State of cat genomics

Stephen J. O'Brien[1], Warren Johnson[1], Carlos Driscoll[1], Joan Pontius[2], Jill Pecon-Slattery[1] and Marilyn Menotti-Raymond[1]

[1] Laboratory of Genomic Diversity, National Cancer Institute, Frederick, MD 21702, USA
[2] Laboratory of Genomic Diversity, Basic Research Program, SAIC-Frederick, Inc., NCI-Frederick, Frederick, MD 21702, USA

Our knowledge of cat family biology was recently expanded to include a genomics perspective with the completion of a draft whole genome sequence of an Abyssinian cat. The utility of the new genome information has been demonstrated by applications ranging from disease gene discovery and comparative genomics to species conservation. Patterns of genomic organization among cats and inbred domestic cat breeds have illuminated our view of domestication, revealing linkage disequilibrium tracks consequent of breed formation, defining chromosome exchanges that punctuated major lineages of mammals and suggesting ancestral continental migration events that led to 37 modern species of Felidae. We review these recent advances here. As the genome resources develop, the cat is poised to make a major contribution to many areas in genetics and biology.

Introduction

Cats and their wild progenitors have existed for a long time – fossil records suggest they first appeared ∼35 million years ago (MYA). Although the great saber-tooth tigers of those times have long been extinct, modern cats are descended from a midsize cat that existed ∼10–11 MYA [1–3]. Humans seem to be fascinated with the feline species – a rich literature and artistic imagery of lions, tigers, cheetahs and leopards testifies to this – and domestic cats rank among the world's most venerated companion animals.

Our affection toward cats, complemented by accounts of cats' behavior and biology, has resulted in their contribution to a diverse range of disciplines (ranging from disease to conservation studies) that are suitable for genetic and genomic enquiry. For example, cats, like dogs, enjoy extensive veterinary medical surveillance (second only to human medicine) and have ∼250 genetic diseases analogous to human disorders [4–8]. Feline infectious agents offer powerful natural models to human diseases including HIV-AIDS [feline immunodeficiency virus (FIV)], SARS (feline coronavirus-FCoV), avian influenza, neurotropic viruses [canine distemper virus (CDV)] and cancers [feline leukemia virus-feline sarcoma virus (FeLV/FeSV)] [5–16]. Cats are a domesticated representative of the Felidae family, which includes some of the most successful, but now the most threatened, predator species on earth [1,2,17].

With a rapidly growing literature on cat behavior, coat colors, reproductive advances and breed development, the domestic cat is poised to benefit from the recently completed draft whole genome sequence of an Abyssinian cat named Cinnamon [8,18]. Cats present a powerful model in development, physiology and neuroscience (some of which are listed in Table 1). Here we review the insights achieved from the cat genome sequence and discuss the wealth of knowledge about genomic organization that is being shown through comparative genomics. Furthermore, we trace the natural history of the wild cats, discuss the origins of cat domestication and review the many aspects of cats that make them a useful animal 'model' for human diseases and for comparative genomics.

Developing and applying a whole genome sequence for cats

The study of feline genetics had its roots in early gene-mapping strategies using linkage and physical mapping approaches [19–22]. Over the past few decades, several important biological and informatics resources required for genetic investigations in cats were developed. The resources are comparable to those developed in other genome projects and include the following: (i) somatic cell and radiation hybrid panels used to build a framework physical map of comparative anchor tagged sequence (C-ATS) markers for cats [19,23–28]; (ii) three large cat pedigrees used to build a cat linkage map [29–32]; (iii) several genomic libraries (BAC, PAC, cosmid and fosmid) for gene discovery and genome assembly validation [33–35] and (iv) finished DNA sequence of three subregions of the cat genome [i.e. FLA, the major histocompatibility complex (3.3 Mbp), mitochondrial DNA and counterpart nuclear mitochondrial DNA (mtDNA; numt) sequence; and 30 Mbp of the ENCODE Project cat genome sequence (http://www.genome.gov/10005107, [33–40]). The application of these important tools led to significant advances in cat genetics including gene discovery and regional gene annotation and to anticipation for a whole genome sequence.

In 2005, the cat was included in a selection of 24 mammals for whole genome sequencing by National Human Genome Research Institute, to facilitate interpretation of the finished human genome sequence. The species were chosen to capture the evolutionary divergence across the 4500 living species of mammals initially with a 'light' (i.e. twofold) genome coverage [18,41]. The principal criteria for selecting these species were (i) to discover short conserved sequence regions among mammals that would include gene regulatory motifs; (ii) to provide a platform for reconstruction of ancestral genomes that preceded the divergence nodes of the mammalian radiations and (iii) to stimulate the development and application of new

Table 1. The cat as a useful model species

Application	Evidence
Medical models of human diseases	There is a rich literature of veterinary disease in cats; >250 hereditary diseases in cats are analogous to human genetic diseases. Eighteen of these have a known gene mutation in cats suitable for pathogenesis and therapeutic study [4–8,32].
Infectious agents	Cats have well-described models for several deadly human viral diseases, notably HIV-AIDS-feline immunodeficiency virus, which is endemic in 14 free ranging species of Felidae including domestic cats [11]. Cats also harbor feline leukemia virus and feline sarcoma virus, which have laid groundwork for oncogene discoveries and a virulent model for SARS, feline coronavirus, and many others [10–16].
Neuroscience and physiology	Cat are traditional subjects for neurophysiology studies that have led to important insight on ocular and neural physiological processes [43,45,97–102].
Behaviors	Cats have many curious behaviors associated with nurturing, defense, allusiveness and tameness that seem to have genetic influences. Social organization in lions and in domestic cats seems to involve co-adaptation of behavior and reproductive strategies [5,88].
Reproduction	The reproductive physiology of cats is relatively advanced; for example, artificial insemination, *in vitro* fertilization, embryo transfer, chineric embryos and nuclear cloning have been performed in cats. Promising advances have occurred in embryonic stem cell isolation, leading to hopes of gene knockouts and gene replacement in laboratory settings.
Domestication	The process of cat domestication represents one of the more fascinating natural experiments. It seems to have happened in a single locale in the Fertile Crescent when humans began agriculture. Cat domestication has led to the growth of 600 million domestic cats worldwide. The adaptive aspects of this process in genetic terms await resolution [42].
Breed development	Approximately 68 certified cat breeds current exist; most are younger than a few hundred years. Each was selected artificially for appearances and behavior, raising the question of what was selected and how well [5,88].
Coat-color variation	The cat breeds are fixed in different combinations for some ~12 coat color genes that have been described and tracked in pedigrees. Resolution of their genetic bases will lead not only to more precise breed improvement but also to a better understanding of pigmentation, hair development and ocular albinism. We know the basis for ten coat color hair length genes, but others (such as *Orange*) remain elusive [31,46–51,88].
Forensic development	Cats have led the way in establishing the legal precedent for introduction of short tandem repeat–based cat individual identification of hairs and other biospecimens found at crime scenes. Cats hairs are easily picked up, so suspects with cats or dogs can implicate themselves through their pets [91,103,104].
Felidae evolution, adaption and natural history	The cat family Felidae has shown much in the application of sophisticated tools of molecular evolution to phylogeny reconstruction. Multidisciplinary interpretation of felid phylogeny, geography, paleontology and geology has allowed inference around their historic continental migrations and origins [1,2,17].
Comparative genomics	The highly conserved synteny of the cat genome map with that of human, dolphin and other mammal species has given us a glimpse of the ancestral genome organization of all mammals. Recent sequence analyses of cat and other mammal genomes are beginning to reveal a view of the pattern of genome organization that has punctuated the mammalian radiations [8,22,60].
Felidae conservation	Humans' fascination with the cat species has produced a plethora of ecological and behavioral descriptions of the plight of the many endangered Felidae species. Genetics in conservation became widely accepted with the finding of the cheetahs' genetic uniformity and progressed to studies on leopards, pumas, tigers, wildcats, lynxes and other free-ranging species [17].

animal models for human medicine and biology. In 2006, the whole genome sequence of a female Abyssinian cat named Cinnamon was determined by Agencourt Bioscience (http://www.agencourt.com); Abyssinian cats are the most inbred, making genome assembly easier. A consortium of scientists assembled, mapped and annotated the cat genome sequence using a comparative approach that involved cross-reference to NCBI-annotated genome assembles of six index mammals (human, chimpanzee, mouse, rat, dog and cow) [8]. For details of the annotation strategy and highlights of genomic features that were discovered, see Box 1.

