Li G, Roelke-Parker M, Murphy WJ. 2015. Mechanisms underlying mammalian hybrid sterility in two feline interspecies models. *Mol Biol Evol* **32**: 2534-2546.

Decker JE, Pires JC, Conant GC, Mckay SD, Heaton MP, Chen KF, Cooper A, Vilkki J, Seabury CM, Caetano AR et al. 2009. Resolving the evolution of extant and extinct ruminants with high-throughput phylogenomics. *Proc Natl Acad Sci USA* **106**:18644-18649.

Durand EY, Patterson N, Reich D, Slatkin M. 2011. Testing for ancient admixture between closely related populations. *Mol Biol Evol* **28**: 2239-2252.

Eizirik E, Yuhki N, Johnson WE, Menotti-Raymond M, Hannah SS, O'Brien SJ. 2003. Molecular genetics and evolution of melanism in the cat family. *Curr Biol* **13**: 448-453.

Ellegren H, Smeds L, Burri R, Olason PI, Backström N, Kawakami T, Künstner A, Mäkinen H, Nadachowska-Brzyska K, Qvarnström A, et al. 2012. The genomic landscape of species divergence in Ficedula flycatchers. *Nature* **491**: 756-760.

Garrigan D, Kingan SB, Geneva AJ, Andolfatto P, Clark AG, Thornton KR, Presgraves DC. 2012. Genome sequencing reveals complex speciation in the *Drosophila simulans* clade. *Genome Res* **22**: 1499-1511.

Good JM, Hird S, Reid N, Demboski JR, Steppan SJ, Martin-Nims TR, Sullivan J. 2008. Ancient hybridization and mitochondrial capture between two species of chipmunks. *Mol. Ecol.* **17**: 1313-1327.

Gray AP. 1972. *Mammalian hybrids, a check-list and bibliography*. Commonwealth Agricultural Bureaux, England.

Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, Patterson N, Li H, Zhai W, Fritz MH et al. 2010. A draft sequence of the Neandertal genome. *Science* **328**: 710-722.

Haldane JBS. 1922. Sex ratio and unisexual sterility in hybrid animals. *Journal of Genetics* **12**: 101-109.

Haq BU, Hardenbol J, Vail PR. 1987. Chronology of fluctuating sea levels since the Triassic. *Science* **235**: 1156-1167.

Hillis DM, Bull JJ. 1993. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. *Syst Biol* **42**: 182-192.

Hobolth A, Christensen OF, Mailund T, Schierup MH. 2007. Genomic relationships and speciation times of human, chimpanzee, and gorilla inferred from a coalescent hidden Markov model. *PLoS Genet* **3**: e7.

Hobolth A, Dutheil JY, Hawks J, Schierup MH, Mailund T. 2011. Incomplete lineage sorting patterns among human, chimpanzee, and orangutan suggest recent orangutan speciation and widespread selection. *Genome Res* **21**: 349-56.

Homyack JA, Vashon JH, Libby C, Lindquist EL, Loch S, McAlpine DF, Pilgrim KL, Schwartz MK. 2008. Canada lynx-bobcat (*Lynx canadensis* x *L. rufus*) hybrids at the southern periphery of lynx range in Maine, Minnesota and New Brunswick. *Am Mid Nat* **159**: 504-508.

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* **311**: 73-77.

Johnson WE, Slattery JP, Eizirik E, Kim JH, Raymond MM, 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**: S79-S94.

Jones CS, Tegelstrom H, Latchman DS, Berry RJ. 1988. An improved rapid method for mitochondrial DNA isolation suitable for use in the study of closely related populations. *Biochem Genet* **26**: 83-88.

Katoh K, Toh H. 2010. Parallelization of the MAFFT multiple sequence alignment program. *Bioinformatics* **26**: 1899-1900.

Khan FAA, Phillips CD, Baker RJ. 2014. Timeframes of speciation, reticulation, and hybridization in the bulldog bat explained through phylogenetic analyses of all genetic transmission elements. *Syst Biol* **63**: 96-110.

Koen EL, Bowman J, Lalor JL, Wilson DE. 2014. Continental-scale assessment of the hybrid zone between bobcat and Canada lynx. *Biol Conserv* **178**: 107-115.

Koufos GD, Kostopoulos DS, Vlachou TD. 2005. Neogene/Quaternary mammalian migrations in Eastern Mediterranean. *Belgian J Zool* **135**: 181-190.

Kutschera VE, Bidon T, Hailer F, Rodi JL, Fain SR, Janke A. 2014. Bears in a forest of gene trees: phylogenetic inference is complicated by incomplete lineage sorting and gene flow. *Mol Biol Evol* **31**: 2004-2017.

