

Research article

Open Access

Analysis of complete mitochondrial genome sequences increases phylogenetic resolution of bears (Ursidae), a mammalian family that experienced rapid speciation

Li Yu*^{†1}, Yi-Wei Li^{†1}, Oliver A Ryder³ and Ya-Ping Zhang*^{2,1}

Address: ¹Laboratory for Conservation and Utilization of Bio-resource, Yunnan University, Kunming 650091, China, ²State Key Laboratory of Genetic Resource and Evolution, Kunming Institute of Zoology, Kunming 650223, China and ³Conservation and Research for Endangered Species, Zoological Society of San Diego, PO Box 551, San Diego, CA 92112, USA

Email: Li Yu* - yuli1220@yahoo.com.cn; Yi-Wei Li - azalee713@gmail.com; Oliver A Ryder - oryder@ucsd.edu; Ya-Ping Zhang* - zhangyp1@263.net.cn

* Corresponding authors †Equal contributors

Published: 24 October 2007

Received: 9 February 2007

BMC Evolutionary Biology 2007, **7**:198 doi:10.1186/1471-2148-7-198

Accepted: 24 October 2007

This article is available from: <http://www.biomedcentral.com/1471-2148/7/198>

© 2007 Yu et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: Despite the small number of ursid species, bear phylogeny has long been a focus of study due to their conservation value, as all bear genera have been classified as endangered at either the species or subspecies level. The Ursidae family represents a typical example of rapid evolutionary radiation. Previous analyses with a single mitochondrial (mt) gene or a small number of mt genes either provide weak support or a large unresolved polytomy for ursids. We revisit the contentious relationships within Ursidae by analyzing complete mt genome sequences and evaluating the performance of both entire mt genomes and constituent mtDNA genes in recovering a phylogeny of extremely recent speciation events.

Results: This mitochondrial genome-based phylogeny provides strong evidence that the spectacled bear diverged first, while within the genus *Ursus*, the sloth bear is the sister taxon of all the other five ursines. The latter group is divided into the brown bear/polar bear and the two black bears/sun bear assemblages. These findings resolve the previous conflicts between trees using partial mt genes. The ability of different categories of mt protein coding genes to recover the correct phylogeny is concordant with previous analyses for taxa with deep divergence times. This study provides a robust Ursidae phylogenetic framework for future validation by additional independent evidence, and also has significant implications for assisting in the resolution of other similarly difficult phylogenetic investigations.

Conclusion: Identification of base composition bias and utilization of the combined data of whole mitochondrial genome sequences has allowed recovery of a strongly supported phylogeny that is upheld when using multiple alternative outgroups for the Ursidae, a mammalian family that underwent a rapid radiation since the mid- to late Pliocene. It remains to be seen if the reliability of mt genome analysis will hold up in studies of other difficult phylogenetic issues. Although the whole mitochondrial DNA sequence based phylogeny is robust, it remains in conflict with phylogenetic relationships suggested by analysis of limited nuclear-encoded data, a situation that will require gathering more nuclear DNA sequence information.

Background

The Ursidae is a major family of the order Carnivora, comprising eight species. They are generally classified into three genera: *Ailuropoda* (giant panda), *Tremarctos* (spectacled bear), and *Ursus* (brown, polar, sloth, sun, and Asiatic and American black bears) [1-3]. Despite the small number of ursid species, bear phylogeny has long been a focus of study due to their conservation value, as all bear genera have been classified as endangered at either the species or subspecies level. The ultimate recognition of the giant panda *Ailuropoda melanoleuca* as the most basal offshoot in the bear family, after more than a century of debate since this species' discovery in 1864, was one of the most spectacular events in the history of modern phylogenetics [4-8]. This achievement also sparked controversy regarding the relationships among the other members of Ursidae and the possibility of an extremely recent radiation for genus *Ursus* [6,8,9] (6 million years ago). Up to now, only the subsequent divergence of *Tremarctos ornatus* (the spectacled bear) to the giant panda, and the sister grouping of *Ursus arctos* (the brown bear) and *Ursus maritimus* (the polar bear) have been unambiguously accepted. All other relationships within the genus *Ursus* remained an unresolved polytomy (Figure 1).