About 327 037 single nucleotide polymorphism (SNP) variants were heterozygous in Cinnamon's genome, and a sampling of 200 SNPs validated ~85% of them as being variable in additional domestic cats from various breeds. Cinnamon's genomic history was molded by three episodes of historic inbreeding; the original domestication event for cats (~10 000 years ago), Abyssinian breed formation and disease (rdAc) pedigree establishment [42,43]. Her genome is a patchwork of short regions of SNP homozygosity interspersed with regions of heterozygosity (more than one SNP per 600 bp), perhaps a reflection of historic inbreeding. In total, there are 275 alternating homozygous/heterozygous

segments in Cinnamon's genome, with 57% of her genome being largely homozygous. Similar alternating homozygous genome segments occur in dogs, likely for the same reason [44].

Indeed, the potential of cat phenotypic and medical surveillance in the genomic era is beginning to be realized. Currently there are 38 cat genes for which mutational variants responsible for feline metabolic diseases or for morphological phenotypes are described [8]. This includes 18 hereditary diseases, where the mutation in cat occurs in the homologous gene to the analogous human syndrome. The single exception, spinal muscular atrophy (SMA), a leading cause of human infant mortality, involved a large 140-kb deletion of the *LIX1* gene in the cat model [45]. The *LIX1* gene is located on chromosome A1 within 25 cm (Mbp) of the *SMN1* gene, whose human homolog carries causal mutations in 97% of human SMA patients. In addition, the resolution of the molecular genetic bases of ten coat color (or hair length) variants now establishes cats as important species for study of pigmentation, hair development and ocular albinism [8,21,31,46–51].

Recently, the NHGRI Committee for the Annotation of the Human Genome promoted the cat to an additional sixfold genome coverage, due to be completed in late

Box 1. Annotation of the cat whole genome sequence

Here we give an overview of the strategy and main conclusions from the cat genome sequence study; full details can be found in Ref. [8]. The 'light' (1.9-fold coverage) of Cinnamon was assembled, mapped and annotated using a comparative approach to finished whole genomes sequence/maps of six mammals (human, chimpanzee, mouse, rat, dog and cow). Briefly, 8 027 672 sequence reads (84% plasmid and 16% fosmid) were assembled to 817 956 overlapping contigs and 217 790 scaffolds using the ARACHNE and PHUSION genome assembly algorithms [93–95]. The contigs were aligned to the index mammals' sequence maps and anchored with 1682 ordered markers from the cat radiation–hybrid map, assuming that the order of reciprocal best match (RBM) sequences between the radiation hybrid (RH) anchor markers for cats was the same as for a their dog and human counterparts. The cat genome sequence is accessible on a web-based genome sequence browser, GARFIELD http://lgd.abcc.ncifcrf.gov [8,96], which includes the following annotated features:

o 817 956 contigs and 217 790 scaffolds across 2.7 Gbp of cat genome sequence
o definition and mapping of >1 million RBM alignments between cat and six index mammals, revealing a set of 133 499 conserved sequence blocks (CSBs) present in all seven genomes. CSBs and RBMs were used to annotate cat genes and homologous synterny blocks (HSBs)
o identification of 20 285 cat genes based on alignment and conserved syntenic orthology with orthologous genes in index mammals
o sequence, iteration and mapping of cat repeat sequence families
o detection and mapping of 201 microRNA loci
o definition of 2814 HSBs across six index mammal genomes
o detection of previously undiscovered nuclear mitochondrial DNA (*numt*) sequences distributed across cat chromosomes
o description of undiscovered endogenous retroviral sequences and phylogenetic lineages, ten times more abundant than previously known FeLV and RD114 endogenous retroviral sequences
o detection of 327 037 single nucleotide polymorphisms, 34 850 deletion insertion polymorphisms and 200 177 short tandem repeat (or microsatellite) loci

2008 [18]. The planned supplement of SNP discovery using 'light' resequence coverage of 30 cats from ten cat breeds assures a large store of SNPs and short tandem repeats (STRs) suitable for gene association discovery using SNP genotyping arrays. The decay of linkage disequilibrium (LD) already measured among 24 cat breeds predicts that a SNP genotypic array including ∼45 000 evenly placed SNPs is appropriate for gene association discovery in cats, similar to the approach successfully applied to dogs [8,44,52,53]. However, cats have shorter LD stretches than dogs (because they were domesticated more recently than dogs; see below) and require three times more SNPs for gene association studies. We anticipate that these efforts will lead to the discovery of 2 million new SNPs, which will aid the development of an eagerly awaited genotype array chip of 80 000–100 000 SNPs, suitable for genome-wide association studies. The dog genome sequence led to scores of new gene discoveries [44,52,53]; the availability of these promising cat genome resources portends that cat advances should be close behind.

Comparative genome organization in mammals

Cat genomics has benefited from and also played a key role in the emerging field of comparative genomics. The earliest comparisons of gene order in mammalian chromosomes revealed long stretches of conserved synteny between human chromosomes and homologous chromosome segments in developing gene maps of mouse, cat and cow [9,19,54,55]. Conserved syntenic region homologies between mammal species were confirmed by precise G-banding pattern identities across chromosome arms and in some cases over entire chromosomes [20,56,57]. These large segments of chromosome homology were explicitly defined by florescent *in situ* hybridization (FISH) or chromosome painting, whereby metaphase chromosomes of various mammals were decorated by flow-sorted and fluorescent-labeled whole chromosome probes [9,22,58,59].

Chromosome painting studies gave us the first look at the details of genome rearrangements that characterized the divergence of mammals. With the caveat that intrachromosomal rearrangements (inversions) were invisible by these methods, a paradigm emerged that chromosomal exchange followed a dichotomous mode in different mammal groups [22,58]. In most lineages, a low or 'default' rate was observed where species groups displayed few translocation exchanges, on the order of one new rearrangement every 10–15 MY (exemplified by cat, human, dolphin, pig and other species). These default (slow or conserved) lineages were in sharp contrast to lineages whose chromosomes were reshuffled extensively (examples of reshuffled lineages were gibbons, new world monkeys, murid rodents, canids and ursids) [22,56–58]. The global reorganization in reshuffled lineages was three to five times more extensive than was seen in the slower 'default' lineages for inexplicable reasons.

Assembled whole genome sequences, including the cat genome, have allowed the detection of intrachromosomal inversions, because ∼100 000 conserved sequence block (CSB) markers of homology between species (see Box 1) enabled an extremely high-density genome resolution of conserved syntenic segments [8,60,61]. Using sophisticated breakpoint parsimony algorithms [genome rearrangements in man and mouse (GRIMM) synteny and multiple genome rearrangement algorithm (MGR)] [62–64] that resolved the breakpoint coordinates joining 500 homologues synteny blocks [HSBs; previously called smallest combined evolutionary unit segments (SCEUS); Figure 1], the paradigm of dichotomous rates of genome exchange was unseated [8,26]. The wrinkle came from the occurrence of numerous intrachromosomal rearrangements but at different tempos in the slow default translocation lineages versus the reshuffled lineages (Figure 1). A preliminary analysis of six species genome sequences showed that lineages with few translocation exchanges display considerably more intrachromosomal inversions, whereas species with reshuffled and increased translocations showed fewer inversions (Table 2). The total number of translocations plus inversions is similar in all lineages, because in the slow 'default' lineages (human and cat), inversions are increased, whereas in the shuffled translocation-rich lineages (murids and dogs), inversions are infrequent. It almost seems that interchromosomal and intrachromosomal exchanges are compensating in different lineages.

Additional fascinating observations have emerged from comparing available mammal species genomes [60]. First, the overall rates of chromosome exchanges within
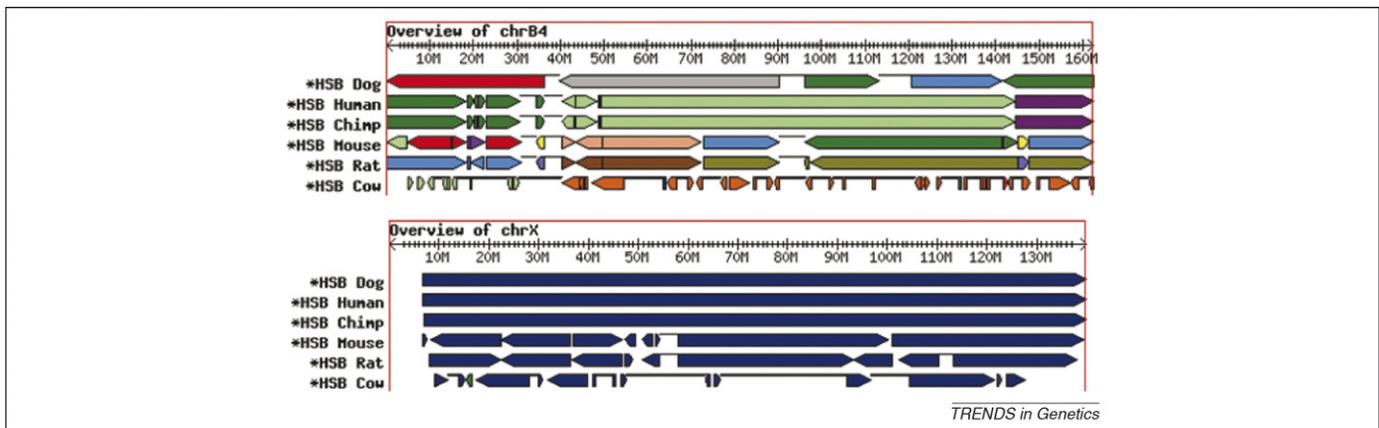
Figure 1. Homologous synteny blocks (HSBs) relative to cat chromosomes B4 and X defined by conserved sequence blocks across six index mammalian species' genome sequences [8,26,60]. HSBs reflect the chromosome segments that are retained across divergent mammals, allowing one to reconstruct the chromosome exchanges and breakpoint that punctuate genome organization throughout the mammalian radiation (see main text and Table 2 for more details).