Le Roux JJ, Foxcroft LC, Herbst M, MacFadyen S. 2015. Genetic analysis shows low levels of hybridization between African wildcats (*Felis silvestris lybica*) and domestic cats (*F.s.catus*) in South Africa. *Ecol Evol* **5**: 288-299.

Li G, Janecka JE, Murphy WJ. 2011. Accelerated evolution of *CES7*, a gene encoding a novel major urinary protein in the cat family. *Mol Biol Evol* **28**: 911-920.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**: 1754-1760.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Proc GPD. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078-2079.

Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y et al. 2012. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience* **1**: 18.

Luo SJ, Zhang Y, Johnson WE, Miao L, Martelli P, Antunes A, Smith JL, O'Brien SJ. 2014. Sympatric Asian felid phylogeography reveals a major Indochinese-Sundaic divergence. *Mol Ecol* **23**: 2072-2092.

Ma J, Iannuccelli N, Duan Y, Huang W, Guo B, Riquet J, Huang L, Milan D. 2010. Recombinational landscape of porcine X chromosome and individual variation in female meiotic recombination associated with haplotypes of Chinese pigs. *BMC Genomics* **11**: 159.

Martin SH, Dasmahapatra KK, Nadeau NJ, Salazar C, Walters JR, Simpson F, Blaxter M, Manica A, Mallet J, Jiggins CD. 2013. Genome-wide evidence for speciation with gene flow in Heliconius butterflies. *Genome Res* **23**: 1817-1828.

McCue ME, Bannasch DL, Petersen JL, Gurr J, Bailey E, Binns MM, Distl O, Guerin G, Hasegawa T, Hill EW et al. 2012. A high density SNP array for the domestic horse and

extant Perissodactyla: utility for association mapping, genetic diversity, and phylogeny studies. *PLoS Genet* **8**: e1002451.

Montague MJ, Li G, Gandolfi B, Khan R, Aken BL, Searle SM, Minx P, Hillier LW, Koboldt DC, Davis BW et al. 2014. Comparative analysis of the domestic cat genome reveals genetic signatures underlying feline biology and domestication. *Proc Natl Acad Sci USA* **111**: 17230-17235.

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**: 406.

Murphy WJ, Sun S, Pecon-Slattery J, Chen ZQ, O'Brien SJ. 1999. Extensive conservation of sex chromosome organization between cat and human revealed by parallel radiation hybrid mapping. *Genome Res* **9**: 1223-1230.

Nussberger B, Greminger MP, Grossen C, Keller LF, Wandeler P. 2013. Development of SNP markers identifying European wildcats, domestic cats, and their admixed progeny. *Mol Ecol Resour* **13**: 447-60.

O'Fallon B. 2013. Purifying selection causes widespread distortions of genealogical structure on the human X chromosome. *Genetics* **194**: 485-492.

Patterson N, Richter DJ, Gnerre S, Lander ES, Reich D. 2006. Genetic evidence for complex speciation of humans and chimpanzees. *Nature* **441**: 1103-1108.

Petit RJ, Excoffier L. 2009. Gene flow and species delimitation. *Trends Ecol Evol* **24**: 386-393.

Pontius JU, Mullikin JC, Smith DR, Lindblad-Toh K, Gnerre S, Clamp M, Chang J, Stephens R, Neelam B, Volfovsky N et al. 2007. Initial sequence and comparative analysis of the cat genome. *Genome Res* **17**: 1675-1689.

Posada D, Crandall KA. 1998. MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14**: 817-818.

Qiu ZX. 2003. Dispersals of Neogene carnivorans between Asia and North America. *Bull Am Mus Nat Hist* **279**: 18-31.

Roca AL, Georgiadis N, O'Brien SJ. 2005. Cytonuclear genomic dissociation in African elephant species. *Nat Genet* **37**: 96-100.

Schmidt-Kuntzel A, Nelson G, David VA, Schaffer AA, Eizirik E, Roelke ME, Kehler JS, Hannah SS, O'Brien SJ, Menotti-Raymond M. 2009. A domestic cat X chromosome linkage map and the sex-linked orange locus: mapping of orange, multiple origins and epistasis over nonagouti. *Genetics* **181**: 1411-1425.

Schwartz MK, Pilgrim KL, McKelvey KS, Lindquist EL, Claar JJ, Loch S, Ruggiero LF. 2004. Hybridization between Canada lynx and bobcats: Genetic results and management implications. *Conserv Genet* **5**: 349-355.

Shimodaira H, Hasegawa M. 2001. CONSEL: for assessing the confidence of phylogenetic tree selection. *Bioinformatics* **17**: 1246-1247.