Finding a valid genetic marker that offers sufficient variation to distinguish among recently divergent species posed a major challenge to advancing the understanding of ursid phylogeny. Previous investigations of phylogenetic relationships among the bear species mainly utilized analysis of portions of a single mitochondrial (mt) gene or a small number of mt genes [8,10-13]. In general, mtDNA accumulates mutations at a relatively faster rate and has a shorter expected coalescence time than other types of sequence data, e.g. nuclear DNA, thus making it particularly useful for revealing closely spaced branching events [14-17]. However, none of the previous mt analyses provided conclusive resolution for this low-level phylogeny. Additionally, analyses of different genes within the mt genome have resulted in inconsistent branching patterns being reported in the Ursidae (Figure 1A-E). Recently, several lines of evidence have demonstrated that using sufficiently large amounts of mtDNA sequence data, e.g. the whole mt genome, is a powerful way to ameliorate the discordances and poor resolution that plague analyses based on single genes or segments [16,18-22]. Thus, as a further step toward the understanding of Ursidae phylogeny, it was highly desirable to address this evolutionary question from a mitogenomic perspective.

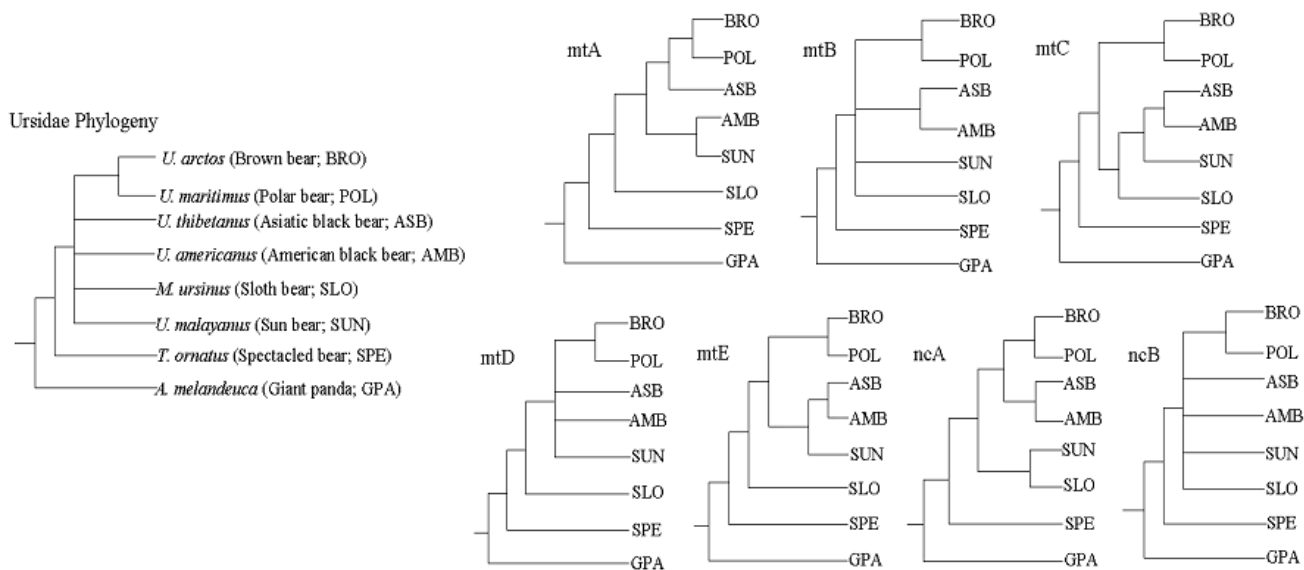


Figure 1

Long-standing unresolved Ursidae phylogeny (tree in the left) and competing hypotheses proposed based on previous sequence data (mtA-E and nuA-B). Trees were constructed from (mtA) combined analysis of partial control region, 12S_rRNA, CYTB, tRNA^{Pro}, and tRNA^{Thr} mt genes [11], (mtB) MP analysis of complete CYTB, tRNA^{Pro}, and tRNA^{Thr} mt genes [12], (mtC) NJ analysis of complete CYTB, tRNA^{Pro}, and tRNA^{Thr} mt genes [12], (mtD) combined analysis of partial control region, CYTB, ND4, ND5, COII, and 16S_rRNA mt genes [8], (mtE) combined analysis of partial control region, 12S_rRNA, complete CYTB, tRNA^{Pro}, and tRNA^{Thr} mt genes [13], (nuA) combined analysis of interphotoreceptor retinoid binding protein (IRBP) exon I and transthyretin (TTR) intron I nuclear genes [13], and (nuB) combined analysis of four type I sequence-tagged sites (STS) and IRBP exon I nuclear genes [60].

