Molecular Phylogenetics and Evolution xxx (2010) xxx–xxx

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Supermatrix and species tree methods resolve phylogenetic relationships within the big cats, *Panthera* (Carnivora: Felidae)

Brian W. Davis, Gang Li, William J. Murphy *

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

ARTICLE INFO

Article history:
Received 10 August 2009
Revised 28 January 2010
Accepted 29 January 2010
Available online xxxx

Keywords:
Felidae
Phylogeny
Panthera
Y chromosome
Cats
numt
BEST

ABSTRACT

The pantherine lineage of cats diverged from the remainder of modern Felidae less than 11 million years ago and consists of the five big cats of the genus *Panthera*, the lion, tiger, jaguar, leopard, and snow leopard, as well as the closely related clouded leopard. A significant problem exists with respect to the precise phylogeny of these highly threatened great cats. Despite multiple publications on the subject, no two molecular studies have reconstructed *Panthera* with the same topology. These evolutionary relationships remain unresolved partially due to the recent and rapid radiation of pantherines in the Pliocene, individual speciation events occurring within less than 1 million years, and probable introgression between lineages following their divergence. We provide an alternative, highly supported interpretation of the evolutionary history of the pantherine lineage using novel and published DNA sequence data from the autosomes, both sex chromosomes and the mitochondrial genome. New sequences were generated for 39 single-copy regions of the felid Y chromosome, as well as four mitochondrial and four autosomal gene segments, totaling 28.7 kb. Phylogenetic analysis of these new data, combined with all published data in GenBank, highlighted the prevalence of phylogenetic disparities stemming either from the amplification of a mitochondrial to nuclear translocation event (numt), or errors in species identification. Our 47.6 kb combined dataset was analyzed as a supermatrix and with respect to individual partitions using maximum likelihood and Bayesian phylogenetic inference, in conjunction with Bayesian estimation of species trees (BEST) which accounts for heterogeneous gene histories. Our results yield a robust consensus topology supporting the monophyly of lion and leopard, with jaguar sister to these species, as well as a sister species relationship of tiger and snow leopard. These results highlight new avenues for the study of speciation genomics and understanding the historical events surrounding the origin of the members of this lineage.

1. Introduction

Nearly every one of the 38 living cat species (Carnivora: Felidae) is denoted as endangered or threatened by international monitoring bodies such as the Convention of International Trade of Endangered Species of Wild Fauna and Flora (CITES), the International Union for Conservation of Nature (IUCN), and the U.S. Endangered Species Act. There is at least one population from every species of the Felidae on Appendix I or II of CITES or on the IUCN Red List of threatened or endangered species. Entire species are on one or both of these lists (Baillie and Groombridge, 1996). This case is especially pronounced for the great pantherine cats, all of which possess protected status. This clade consists of the five big cats of the genus *Panthera, P. leo* (lion), *P. tigris* (tiger), *P. onca* (jaguar), *P. pardus* (leopard), and *P. uncia* (snow leopard), as well as the closely related *Neofelis nebulosa* (clouded leopard), which diverged from *Panthera* approximately 6 million years ago (Buckley-Beason et al., 2006; Johnson et al., 2006). It is well known that mankind has had a large influence on the dwindling numbers of these wild cats, and conservationists are increasingly utilizing genetic data to formulate conservation action plans for both land and marine mammals (Brooks et al., 1992; Schipper et al., 2008). Members of the Felidae, particularly those within genus *Panthera*, are often the top predator in an ecosystem, existing in comparatively low density to other species, thus requiring larger territories. In an ever-shrinking global ecosystem they are under increasingly consistent threat of eradication as human expansion constrains their range.

Despite their highly threatened status, the evolutionary history of the big cats has been largely obscured by a poor fossil record, their recent and rapid radiation during the Pliocene, individual speciation events occurring within less than 1 million years, and probable introgression between lineages following their divergence (Johnson et al., 2006). Multiple groups have attempted to resolve this problem using morphological (Christiansen, 2008;

* Corresponding author. Fax: +1 979 845 9972.
E-mail address: wmurphy@cvm.tamu.edu (W.J. Murphy).

1055-7903/$ - see front matter © 2010 Published by Elsevier Inc.
doi:10.1016/j.ympev.2010.01.036

Hemmer, 1978; Herrington, 1986; Salles, 1992) as well as biochemical and molecular characters (Bininda-Emonds et al., 2001, 1999; Jae-Heup et al., 2001; Janczewski et al., 1995; Johnson et al., 1996, 2006; Johnson and O'Brien, 1997; Mattern and McLennan, 2000; Pecon-Slattery et al., 2004; Wei et al., 2009; Yu and Zhang, 2005). Fig. 1 illustrates the conflicting hypotheses and lack of corroboration between published molecular studies. The disparity between these studies may stem from limited phylogenetic signal, systematic errors such as long-branch attraction and star-shaped phylogenies, as well as the prevalence of mitochondrial to nuclear translocation events (numt), and heavy reliance upon such mitochondrial DNA (mtDNA) markers as true cytoplasmic mitochondrial (cymt) sequences without experimental verification. A resolved phylogenetic tree provides a strong historical foundation for future population genetic and phylogeographic studies, opening up new avenues for the study of speciation genomics and understanding the biogeographic events surrounding the origin of the members of this lineage. Ultimately, phylogenetic and character-based approaches will allow researchers to track the evolution of clade- and species-specific traits that contribute to the success of these graceful, yet powerful apex predators.

Previous phylogenetic studies of felid relationships showed the Y chromosome has a very low level of homoplasy in the form of convergent, parallel, or reversal substitutions, rendering the vast majority of substitutions phylogenetically informative (Pecon-Slattery et al., 2004). The constitutively haploid Y chromosome has a

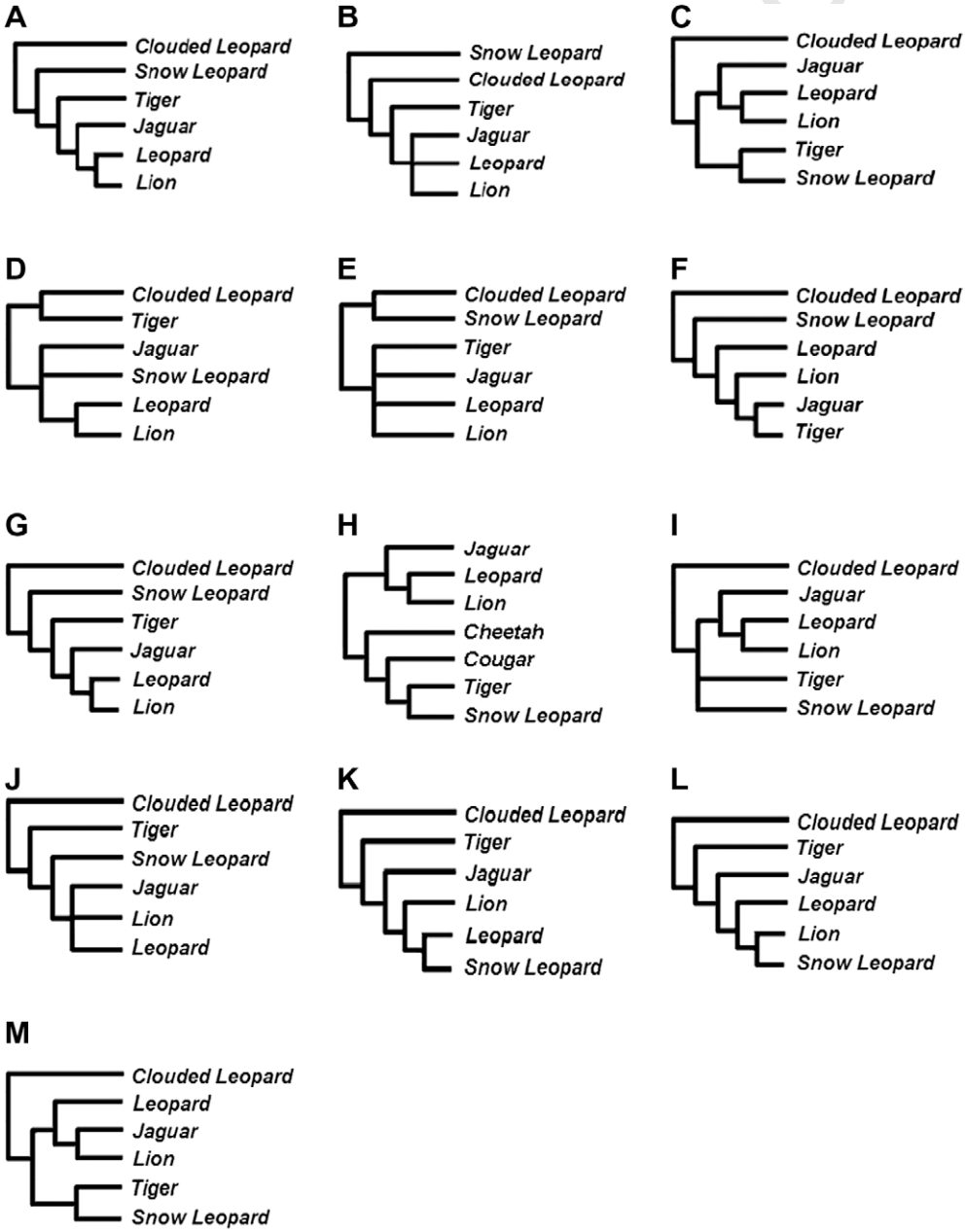


Fig. 1. Prior phylogenetic hypotheses of the genus *Panthera*. (A and B) based solely on morphological characters, (C–M) from biochemical or molecular studies. (A) Hemmer (1978), Herrington (1986), Christiansen (2008). (B) Salles (1992), (C) RFLP of complete mtDNA genomes using 28 restriction endonucleases (Johnson et al., 1996), (D) 2 mtDNA genes [647 bp] (Janczewski et al., 1995), (E) 2 mtDNA genes [697 bp] (Johnson and O'Brien, 1997), (F) 4 mtDNA genes [1435 bp] (Mattern and McLennan, 2000), (G) 40 source trees, 282 elements (Bininda-Emonds et al., 1999), (H) 1316 chemical characters (Bininda-Emonds et al., 2001), (I) Variation within the hypervariable mtDNA CR and RS2 (Kim et al. 2001; Jae-Heup et al., 2001), (J) 3 Y-linked [3604 bp] (Pecon-Slattery et al., 2004), (K) 6 mtDNA and 3 Autosomal genes [6500 bp] (Yu and Zhang, 2005), (L) 7 mtDNA genes [3816 bp] (Wei et al., 2009), (M) 19 Autosomal, 5 X, 4 Y, 6 mtDNA genes [23,920 bp] (Johnson et al., 2006).

Please cite this article in press as: Davis, B.W., et al. Supermatrix and species tree methods resolve phylogenetic relationships within the big cats, *Panthera* (Carnivora: Felidae). Mol. Phylogenet. Evol. (2010), doi:10.1016/j.ympev.2010.01.036

uniparental, male-specific inheritance, passing only from father to son. The majority of this chromosome in mammals is male-specific (termed MSY for male-specific region on the Y (Skaletsky et al., 2003)) and is almost totally unaffected by meiotic recombination events (Jobling and Tyler-Smith, 2003). The exception is the pseudoautosomal (PAR) region, a terminal segment which synapses with a homologous region on the X to facilitate meiotic crossover and to ensure accurate sex chromosome segregation in males. The escape of MSY genes from recombination is of primary importance for phylogeny in that Y-specific haplotypes will typically pass intact through generations, changing only by mutation, therefore preserving a simpler record of patrilineal evolutionary history (Pecon-Slattery et al., 2004). The combination of these properties makes this an effective region for phylogenetic reconstruction.