Smith JM, Haigh J. 1974. The hitch-hiking effect of a favourable gene. *Genet Res* **89**: 391-403.

Stamatakis A. 2006. RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**: 2688-2690.

Sullivan J, Demboski JR, Bell KC, Hird S, Sarver B, Reid N, Good JM. 2014. Divergence with gene flow within the recent chipmunk radiation (*Tamias*). *Heredity* **113**: 185-194.

Sunquist M, Sunquist F. 2002. Wild cats of the World. Univ of Chicago Press, Chicago.

Toews DPL, Brelsford A. 2012. The biogeography of mitochondrial and nuclear discordance in animals. *Mol Ecol* **21**: 3907-3930.

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**: 4317-4333.

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**: 2528-2533.

Trigo TC, Tirelli FP, 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**: e108469.

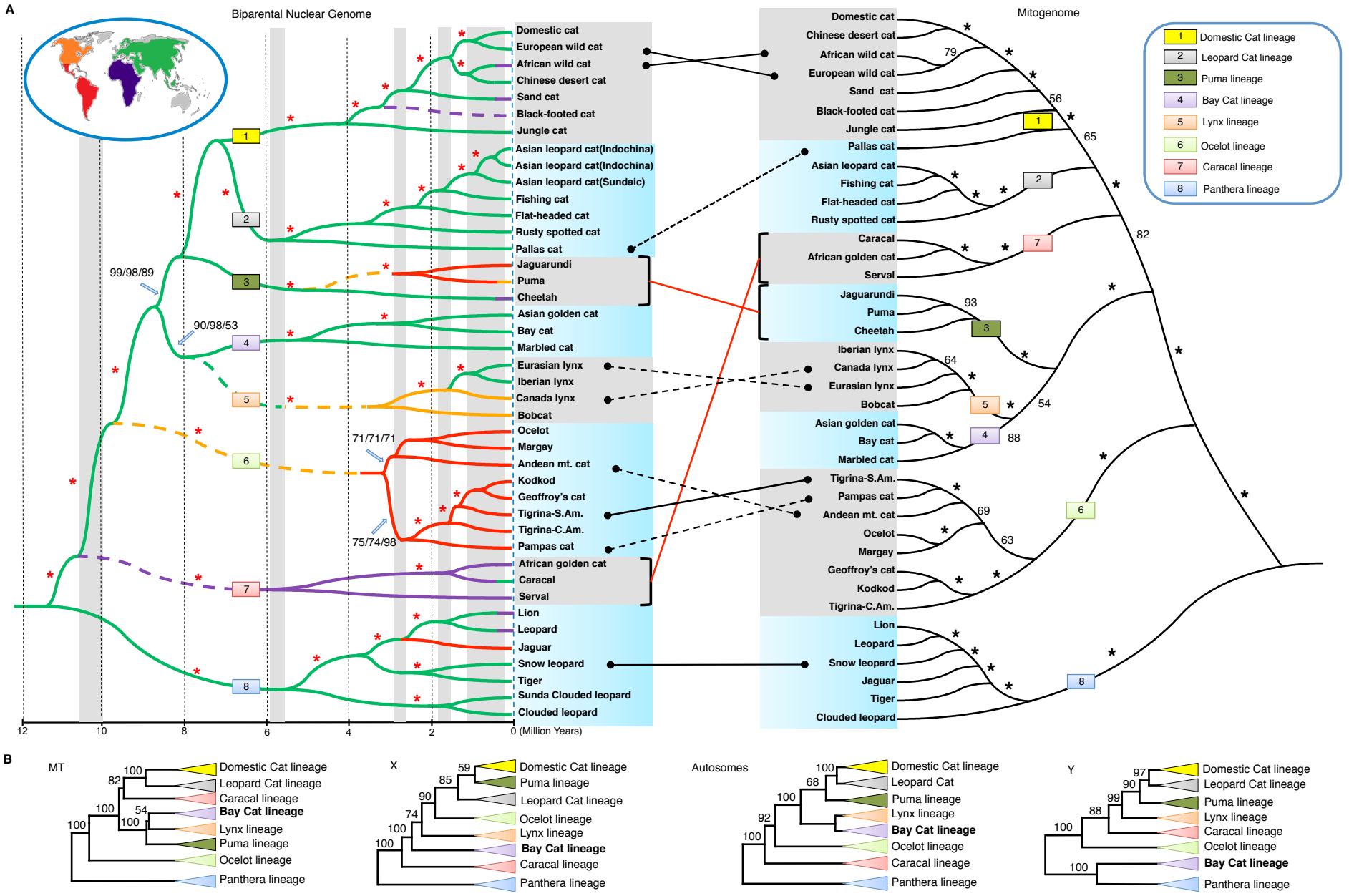
Wurster-Hill DH, Centerwall WR. 1982. The interrelationships of chromosome-banding patterns in canids, mustelids, hyena, and felids. *Cytogenet Cell Genet* **34**: 178-192.