On the basis of the most comprehensive molecular data set assembled to date for the Ursidae, the present work revisits the contentious relationships within genus *Ursus* by analyzing complete mt genome sequences from all representatives of bears and, for the first time, evaluates the performance of both entire mt genomes and constituent mtDNA genes in recovering a phylogenetic tree within a rapid, recent radiation. The improved reconstruction of ursid relationships utilizing the entire mt genome also permitted a refined dating of evolutionary divergence among these bear species. Our research thus not only generates a strong Ursidae phylogenetic framework for future validation using additional evidence but, most significantly, provides a model system against which to examine the usefulness of mt genomes for resolving difficult phylogenies with rapid species radiation.

Results

Sequence Characteristics

The general characteristics of eight bear mt genomes are summarized in Table 1. These complete mt genomes range from 16,746–17,020 bp in size. Length differences are largely due to the variation in copy number of tandem repeated sequences in the conserved sequence block (CSB) domains of the mt control region. All genomes share not only 13 protein-coding genes, 22 tRNAs genes, 2 rRNAs, and a control region, but also the same gene order. The overall average nucleotide composition of bear mt genomes is A = 30.9%, C = 25.0%, G = 15.6%, and T = 28.5%. The constancy of nucleotide base composition was examined using the chi-square test with PUZZLE for different subsets of mt genome (defined in Methods), and heterogeneous nucleotide composition ($p < 0.05$) was observed in the combined protein-coding gene dataset and the complete dataset (as described below). The K2P distances [23] among seven ingroup taxa calculated using MEGA3 [24] range from 2.3 to 19.9% for the protein-coding dataset (average 12.8%), from 1.2 to 10.1% for the rRNA dataset (average 6.1%), from 1.3 to 8.7% for the

tRNA dataset (average 5.2%), from 5.0 to 18.9% for the control region (average 10%), and from 2.2 to 17.0% for the complete dataset (average 10.7%).

Reconstructing the Phylogenetic Relationships

The combined data set of 12 protein-coding genes (10882 aligned nucleotide sites; 3534 variable and 1970 parsimony-informative) produced a single most-parsimonious tree of 5512 steps without MP weighting (Figure 2A). In this tree, the spectacled bear diverged earliest (MP BS = 100%), followed by the sloth bear (MP BS = 67%). Sister-group relationships were indicated between the Asiatic black and the American black bear (MP BS = 86%), as well as between the brown and the polar bear (MP BS = 100%), while the sun bear clustered with the two black bears (MP BS = 65%). Because the giant panda sequence deviates significantly from the mean nucleotide frequency on the 3rd codon position by a 5% chi-square test ($p < 0.05$), the phylogenetic analysis was therefore also conducted under different weighting schemes, including P12 and RY-coding methods (see Methods). In all cases, an identical topology to that of unweighted analysis was obtained, but the close relatedness between the Asian black and American black bears was less well supported (MP BS < 50%; Figure 2A). Generally, better resolution and stronger bootstrap supports were obtained from DNA datasets that included all substitutions than from those subjected to the weighting. As an alternative attempt to evaluate the effect of compositional bias on the reconstructed tree, we reanalyzed the data without the giant panda and used spectacled bear for rooting. This approach provided a data set without significant base composition variation ($p > 0.05$). Interestingly, the resulting tree topology remained constant but there was a noticeable effect on the nodal support, where all relationships in the tree were robustly identified (MP BS > 85%; Figure 2A). Particularly, BS for the positions of the sloth bear and the sun bear increased to 90% and 87%, respectively. ML and partitioned Bayesian analyses (using distinct models and rates for each