Here, we provide an alternative evaluation of the evolutionary history of the pantherine lineage using intronic sequences contained within single-copy genes on the felid Y chromosome. This information was combined with previously published data (Johnson et al., 2006), and newly generated sequence for four mitochondrial and four autosomal genes, highlighting areas of phylogenetic incongruence. *In silico* evaluation, identification, and removal of putative numt sequences, together with a thorough phylogenetic exploration of the complete dataset provided a highly supported topology, consistent with several biochemical and morphological character sets (Bininda-Emonds et al., 2001; Johnson et al., 1996). The results of these comprehensive analyses are summarized and compared to outline the complex evolutionary history of *Panthera*.

2. Materials and methods

2.1. Marker selection

The complete domestic cat cDNA sequence for each Y-linked, X-degenerate, single-copy gene was obtained from GenBank: *SMCY* (EU879977), *EIF1AY* (EU879973), *EIF2S3Y* (EU879975), *DDX3Y* (EU879971), *USP9Y* (EU879980), *UBE1Y* (DQ329521), *UTY* (EU879982), and *ZFY* (EU879984) (Murphy et al., 2006; Pearks-Wilkerson et al., 2008). The putative exon–intron boundary was defined using BLAST (Altschul et al., 1990) queried against the *Mus musculus* Y chromosome assembly in build 37.1 (NCBI, 2008b) for *Eif2s3y* and *Ube1y*, and *Homo sapiens* build 36.3 (NCBI, 2008a) for the remaining genes. Gaps in the feline cDNA sequence alignment relative to the human or mouse genomic sequence were used to define approximate locations of intron–exon boundaries. As intron length tends to covary with genome size across species (Ogata et al., 1996), the observed length in the human and mouse genomes was used as an estimate of felid intron size, to determine optimal PCR protocols.

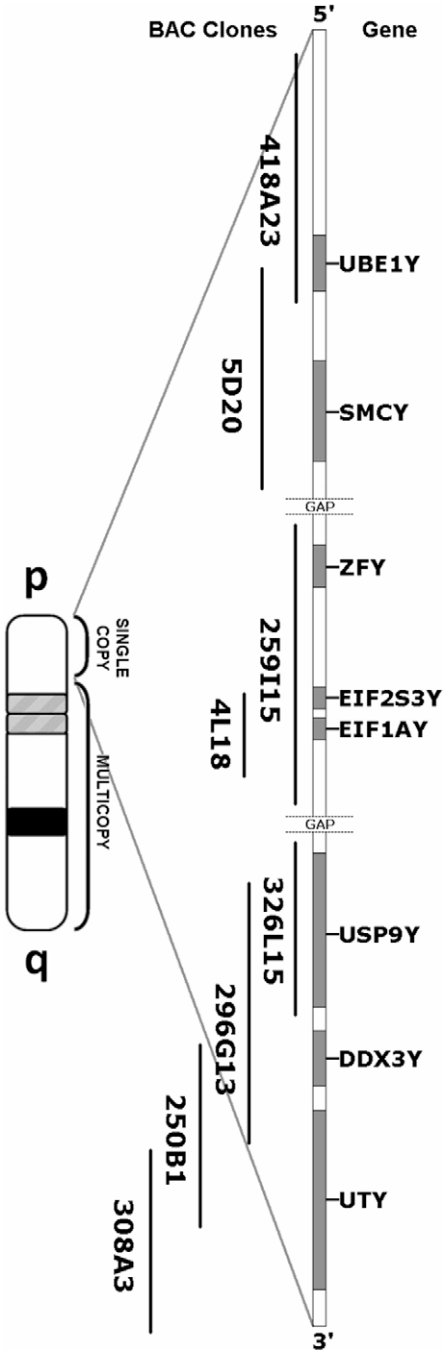


Fig. 2. Physical map of the single-copy X-degenerate region of the domestic cat Y chromosome, and selected RPCI-86 BAC library clones used in this study.

2.2. Primers and sequencing

EPIC (exon-primed, intron crossing) primers were designed using Primer3 (Rozen and Skaletsky, 2000) to target the BLAST-defined exonic flanks and extend into each intronic region of the eight Y-linked genes. Each 25 μL PCR reaction contained 2.5 μL 10 × PCR buffer (Invitrogen), 0.75 μL 50 mM $MgCl_2$, 2 μL 10 mM dNTPs (Applied Biosystems Inc.), and 2 μL of each 5 μM forward and reverse primer. 10–30 ng of template DNA isolated from domestic cat Y chromosome BAC clones was used to ensure amplification of Y chromosome DNA sequences, rather than priming the exons of highly similar X-linked gametologs (Pearks-Wilkerson et al., 2008). All BAC clones derive from the male RPCI-86 10X BAC library. Fig. 2 depicts a physical map of the domestic cat single-copy Y-linked gene region and the corresponding BAC clone DNAs used for PCR amplification. Amplicons smaller than 6 kb were amplified with Platinum *Taq* DNA polymerase, while amplicons from 6 kb to 11 kb were amplified with AccuPrime High Fidelity DNA polymerase. PCR was performed with an Applied Biosystems GeneAmp 9700 thermal cycler with optimal conditions for each reaction of: denaturation 94 °C (15 s), annealing 58 °C (30 s), extension 72 °C (1 min per 1 kb estimated amplicon size) for 35 cycles; with initial denaturation 94 °C (2 min) and final extension 72 °C (5 min).

Primer pairs that produced robust, single banded, amplicons were sequenced using an Applied Biosystems 3730 DNA Analyzer, to provide intronic sequence for subsequent Y-specific primer design. Each Y-specific primer was initially tested in male and feline genomic DNAs from both jaguar and clouded leopard to assess

Please cite this article in press as: Davis, B.W., et al. Supermatrix and species tree methods resolve phylogenetic relationships within the big cats, *Panthera* (Carnivora: Felidae). Mol. Phylogenet. Evol. (2010), doi:10.1016/j.ympev.2010.01.036

cross-species, Y-specificity. Successful amplicons were sequenced, and those primers producing high quality sequence reads from both species, determined in Sequencher 4.7 (2008), were amplified and sequenced in the remaining four *Panthera* species.

2.3. Matrix construction

To build upon previous published work and increase the amount of data available for supermatrix analysis, previously published sequences available across all six species were obtained from GenBank. All GenBank sequences were the same length and averaged less than one small gap per gene segment, so were aligned by eye.

2.3.1. Nuclear genes

The Johnson et al. (2006) dataset utilized previously published Y chromosome sequences (King et al., 2007; Pecon-Slattery et al., 2004) and generated new sequences for the X chromosome and autosomal partitions (Janczewski et al., 1995; Johnson and O'Brien, 1997), resulting in 19 autosomal (11,030 bp), 5 X-linked (3223 bp), and 4 Y-linked (4540 bp) gene segments for pantherines. The Yu and Zhang (2005) study also utilized previously published datasets (Janczewski et al., 1995; Johnson and O'Brien, 1997; Masuda et al., 1996; Yu et al., 2004) and contributed three autosomal (2767 bp) gene segments for all taxa except jaguar. We resequenced these three gene segments, *IRBP* (interphotoreceptor retinoid-binding protein), exon 1; *FGB* (β fibrinogen), intron 7; and *TTR* (transthyretin) intron 1, for all six pantherine species using the same reagents and thermocycler protocols as the sub-6 kb EPIC amplifications. An additional autosomal locus with high variability within *Panthera*, *CES7*, was also sequenced in all pantherines. New and published (Johnson et al., 2006) sequences were assembled into the supermatrix, with redundancies represented by new sequences.

2.3.2. Mitochondrial DNA sequences

Because the majority of all previous phylogenetic studies of *Panthera* relied at least partially on mitochondrial data, the published existence of *Panthera* numts raised the possibility of numt amplification for these sequences (Johnson et al., 1996; Kim et al., 2006). Two recent publications, Yu and Zhang (2005) , and Johnson et al. (2006) included virtually all mitochondrial gene segments used in prior molecular phylogenies for *Panthera* (Janczewski et al., 1995; Johnson and O'Brien, 1997; Mattern and McLennan, 2000; Wei et al., 2009). Therefore these sequences were extensively scrutinized prior to inclusion in the final supermatrix. Accession numbers for mtDNA segments and publication are listed in Supplementary Table 2.

All published mtDNA sequences were obtained from GenBank for each of the six taxa (Supplementary Table 1), including the six gene segments from Johnson et al. (2006) (3936 bp) and six from Yu and Zhang (2005) (3472 bp). Five complete pantherine mitochondrial genome sequences available in GenBank are reported as being derived from PCR products amplified from an enriched mtDNA fraction, reducing the likelihood that these sequences represent numts (Kim et al., 2006; Wei et al., 2009; Wu et al., 2007). Supplementary Table 2 lists the accession numbers for all sequences used in our analyses. RAxML-VI-HPC 7.0.4 (Stamatakis, 2006) was used to generate a ML tree for each mtDNA gene segment (Supplementary Fig. 1). By including all published and novel (i.e. those identified in this study) numts, as well as the complete mtDNA genome sequences that are putatively cytoplasmic in origin, we readily identified and removed numts from the dataset (See Fig. 3) (Barnett et al., 2009).

Because all published jaguar and lion sequences are based on PCR amplicons derived from genomic DNA, it was difficult to verify their cytoplasmic origin. Therefore, we only included sequences that did not cluster within a numt clade, and whose species identification was corroborated by at least one other sequence from a separate study. Once these criteria were met, the sequences with the greatest length were selected first, followed by those with the most recent publication date (Table 1). All sequences were translated to verify that they did not contain stop codons, and base frequencies were computed to quantify anti-guanine bias characteristic of mammalian mitochondrial genomes.

2.3.3. Homoplasy, combinability, and topological support metrics

Homoplasy indices were computed for each partition. An incongruence-length difference test (ILD) (Farris et al., 1995) (1000 replicates) both within and between partitions followed the precedent set by Sullivan (1996) and Cunningham (1997) in implementing a significance threshold of $\alpha = 0.01$. A Shimodaira–Hasegawa (SH) test was performed using 10,000 RELL bootstrap replicates (Shimodaira and Hasegawa, 1999). A jacknifing approach investigated the effect of removing each gene segment from the concatenated alignment on ML bootstrap support. Each metric was calculated using PAUP*.

2.4. Sequence alignment and phylogenetic analyses

Sequence alignments were performed using ClustalX 2.0.3 (Larkin et al., 2007) (gap opening penalty of 10, gap extension penalty of 0.2) with subsequent by-eye verification and manual editing done with BioEdit 7.0.9.0 (Hall, 1999). Data was partitioned based on the mode of inheritance: (1) Y chromosome, (2) autosomal, (3) X chromosome, and (4) mitochondrial. Combinations of partitions were also defined as (5) nuclear (Y chromosome, X chromosome, and autosomes) and (6) uniparental (Y chromosome and mitochondria).

2.4.1. Maximum likelihood

Exhaustive maximum likelihood (ML) tree searches were performed using PAUP 4.0b10 (Swofford, 2002) based on the parameter vales obtained utilizing the Akaike Information Criteria (AIC) hierarchical test statistic in ModelTest (Posada and Buckley, 2004; Posada and Crandall, 1998) (Supplementary Table 3). Gene trees were estimated independently for each locus by exhaustive ML searches in PAUP* using the models in Supplementary Table 3, as was done for the combined partitioned and the supermatrix trees. Bootstrapping was performed using 1000 iterations with TBR branch-swapping.