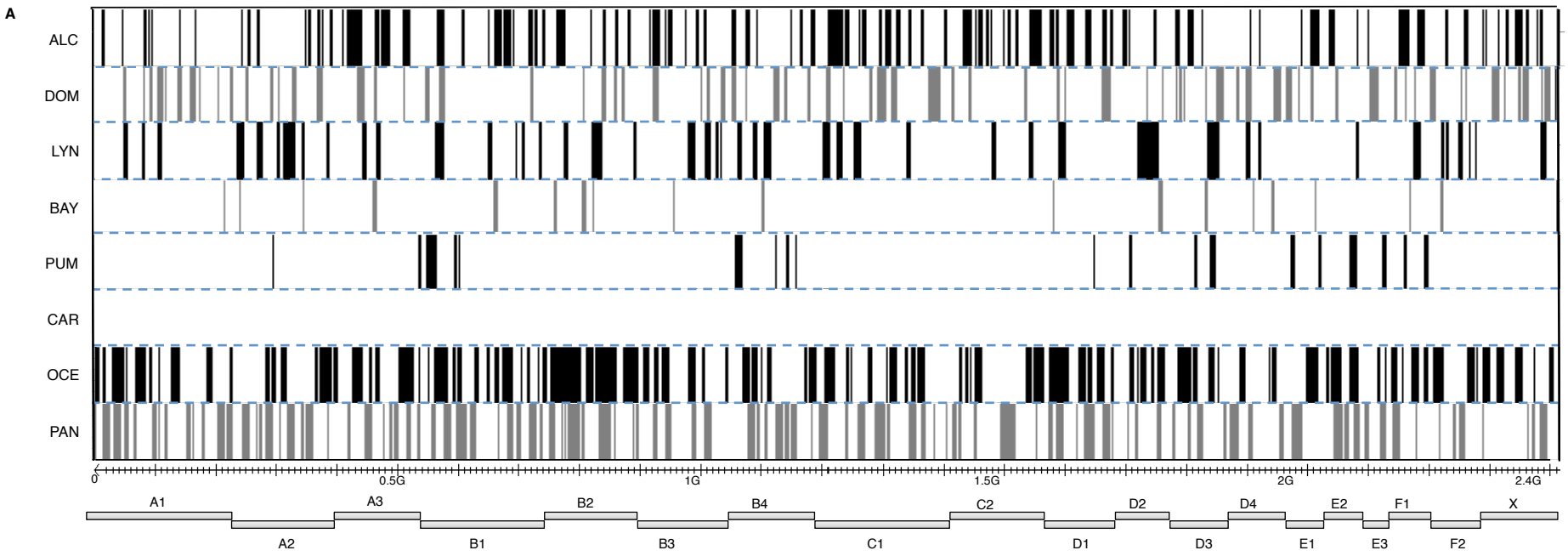
Yang ZH. 2007. PAML 4: Phylogenetic analysis by maximum likelihood. *Mol Biol Evol* **24**: 1586-1591.

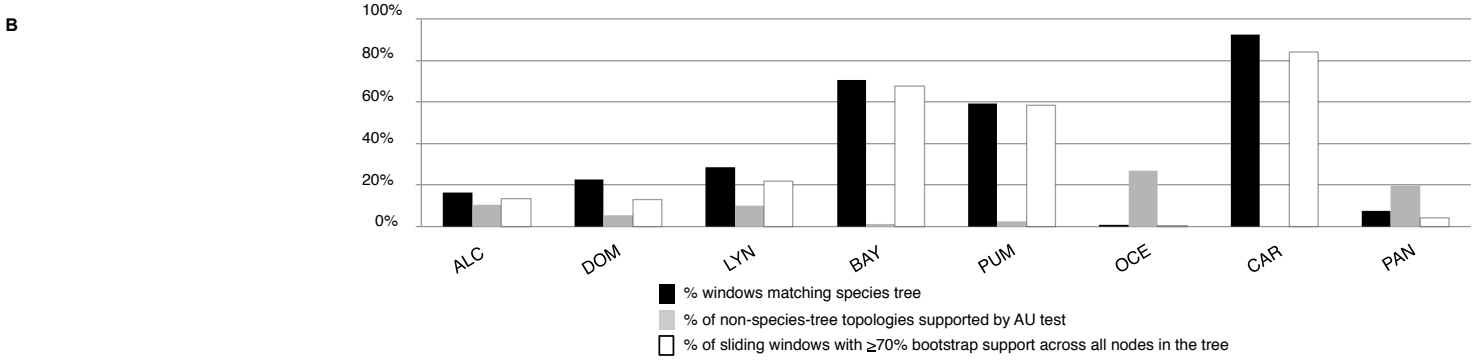
Table 1. Measures of phylogenetic discordance and admixture within felid lineages.