Table 1: General characteristics of eight bear mt genomes

Taxa				Genome Length(bp)					G+C nucleotide content (%)				
Genus	Scientific Name	Sample Source	Accession No.	total	protein coding	rRNAs	tRNAs	control region	total	protein-coding	rRNAs	tRNAs	control region
<i>Ursus</i>	<i>U. arctos</i>	----	AF303110 [38]	17,020	10,882	2,541	1,511	1,578	40.8	41.5	40.5	35.9	41.6
	<i>U. maritimus</i>	----	AF303111 [38]	17,017	10,882	2,542	1,511	1,575	40.8	41.4	40.6	36.2	42.0
	<i>U. thibetanus</i>	Yunnan Province, China	EF196661 (this study)	16,795	10,877	2,548	1,512	1,346	40.7	41.4	40.3	35.9	40.6
	<i>U. americanus</i>	----	AF303109 [38]	16,841	10,882	2,545	1,512	1,396	40.4	41.0	40.2	36.4	40.3
	<i>U. malayanus</i>	Yunnan Province, China	EF196664 (this study)	16,783	10,882	2,546	1,512	1,337	40.7	41.4	40.2	35.9	40.2
	<i>U. ursinus</i>	San Diego Zoo, USA	EF196662 (this study)	16,817	10,882	2,545	1,511	1,371	41.6	42.3	40.9	36.9	41.7
<i>Tremarctos</i>	<i>T. ornatus</i>	San Diego Zoo, USA	EF196665 (this study)	16,766	10,882	2,553	1,510	1,315	41.1	41.9	40.1	35.7	43.0
<i>Ailuropoda</i>	<i>A. melandeuca</i>	Sichuan Province, China	EF196663 (this study)	16,746	10,880	2,551	1,515	1,286	38.5	38.3	39.0	37.0	41.1

protein gene) of the dataset, whether the giant panda or the spectacled bear was used as outgroup, showed the same tree topology as Figure 2A. All nodes in the ML and Bayesian trees received high BP ($\geq 85\%$) and PP (≥ 0.99) except for the position of the sloth bear. Tables in Figure 2A illustrate the confidence level of the nodal relationships under all analytical approaches.

Both of the combined RNA data sets (rRNAs and tRNAs) demonstrated reduced resolving power for phylogenetic inference compared to protein-coding gene analysis (Figure 2B–D). The aligned rRNA sequences (combined 12S and 16SrRNA genes) were 2574 bp in length, of which 480 nucleotide sites were variable and 222 were parsimony-informative. MP analysis yielded two equally most-parsimonious trees of 683 steps. One of them is topologically identical to the protein-coding gene tree shown in Figure 2A. Although most relationships collapsed on the

50% majority-rule consensus of the two parsimonious trees (Figure 2B), the two black bears and the sun bear formed a clade on the ML (ML BS = 62%) and partitioned Bayesian trees (distinct models and rates for two rRNA genes; PP = 0.95; Figure 2C), a relationship also supported in the protein-coding gene analysis. Interestingly, when stem-loop secondary structures were considered, the single most-parsimonious tree (495 steps) based on the loop region (1283 bp) identified the same topology as Figure 2A. In contrast, 50% majority-rule consensus of three equally most-parsimonious trees (186 steps) based on the stem region (1291 bp) recovered a tree topology that differed in placing the Asiatic black bear as basal to the rest of *Ursus*, while joining the American black bear, the sun bear, and the sloth bear on a common branch (data not shown). However, nodal supports for most relationships in the stem and loop trees are below 50%, with the exception of the earliest branching of the spectacled bear among the in-group and the close association of the brown and polar bears (MP BS = 100%).

The tRNA data set (combined 22 tRNA genes) contained 1518 bp of aligned sites, of which 256 were variable and 100 were parsimony-informative. Parsimony analysis produced a most-parsimonious tree of 341 steps with a different topology from those of protein-coding and rRNA gene analyses. According to this tree, the sloth bear was grouped with the Asiatic black bear, and they are placed as a clade sister to the lineage leading to the American black bear and the sun bear. However, in this tree only the basal position of the spectacled bear, and the close association of the brown bear and the polar bear, were convincingly supported (MP BS = 100%; Figure 2D); none of the other relationships received MP BS larger than 50%. ML and Bayesian analyses produced a similar tree topology and nodal support as the MP tree.