2.4.2. Bayesian phylogenetic inference

Bayesian infethe rence (BI) was implemented using MrBayes 3.0.4 (Ronquist and Huelsenbeck, 2003) with the models deduced by AIC in MrModeltest v2 (Supplementary Table 4). For individual gene segments, MrBayes ran for 1,500,000 generations, saving every 100th tree, discarding the first 250,000 as burn-in. For the 6 partitioned matrices, the MCMC algorithm ran for 3,000,000 generations, with every 100th tree saved and the first 750,000 generations discarded as burn-in. For both BI analyses, one cold and seven hot chains were used to explore treespace.

2.4.3. Bayesian estimation of species trees

The Bayesian estimation of species trees (BEST) method was used to construct a species tree from individual Bayesian gene trees in BEST, a modified MrBayes package (Liu and Pearl, 2007; Liu et al., 2008), using the locus-specific models detailed in Supplementary Table 4. The data was partitioned into 29 "genes" with the mitochondria and the Y chromosome genes combined into a single partition, respectively, since they do not undergo recombination and are inherited as a complete unit. Parameters specifically implemented in the BEST analysis involved specifying the haploid nature

Please cite this article in press as: Davis, B.W., et al. Supermatrix and species tree methods resolve phylogenetic relationships within the big cats, *Panthera* (Carnivora: Felidae). Mol. Phylogenet. Evol. (2010), doi:10.1016/j.ympev.2010.01.036

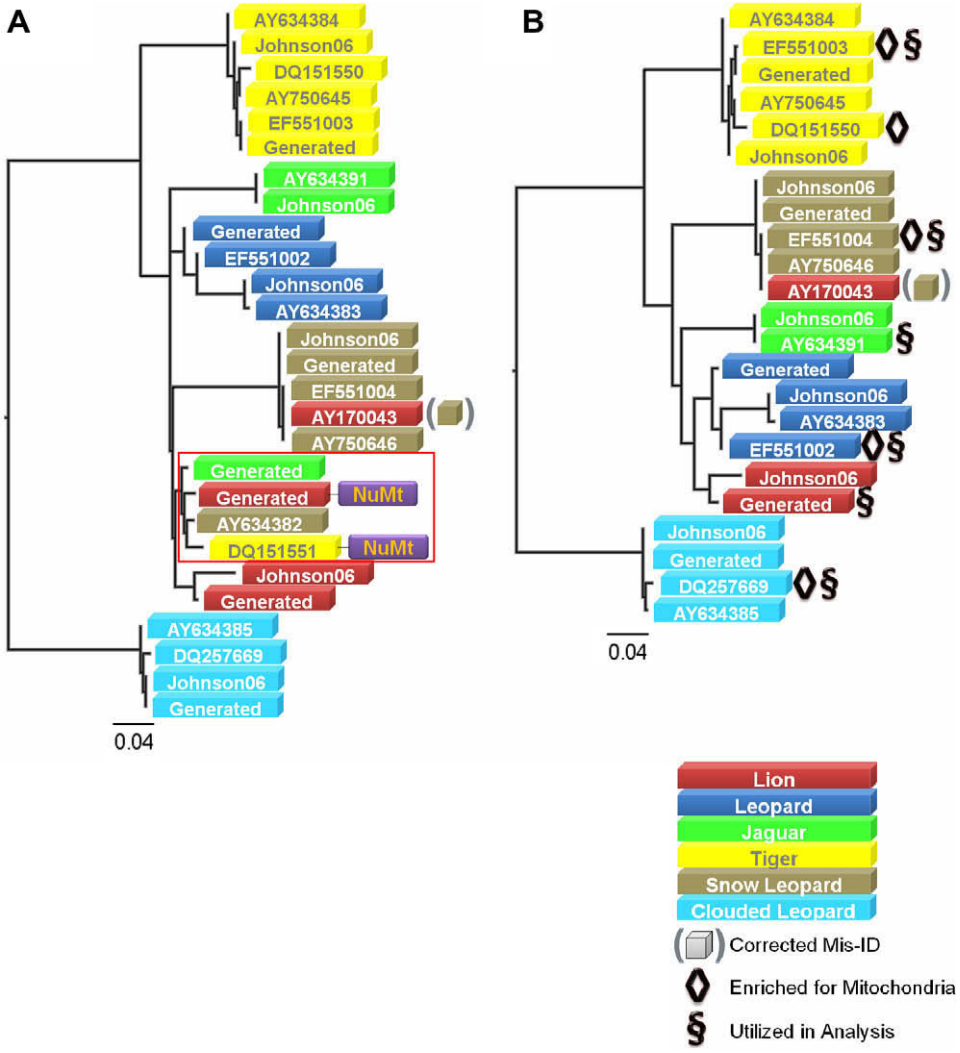


Fig. 3. Maximum likelihood tree (GTR + Γ) of all published and new *ND2* mtDNA gene segments. (A) Clear evidence of numt amplification within the red box. (B) ML tree with numts removed. Evidence for species misidentification is shown in (B) for the "lion" sequence that groups within the snow leopard clade. Corrected species identification is indicated in parentheses.

Table 1
Q4 Sources of mtDNA gene segment sequences used for phylogenetic analyses.

	12S	*16S*	*CYTB*	*ND1*	*ND2*	*ND4*	*ND5*
Lion	Newly generated	AF006457	Newly generated	Johnson (2006)	Newly generated	Newly generated	Johnson (2006)
Leopard	EF551002	EF551002	EF551002	EF551002	EF551002	EF551002	EF551002
Tiger	EF551003	EF551003	EF551003	EF551003	EF551003	EF551003	EF551003
Jaguar	Newly Generated	AF006441	Newly Generated	Johnson (2006)	AY634391	AY634403	Johnson (2006)
Snow leopard	EF551004	EF551004	EF551004	EF551004	EF551004	EF551004	EF551004
Clouded leopard	DQ257669	DQ257669	DQ257669	DQ257669	DQ257669	DQ257669	DQ257669

of the mitochondrial and MSY segments and setting the bounds for the prior distribution of the mutation rate across loci to between 0.2 and 2.0, with the average as 1 as suggested by the documentation. The MCMC algorithm ran for 10,000,000 generations, saving every 1000th gene tree, sampling 1000 species trees, and discarding the first 1,000,000 gene trees as burn-in. The result was a final species tree topology with support values listed as posterior probabilities.

2.5. Molecular dating

A relative rate test was performed in PAUP for each partition (df = 4, $\alpha = 0.05$) to test for clocklike nucleotide substitution behavior. To infer divergence times we derived ML estimates of sequence divergence by the ESTBRANCHES (partitioned into 29 "genes") component of PAML 3.15 (Yang, 2007) followed by Bayesian relaxed clock dating in MULTIDIVTIME (Thorne and Kishino, 2002). We used three fossil-based calibrations: (1) a minimum of 1.6 MYA (Janczewski et al., 1995; Kurten and Anderson, 1980) for the base of the lion–leopard–jaguar clade; (2) a minimum of 1.8 MYA (Janczewski et al., 1995; Neff, 1982) for the base of the tiger–snow leopard clade; and (3) the minimum for earliest *Panthera* species based on leopard fossils from African Villafranchian deposits was 3.8 MYA (Johnson et al., 2006; Werdelin and Lewis, 2005). To evaluate the sensitivity of divergence estimates to the removal of constraints, the MULTIDIVTIME analysis was performed in four replicates, each time removing one constraint.

Please cite this article in press as: Davis, B.W., et al. Supermatrix and species tree methods resolve phylogenetic relationships within the big cats, *Panthera* (Carnivora: Felidae). Mol. Phylogenet. Evol. (2010), doi:10.1016/j.ympev.2010.01.036

3. Results

3.1. Matrix analysis

We successfully sequenced 39 Y chromosome intron regions from all six species, totaling 15,392 bp, in addition to 8872 bp from four autosomal genes (*IRBP*, *FGB*, *TTR*, *CES7*), and 4510 bp from 4 mitochondrial DNA segments (*ND2*, *ND4*, *CYTB*, *12S*). The fully assembled pantherine dataset samples 43 loci for the five species of *Panthera*, and a clouded leopard (*Neofelis nebulosa*) as the outgroup taxon. The final supermatrix consisted of 47,628 nucleotides (974 sites excluded as either gaps or ambiguous) partitioned as follows: Y chromosome (19,140 bp), autosomes (19,124 bp), X chromosome (3223 bp) and mitochondria (6141 bp). The Y chromosome partition was significantly less homoplastic than all other regions of the genome (Supplementary Table 5), confirming the conclusions of Pecon-Slattery et al. (2004) that there is a very low amount of convergent, parallel, or reversal substitutions and that the vast majority of substitutions are phylogenetically informative.

In order to compare the similarities between homologous sequences generated independently by both publications, the DNAdist component of PHYLIP 3.67 (Felsenstein, 2004) was used to compute a LogDet distance matrix between each species for *ND2*. As shown graphically in Fig. 4 discrepant phylogenetic relationships were indicated by the high level of intraspecies dissimilarity observed between both the lion and snow leopard sequences from different publications. Conversely, the pairwise interspecies difference between Yu and Zhang's lion *ND2* sequence and the Johnson et al. snow leopard sequence was 0.01 (Supplementary Table 6). This value is much lower than all other interspecies comparisons (avg = 0.128), and at a level consistent with other pantherid intraspecies distance calculations (avg = 0.006).

3.2. Transthyretin

A preliminary analysis of published pantherid *TTR* genes (Flynn et al., 2005; Flynn and Nedbal, 1998; Johnson et al., 2006; Yu and Zhang, 2005) revealed similar discordant levels of intra- and interspecies dissimilarity involving snow leopard and lion sequences. Therefore, we amplified and sequenced intron 1 using primers designed from a consensus sequences derived from previous studies. These new *TTR* intron 1 sequences were combined with the Johnson et al. (2006) and Yu and Zhang (2005) (Edwards et al., 2007) sequences as well as one tiger sequence from other phylogenetic studies (Flynn et al., 2005; Flynn and Nedbal, 1998). There was no jaguar sequence for *TTR* in the Yu and Zhang publication, and the lion *TTR* intron 1 sequence generated for reference Mattern and McLennan (2000) was used by Yu and Zhang in their analysis. The ML analysis not only indicated species misidentification of the Yu and Zhang snow leopard sequence, but produced a unique topology compared to all other gene segments in the dataset (Fig. 5). Specifically, *TTR* intron 1 appeared to track the divergence of the lion–leopard–jaguar and tiger–snow leopard clades, with ten substitutions distinguishing the two clades, and no changes distinguishing each species since that event (i.e. all nucleotide changes occur on the internal branches leading to these clades and the sequences within these two clades were 100% identical). These results were similar across studies (Fig. 5b–d).

3.3. Mitochondrial DNA analysis

To further determine whether the phylogenetic discrepancies observed in the *ND2* gene between the two independent publications could be attributed to species misidentification, or the amplification of a mitochondrial pseudogene present in the nuclear genome (numt), we sequenced the *12S*, *CYTB*, *ND2*, and *ND4* gene segments using in-house DNAs with previously described reagent and thermal cycler protocols. Direct sequencing of the lion *ND2* PCR amplicon produced sequence traces with a region of superimposed chromatograms (i.e. multiple peaks), suggesting multiple amplifications sequenced in a single reaction. Subsequent cloning and sequencing of this PCR product showed one clone to have a similar length sequence as all other pantherines, and the second possessed a 4 bp insertion and a 70 bp deletion, confirming a numt co-amplification. An anti-guanine bias was apparent for all mtDNA segments, indicating true mtDNA amplification (Supplementary Table 7).