mammalian orders have increased (more than tenfold) since the Cretaceous–Tertiary boundary (∼65 MYA), coincident with the advancement and diversification of most mammalian orders at the time. Second, ∼20% of chromosome breakpoint regions defined by comparative alignments were not only reused more than once during mammalian evolution, but the reused sites are clustered nonrandomly around centromeres. Third, chromosomal breakpoints defined between mammals show a marked enrichment in gene density compared with the genome-wide average; they seem to cluster in gene-rich areas. Fourth, primate-defined breakpoints (translocation and inversion) occur in regions of human segmental duplications, suggesting that such local aneuploidy plays a part in chromosomal exchange. Finally, evolutionary breakpoints in mammal lineages seem to coincide with breakpoints that occur in human cancers, suggestive of a common underlying mechanism.

These conclusions emerged from comparisons of species that have a whole genome sequence and a dense radiation hybrid (RH) map. The inferences must be affirmed by future comparisons, particularly with the additional 24 species selected for whole genome sequence [18]. The puzzle is not completely solved, but the prospect of richer whole genome sequence data across mammals in the future holds high anticipation for both resolving the effective chromosomal dynamics and also for inferring the genome organization of the ancestral bears, felids, at other mammal groups in high definition.

Genomic natural history of the world's cats

Nestled in the genome of every individual are cryptic footprints of historic episodes encountered by their ancestral progenitors. The emerging field of genomic archeology (or genomic prospecting) aims to mine the ancient DNA sequence to enable better interpretation of the divergence, migration and demographic perturbations that occurred in the silence of prehistory [65–68]. The cat family Felidae – an assemblage of 37 successful predatory carnivore species on five continents (Figure 2) – provides a cogent example of how genomic data can complement geographic, paleontological and geological aspects to characterize a species group's natural history [1,2] (Figure 3).

The first step in reconstructing cat origins was to establish a robust molecular phylogeny, which informed the hierarchical relationship and divergence nodes that define the Felidae radiation. This was accomplished by selecting 35 cat genes (a total of 22 789 bp chosen from autosomal, mitochondrial, X- and Y-linked loci) that were sequenced in all cat species [1]. Available tree-building algorithms converged on a phylogenetic tree on which the nodes could be dated with a dozen well-calibrated fossils. The earliest predecessor of living cats, *Pseudaelurus*, lived in Eurasia during the Miocene and spawned the Asian ancestor of modern cats ∼11 MYA. Molecular phylogeny defined eight principal Felidae lineages (Figure 3), descended from major Miocene divergence events that form the axiom for Felidae genus recognition [69]. Armed with robust and high resolution molecular phylogeny, the fossil-calibrated dates for the branches, the current and historic locations of cat species and a record of global sea level changes and continental movements, we postulated a plausible sequence of 11 historic felid migrations to explain available data [1,2]. This scenario describes our current understanding of how cats came to be (Figure 3).

Table 2. Counting ancestral translocations and inversions (breakpoints) that discriminate the genome of cat from index mammals

Counts	Number of chromosome breakpoint events in comparing cat to				
	Human	**Chimpanzee**	**Mouse**	**Rat**	**Dog**
Breakpoints (BP)	135	136	256	258	100
Translocations	29	29	110	106	58
Inversions	106	107	146	152	42
Translocation:inversion	0.27	0.27	0.75	0.70	1.38
BP/MY	0.71	0.74	1.36	1.4	0.89

[a]The number of interchromosomal exchanges (translocations) between the genome of cat versus primates is two to five times lower than that of cat versus 'shuffled' taxa mouse, rat and dog, whereas intrachromosomal exchanges (inversions) are fourfold higher in cat versus primates than translocations. This difference is apparent in low translocation:inversion ratios in cat-primate versus cat-rodent or cat-dog, and results in a balanced overall rate (BP/MY) among all comparisons (see Refs. [8,26,60]).
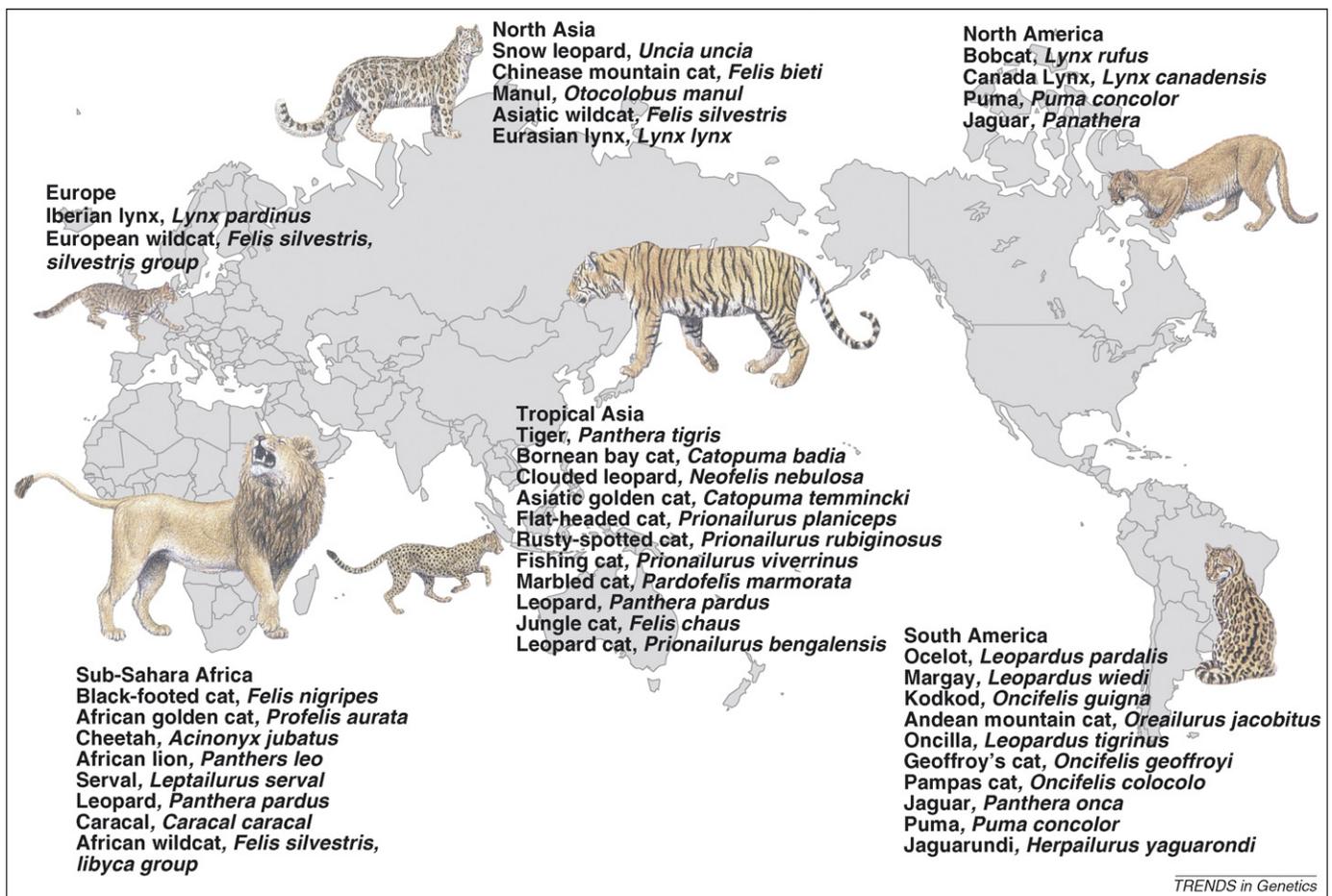
Figure 2. Geographic distribution and Latin names for the 37 living species of the Felidae family. All these species except domestic cats are considered threatened or endangered by the conservation organization of the world.