In the control region of the mt genome, tandem repeated sequences and ambiguously aligned regions were excluded, leaving 1008 positions in the phylogenetic analysis. 316 were variable and 124 were parsimony-informative. MP reconstruction yielded a most-parsimonious tree of 463 steps. The striking topological difference from protein-coding and RNA gene analyses was the positioning of the sun bear, Asian black bear, and American bears as the successive sister taxa to the clade consisting of the brown and polar bears. However, only the branch that separates the spectacled bear from *Ursus* and the basal diverging of the sloth bear within *Ursus*, and the affinity of the brown and polar bears, received $>50\%$ bootstrap support (Figure 2E). ML and Bayesian analyses of the same data set found a large unresolved polytomy leading to most bear species, and only the sister-grouping of the brown and polar bears was strongly supported (ML BS = 99% and PP = 1.00; Figure 2F).

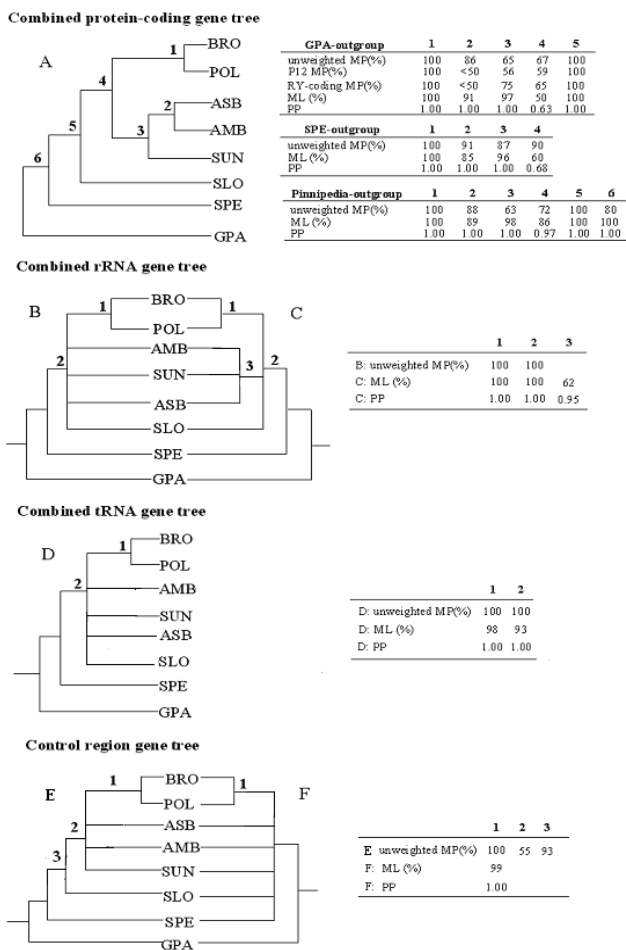


Figure 2 Phylogenetic trees and nodal supports (groups that received more than 50% BS or 0.6 PP were retained) based on analyses of different subsets of mt genome.

To maximize the amount of phylogenetic information, we also pooled all mt protein-coding genes, RNA genes, and control regions to form a single data set with a length of 15982 aligned sites for "genome phylogeny" reconstruction, of which 4586 were variable and 2416 were parsimony-informative. A unique topology exactly identical to that produced by the combined protein-coding gene analysis, but with higher statistical nodal supports, was obtained by all three analytical approaches (Figure 3; 70–100% BS and 0.85–1.00 PP). The tree topology was not affected by exclusion of the giant panda sequence ($p < 0.05$ with PUZZLE) and moreover, by rooting the tree with the spectacled bear, the supports for most branches increased to convincing statistical significance (≥ 85 BS and ≥ 0.95 PP), increasing our confidence in the result. The complete mtDNA genome-based phylogeny of the Ursidae incorporates the largest amount of phylogenetically informative sequence-based characters and provides the most robust tree for this carnivore family in terms of resolved topology and support for nodes.

Assessing the Performance of Individual Genes

The determination of complete mt sequences from all bear species affords the opportunity not only to utilize these data to estimate the phylogeny of the Ursidae but also, to examine characteristics of the evolution of mitochondrial protein coding and non-coding genes. It is of interest to evaluate their individual performance in supporting the complete mtDNA-based phylogeny within such a recent radiation involving specializations to a variety of habitats and reproductive patterns. For this reason, unweighted MP analysis was also performed on protein-coding and rRNA genes individually (Figure 4; only >50% MP BS are indicated on the branches). The earliest split of the spectacled bear and the grouping of the brown bear and polar bear, which are the only two strongly supported hypotheses in Ursidae phylogeny to date, are observed in all of these single-gene trees (61–100% and 76–100% MP BS, respectively) whereas the position of the other bear

species in *Ursus* was either not recovered at all or varied considerably with little or no nodal support. Among these single gene trees, only the CYTB and ND5 trees have the same branching order as that from the combined all gene analysis (Figure 3).