3.4. Phylogenetic reconstruction

Maximum likelihood, Bayesian phylogenetic inference, and BEST all produced identical rooted topologies for the complete supermatrix. A summary of tree statistics for the PAUP* and MrBayes analyses for all six partitions and the supermatrix is given in Table 2. The support values for ML, Bayesian phylogenetic inference, and BEST are indicated on the maximum likelihood topology shown in Fig. 6 and individual gene tree topologies with bootstrap support in Supplementary Fig. 2.

3.4.1. ILD tests and data partitioning

Combinability between and within the defined partitions (ILD, $\alpha = 0.01$) (Cunningham, 1997; Sullivan, 1996) indicated sufficient congruence for individual gene segments within each partition,

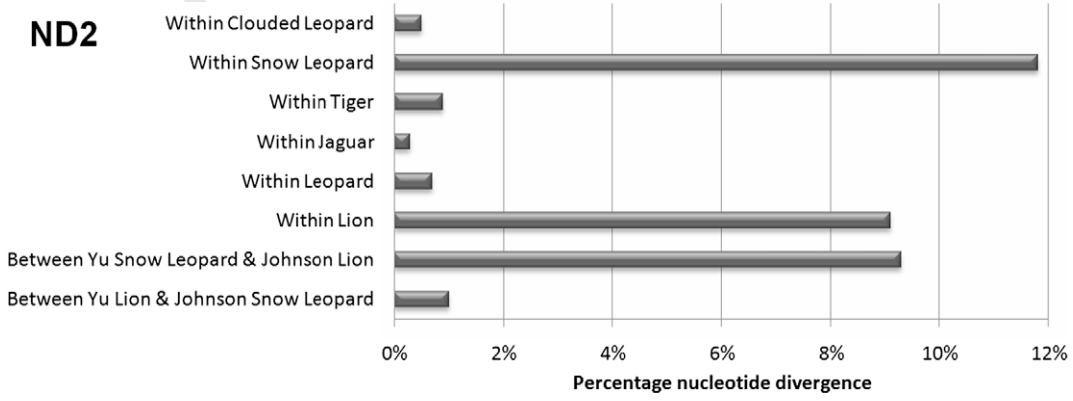


Fig. 4. LogDet pairwise distances between the *ND2* sequences generated by the Yu and Zhang (2005) and Johnson et al. (2006) publications show a very high disparity within snow leopard and lion sequences. There is more interspecies similarity between Yu's lion sequence and Johnson's snow leopard sequence than within each species.

Please cite this article in press as: Davis, B.W., et al. Supermatrix and species tree methods resolve phylogenetic relationships within the big cats, *Panthera* (Carnivora: Felidae). Mol. Phylogenet. Evol. (2010), doi:10.1016/j.ympev.2010.01.036

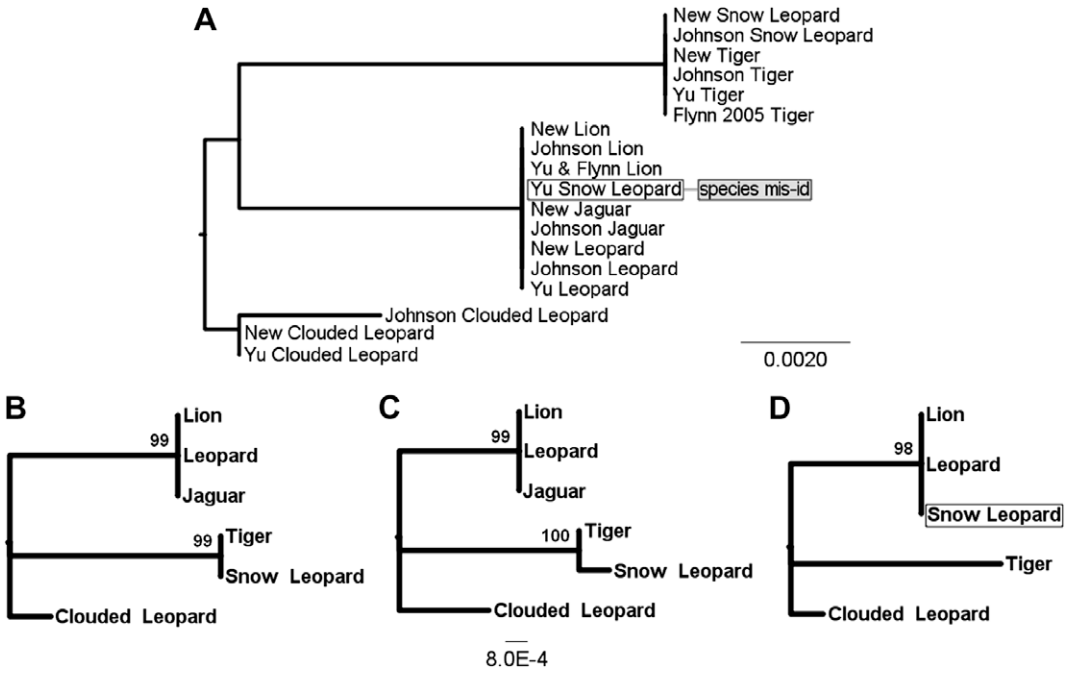


Fig. 5. (A) Maximum likelihood topology for *TTR* shows putative misidentification of the snow leopard sequence in Ref (Yu and Zhang, 2005) (boxed, arrow indicates taxon placement if species ID was accurate). Note the unique phylogenetic topology tracking the divergence of the tiger–snow leopard and lion–leopard–jaguar clade with a lack of any subsequent changes since that event. Separate topologies with nonparametric bootstrap support (1000 replicates) indicated in red for (B) newly generated sequences (C) Johnson et al. (2006) (D) Yu and Zhang (2005) with putative misidentification boxed.

Table 2
Support for species relationships within genus Panthera for supermatrix and partitioned analyses. Maximum likelihood nonparametric bootstrap values (ML), Bayesian posterior probabilities (BPP), and BEST weighted posterior probabilities are shown. Rooted analyses included all six taxa. Unrooted analyses performed without clouded leopard.

		Tiger Snow	Lion Leopard	Lion Jaguar	Lion Leopard Jaguar	Tiger Jaguar	Jaguar Snow	Tiger Clouded	Lion Jaguar Clouded
Rooted supermatrix43 partitions	BEST	0.91	0.63	0.35	–	–	–	–	–
	ML	93.7	100	–	100	–	–	6.0	–
	BPP	1.00	1.00	–	1.00	–	–	–	–
29 Partitions	BPP	1.00	1.00	–	1.00	–	–	–	–
4 Partitions	BPP	0.70	1.00	–	1.00	–	–	–	–
Unrooted supermatrix	ML	100	98.9	–	100	–	–	N/A	N/A
	BPP	1.00	1.00	–	1.00			N/A	N/A
Rooted nuclear	ML	100	62.5	37.5	100	–	–	–	–
	BPP	1.00	–	0.97	1.00	–	–	–	–
Unrooted nuclear	ML	100	55.5	44.4	100	–	–	N/A	N/A
	BPP	1.00	–	0.99	1.00			N/A	N/A
Rooted autosomal	ML	93.4	66.6	33.4	99.8	–	–	–	–
	BPP	1.00	0.41	0.59	1.00	–	–	–	–
Unrooted autosomal	ML	100	54.8	45.0	100	–	–	N/A	N/A
	BPP	1.00	0.21	0.79	1.00	–	–	N/A	N/A
Rooted uniparental	ML	56.4	100	–	100	–	–	34.0	–
	BPP	0.80	1.00	–	1.00	–	–	–	–
Unrooted uniparental	ML	99.8	100	–	100	–	–	N/A	N/A
	BPP	1.00	1.00	–	1.00	–	–	N/A	N/A
Rooted Y chromosome	ML	100	70	–	100	–	–	–	–
	BPP	1.00	0.87	–	1.00	–	–	–	–
Unrooted Y chromosome	ML	100	82.2	–	–	–	–	N/A	N/A
	BPP	1.00	0.99	–	1.00	–	–	N/A	N/A
Rooted X chromosome	ML	64.3	–	61.1	–	–	–	–	66.0
	BPP	0.98	–	0.97	–	–	–	–	0.98
Unrooted X chromosome	ML	64.3	–	85.4	–	–	–	N/A	N/A
	BPP	0.98	–	1.00	–	–	–	N/A	N/A
Rooted MtDNA	ML	–	99.8	–	71.2	–	–	94.3	–
	BPP	–	1.00	–	0.86	–	0.07	1.00	–
Unrooted MtDNA	ML	48.4	99.3	–	–	36.9	12.6	N/A	N/A
	BPP	0.38	1.00	–	0.39	–	0.23	N/A	N/A

with the exception of the autosomal partition (ILD = 0.002) (Supplementary Table 8). This partition was incongruent with the Y chromosome (ILD = 0.001), and was nearly incongruent with the X chromosome (ILD = 0.016). Within the nuclear partition there

Please cite this article in press as: Davis, B.W., et al. Supermatrix and species tree methods resolve phylogenetic relationships within the big cats, *Panthera* (Carnivora: Felidae). Mol. Phylogenet. Evol. (2010), doi:10.1016/j.ympev.2010.01.036

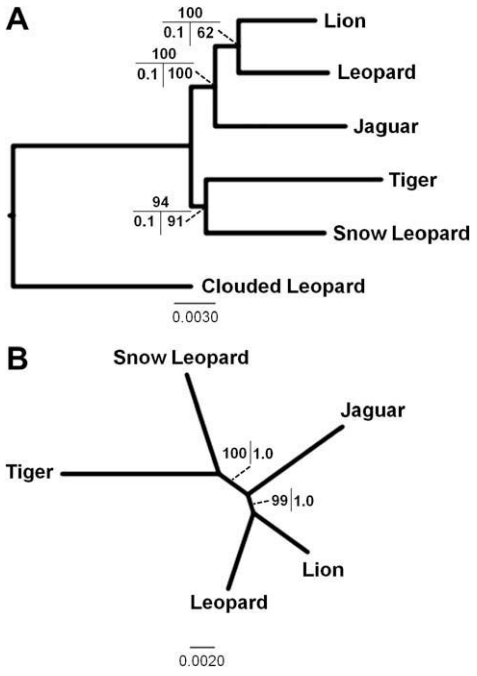


Fig. 6. Maximum likelihood (ML) tree based on analysis of the complete supermatrix. (A) Rooted with clouded leopard as outgroup. 1000 ML bootstrap replicate percentages depicted on the top, Bayesian posterior probabilities (BPP) on the bottom left, and BEST posterior probabilities on the bottom right. (B) Unrooted topology with ML bootstrap percentages on the left and BPP on the right.

was also significant incongruence between gene segments (ILD = 0.008). When the mtDNA partition was added to any other partition combination, the ILD score increased above statistical significance, most likely due to the heterogeneity of the phylogenetic signal within this partition. The complete supermatrix passed ILD when partitioning each gene segment separately (43 partitions) and with the mitochondria and Y chromosome as individual partitions respectively (29 partitions). However, with four partitions (autosomes, mitochondria, X chromosome and Y chromosome) the dataset was statistically incongruent. Therefore the dataset was analyzed with multiple phylogenetic methods and varying partitioning schemes to ensure an accurate species tree topology.