The first split from the *Pseudaelurus* occurred ~10.8 MYA in Asia, producing the five great roaring cats of the genus *Panthera* (lion, tiger, leopard, jaguar and snow leopard) and the earliest offshoot (the two clouded leopard species). Interestingly, a novel species of clouded leopards, *Neofelis diardi*, the first new Felidae species described in a century, was defined recently on the basis of genomic analysis [70,71]. A second major separation, also in Asia, led to the Bornean bay cat lineage, a group that evolved, speciated and currently resides in Southeast Asia. The next divergence and first intercontinental migration founded the caracal lineage, three African cat species whose progenitors crossed into Africa 6–10 MYA (migration route A; Figure 3). During the Miocene, sea levels were ~60 m below current levels such that Africa and the Arabian Peninsula were connected by land bridges across the Red Sea, facilitating the first felid migration to Africa. During the same period, cats dispersed across Asia and traversed the Bering Straits to Alaska (migration route B; Figure 3). The earliest progenitors of three subsequent cat lineages, (the ocelot, lynx and puma) were found in North America. Thus, cats were present in Asia, Europe, Africa and North America. Subsequently the sea levels rose, separating the continents; the changing habitats of the now 'isloated' cats contributed to the evolution of new cat species we are familiar with today.

A dynamic evolutionary process in North America produced the lynx and puma lineages 8.0 and 7.2 million years ago, respectively, to generate pumas, jaguarundi, American cheetahs, three lynx species and bobcats, species whose fossil remains in American deposits affirm their Western Hemisphere origins. Eurasian lynx species' progenitors and American cheetahs would later migrate back to Asia during the Pliocene (3–4 MYA) when the sea levels dropped once more (migration route D; Figure 3).

Toward the end of the Pliocene (2–3 MYA), the oceans receded once again, and the North and South American continents were connected by the Isthmus of Panama. North American cat species were able to migrate south via the Isthmus of Panama (migration route C; Figure 3), where they encountered a continent with no placental carnivores (i.e. no bears, dogs, cats, skunks) [72,73]. South America had been isolated from northern land masses for tens of millions of years (since Australia, Africa and South American parted from the southern supercontinent Gondwanaland > 100 MYA). Several marsupial species, including a few successful carnivorous varieties, had evolved in South America. These carnivorous marsupials were no match for the newly arrived cats [72,73]. Outcompeted in their habitat and ecological niches, nearly all marsupials were quickly replaced by migrant carnivores such as the cats from the ocelot lineage, a divergent group of seven feline species that survive in South America.
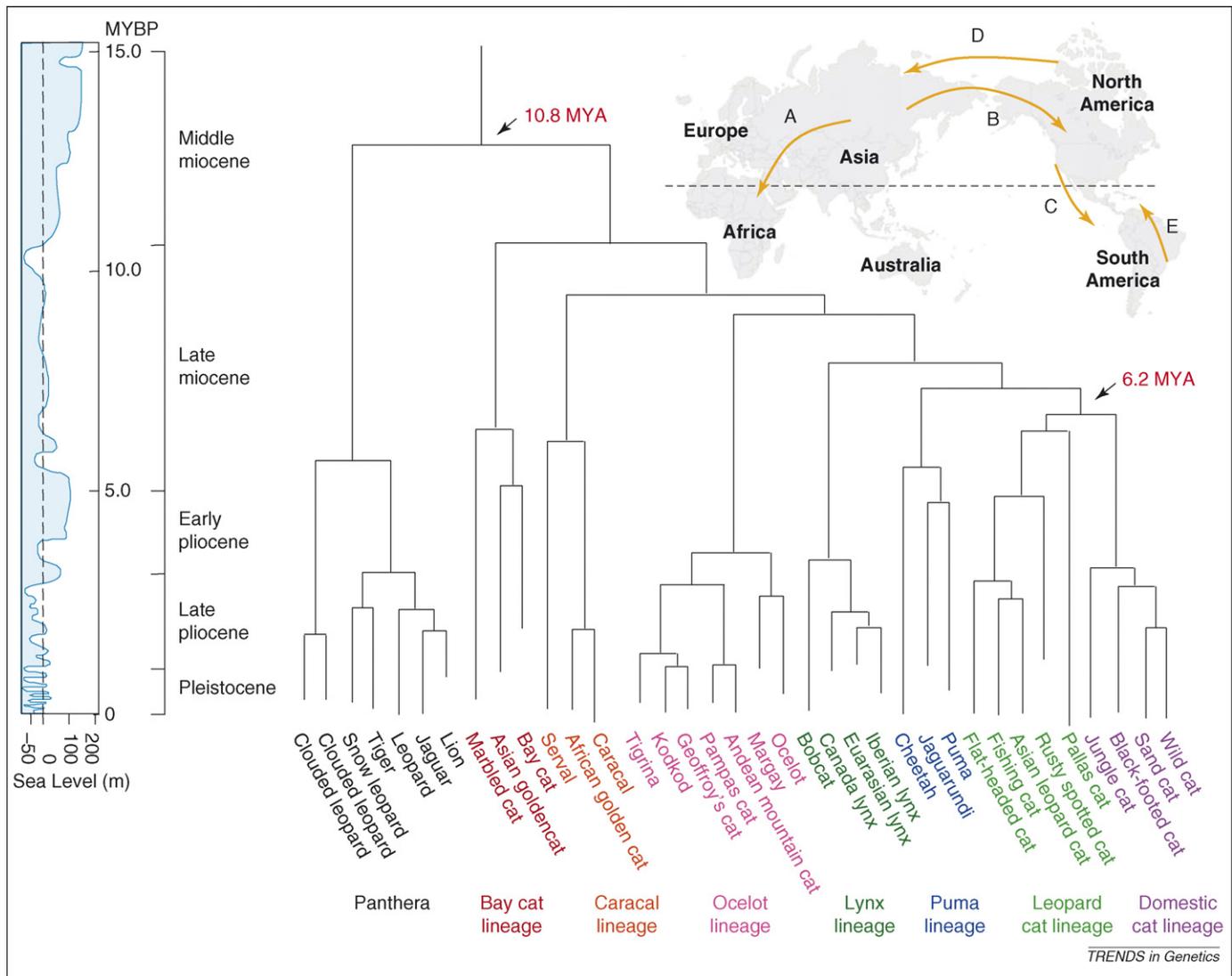
Figure 3. Molecular phylogeny of Felidae [1,2]. The branching hierarchy of the cat family as discerned from an analysis of 30 genes from members of each species. Modern cat species assort into eight lineages, each representing a close relative relationship among species within each group. The time scale is imputed from several dates of fossils that define certain portions of the tree (arrows). The geological periods during the 11 million year interval of cat evolution are presented beside the time scale. The global sea levels relative to current sea level are shown in meters. Arrows A–E depict five imputed ancestral migration routes discussed in text that we infer were traversed by ancestors of modern cat species.

Contemporaneous with the cats arrival in South America, North American dogs took the empty southern continent as a terrific opportunity to exploit. New South American Canidae (dog family) species rapidly evolved and, similar to the Felidae, they would display a wide breadth of morphological and physiological adaptations [44,74].

More recently, cheetahs and pumas became extinct in North America, after the last ice age ∼12 000 years ago. A major Pleistocene extinction of mammals eliminated 40 species of mammals from North America, including 75% of the large vertebrates living there [75,76]. Mammoths, mastodons, dire wolves, massive short-faced bears, giant ground sloth, American lions, saber-tooth cats, puma and cheetahs abruptly disappeared from North America. Much earlier, ancestral populations of cheetahs living in Asia would migrate to Africa where they survive today. Pumas avoided complete annihilation as some had moved to South America and returned to North America many generations later (migration route E; Figure 3) [77]. The other large mammal species would never return. The series of

migration events needs to be confirmed by more extensive paleontological and geographical precision in the future; if it is, it would provide one of the most detailed accounts of natural history dispersal for any mammalian species group.

Origins of cat domestication

The next act in the cat's journey, from the wild to human settlements, occurred ∼10 000 years ago in Southwest Asia. At that time, a handful of diminutive cat species intermingled within dense forests of the Mediterranean basin: jungle cat, desert cat and a ubiquitous wildcat species, *Felis silvestris*, which had three recognized wildcat subspecies – *F. silvestris silvestris* (European wildcat), *F.s. ornata.*(Middle Eastern wildcat) and *F.s. lybica* (Asian and Near Eastern wildcat). From one of these progenitors, perhaps the most successful experiment in cat natural history began – that of cat domestication.