To better assess which grouping in whole-genome phylogeny was supported by different parts of the mt genome (i.e., protein-coding, RNA genes, and control region), we performed partitioned Bremer support (PBS) analysis and the result is shown in Figure 5. The partial Bremer index for each data partition is determined by subtracting the number of steps for that partition in the most parsimonious tree from the number of steps for that partition in the shortest tree lacking the node in question [25]. Partial decay indices may be either positive or negative for an individual data partition. PBS analysis indicated that among the 16 gene partitions examined (Figure 5), the ND5 gene provides the greatest contribution to the complete gene tree resolution ($124/750 = 16.53\%$) while the ND3 gene contributes the least ($9/750 = 1.30\%$). From the results of PBS analysis, we find that all genes allocated Bremer supports predominately in the traditionally well-established relationships, i.e., branches 1 and 5 in the tree (94.93% in total; data not shown). Leaving these two branches out of the analysis, the ND5 gene still provided the largest proportion of PBS values ($19/40 = 47.50\%$) compared to those of the remaining branches (Figure 5). In contrast, the ND1 gene provided the highest conflict values ($-14/40 = -35.00\%$). Whether branches 1 and 5 were included or not, PBS analyses give the rough appearance of relatively superior performance of ND5, ND4, CYTB, 16SrRNA, and ND2 genes, and on the other hand, of the poor utility of ND3, ND4L and ND1 genes. Combined tRNAs, COX2, 12SrRNA genes, and control region are intermediate (Figure 5).

Dating the Evolutionary Divergences

Results of a likelihood ratio test of the molecular-clock hypothesis for the five data sets, as defined for the phylogenetic analysis, are presented in Table 2. The combined protein-coding data set was used in estimating divergence times of Ursidae radiation, in view of its relatively good tree resolution and constant evolutionary rates across all taxa. By applying a minimum paleontological date of separation between the giant panda and the other bear species of 12 million years ago [26,27] (MYA), which has also been used as an established reference in all previous studies of Ursidae [8,12,13], we inferred that the divergence of the spectacled bear from the ursine clade occurred at 10.91 MYA (95% confidence intervals = 9.93–11.89 MYA). The six closely related bears in genus *Ursus* began their recent radiation from a common ancestor at 6.34 MYA (95% confidence intervals = 5.95–6.73 MYA; node 5 in Figure 3). Similar thinking applied to the genus

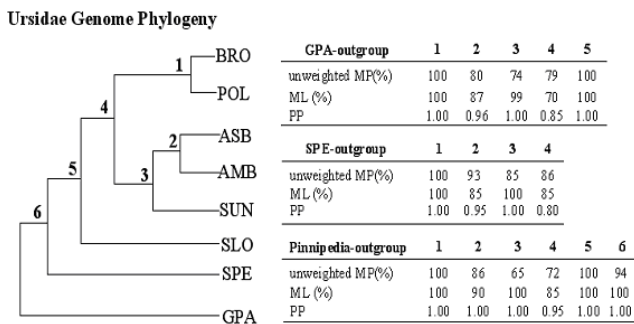


Figure 3
Mt genome tree and nodal supports.