3.4.2. Partition topology and support

Maximum likelihood analysis produced identical topologies for the supermatrix, Y chromosome, autosomal, uniparental, and nuclear partitions (Fig. 6), placing lion and leopard as sister taxa and jaguar as the basal member of this clade. Tiger was placed as sister to snow leopard in a separate monophyletic group, with clouded leopard as the outgroup to *Panthera*. Autosomal and nuclear BI topologies grouped lion and jaguar as sister taxa rather than lion and leopard (Supplementary Fig. 3). ML bootstrap values and Bayesian posterior probabilities for lion–leopard monophyly were high for the rooted uniparental, and mtDNA partitions, and moderate for the Y chromosome partition (Table 2; Supplementary Fig. 4) with support increasing when unrooted. Individual maximum likelihood topologies with branch lengths and clade support values are included for all partitions in Supplementary Fig. 5. The varying histories of autosomal loci can be seen in the lower ML bootstrap support for lion–leopard monophyly in the autosomal and nuclear partitions and in the reconstruction of lion–jaguar monophyly by the Bayesian results. The X chromosome partition also recapitulated lion–jaguar monophyly with significantly greater support using BI than ML, consistent with the non-conservative nature of the Bayesian method (Suzuki et al., 2002). A summary of phylogenetically informative sites revealed lion–leopard monophyly was supported by roughly twice the total number of characters as lion–jaguar monophyly in the supermatrix, mitochondrial, and Y chromosome partitions (Supplementary Table 9) with less support in the autosomes. A complete listing of phylogenetically informative sites for each partition is contained in Supplementary Tables 10–12. There was complete support for tiger–snow leopard monophyly from the Y chromosome and nuclear partitions, high support from the autosomal partition and uniparental partitions, and moderate support from the X chromosome partition. Comparison of partition-specific topologies using the Shimodaira–Hasegawa test ($\alpha = 0.05$) showed significantly discordant topologies for the X chromosome partition when compared to other partition topologies (Supplementary Table 13). The rooted mtDNA partition did not support tiger–snow leopard monophyly with low support when unrooted (Supplementary Fig. 6).

3.4.3. Bayesian estimation of species trees

The BEST method was used to construct a species tree from the individual gene trees generated using Bayes (Liu and Pearl, 2007; Liu et al., 2008). In this way we were able to estimate the posterior distribution of species trees using the multilocus dataset and locus-specific models (Supplementary Table 4), allowing for heterogeneous gene trees among loci under the multi-species coalescent model (Edwards et al., 2007; Wu et al., 2007). The BEST method was implemented on the total matrix and reconstructed the same topology as the ML and BI supermatrix approaches (Fig. 6), however, with lower support for lion–leopard monophyly (ML: 100, BPP: 1.0, BEST: 0.63) and tiger–snow leopard monophyly (ML: 100, BPP: 94, BEST: 91). Convergence was reached with 10,000,000 generations (standard deviation of the split frequencies < 0.01). Increasing the number of genes to 43 by separating individual Y-linked and mitochondrial loci slightly decreased the BEST support values for these nodes. Decreasing the gene number to 4 (Mitochondrial, Autosomal, X, Y) increased BEST support, but disallows consideration of heterogeneous gene histories in the X and autosomal partitions (data not shown).

3.4.4. Gene jacknifing

We investigated the signal contribution of each gene segment relative to each phylogenetic grouping by removing each gene segment partition, and recording the resulting changes in topology and bipartition support (i.e. gene jacknifing, or GJ), (Supplementary Table 14). The support for relationships was largely unaffected by removal of individual gene segments, with the exception of *TTR* and *CES7*. When *TTR* (885 bp) was removed from the autosomal partition, we observed a pronounced decrease in the ML bootstrap support for snow leopard–tiger monophyly (~70–9%) and a decrease for lion–leopard–jaguar monophyly (~90–45%) (Fig. 7). The effects of removing *CES7* were much more pronounced due to its longer length and increased phylogenetic signal (Supplementary Fig. 7): support for tiger–snow leopard monophyly dropped from ~95% to 70%, and lion–leopard monophyly dropped from ~70% to 6%, with a corresponding increase in support for lion–jaguar monophyly (~30–94%). Changes observed when removing *TTR* were more evident when *CES7* was not included in the autosomal matrix (see Supplementary Figs. 8–11 for the remainder of the gene jacknifing results with tabulated bootstrap support values listed in Supplementary Tables 15–19). Rooted BEST support for snow leopard–tiger monophyly dropped to 0.75 when *TTR* was removed from the supermatrix. When both *TTR* and *CES7* were removed from the supermatrix, little change in unrooted support was observed for lion–leopard (ML: 95, BPP: 1.0) and tiger–snow leopard monophyly (ML: 100, BPP: 1.0). Thus, they were included in the dataset for phylogenetic analysis.

Please cite this article in press as: Davis, B.W., et al. Supermatrix and species tree methods resolve phylogenetic relationships within the big cats, *Panthera* (Carnivora: Felidae). Mol. Phylogenet. Evol. (2010), doi:10.1016/j.ympev.2010.01.036

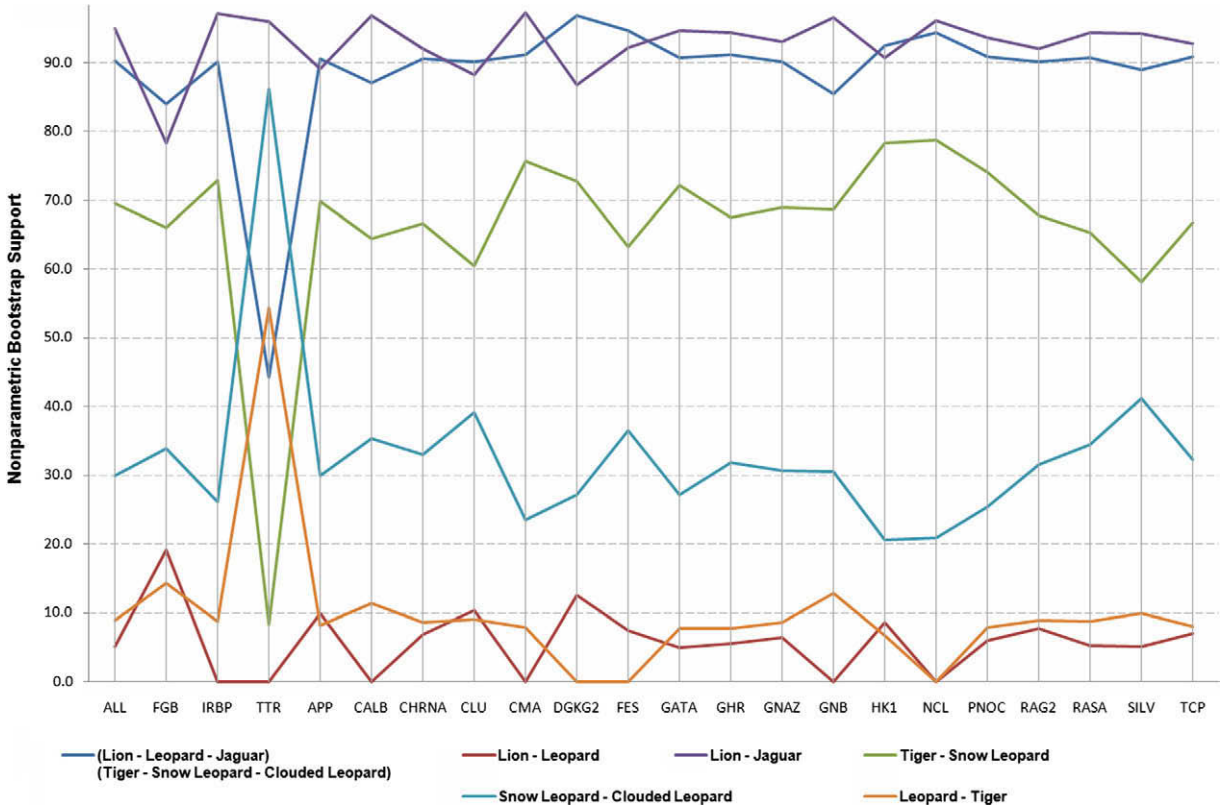


Fig. 7. Bootstrap bipartition support for the autosomal partition (*CES7* excluded) with each gene segment jacknifed out. *Y*-axis is bootstrap percentages from 1000 replicates. The difference in bipartition support when *TTR* is removed from the dataset is evident, indicating topological change.

3.5. Molecular dating

A relative rate test shows that at $\alpha = 0.05$ no partitions behave in a clocklike manner (Supplementary Table 20). Thus, in lieu of using a strict molecular clock approach to date the divergences within *Panthera*, a Bayesian relaxed clock approach was implemented with MULTIDIVTIME using 3 fossil calibrations (Fig. 8). The 95% Bayesian credibility interval for the basal divergence time of *Panthera* was 3.80–4.31 MYA. The jaguar divergence from the lion–leopard lineage was between 2.56 and 3.66 MYA, while the lion and leopard diverged 1.95–3.10 MYA. The snow leopard and tiger diverged from one another roughly 2.70–3.70 MYA. The removal of each individual internal calibration point did not significantly affect the divergence times or the 95% credibility intervals. However, removing the minimum constraint for the base of *Panthera* reduced the divergence times at each node by roughly 50% (Fig. 8).

4. Discussion

4.1. Supermatrix and partitioned analyses

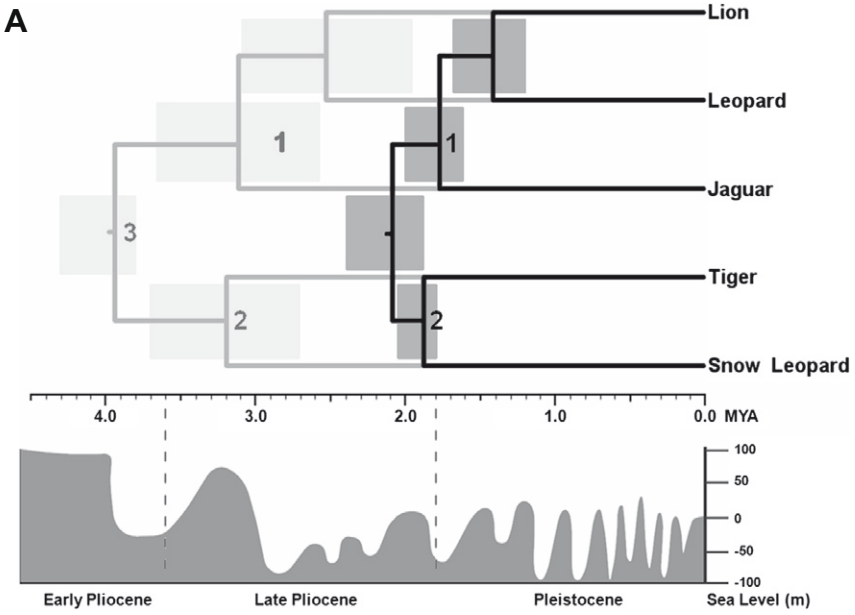
In this study, we provide an independent assessment of the pantherine phylogeny by supplementing previously published datasets with newly generated sequences from 39 Y-linked segments, three autosomal genes, and four mtDNA genes. Phylogenetic inference provided by maximum likelihood, Bayesian phylogenetic inference, and BEST all reconstruct the same topology (Fig. 6). This result is derived from the largest and most comprehensive dataset for *Panthera*, spanning the largest number of genomic regions, and the strictest *in silico* vetting process for determining bonafide mtDNA genes, as opposed to numt paralogs. We therefore believe that the topology shown in Fig. 6 is the current best estimate of species relationships within *Panthera*.