A recent phylogeographic study of wildcats and domestic cats (N = 979) from three continents (Figure 4)
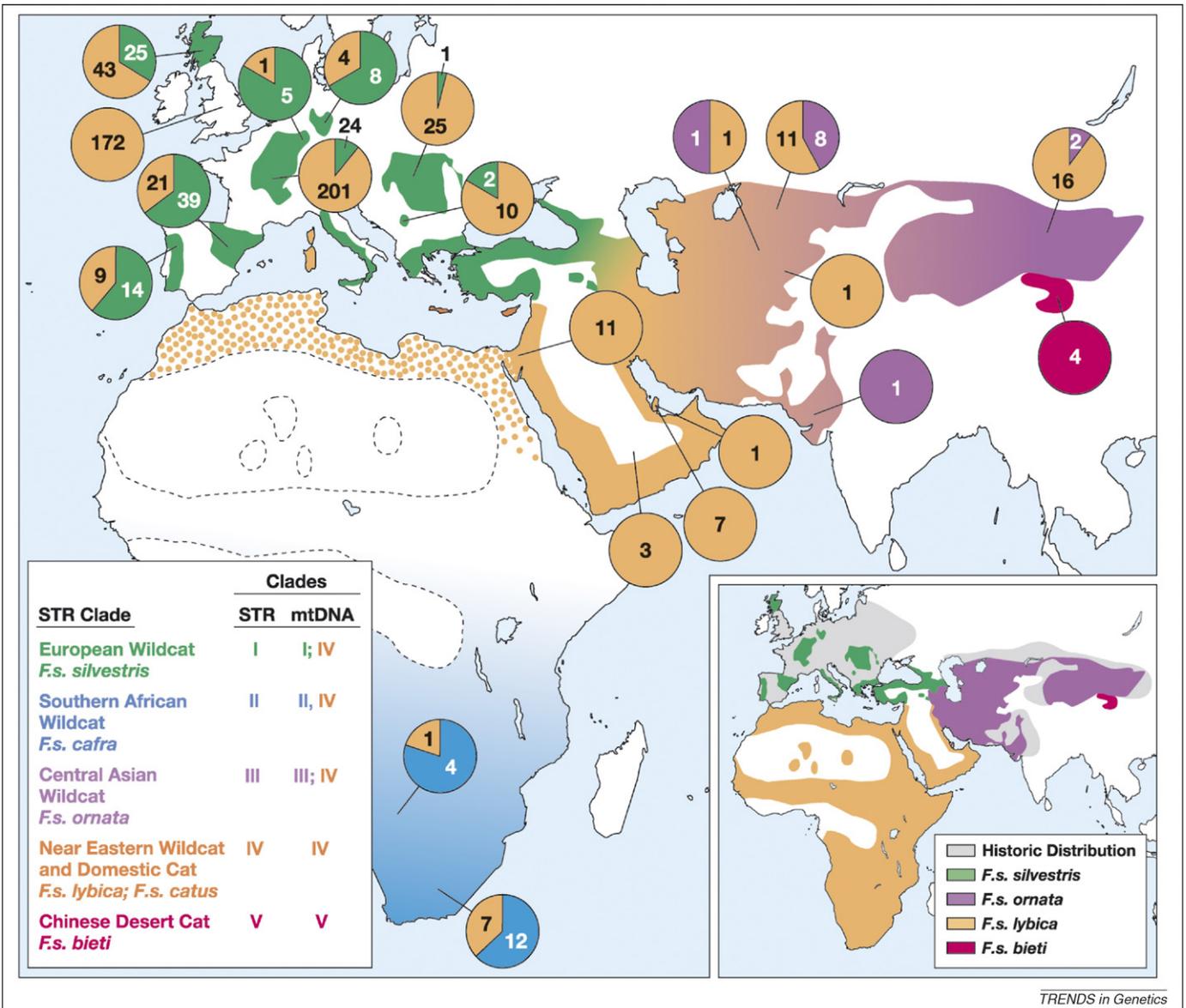
Figure 4. The current range of *Felis silvestris* and areas of sample collection are shown [42]. The colored regions reflect the location of capture of individuals with different short tandem repeat and mitochondrial DNA (mtDNA) clade genotypes (defined in the bottom left). mtDNA haplotype frequencies are indicated in pie charts specifying the number of specimens carrying each mtDNA haplotype clade. Domestic cats, *F. s. catus,* are distributed world wide and overwhelmingly carry Clade IV mtDNA haplotypes (beige). The inset on the right shows the current and historic range of *F. silvestris* subspecies on the basis of traditional morphology-based taxonomy. The Chinese desert cat is considered a wildcat subspecies, *F. silvestris bieti* as supported by data presented in Ref. [42].

used 36 STR loci and 2604 bp of mitochondrial DNA to resolve subspecies structure of *F. silvestris* and their relationship to domestic cats [42]. Although it was well known that modern wildcat populations often include ongoing admixture of domestic cats and wildcats [42,78–81], subspecies-specific nuclear and mtDNA lineages in each population were clearly resolved. For example, in sub-Saharan Africa, two versions of mtDNA genotypes were apparent (domestic and African wildcat; beige and blue pie charts, respectively, in Figure 4). The derivative molecular phylogeny for five geographically distinct subspecies (*F.s. silvestris, F. cafra, F.s. ornata, F.s.lybica* and *F.s. bieti*) were monophyletic and well resolved using both mtDNA sequence (Figure 5a) and composite STR genotypes [42]. The analysis also showed that a large assemblage of domestic cats from across the globe (including 112 fancy breed cats)

had genotypes that were indistinguishable from the clades defined by the Near East subspecies *F.s. lybica*. Furthermore, a STRUCTURE-based population genetic analysis identified a discrete population of wildcats from the Near East (Israel, United Arab Emirates and Saudi Arabia; Figure 5b) that probably reflects the ancestral founder population for the world's domestic cats.