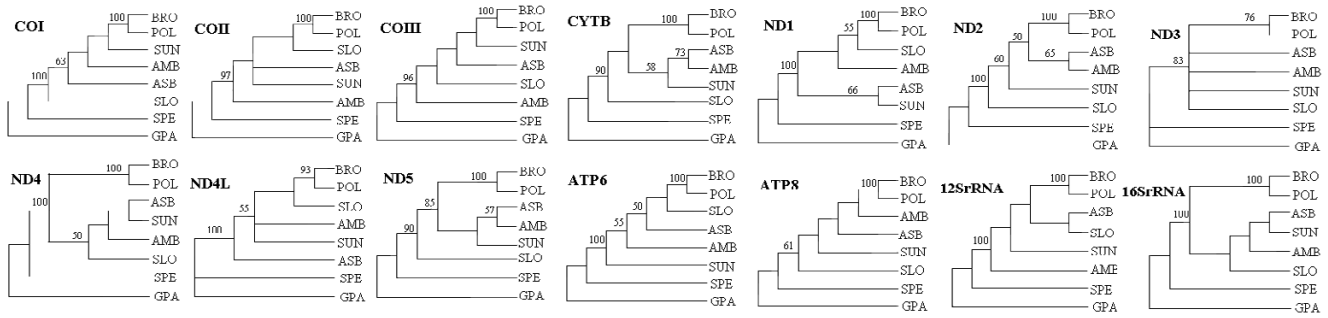


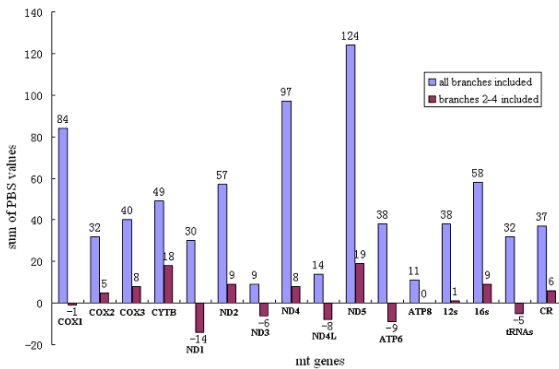
Figure 4
Phylogenetic trees based on individual mt genes. Only >50% MP BS are indicated on the branches.

Ursus suggests a divergence time between the brown and polar bears and the two black bears/sun bear of 6.13 MYA (95% confidence intervals = 5.54–6.72 MYA; node 4), and that between the sun bear and the two black bears of 5.673 MYA (95% confidence intervals = 5.09–6.26 MYA; node 3). The dating for the divergence between the Asian black and American black bear was 5.19 MYA (95% confidence intervals = 4.6–5.78 MYA; node 2), and that between the brown bear and polar bear 1.32 MYA (95% confidence intervals = 0.93–1.71 MYA; node 1). Table 3 shows these divergence time evaluations and also those in previous studies for comparison.

Discussion

Among mammalian phylogenies, those characterized by rapid species radiations have long been one of the most plaguing and challenging problems in species tree reconstruction [28]. This is the first study utilizing data from whole mitochondrial genome sequences from ursids, an approach that allows increased phylogenetic resolution of the Ursidae family, whose origin can be traced back to the extremely recent mid-Miocene [6,9] (15–20 MYA). Previous molecular studies relevant to Ursidae phylogeny provided either an inconsistent view on the issue or weak statistical support for discriminating alternative hypotheses (Figure 1). The branching event following the divergence of the spectacled bear has long been a large unresolved polytomy leading to six *Ursus* species, of which only the sister-relationship between the brown bear and polar bear was unanimously favored. The close relatedness of brown bear and polar bear, as well as the paraphyletic association between mtDNA of these two bears has been upheld in previous studies [29–31]. More sequences of brown bear and polar bear included in the future research will help test further the earlier observations. In sum, the long-standing lack of full resolution within Ursidae may be primarily due to the low level of variation harbored in much shorter sequences than those used in this study.

Based on the largest available mt data set from Ursidae, our genome phylogeny provides strong evidence that within genus *Ursus*, the sloth bear is the sister taxa of all the other five ursines, and that the latter group is divided into the brown bear/polar bear and the two black bears/sun bear assemblages, upholding and strengthening the hypothesis drawn by our previous analysis of the five fragments of mtDNA [13] (Figure 1E). Alternative hypotheses for a mitochondrial sequence based phylogeny are not supported when the entire mitochondrial DNA sequence information is utilized. In particular, when nucleotide base compositional bias introduced by the giant panda outgroup was removed from our analysis and the specta-



	cox1	cox2	cox3	cytb	nd1	nd2	nd3	nd4	nd4l	nd5	atp6	atp8	12s	16s	tRNAs	CR
Zardoya and Meyer (1996)	G	M	M	G	M	G	P	G	P	G	P	P	-	-	-	-
Russo, Takezaki, and Nei (1996)	-	P	-	G	P	-	G	P	G	P	G	-	-	-	-	-
this study	G/M	M	M/G	G	P	G	P	G	P	G	M/P	P/M	M	G	M	M

G, M, P stand for good, medium, poor performers
 G/M, M/G, M/P, P/M stand for classifications when all branches/branches 2-4 included in this study
 - stands for no performance evaluation in the study
 CR stands for control region, tRNAs stands for combined tRNAs

Figure 5
Results of partitioned Bremer support (PBS) analyses to each node on the mt genome tree and comparisons of phylogenetic performance of mt genes among the studies.