Our analysis demonstrates that discordance between published estimates of *Panthera* phylogeny can be attributed to a number of factors, most prominently sample/sequence misidentification, and inclusion of numts in published datasets. For example, comparison of the pairwise LogDet distances between mitochondrial *ND2* gene segments from the Yu and Zhang (2005) and Johnson et al. (2006) datasets show interspecies divergences of 10–12% and intraspecies divergences of less than 1%. This holds true for all taxa except lion and snow leopard, which have intraspecies divergences of 9.1% and 11.8%, respectively, values comparable to interspecies divergences observed between other pantherid taxa, and nearly 10-fold greater than the 1% interspecies divergence observed between the two species. By analyzing all published sequences in a single analysis, we were able to identify examples of species misidentification by consensus.

Previous research has shown that species barcoding initiatives based on mtDNA sequence can result in overestimation of species diversity, in spite of attempts to remove numts (Song et al., 2008). Evaluation of sequences generated from genomic DNA-derived amplicons using criteria such as presence of indels, in-frame stop codons, and atypical nucleotide composition will minimize, but not eliminate, the presence of numts (Song et al., 2008). We observed similar issues whereby potentially bonafide *Panthera* mtDNA segments were excluded on the basis that they clustered within a clade of numts, and outside of the majority of other sequences for a given species. The only proven way to ensure mitochondrial amplification is to enrich for mtDNA by centrifugation in a gradient medium. Therefore, until such methods are used to generate mtDNA sequences from all members of *Panthera*, published mitochondrial gene segments should be used with caution.

Please cite this article in press as: Davis, B.W., et al. Supermatrix and species tree methods resolve phylogenetic relationships within the big cats, *Panthera* (Carnivora: Felidae). Mol. Phylogenet. Evol. (2010), doi:10.1016/j.ympev.2010.01.036

A



B

	Date (MYA)	Standard Error	95% CI	Constraints
Leopard – Lion	2.51791	0.29980	1.94543 – 3.09747	None
(Leopard-Lion) – Jaguar	3.11937	0.28788	2.55964 – 3.65988	(1) Min = 1.6
Tiger – Snow Leopard	3.19020	0.25875	2.69694 – 3.70713	(2) Min = 1.8
(Tiger- Snow Leopard) – (Leopard-Lion-Jaguar)	3.93854	0.14132	3.80334 – 4.31548	(3) Min = 3.8
Leopard – Lion	1.42900	0.12029	1.21738 – 1.68751	None
(Leopard-Lion) – Jaguar	1.77558	0.10976	1.61313 – 2.02800	(1) Min = 1.6
Tiger – Snow Leopard	1.87336	0.07335	1.80196 – 2.06906	(2) Min = 1.8
(Tiger- Snow Leopard) – (Leopard-Lion-Jaguar)	2.08457	0.13506	1.87767 – 2.39847	No Min
Leopard – Lion	2.52895	0.29429	1.95579 – 3.09698	None
(Leopard-Lion) – Jaguar	3.12846	0.27900	2.56703 – 3.65990	(1) Min = 1.6
Tiger – Snow Leopard	3.19831	0.25813	2.70868 – 3.71997	No Min
(Tiger- Snow Leopard) – (Leopard-Lion-Jaguar)	3.93983	0.14295	3.80351 – 4.32539	(3) Min = 3.8
Leopard – Lion	2.52864	0.29452	1.96272 – 3.10415	None
(Leopard-Lion) – Jaguar	3.12868	0.28137	2.57948 – 3.67079	No Min
Tiger – Snow Leopard	3.19702	0.25972	2.70633 – 3.72084	(2) Min = 1.8
(Tiger- Snow Leopard) – (Leopard-Lion-Jaguar)	3.94055	0.14360	3.80344 – 4.32649	(3) Min = 3.8
Jaguar —Lion	2.06	From Johnson et al. (2006) (Johnson et al., 2006)		
(Jaguar-Lion) – Leopard	2.87			
Tiger – Snow Leopard	2.88			
(Tiger- Snow Leopard) – (Leopard-Lion-Jaguar)	3.72			

Fig. 8. (A) Divergence times estimated using MULTIDIVTIME with fossil calibrations: (**1**) 1.6 MYA minimum (**2**) 1.8 MYA minimum (**3**) 3.8 MYA minimum. Colors indicate node 95% credibility intervals. Black tree represents times estimated with fossil calibration (C) removed and grey tree estimated with fossil calibration (**3**) included. (B) Effects of removing each fossil calibration individually and combined are shown along with the standard error and Bayesian 95% highest posterior densities compared to dates determined by Johnson et al. (2006).

The lack of congruence of the autosomal partition from the ILD test (Supplementary Table 7) and the presence of highly varied support for multiple topologies in the signal quantification for each gene segment (Supplementary Table 12), indicates a large amount of signal heterogeneity. In such instances concatenation in a Bayesian phylogenetic inference framework, as is implemented by MrBayes, may overestimate nodal support based on posterior probabilities (Liu et al., 2008). The BEST analysis was therefore performed to estimate the final species tree from individual gene trees, allowing for heterogeneity among loci. This method has been shown to consistently estimate species trees, even when the species tree is in the "anomaly zone", a class of species trees whose most common gene tree is topologically different due to very short branches in the species tree as measured in coalescent units (Degnan et al., 2008). Species tree approaches are advantageous over strict supermatrix approaches which assume homogeneous tree topologies across loci especially within or near the anomaly zone (Edwards et al., 2007), a region that does not possess such homogeneity (Kubatko and Degnan, 2007).

The results of the BEST analysis are topologically similar, but lower in confidence, to results from ML or BI analysis of the supermatrix. This is most likely the result of post-speciation gene flow (hybridization), or lineage sorting of ancestral polymorphism. Both processes may result in decreased confidence levels in the BEST analysis (Liu and Pearl, 2007). Despite the observed heterogeneity in this dataset, the best supported topology depicted in Fig. 6 is corroborated by results from nearly every inheritable portion of the genome, as well as non-genetic characters.

4.2. Phylogeny

In two separate studies that examined morphological, ethological, and physiological features, Hemmer indicated that *Panthera* appeared to divide into two distinct clades (Hemmer, 1974, 1981). According to his studies, lions, leopards and jaguars share a specific set of common characters that distinguishes them from the second large cat clade containing the tiger. These results support the reciprocal monophyly of the lion–leopard–jaguar and tiger–snow leopard clades observed in our genetic results, a topology also supported by mitochondrial RFLP analysis (Johnson et al., 1996), as well as an analysis of excretory chemical signals (Bininda-Emonds et al., 2001). The only recent study based on nucleotide sequence data to recover this topology was the comprehensive Johnson et al. (2006) supermatrix analysis, though support values were low. All other published molecular phylogenetic studies failed to fully support this two clade distinction, probably be-

Please cite this article in press as: Davis, B.W., et al. Supermatrix and species tree methods resolve phylogenetic relationships within the big cats, *Panthera* (Carnivora: Felidae). Mol. Phylogenet. Evol. (2010), doi:10.1016/j.ympev.2010.01.036

cause they either relied heavily on mtDNA sequences that had not been vetted as true cymt amplifications, suffered from species misidentification, or lacked sufficient phylogenetic signal (Janczewski et al., 1995; Johnson and O'Brien, 1997; Mattern and McLennan, 2000; Pecon-Slattery et al., 2004; Wei et al., 2009; Yu and Zhang, 2005).

The monophyly of lion–leopard–jaguar is well supported by our supermatrix and BEST analyses, consistent with many previous morphological (Christiansen, 2008; Hemmer, 1974, 1976; Hemmer, 1978; Herrington, 1986; Salles, 1992) and molecular studies (Bininda-Emonds et al., 2001, 1999; Jae-Heup et al., 2001; Johnson et al., 1996, 2006; Pecon-Slattery et al., 2004). However, relationships between these three cat species have been difficult to resolve. Our results support lion–leopard monophyly, and are corroborated by multiple studies (Barnett et al., 2009; Bininda-Emonds et al., 2001, 1999; Christiansen, 2008; Hemmer, 1978; Herrington, 1986; Jae-Heup et al., 2001; Janczewski et al., 1995; Johnson et al., 1996), including a morphological analysis of 45 osteological, 13 soft tissue, and behavioral characters (Christiansen, 2008), which places leopard as a closer relative to lion than both the extinct American lion (*P. l. atrox*) and cave lion (*P. l. spelaea*). A recent study by Barnett et al. (2009) utilized a median-joining network analysis of mitochondrial control region and *ATP8* gene segments placed the two extinct lions closer to the extant lion, but also supported monophyly of lion and leopard. Lion–leopard monophyly was also observed in the mitochondrial RFLP study (Johnson et al., 1996), and in a characterization of the variability of the mitochondrial control region (Jae-Heup et al., 2001). It is noteworthy that these two publications were the only phylogenetic studies for *Panthera* thus far to experimentally control for numts. The Barnett study utilized the *in silico* method implemented in our study. The RFLP analysis isolated mtDNA from the nuclear fraction using a cesium chloride gradient and compared the purified cytoplasmic DNA preparations with whole cell DNA preparations to identify numts. A recent supermatrix study of felid relationships recovered lion–jaguar monophyly, the only study to support this relationship. Though this result was based on the largest number of autosomal gene segments prior to the current study, it did not analytically account for the heterogeneous gene histories for each gene segment within the autosomal partition.

The sister relationship of tiger and snow leopard is highly supported throughout most partitions, with the exception of the mtDNA partition. This exception is not surprising given the large amount of homoplasy in the mtDNA partition (Supplementary Table 5): phylogenetically informative sites can be found supporting virtually every interspecies relationship (Supplementary Table 10), and different gene segments produce different topologies (Supplementary Fig. 12). The alternative mtDNA topology, rooting on the tiger branch, appears to be the product of long-branch attraction (LBA) between an accelerated tiger lineage and the divergent outgroup, the clouded leopard (Andersson and Swofford, 2004) (Supplementary Fig. 13).

4.3. Molecular dating

Divergence times inferred by this study are consistent with an early, rapid radiation of the big cats occurring within the past ~3–4 MYA (Fig. 8) (Johnson et al., 2006). Though individual removal of each internal (within-*Panthera*) calibration point did not significantly affect the divergence times or the 95% credibility intervals, removal of the 3.8 MYA minimum constraint for the base of *Panthera* decreased divergence times at every node by roughly 40–50%. This places the divergence of lion and leopard in the Pleistocene, and places the split of jaguar from lion/leopard, as well as the divergence of snow leopard and tiger, near the Plio-Pleistocene boundary. This result calls into question the reliability of the assignment of a 3.8 MYA pantherid-like fossil to crown group *Panthera*, whose taxonomic placement is considered unclear (NCBI, 2008b), further suggesting this is an unreliable minimum constraint.