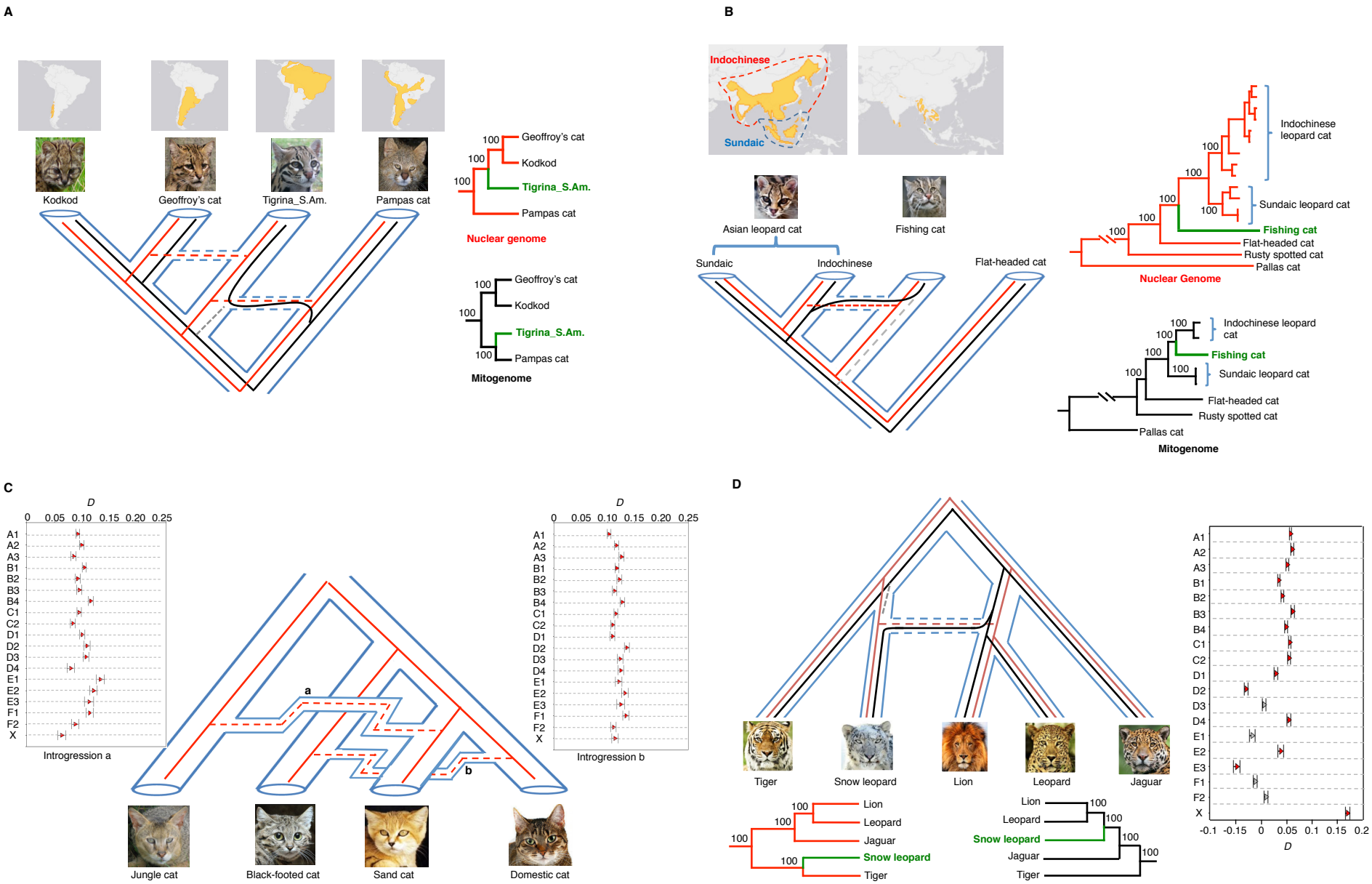
Lineage	% of windows matching species tree	% of non-species-tree topologies supported by likelihood ratio tests	Species Tree (P1,P2) P3	Outgroup	ABBA/BABA Results				
					D	*Z*	*P*	ABBA	BABA
Ocelot	1.5	26.0	(Kodkod, Tigrina_S.Am.), Pampas cat [a]	Serval	0.102 ± 0.057	1.782	0.0748	173	141
			(Kodkod, Geoffroy's cat), Tigrina (S.Am.)[a]	**Serval**	**0.146 ± 0.048**	**3.064**	**0.0022**	**232**	**173**
			(Ocelot, Margay), Pampas cat	Serval	0.063 ± 0.063	1.466	0.1427	271	239
			(Ocelot, Margay), Tigrina (S.Am.)	Serval	0.029 ± 0.045	0.079	0.9368	293	295
Panthera	14	14.3	(Leopard, Lion), Jaguar	Sunda Clouded leopard	0.013 ± 0.043	0.310	0.7565	264	257
			(Lion, Leopard), Tiger	Sunda Clouded leopard	0.031 ± 0.048	0.652	0.5144	215	202
			(Snow leopard, Tiger), Leopard	Sunda Clouded leopard	0.056 ± 0.042	1.345	0.1788	291	260
			(Tiger, Snow leopard), Lion[a,b]	**Domestic cat**	**0.041 ± 0.001**	**33.39**	**0**	**357,875**	**329,510**
Lynx	37.3	8.7	**(Iberian lynx, Canada lynx), Bobcat [a]**	**Asian Leopard cat**	**0.176 ± 0.056**	**3.182**	**0.0015**	**204**	**143**
			(Iberian lynx, Eurasian lynx), Canada lynx	**Asian Leopard cat**	**0.199 ± 0.047**	**6.112**	**1.9e-05**	**274**	**183**
Leopard Cat	26.6	7.6	**(Sunda ALC, Indochinese ALC), Fishing cat [a]**	**Pallas cat**	**0.179 ± 0.047**	**3.847**	**0.0001**	**277**	**193**
Domestic Cat	33.1	4.4	**(Sand cat, Domestic cat), Black-footed cat**	**Jungle cat**	**0.152 ± 0.023**	**6.630**	**3.4e-11**	**1,033**	**761**
			(Black-footed cat, Sand cat), Jungle cat	**Bobcat**	**0.090 ± 0.028**	**3.223**	**0.0013**	**690**	**577**
			(Domestic cat, Sand cat), Jungle cat	**Bobcat**	**0.186 ± 0.024**	**7.626**	**2.4e-14**	**926**	**636**
Puma	65.7	1.5	(Jaguarundi, Puma), Cheetah	Serval	0.024 ± 0.044	0.546	0.5850	257	245
Bay Cat	80.4	1.0	(Bay cat, Asian golden cat), Marbled cat	Tiger	0.139 ± 0.042	2.976	0.0029	312	243
Caracal	98.0	0.0	(African golden cat, Caracal), Serval	Asian golden cat	0.058 ± 0.069	0.842	0.3997	109	97