The oldest known archaeological deposits that show co-occurrence of cat and human remains date to 9500 years ago in Cyprus [82], some 5000–6000 years before the ancient Egyptian civilization existed, which had been thought to be the site of cat domestication [83–86]. Alternatively, the combined genomic archaeological data seem to point to a domestication event in the Near East around the same time as the first agricultural village settlements in the Fertile Crescent (∼10 000 years ago).Perhaps the
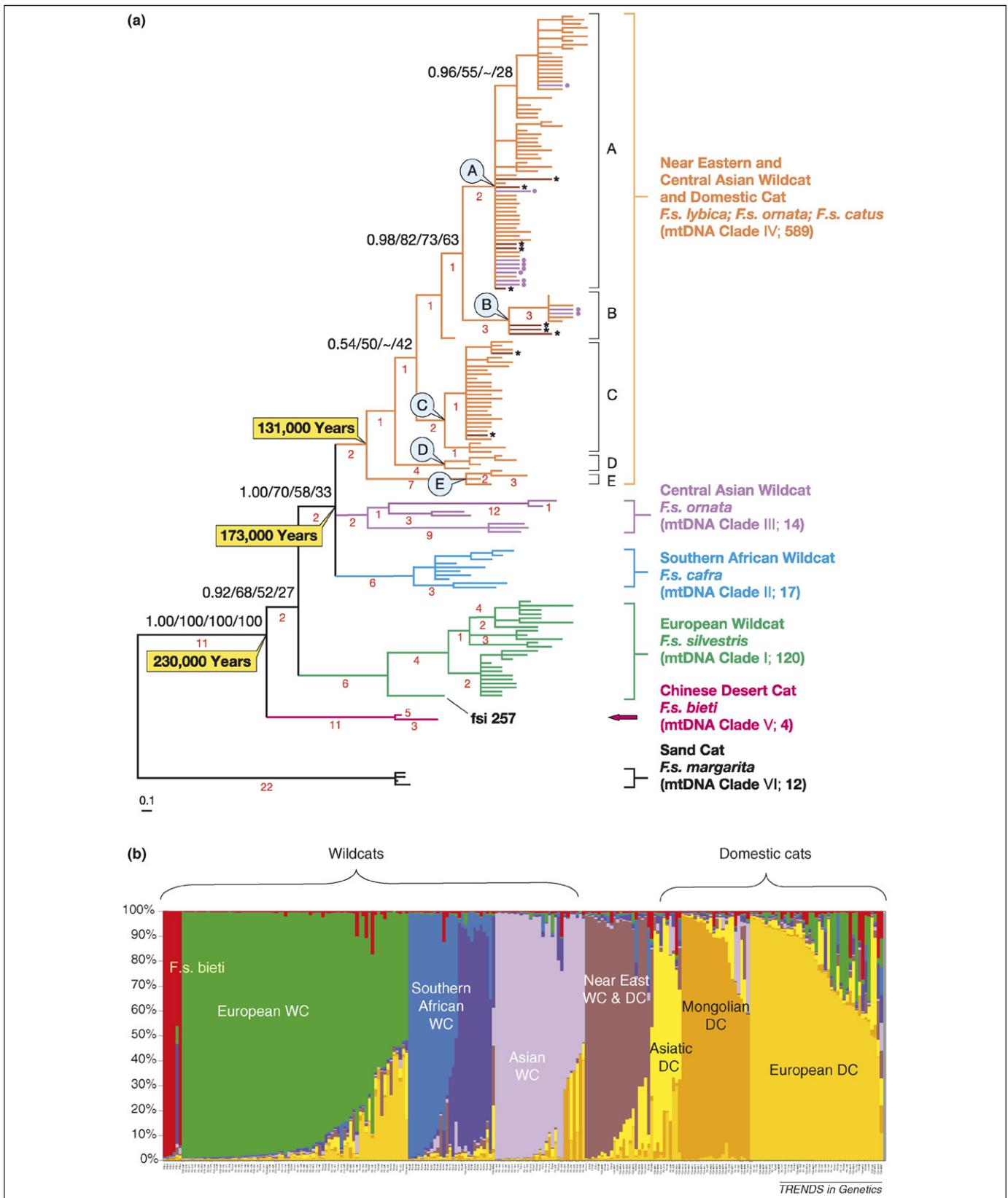
Figure 5. Resolving the origins of cat domestication. **(a)** Phylogenetic tree of mitochondrial DNA sequence (minimum evolution/neighbor joining phylogram of 2604 bp of the genes *NADH5* and *NADH6*) of 176 haplotypes discerned from 742 cats sampled across the range of the domestic cat, European, Asian and African wildcat, Chinese desert cat and sand cat. Trees created from Bayesian, maximum likelihood (ML) and maximum parsimony (MP) methods result in identical topologies for clade groupings. Confidence/ bootstrap values [Bayes/MP/ML/minimum evolution (ME)] are based on 1000 iterations and are adjacent to nodes. The number of single nucleotide differences is indicated in red below the corresponding branch. Clade designations and number of individuals is indicated in parentheses after the corresponding common name and taxonomic trinomial. Beige Clade IV bearing mtDNA haplotypes are found among domestic cats, in wild potentially admixed populations in Europe, Asia, or Africa (see Figure 4) and in Near Eastern wildcats (see main text for further details). Nodes A–E are mtDNA lineages occurring in modern domestic cats that they retain from their wildcat forbearers, *F.s. lybica*; see text. **(b)** STRUCTURE-based populations resolved 851 cats into several wildcat groups, three domestic cat groups and one group (brown) that included both domestic cats and Near East wildcats [42,92]. *y*-axis represent Q-value, the percent representation of resolved populations (colors) within each individual (listed on *x*-axis) [42,92].
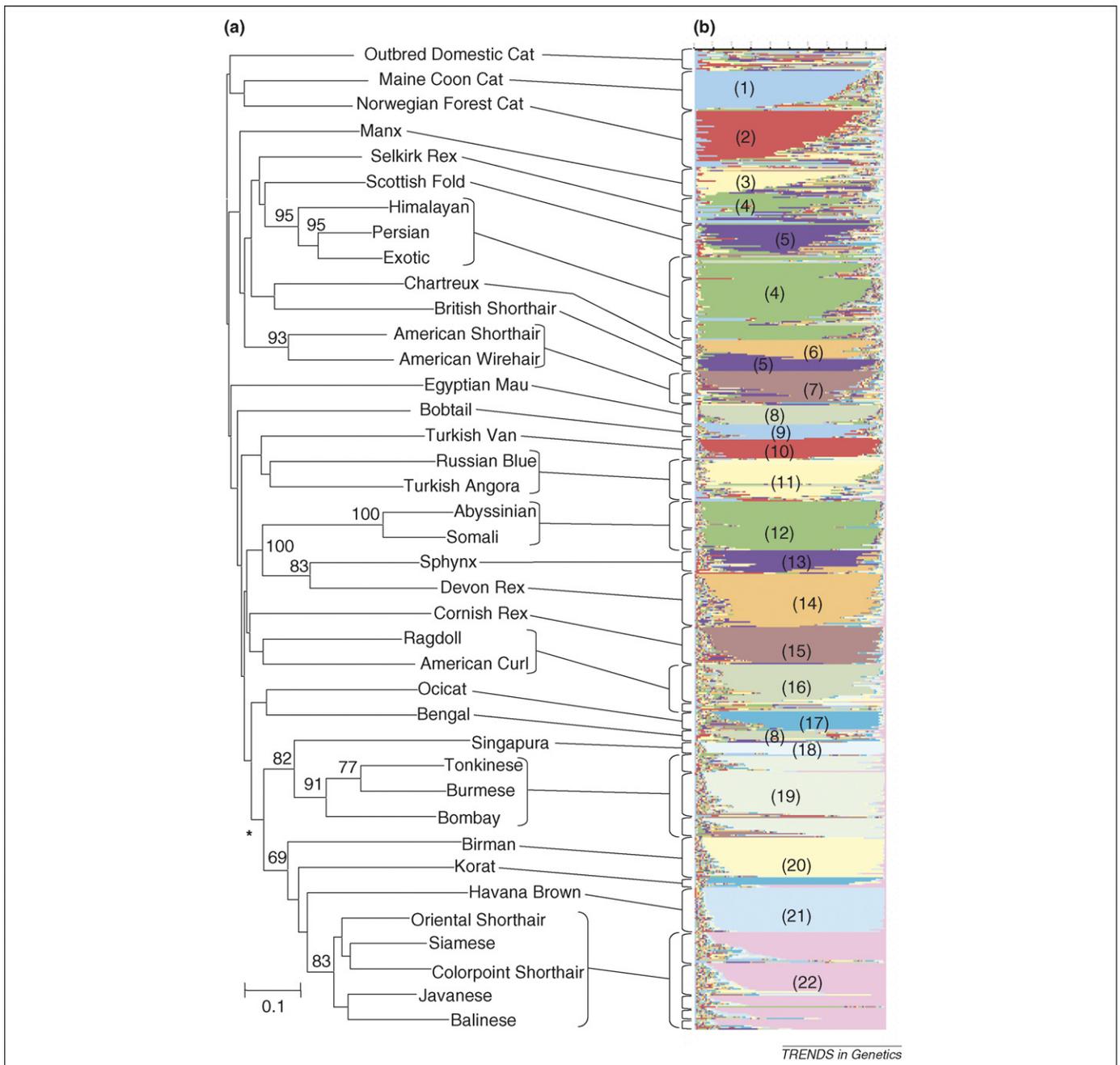
Figure 6. Phylogenetic neighbor-joining tree **(a)** of individuals from 38 cat breeds based on distance matrices generated from proportion of shared alleles algorithm (Dps) from composite genotypes [83]. Bootstrap support for branches that are supported in >60% of 100 replicates are indicated. The asterisk indicates a group of breeds that was derived completely or in part from Southeast Asian ancestors. The histogram **(b)**, generated from STRUCTURE analysis of 1040 cats, shows the proportion of each individual's genome that originated from 22 populations [83,92]. The numbers in colored blocks refer to 22 distinct cluster groups that were resolved. Some populations are composed of multiple breeds [83].

wild cats made themselves useful to early farmers by dispatching rodents from the early grain stores.

Evidence suggests that cats were probably domesticated on multiple occasions in separate locations, because at least five much older (>100 000 years old) mtDNA lineages were found in extant domestic cats (Clades A–E in domestic cat cluster in Figure 5a). From the Near East origins of domestication, subsequent gradual movements of cats with their human companions would spread domestic cats across the globe. The 600 million domestic cats alive today comprise the only Felidae species (*Felis*

catus) that is not considered to be threatened or endangered by the world's conservation bodies [87].

By the time of the industrial revolution (late 18th–early 19th century), pet cat owners were selectively mating their pet tabbies to produce fancy breeds. The American Cat Fanciers Association (http://www.acfacats.com) and the International Cat Association (http://www.tica.org) currently recognize 68 official cat breeds, from Maine Coon, Siamese, Persian to Korat; all of their roots can be traced to the origins of human and feline civilization in the Fertile Crescent.

Modern cat breeds derive from the earliest fancy breeds (Siamese, Persian, Korat, Egyptian Mau, Manx, Turkish Angora and others) established around the 17th century to the most recent (American Curl, Selkirk Rex, Singapora) established during the late 20th century [88]. A comprehensive SNP and STR survey of 38 cat breeds (∼27 individuals per breed, 1040 cats in total) revealed only modest phylogenetic and population genetic partitions caused by the recent timing of breed initiation and (unlike dog breeds) the allowable intercrossing between certain breeds in recent generations [89,90]. Nonetheless, there is a recognizable and diagnostic population structure that permitted confirmation of 96% breed assignment on the basis of a forensic panel of ten highly informative, tetra-nucleotide STR loci [91]. In Figure 6, we present a phylogenetic topology of 34 cat breeds based on composite STR genotype, population genetic–based STRUCTURE algorithm output of the same individuals. The results provide a genomic view of the slight but useful genetic differentiation among modern breeds of domestic cats [89,90].

Concluding remarks and future perspectives

Humans' fascination with cats seems to have spread to the science and genetic community. We have attempted to highlight here some of the genetic advances that provide new opportunities for better understanding cat biology and the evolutionary processes that created this exquisitely successful group of predators. The tools have applications in every aspect of cat biology, and it is exhilarating for us to witness the invigoration of science potential for the species.