There is no evidence from divergence times (with their broad confidence intervals) and historical and present distributions that would suggest an obvious mechanism for speciation by allopatry within *Panthera*. Even today, the colinearity of felid chromosomes (Davis et al., 2009) and low sequence divergence contribute to the hybrid compatibility of the great cats (Gray, 1954), albeit subject to Haldane's rule (Haldane, 1922). The recent *Panthera* divergences estimated here further emphasizes the possibility of both lineage sorting of ancestral polymorphism, as well as post-speciation introgression between the ancestors of lion, leopard, and jaguar, and more recently both within Asia and Africa (lion and leopard). Furthermore, the probable sympatry of lion and jaguar in Pleistocene North America (Barnett et al., 2009, 2006) would be consistent with the heterogeneous gene histories observed in the autosomal partition.

4.4. Transthyretin and speciation

One of the more intriguing results from this study was the unique topology produced from the Transthyretin (*TTR*) gene segment, which produced relationships and branch lengths not reproduced by any of the 42 other gene segments or partitions in the supermatrix, regardless of size or genomic location (Fig. 5). Within the 779 bp region sequenced from *TTR* intron 1 there were a total of 10 phylogenetically informative sites differentiating the lion/leopard/jaguar clade from the snow leopard-tiger clade (5 changes per branch in the rooted topology), and zero sites differentiating snow leopard from tiger, or lion, jaguar, and leopard from each other. This is a significantly greater proportion than is observed in any other gene segment, and is in striking contrast to other gene tree topologies, including the final supermatrix topology (Fig. 6, Supplemental Figs. 5 and 12), where most substitutions accumulate on the terminal branches, rather than on the short internal branches prior to individual speciation events. To put this in the context of the divergence times displayed in Fig. 8, we would infer that the substitution rate changed from a remarkable 25 substitutions/MY within the ~200,000–300,000 years following the initial *Panthera* radiation, to zero substitutions/MY for the remaining 1.8 million years. Within the big cats, this gene segment appears to precisely follow the divergence of the two clades supported by phylogenetic signal within the other data partitions, with no interspecies divergence within each clade. The jacknifing analysis showed that when the *TTR* segment was excluded from the autosomal partition, the bipartition support for relationships changed significantly (Fig. 7). *TTR* was the only gene segment that had such a profound effect on topological rearrangement when removed from the autosomal partition (without the much longer *CES7* sequence).

The transthyretin protein has been identified as a major urinary protein (MUP) in the urine fraction of the marking fluid of the male Bengal tiger (Burger et al., 2008), and presumably other big cats, though it has not been documented in domestic cat urine (Burger et al., 2008). It has been characterized as a carrier protein for many different molecules, including retinol and thyroxine in humans (Monaco, 2000), and it is not present in urine. MUPs are involved in chemical communication in some species of mammals (Logan and Lisa Stowers, 2008; Hurst et al., 1998; Sharrow et al., 2002), including cats. Excreted proteins in the urinary fraction of the domestic cat, namely the *CES7* gene utilized here, have been shown to perform an enzymatic role in the synthesis of putative pheromone precursor proteins (Miyazaki et al., 2006). It seems probable that *TTR* serves as a carrier molecule in the urine of big cats, and

Please cite this article in press as: Davis, B.W., et al. Supermatrix and species tree methods resolve phylogenetic relationships within the big cats, *Panthera* (Carnivora: Felidae). Mol. Phylogenet. Evol. (2010), doi:10.1016/j.ympev.2010.01.036

may be involved in their territorial markings. The strong phylogenetic signal tracking the early divergence between lion–leopard–jaguar and tiger–snow leopard could be explained if this locus, or one nearby, was a speciation gene that contributed to the initial divergence of these two clades, which are independently supported by the Y chromosome, mtDNA, and supermatrix partitions.

With a reliable species tree for *Panthera* in hand, future research can focus on the genomic landscape of speciation within this recent species radiation. Chromosome- and genome-wide scans of variation and polymorphism, and identification of differences in phylogenetic topologies or molecular ages that are discordant with the ages determined here for the species tree, may identify genomic regions and genes involved in recent introgression events, as well as those regions that resist gene flow and might ultimately be involved in reproductive isolation (Kulathinal et al., 2009). Detailed analysis of these regions will allow a much more detailed story of the evolutionary history of *Panthera*, and provide a new case study for examining the genetic mechanisms behind speciation.

Acknowledgments

We would like to thank Dr. Oliver Ryder of the San Diego Zoological Society/CRES for providing male DNA samples (Supplementary Table 21); Jan Janecka, Alison Pearks-Wilkerson, and Emma Teeling provided advice, technical support and/or helpful comments on this manuscript. This work was supported in part by Grants from the National Science Foundation (EF0629849) and the Morris Animal Foundation (DO6FE-063) to W.J.M.

Appendix A. Supplementary data

Q2 Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ympev.2010.01.036.

References

Sequencher©, 2008. Gene Codes Corporation, Ann Arbor, MI.
Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–410.
Andersson, F.E., Swofford, D.L., 2004. Should we be worried about long-branch attraction in real data sets? Investigations using metazoan 18S rDNA. Mol. Phylogenet. Evol. 33, 440–451.
Baillie, J., Groombridge, B., 1996. IUCN red list of threatened animals. International Union for Conservation of Nature and Natural Resources, Gland, Switzerland.
Barnett, R., Shapiro, B., Barnes, I., Ho, S.Y.W., Burger, J., Yamaguchi, N., Higham, T.F.G., Wheeler, H.T., Rosendahl, W., Sher, A.V., Sotnikova, M., Kuznetsova, T., Baryshnikov, G.F., Martin, L.D., Harington, C.R., Burns, J.A., Cooper, A., 2009. Phylogeography of lions (*Panthera leo* ssp.) reveals three distinct taxa and a late Pleistocene reduction in genetic diversity. Mol. Ecol. 18, 1668–1677.
Barnett, R., Yamaguchi, N., Barnes, I., Cooper, A., 2006. The origin, current diversity and future conservation of the modern lion (Panthera leo). P. Roy. Soc. Lond. B. Bio. 273, 2119–2125.
Bininda-Emonds, O.R.P., Decker-Flum, D.M., Gittleman, J.L., 2001. The utility of chemical signals as phylogenetic characters: an example from the Felidae. Biol. J. Linn. Soc. 72, 1–15.
Bininda-Emonds, O.R.P., Gittleman, J.L., Purvis, A., 1999. Building large trees by combining phylogenetic information: a complete phylogeny of the extant carnivora (mammalia). Biol. Rev. 74, 143–175.
Brooks, D.R., Mayden, R.M., McLennan, D.A., 1992. Phylogeny and biodiversity: conserving our evolutionary legacy. Trends Ecol. Evol. 7, 55–59.
Buckley-Beason, V.A., Johnson, W.E., Nash, W.G., Stanyon, R., Menninger, J.C., Driscoll, C.A., Howard, J., Bush, M., Page, J.E., Roelke, M.E., Stone, G., Martelli, P.P., Wen, C., Ling, L., Duraisingam, R.K., Lam, P.V., O'Brien, S.J., 2006. Molecular evidence for species-level distinctions in clouded leopards. Curr. Biol. 16, 2371–2376.
Burger, B.V., Viviers, M.Z., J.P.I.B., Roux, M.l., Fish, N., Fourie, W.B., Weibchen, G., 2008. Chemical characterization of territorial marking fluid of male Bengal tiger, Panthera tigris. J. Chem. Ecol. 34, 659–671.
Christiansen, P., 2008. Phylogeny of the great cats (Felidae: Pantherinae), and the influence of fossil taxa and missing characters. Cladistics 24, 977–992.
Cunningham, C.W., 1997. Can tree incongruence tests predict when data should be combined? Mol. Biol. Evol. 14, 733–740.
Logan, Darren.W., Lisa Stowers, T.F.M., 2008. Species specificity in major urinary proteins by parallel evolution. PLoS One 3, e3280.
Davis, B.W., Raudsepp, T., Wilkerson, A.J.P., Agarwala, R., Schäffer, A.A., Houck, M., Ryder, O.A., Chowdhdary, B.P., Murphy, W.J., 2009. A high-resolution cat radiation hybrid and integrated FISH mapping resource for 2 phylogenomic studies across Felidae. Genomics 299, 304.
Degnan, J.H., DeGiorgio, M., Bryant, D., Rosenberg, N.A., 2008. Properties of consensus methods for inferring species trees from gene trees., arXiv:0802.2355v1 [q-bio.PE].
Edwards, S.V., Liu, L., Pearl, D.K., 2007. High-resolution species trees without concatenation. PNAS 104, 5936–5941.
Farris, J.S., Kallersjo, M., Kluge, A.G., Bult, C., 1995. Testing significance of incongruence. Cladistics 10, 315–319.
Felsenstein, J., 2004. PHYLIP (Phylogeny Inference Package) version 3.6. Department of Genome Sciences. University of Washington, Seattle, WA.
Flynn, J.J., Finarelli, J.A., Zehr, S., Hsu, J., Nedbal, M., 2005. Molecular phylogeny of the carnivora (mammalia): assessing the impact of increased sampling on resolving enigmatic relationships. Syst. Biol. 54, 317–337.
Flynn, J.J., Nedbal, M.A., 1998. Phylogeny of the Carnivora (Mammalia): congruence vs incompatibility among multiple data sets. Mol. Phylogenet. Evol. 9, 414–426.
Gray, A.P., 1954. Mammalian Hybrids. A check-list with bibliography. Commonwealth Agricultural Bureaux, Bucks, England.
Haldane, J.B.S., 1922. Sex ratio and unisexual sterility in animal hybrids. J. Genet. 12, 101–109.
Hall, T.A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp. Ser. 41, 95–98.
Hemmer, H., 1974. Untersuchungen zur stammesgeschichte der Pantherkatzen (Pantherinae). Teil III. Zur Artgeschichte des Löwen, Panthera leo (Linnaeus 1758). Veröffentlichungen der Zoologischen Staatssammlung 17, 167–280.
Hemmer, H., 1976. Fossil history of the living Felidae. In: Eaton, R.L. (Ed.), The World's Cat. Carnivore Research Institute, Burke Museum, Seattle, WA, pp. 1–14.
Hemmer, H., 1978. The evolutionary systematics of living Felidae: present status and current problems. Carnivore 1, 71–79.
Hemmer, H., 1981. Die evolution der Pantherkatzen: Modell zur überprüfung der brauchbarkeit der hennigschen prinzipien der phylogenetischen systematik für wirbeltierpaläontologische studien. Paläontologische Zeitschrift 109, 116.
Herrington, S.J., 1986. Phylogenetic relationships of the wild cats of the world. PhD dissertation. University of Kansas, Lawrence.
Hurst, J.L., Robertson, D.H.L., Tolladay, U., Beynon, R.J., 1998. Proteins in urine scent marks of male house mice extend the longevity of olfactory signals. Anim. Behav. 55, 1289–1297.
Jae-Heup, K., Eizirik, E., O'Brien, S.J., Johnson, W.E., 2001. Structure and patterns of sequence variation in the mitochondrial DNA control region of the great cats. Mitochondrion 1, 279–292.
Janczewski, D.N., Modi, W.S., Stephens, J.C., O'Brien, S.J., 1995. Molecular evolution of mitochondrial 12S RNA and cytochrome b sequences in the pantherine lineage of Felidae. Mol. Biol. Evol. 12, 690–707.
Jobling, M.A., Tyler-Smith, C., 2003. The human Y chromosome: an evolutionary marker comes of age. Nat. Rev. Genet. 4, 598–612.
Johnson, W.E., Dratch, P.A., Martenson, J.S., O'Brien, S.J., 1996. Resolution of recent radiations within three evolutionary lineages of Felidae using mitochondrial restriction fragment length polymorphism variation. J. Mammal. Evol. 2, 97–120.
Johnson, W.E., Eizirik, E., Pecon-Slattery, J., Murphy, W.J., Antunes, A., Teeling, E., O'Brien, S.J., 2006. The late Miocene radiation of modern Felidae: a genetic assessment. Science 311, 73–77.
Johnson, W.E., O'Brien, S.J., 1997. Phylogenetic reconstruction of the Felidae using 16S rRNA and NADH-5 mitochondrial genes. J. Mol. Evol. 44 (Suppl. 1), S98–S116.
Kim, J.-H., Antunes, A., Luo, S.-J., Menninger, J., Nash, W.G., O'Brien, S.J., Johnson, W.E., 2006. Evolutionary analysis of a large mtDNA translocation (numt) into the nuclear genome of the Panthera genus species. Gene 366, 292–302.
King, V., Goodfellow, P.N., Wilkerson, A.J.P., Johnson, W.E., O'Brien, S.J., Pecon-Slattery, J., 2007. Evolution of the male-determining gene SRY Within the cat family Felidae. Genetics 175, 1855–1867.
Kubatko, L.S., Degnan, J.H., 2007. Inconsistency of phylogenetic estimates from concatenated data under coalescence. Syst. Biol. 56, 17–24.
Kulathinal, R.J., Stevison, L.S., Noor, M.A.F., 2009. The genomics of speciation in drosophila: diversity, divergence, and introgression estimated using low-coverage genome sequencing. PLoS Genet. 5, e1000550.
Kurten, B., Anderson, E., 1980. Pleistocene mammals of North America. Columbia University Press, New York.
Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., Higgins, D.G., 2007. Clustal W and Clustal X version 2.0. Bioinformatics 23, 2947–2948.
Liu, L., Pearl, D.K., 2007. Species trees from gene trees: reconstructing Bayesian posterior distributions of a species phylogeny using estimated gene tree distributions. Syst. Biol. 56, 504–514.
Liu, L., Pearl, D.K., Brumfield, R.T., Edwards, S.V., 2008. Estimating species trees using multiple-allele DNA sequence data. Evolution 62, 2080–2091.
Masuda, R., Lopez, J.V., Slattery, J.P., Yuhki, N., O'Brien, S.J., 1996. Molecular phylogeny of mitochondrial cytochrome *b* and 12S rRNA sequences in the Felidae: ocelot and domestic cat lineages. Mol. Phylogenet. Evol. 6, 351–365.