[a]Species comparisons with documented mitochondrial introgression/hybridization.
[b]Whole genome SNV alignment results
Boldface indicates statistically significant *D*-statistics ($Z \geq 3.0$) with SNP and/or SNV data.

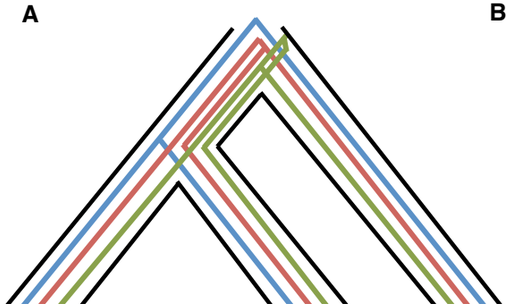

A
Biparental Nuclear Genome
Mitogenome
Domestic cat
European wild cat
African wild cat
Chinese desert cat
Sand cat
Black-footed cat
Jungle cat
Asian leopard cat(Indochina)
Asian leopard cat(Indochina)
Asian leopard cat(Sundaic)
Fishing cat
Flat-headed cat
Rusty spotted cat
Pallas cat
Jaguarundi
Puma
Cheetah
Asian golden cat
Bay cat
Marbled cat
Eurasian lynx
Iberian lynx
Canada lynx
Bobcat
Ocelot
Margay
Andean mt. cat
Kodkod
Geoffroy's cat
Tigrina-S.Am.
Tigrina-C.Am.
Pampas cat
African golden cat
Caracal
Serval
Lion
Leopard
Jaguar
Snow leopard
Tiger
Sunda Clouded leopard
Clouded leopard
99/98/89
90/98/53
71/71/71
75/74/98
12
10
8
6
4
2
0 (Million Years)
1 Domestic Cat lineage
2 Leopard Cat lineage
3 Puma lineage
4 Bay Cat lineage
5 Lynx lineage
6 Ocelot lineage
7 Caracal lineage
8 Panthera lineage
B
MT
Domestic Cat lineage
Leopard Cat lineage
Caracal lineage
Bay Cat lineage
Lynx lineage
Puma lineage
Ocelot lineage
Panthera lineage
X
Domestic Cat lineage
Puma lineage
Leopard Cat lineage
Ocelot lineage
Lynx lineage
Bay Cat lineage
Caracal lineage
Panthera lineage
Autosomes
Domestic Cat lineage
Leopard Cat
Puma lineage
Lynx lineage
Bay Cat lineage
Ocelot lineage
Caracal lineage
Panthera lineage
Y
Domestic Cat lineage
Leopard Cat lineage
Puma lineage
Lynx lineage
Caracal lineage
Ocelot lineage
Bay Cat lineage
Panthera lineage