Whether pursuing genetic diseases, behavior or species conservation, the cat has provided powerful examples and lessons of evolutionary, developmental and adaptive processes. We anticipate a bright future for the cat's entry among the company of major animal models of genetic biology.

References

1 O'Brien, S.J. and Johnson, W.E. (2005) Big cat genomics. *Annu. Rev. Genomics Hum. Genet.* 6, 407–429

2 O'Brien, S.J. and Johnson, W.E. (2007) The evolution of cats. Genomic paw prints in the DNA of the world's wild cats have clarified the cat family tree and uncovered several remarkable migrations in their past. *Sci. Am.* 297, 68–75

3 Savage, D.E. and Russell, D.E. (1993) *Mammalian Paleofaunas of the World,* Addison-Wesley

4 Lenffer, J. *et al.* (2006) OMIA (Online Mendelian Inheritance in Animals): an enhanced platform and integration into the Entrez search interface at NCBI. *Nucleic Acids Res.* 34, D599–D601

5 Griffin, B. and Baker, H.J. (2000) *Domestic Cats as Laboratory Animals,* Academic Press

6 O'Brien, S.J. *et al.* (2002) The feline genome project. *Annu. Rev. Genet.* 36, 657–686

7 O'Brien, S.J. (2004) *Cats. Curr. Biol.* 14, R988–R989

8 Pontius, J.U. *et al.* (2007) Initial sequence and comparative analysis of the cat genome. *Genome Res.* 17, 1675–1689

9 O'Brien, S.J. *et al.* (1997) Comparative genomics: lessons from cats. *Trends Genet.* 13, 393–399

10 Willett, B.J. *et al.* (1997) FIV infection of the domestic cat: an animal model for AIDS. *Immunol. Today* 18, 182–189

11 Troyer, J.L. *et al.* (2005) Seroprevalence and genomic divergence of circulating strains of feline immunodeficiency virus among Felidae and Hyaenidae species. *J. Virol.* 79, 8282–8294

12 Brown, M.A. *et al.* (2008) Genetic characterization of feline leukemia virus from Florida panther. *Emerg. Infect. Dis.* 14, 252–259

13 Pearks-Wilkerson, A.J. *et al.* (2004) Coronavirus outbreak in cheetahs: lessons for SARS. *Curr. Biol.* 14, R227–R228

14 Hardy, W.D. *et al.* (1980) *Feline Leukemia Virus,* Elsevier

15 Roelke-Parker, M.E. *et al.* (1996) A canine distemper virus epidemic in Serengeti lions (Panthera leo). *Nature* 379, 441–445

16 Kuiken, T. *et al.* (2004) Avian H5N1 influenza in cats. *Science* 306, 241

17 Sunquist, M. and Sunquist, F. (2002) *Wild Cats of the World,* University of Chicago Press

18 Green, P. (2007) 2x genomes–does depth matter? *Genome Res.* 17, 1547–1549

19 O'Brien, S.J. and Nash, W.G. (1982) Genetic mapping in mammals: chromosome map of domestic cat. *Science* 216, 257–265

20 Nash, W.G. and O'Brien, S.J. (1982) Conserved regions of homologous G-banded chromosomes between orders in mammalian evolution: carnivores and primates. *Proc. Natl. Acad. Sci. U. S. A.* 79, 6631–6635

21 O'Brien, S.J. *et al.* (1986) Chromosomal mapping of beta-globin and albino loci in the domestic cat. A conserved mammalian chromosome group. *J. Hered.* 77, 374–378

22 O'Brien, S.J. *et al.* (1999) The promise of comparative genomics in mammals. *Science* 286, 458-462, 479–481

23 O'Brien, S.J. *et al.* (1997) Comparative gene mapping in the domestic cat (*Felis catus*). *J. Hered.* 88, 408–414

24 Murphy, W.J. *et al.* (1999) Development of a feline whole genome radiation hybrid panel and comparative mapping of human chromosome 12 and 22 loci. *Genomics* 57, 1–8

25 Menotti-Raymond, M. *et al.* (2003) Radiation hybrid mapping of 304 novel microsatellites in the domestic cat genome. *Cytogenet. Genome Res.* 102, 272–276

26 Murphy, W.J. *et al.* (2007) A 1.5-Mb-resolution radiation hybrid map of the cat genome and comparative analysis with the canine and human genomes. *Genomics* 89, 189–196

27 Lyons, L.A. *et al.* (1999) Development of comparative anchor tagged sequences (CATS) for canine genome mapping. *J. Hered.* 90, 15–26

28 Murphy, W.J. and O'Brien, S.J. (2007) Designing and optimizing comparative anchor primers for comparative gene mapping and phylogenetic inference. *Nat. Protoc.* 2, 3022–3030

29 Menotti-Raymond, M. *et al.* (1999) A genetic linkage map of microsatellites in the domestic cat (Felis catus). *Genomics* 57, 9–23

30 Menotti-Raymond, M. *et al.* (2003) Second-generation integrated genetic linkage/radiation hybrid maps of the domestic cat (Felis catus). *J. Hered.* 94, 95–106

31 Eizirik, E. *et al.* (2003) Molecular genetics and evolution of melanism in the cat family. *Curr. Biol.* 13, 448–453

32 Menotti-Raymond, M. and O'Brien, S.J. (2008) The domestic cat, *Felis catus*, as a model of hereditary and infectious disease. In *Source Book for Models for Biomedical Research* (Conn, P.M., ed.), pp. 221–232, Humana Press

33 Beck, T.W. *et al.* (2005) The feline major histocompatibility complex is rearranged by an inversion with a breakpoint in the distal class I region. *Immunogenetics* 56, 702–709

34 Beck, T.W. *et al.* (2001) Comparative feline genomics: a BAC/PAC contig map of the major histocompatibility complex class II region. *Genomics* 71, 282–295

35 Yuhki, N. *et al.* (2003) Comparative genome organization of human, murine, and feline MHC class II region. *Genome Res.* 13, 1169–1179

36 Lopez, J.V. *et al.* (1996) Complete nucleotide sequences of the domestic cat (Felis catus) mitochondrial genome and a transposed mtDNA tandem repeat (Numt) in the nuclear genome. *Genomics* 33, 229–246

37 Lopez, J.V. *et al.* (1994) Numt, a recent transfer and tandem amplification of mitochondrial DNA to the nuclear genome of the domestic cat. *J. Mol. Evol.* 39, 174–190

38 Kim, J.H. *et al.* (2006) Evolutionary analysis of a large mtDNA translocation (numt) into the nuclear genome of the Panthera genus species. *Gene* 366, 292–302

39 ENCODE Project Consortium (2004)The ENCODE (encyclopedia of DNA elements) project. *Science* 306, 636–640

40 Guigó, R. *et al.* (2006) EGASP: the human ENCODE genome annotation assessment project. *Genome Biol.* 7, 1–31

41 O'Brien, S.J. *et al.* (2001) On choosing mammalian genomes for sequencing. *Science* 292, 2264–2266

42 Driscoll, C.A. *et al.* (2007) The Near Eastern origin of cat domestication. *Science* 317, 519–523

43 Menotti-Raymond, M. *et al.* (2007) Mutation in CEP290 discovered for cat model of human retinal degeneration. *J. Hered.* 98, 211–220

44 Lindblad-Toh, K. *et al.* (2005) Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature* 438, 803–819

45 Fyfe, J.C. *et al.* (2006) An approximately 140-kb deletion associated with feline spinal muscular atrophy implies an essential LIX1 function for motor neuron survival. *Genome Res.* 16, 1084–1090

46 Imes, D.L. *et al.* (2006) Albinism in the domestic cat (Felis catus) is associated with a tyrosinase (TYR) mutation. *Anim. Genet.* 37, 175–178

47 Schmidt-Kuntzel, A. *et al.* (2005) Tyrosinase and tyrosinase related protein 1 alleles specify domestic cat coat color phenotypes of the albino and brown loci. *J. Hered.* 96, 289–301

48 Giebel, L.B. *et al.* (1991) A tyrosinase gene missense mutation in temperature-sensitive type I oculocutaneous albinism. A human homologue to the Siamese cat and the Himalayan mouse. *J. Clin. Invest.* 87, 1119–1122

49 Lyons, L.A. *et al.* (2005) Tyrosinase mutations associated with Siamese and Burmese patterns in the domestic cat (Felis catus). *Anim. Genet.* 36, 119–126

50 Ishida, Y. *et al.* (2006) A homozygous single-base deletion in MLPH causes the dilute coat color phenotype in the domestic cat. *Genomics* 88, 698–705

51 Kehler, J.S. *et al.* (2007) Four independent mutations in the feline fibroblast growth factor 5 gene determine the long-haired phenotype in domestic cats. *J. Hered.* 98, 555–566

52 Karlsson, E.K. *et al.* (2007) Efficient mapping of mendelian traits in dogs through genome-wide association. *Nat. Genet.* 39, 1321–1328