Mattern, M.Y., McLennan, D.A., 2000. Phylogeny and speciation of felids. Cladistics 16, 232–253.

Miyazaki, M., Yamashita, T., Suzuki, Y., Saito, Y., Soeta, S., Tiara, H., Suzuki, A., 2006. Major urinary protein of the domestic cat regulates the production of felinine, a putative pheromone precursor. Chem. Biol. 13, 1071–1079.

Monaco, H.L., 2000. The transthyretin–retinol-binding protein complex. Biochim. Biophys. Acta 1482, 65–72.

Murphy, W.J., Pearks-Wilkerson, A.J., Raudsepp, T., Agarwala, R., Schäffer, A.A., Stanyon, R., Chowdhary, B.P., 2006. Novel gene acquisition on carnivore Y chromosomes. PLoS Genet. 2, e43.

NCBI, 2008a. Homo sapiens (human) genome build 36.3. Available at http://www.ncbi.nlm.nih.gov/mapview/map_search.cgi?chr=hum_chr.inf&query=.

NCBI, 2008b. Mus musculus (laboratory mouse) genome build 37.1. Available at http://www.ncbi.nlm.nih.gov/mapview/map_search.cgi?taxid=10090.

Neff, N.A., 1982. The Big Cats: Paintings by Guy Coheleach. Abrams, New York.

Ogata, H., Fujibuchi, W., Kanehisa, M., 1996. The size differences among mammalian introns are due to the accumulation of small deletions. FEBS Lett. 390, 99–103.

Pearks-Wilkerson, A.J., Raudsepp, T., Graves, T., Albracht, D., Warren, W., Chowdhary, B.P., Skow, L.C., Murphy, W.J., 2008. Gene discovery and comparative analysis of X-degenerate genes from the domestic cat Y chromosome. Genomics 92, 329–338.

Pecon-Slattery, J., Pearks-Wilkerson, A.J., Murphy, W.J., O'Brien, S.J., 2004. Phylogenetic assessment of introns and SINEs within the Y chromosome using the cat family Felidae as a species tree. Mol. Biol. Evol. 21, 2299–2309.

Posada, D., Buckley, T.R., 2004. Model selection and model averaging in phylogenetics: advantages of the AIC and Bayesian approaches over likelihood ratio tests. Syst. Biol. 53, 793–808.

Posada, D., Crandall, K.A., 1998. Modeltest: testing the model of DNA substitution. Bioinformatics 14, 817–818.

Ronquist, F., Huelsenbeck, J., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572–1574.

Rozen, S., Skaletsky, H.J., 2000. Primer3 on the WWW for general users and for biologist programmers. In: Krawetz, S., Misener, S. (Eds.), Bioinformatics Methods and Protocols: Methods in Molecular Biology. Humana Press, Totowa, NJ, pp. 365–386.

Salles, L.O., 1992. Felid phylogenetics: extant taxa and skull morphology (Felidae, Aeluroidea). Am. Mus. Novit. 3047, 1–67.

Schipper, J., Chanson, J.S., Chiozza, F., Cox, N.A., Hoffmann, M., Katariya, V., Lamoreux, J., Rodrigues, A.S.L., Stuart, S.N., Temple, H.J., Baillie, J., Boitani, L., Lacher Jr., T.E., Mittermeier, R.A., Smith, A.T., Absolon, D., Aguiar, J.M., Amori, G., Bakkour, N., Baldi, R., Berridge, R.J., Bielby, J., Black, P.A., Blanc, J.J., Brooks, T.M., Burton, J.A., Butynski, T.M., Catullo, G., Chapman, R., Cokeliss, Z., Collen, B., Conroy, J., Cooke, J.G., da Fonseca, G.A.B., Derocher, A.E., Dublin, H.T., Duckworth, J.W., Emmons, L., Emslie, R.H., Festa-Bianchet, M., Foster, M., Foster, S., Garshelis, D.L., Gates, C., Gimenez-Dixon, M., Gonzalez, S., Gonzalez-Maya, J.F., Good, T.C., Hammerson, G., Hammond, P.S., Happold, D., Happold, M., Hare, J., Harris, R.B., Hawkins, C.E., Haywood, M., Heaney, L.R., Hedges, S., Helgen, K.M., Hilton-Taylor, C., Hussain, S.A., Ishii, N., Jefferson, T.A., Jenkins, R.K.B., Johnston, C.H., Keith, M., Kingdon, J., Knox, D.H., Kovacs, K.M., Langhammer, P., Leus, K., Lewison, R., Lichtenstein, G., Lowry, L.F., Macavoy, Z., Mace, G.M., Mallon, D.P., Masi, M., McKnight, M.W., Medellin, R.A., Medici, P., Mills, G., Moehlman, P.D., Molur, S., Mora, A., Nowell, K., Oates, J.F., Olech, W., Oliver, W.R.L., Oprea, M., Patterson, B.D., Perrin, W.F., Polidoro, B.A., Pollock, C., Powel, A., Protas, Y., Racey, P., Ragle, J., Ramani, P., Rathbun, G., Reeves, R.R., Reilly, S.B., Reynolds III, J.E., Rondinini, C., Rosell-Ambal, R.G., Rulli, M., Rylands, A.B., Savini, S., Schank, C.J., Sechrest, W., Self-Sullivan, C., Shoemaker, A., Sillero-Zubiri, C., De Silva, N., Smith, D.E., Srinivasulu, C., Stephenson, P.J., van Strien, N., Talukdar, B.K., Taylor, B.L., Timmins, R., Tirira, D.G., Tognelli, M.F., Tsytsulina, K., Veiga, L.M., Vie, J.-C., Williamson, E.A., Wyatt, S.A., Xie, Y., Young, B.E., 2008. The status of the world's land and marine mammals: diversity, threat, and knowledge. Science 322, 225–230.

Sharrow, S.d., Vaughn, J.L., Zidek, L., Novotny, M.V., Stone, M.J., 2002. Pheromone binding by polymorphic mouse major urinary proteins. Protein Sci. 11, 2247–2256.

Shimodaira, H., Hasegawa, M., 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol. Biol. Evol. 16, 1114–1116.

Skaletsky, H., Kuroda-Kawaguchi, T., Minx, P.J., Cordum, H.S., Hillier, L., Brown, L.G., Repping, S., Pyntikova, T., Ali, J., Bieri, T., Chinwalla, A., Delehaunty, A., Delehaunty, K., Du, H., Fewell, G., Fulton, L., Fulton, R., Graves, T., Hou, S.-F., Latrielle, P., Leonard, S., Mardis, E., Maupin, R., McPherson, J., Miner, T., Nash, W., Nguyen, C., Ozersky, P., Pepin, K., Rock, S., Rohlfing, T., Scott, K., Schultz, B., Strong, C., Tin-Wollam, A., Yang, S.-P., Waterston, R.H., Wilson, R.K., Rozen, S., Page, D.C., 2003. The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. Nature 423, 825–837.

Song, H., Buhay, J.E., Whiting, M.F., Crandall, K.A., 2008. Many species in one: DNA barcoding overestimates the number of species when nuclear mitochondrial pseudogenes are coamplified. PNAS 105, 13486–13491.

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

Sullivan, J., 1996. Combining data with different distributions of among-site variations. Syst. Biol. 45, 375–380.

Suzuki, Y., Glazko, G.V., Nei, M., 2002. Overcredibility of molecular phylogenies obtained by Bayesian phylogenetics. PNAS 99, 16138–16143.

Swofford, D.L., 2002. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods) Version 4. Sinauer Associates, Sunderland, MA.

Thorne, J.L., Kishino, H., 2002. Divergence time and evolutionary rate estimation with multilocus data. Syst. Biol. 51, 689–702.

Wei, L., Wu, X., Jiang, Z., 2009. The complete mitochondrial genome structure of snow leopard Panthera uncia. Mol. Biol. Reports 36, 871–878.

Werdelin, L., Lewis, M.E., 2005. Plio-Pleistocene Carnivora of eastern Africa: species richness and turnover patterns. Zool. J. Linn. Soc. 144, 121–144.

Wu, X., Zheng, T., Jiang, Z., Wei, L., 2007. The mitochondrial genome structure of the clouded leopard (Neofelis nebulosa). Genome 50, 252–257.

Yang, Z., 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. 24, 1586–1591.

Yu, L., Li, Q.W., Ryder, O.A., Zhang, Y.P., 2004. Phylogenetic relationships within mammalian order Carnivora indicated by sequences of two nuclear DNA genes. Mol. Phylogenet. Evol. 33, 694–705.

Yu, L., Zhang, Y.P., 2005. Phylogenetic studies of pantherine cats (Felidae) based on multiple genes, with novel application of nuclear beta-fibrinogen intron 7 to carnivores. Mol. Phylogenet. Evol. 35, 483–495.

Please cite this article in press as: Davis, B.W., et al. Supermatrix and species tree methods resolve phylogenetic relationships within the big cats, *Panthera* (Carnivora: Felidae). Mol. Phylogenet. Evol. (2010), doi:10.1016/j.ympev.2010.01.036