A
ALC
DOM
LYN
BAY
PUM
CAR
OCE
PAN
0
0.5G
1G
1.5G
2G
2.4G
A1
A3
B2
B4
C2
D2
D4
E2
F1
X
A2
B1
B3
C1
D1
D3
E1
E3
F2

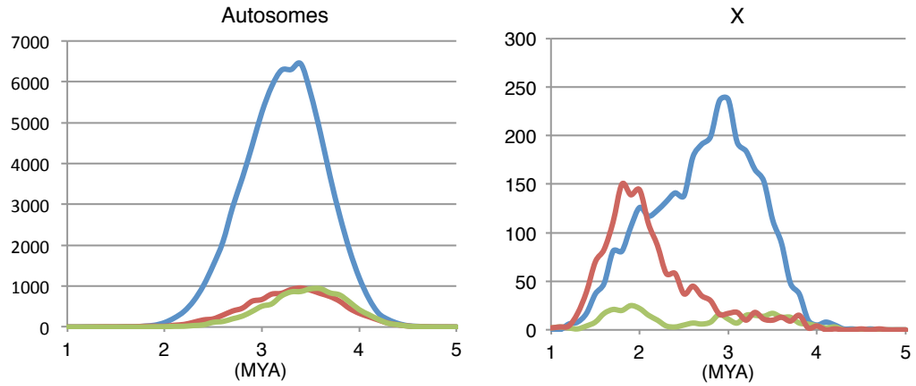

B
100%
80%
60%
40%
20%
0%
ALC
DOM
LYN
BAY
PUM
OCE
CAR
PAN
% windows matching species tree
% of non-species-tree topologies supported by AU test
% of sliding windows with ≥70% bootstrap support across all nodes in the tree


A
Kodkod
Geoffroy's cat
Tigrina_S.Am.
Pampas cat
100
Geoffroy's cat
Kodkod
Tigrina_S.Am.
Pampas cat
Nuclear genome
Mitogenome
B
Indochinese
Sundaic
Asian leopard cat
Fishing cat
Flat-headed cat
Indochinese leopard cat
Sundaic leopard cat
Rusty spotted cat
Pallas cat
Nuclear Genome
Mitogenome
C
D
0 0.05 0.10 0.15 0.20 0.25
A1 A2 A3 B1 B2 B3 B4 C1 C2 D1 D2 D3 D4 E1 E2 E3 F1 F2 X
Introgression a
Introgression b
a
b
Jungle cat
Black-footed cat
Sand cat
Domestic cat
D
Tiger
Snow leopard
Lion
Leopard
Jaguar
Nuclear Genome
Mitogenome
-0.1 -0.15 0 0.05 0.1 0.15 0.2
D