53 Ostrander, E.A. *et al.* (2006) *The Dog and Its Genome,* Cold Spring Harbor Press

54 Lalley, P.A. *et al.* (1978) Conservation of autosomal gene synteny groups in mouse and man. *Nature* 274, 160–163

55 Womack, J.E. and Moll, Y.D. (1986) Gene map of the cow: conservation of linkage with mouse and man. *J. Hered.* 77, 2–7

56 Nash, W.G. and O'Brien, S.J. (1987) A comparative chromosome banding analysis of the Ursidae and their relationship to other carnivores. *Cytogenet. Cell Genet.* 45, 206–212

57 Nash, W.G. *et al.* (2001) The pattern of phylogenomic evolution of the Canidae. *Cytogenet. Cell Genet.* 95, 210–224

58 Graves, J.A.M. and Vanderberg, J.L. (1998) Comparative gene mapping: introduction. *ILAR J.* 39, 47–260

59 O'Brien, S.J. and Stanyon, R. (1999) Phylogenomics. Ancestral primate viewed. *Nature* 402, 365–366

60 Murphy, W.J. *et al.* (2005) Dynamics of mammalian chromosome evolution inferred from multispecies comparative maps. *Science* 309, 613–617

61 Murphy, W.J. *et al.* (2004) Mammalian phylogenomics comes of age. *Trends Genet.* 20, 631–639

62 Bourque, G. and Pevzner, P.A. (2002) Genome-scale evolution: reconstructing gene orders in the ancestral species. *Genome Res.* 12, 26–36

63 Tesler, G. (2002) GRIMM: genome rearrangements web server. *Bioinformatics* 18, 492–493

64 Tesler, G. (2002) Efficient algorithms for multichromosomal genome rearrangements. *J. Comput. Syst. Sci.* 65, 587–609

65 O'Brien, S.J. (1995) Genomic prospecting. *Nat. Med.* 1, 742–744

66 Dominguez, I. *et al.* (2003) Plant genome archaeology: evidence for conserved ancestral chromosome segments in dicotyledonous plant species. *Plant Biotechnol. J.* 1, 91–99

67 Cherry, M. (2003) Genetics to unlock secrets of our African past. *Nature* 422, 460

68 Hoft, D.F. and Belshe, R.B. (2004) The genetic archaeology of influenza. *N. Engl. J. Med.* 351, 2550–2551

69 Eizirik, E., *et al.* Molecular systematics and revised classification of the family Felidae (Mammalia, Carnivora). *J Mammal.* (in press)

70 Buckley-Beason, V.A. *et al.* (2006) Molecular evidence for species-level distinctions in clouded leopards. *Curr. Biol.* 16, 2371–2376

71 Wilting, A. *et al.* (2007) Clouded leopard phylogeny revisited: support for species recognition and population division between Borneo and Sumatra. *Front. Zool.* 4, 15

72 Novacek, M.J. (1986) A biogeographic event: the great American biotic interchange. *Science* 231, 1021–1022

73 Marshall, L.G. *et al.* (1982) Mammalian evolution and the great American interchange. *Science* 215, 1351–1357

74 Macdonald, D.W. and Silero-Zubiri, C. (2004) *The Biology and Conservation of Wild Canids,* Oxford University Press

75 Barnosky, A.D. *et al.* (2004) Assessing the causes of late Pleistocene extinctions on the continents. *Science* 306, 70–75

76 Hofreiter, M. (2007) Pleistocene extinctions: haunting the survivors. *Curr. Biol.* 17, R609–R611

77 Culver, M. *et al.* (2000) Genomic ancestry of the American puma (Puma concolor). *J. Hered.* 91, 186–197

78 Beaumont, M. *et al.* (2001) Genetic diversity and introgression in the Scottish wildcat. *Mol. Ecol.* 10, 319–336

79 Randi, E. *et al.* (2001) Genetic identification of wild and domestic cats (Felis silvestris) and their hybrids using Bayesian clustering methods. *Mol. Biol. Evol.* 18, 1679–1693

80 Lecis, R. *et al.* (2006) Bayesian analyses of admixture in wild and domestic cats (Felis silvestris) using linked microsatellite loci. *Mol. Ecol.* 15, 119–131

81 Macdonald, D.W. *et al.* (2004) *The Scottish Wildcat: Analyses for Conservation and an Action Plan,* Oxford University Press

82 Vigne, J.D. *et al.* (2004) Early taming of the cat in Cyprus. *Science* 304, 259

83 Serpell, J.A. (2000) *The Domestic Cat: the Biology of its Behaviour,* Cambridge University Press

84 Clutton-Brock, J. (1999) *A Natural History of Domesticated Mammals,* Cambridge University Press

85 Clutton Brock, J. (1993) *Cats, Ancient and Modern,* Harvard University Press

86 Zeuner, F.E. (1963) *A History of Domesticated Animals,* Harper & Row

87 Nowell, K. and Jackson, P. (1996) *Status Survey and Conservation Action Plan, Wild Cats,* International Union for Conservation of Nature and Natural Resources

88 Vella, C. and Robinson, R. (1999) *Robinson's Genetics for Cat Breeders and Veterinarians,* Oxford University Press

89 Menotti-Raymond, M. *et al.* (2008) Patterns of molecular genetic variation among cat breeds. *Genomics* 91, 1–11

90 Lipinski, M.J. *et al.* (2008) The ascent of cat breeds: genetic evaluations of breeds and worldwide random-bred populations. *Genomics* 91, 12–21

91 Menotti-Raymond, M.A. *et al.* (2005) An STR forensic typing system for genetic individualization of domestic cat (Felis catus) samples. *J. Forensic Sci.* 50, 1061–1070

92 Pritchard, J.K. *et al.* (2000) Inference of population structure using multilocus genotype data. *Genetics* 155, 945–959

93 Mullikin, J.C. and Ning, Z. (2003) The phusion assembler. *Genome Res.* 13, 81–90

94 Batzoglou, S. *et al.* (2002) ARACHNE: a whole-genome shotgun assembler. *Genome Res.* 12, 177–189

95 Jaffe, D.B. *et al.* (2003) Whole-genome sequence assembly for Mamm. *Genomes: Arachne 2. Genome Res.* 13, 91–96

96 Pontius, J.U. and O'Brien, S.J. (2007) Genome Annotation Resource Fields–GARFIELD: a genome browser for Felis catus. *J. Hered.* 98, 386–389

97 Payne, B.R. and Peters, A. (2002) *The Cat Primary Visual Cortex,* Academic Press

98 Kind, P.C. *et al.* (2002) Correlated binocular activity guides recovery from monocular deprivation. *Nature* 416, 430–433

99 Anderson, J.S. *et al.* (2000) The contribution of noise to contrast invariance of orientation tuning in cat visual cortex. *Science* 290, 1968–1972

100 Henson, M.S. and O'Brien, T.D. (2006) Feline models of type 2 diabetes mellitus. *ILAR J.* 47, 234–242

101 Rah, H. *et al.* (2005) Early-onset, autosomal recessive, progressive retinal atrophy in Persian cats. *Invest. Ophthalmol. Vis. Sci.* 46, 1742–1747

102 Ryugo, D.K. *et al.* (2005) Restoration of auditory nerve synapses in cats by cochlear implants. *Science* 310, 1490–1492

103 Menotti-Raymond, M. *et al.* (1997) Pet cat hair implicates murder suspect. *Nature* 386, 774

104 Menotti-Raymond, M. *et al.* (1997) Genetic individualization of domestic cats using feline STR loci for forensic applications. *J. Forensic Sci.* 42, 1039–1051

Articles of interest in other Cell Press journals

Intranuclear Distribution and Local Dynamics of RNA Polymerase II during Transcription Activation
Jie Yao, M. Behfar Ardehali, Christopher J. Fecko, Watt W. Webb, and John T. Lis
Molecular Cell December 28, 2008

Human Alu RNA Is a Modular Transacting Repressor of mRNA Transcription during Heat Shock
Peter D. Mariner, Ryan D. Walters, Celso A. Espinoza, Linda F. Drullinger, Stacey D. Wagner, Jennifer F. Kugel, and James A. Goodrich
Molecular Cell February 29, 2008

Molecular Dissection of Mammalian RNA Polymerase II Transcriptional Termination
Steven West, Nicholas J. Proudfoot, and Michael J. Dye
Molecular Cell March 14, 2008

Poised Polymerases: On Your Mark...Get Set...Go!
David Price
Molecular Cell April 11, 2008

Genome-wide RNA polymerase II: not genes only!
Frederic Koch, Frederic Jourquin, Pierre Ferrier and Jean-Christophe Andrau
Trends in Biochemical Sciences June 2008

And in this issue of TiG...

Molecular Evolution of RNA Polymerase II CTD
Rob D. Chapman, Martin Heidemann, Corinna Hintermair and Dirk Eick
Trends in Genetics June 2008