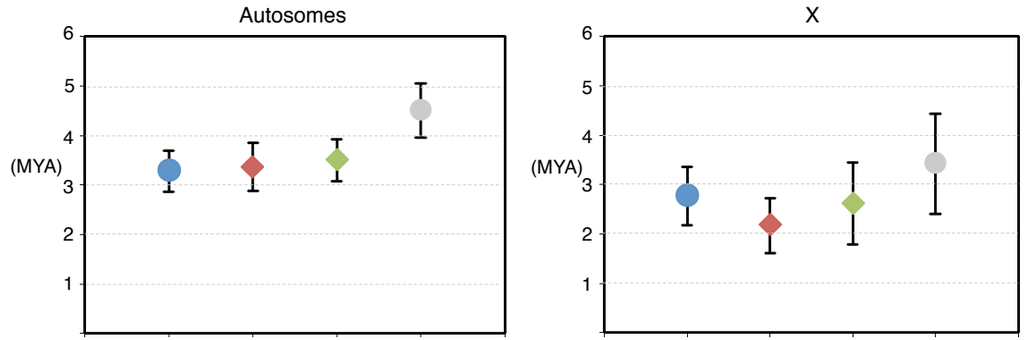
A


Tiger
Snow leopard
Lion

B

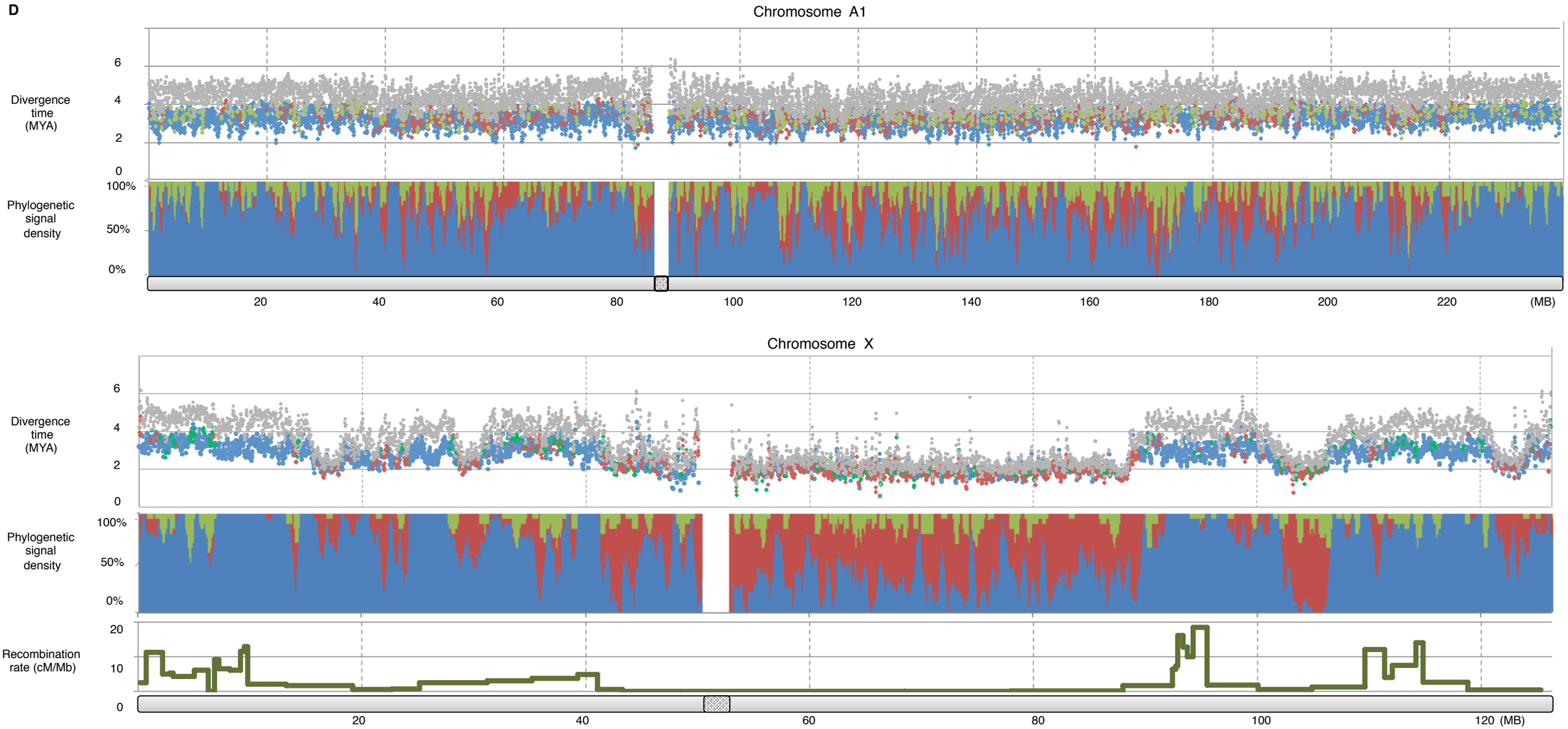

Autosomes
7000
6000
5000
4000
3000
2000
1000
0
1
2
3
4
5
(MYA)
X
300
250
200
150
100
50
0
1
2
3
4
5
(MYA)

C


Autosomes
6
5
4
3
2
1
(MYA)
X
6
5
4
3
2
1
(MYA)


Outgroup(Lion, Tiger, Snow leopard)
Lion(Tiger, Snow leopard)
Tiger(Lion, Snow leopard)
Snow leopard (Lion, Tiger)

D


Chromosome A1
Divergence time (MYA)
6
4
2
0
Phylogenetic signal density
100%
50%
0%
20
40
60
80
100
120
140
160
180
200
220
(MB)
Chromosome X
Divergence time (MYA)
6
4
2
0
Phylogenetic signal density
100%
50%
0%
Recombination rate (cM/Mb)
20
10
0
20
40
60
80
100
120 (MB)