Table 2: Results of likelihood ratio test of the molecular-clock hypothesis

	combined protein-coding gene	combined rRNAs gene	combined tRNAs gene	control region	combined all gene
Log L without clock	-36253.65	-6614.67	-3682.41	-3351.94	-50256.75
Log L with clock	-36258.55	-6616.48	-3685.10	-3431.73	-30265.12
significance level (p)	0.13	0.73	0.49	<0.01	<0.05

cluded bear was used for rooting, the overall Ursidae relationships recovered have the largest statistical support in comparison to all other previously proposed hypotheses. To further examine the sensitivity of our tree topology to outgroup choice, we selected Pinnipedia. This superfamily in Carnivora includes the Otariidae, Phocidae, and Odobenidae families, members of which have been used as alternative roots in earlier studies of bears [8,12].

Combined protein-coding genes (10888 aligned sites, 4522 variable and 3245 parsimony-informative) and all mt genes (15017 aligned sites, 5642 variable and 3926 parsimony-informative; control region was not included due to alignment difficulty) analyses of the eight bear species using available Pinnipedia mt genomes as outgroups (Accession No. [AJ428576](#), [AJ428578](#) and [X63726](#); [32,33]) gave an identical tree topology to that obtained using the giant panda as the outgroup. Support levels for most branches were similar to those estimated with the giant panda outgroup except for an increased ML BS (≥ 85) and PP (≥ 0.95 PP) of placement of the sloth bear (Figure 2A and 3). A chi-square test of composition stationarity showed that both Pinnipedia and the giant panda have deviant base composition with respect to the other bears ($p < 0.05$), a circumstance that might have a negative impact on the branch supports in MP analysis. Thus, our genome phylogeny was robust with either outgroup, though the spectacled bear, exhibiting the least phylogenetic noise, appears a more favorable outgroup for Ursidae in terms of overall support levels. Taken together, the present genome result significantly resolved the conflict between those trees using partial mt genes [8,11-13] and

represents the most probable explanation of bear evolution.

Nevertheless, a more in-depth understanding of the Ursidae relationships will definitely benefit from the addition of independent sequence data, considering that the genome phylogeny obtained here is based on a single and haploid linkage unit. The necessity of including other unlinked genes for phylogenetic resolution of the Ursidae is also illustrated by the fact that our recent study on two nuclear genes, transthyretin (TTR) and interphotoreceptor retinoid binding protein (IRBP), has united the sloth bear and the sun bear as sister taxa with high statistical support [13] (Figure 1 nuA), a relationship not consistent with mt gene analyses, including the present study. In mt trees, the sloth bear was mostly placed as the earliest diverging taxa among the six ursine species [8,11,13, and the present result], and the sun bear closer to the American black bear [11], the brown bear/polar bear clade [MP phylogeny in 12; ML phylogeny in 8], or the clade including the two black bears (NJ phylogeny in 12; 13 and the present result). Notably, the sister relationship between the Asian and American black bears, previously proposed on paleontological and morphological grounds [34,35] was reinforced by consistent recovery from both our mt genome and nuclear analysis. Such a grouping has also been retrieved previously with moderate support by mt analysis of complete CYTB and 2 tRNAs [12], as well as the addition of the partial D-loop region [13]. Thus, the placements of the sun bear and the sloth bear represent the most obvious discrepancy observed in the mt and nuclear trees comparisons. Our genome analysis has established a

Table 3: Divergence time evaluation of Ursidae family (Mya)

	fossil records ^a	protein electrophoresis analysis from Goldman, Giri and O'Brien (1989) [6]	partial mt gene analysis from Talbot and Shields (1996) [12]	two nuclear gene analysis from Yu et al (2004) [13]	mt genome analysis from this study
<i>T. ornatus</i> separation	5-7	10-15	12-13	6-8	10.91
six <i>Ursine</i> bears (or <i>U. ursinus</i>) divergence (node 5)	4-6	4-8	5-7	2-5	6.34
five <i>Ursine</i> bears (except <i>U. ursinus</i>) divergence (node 4)	-	-	6	-	6.13
<i>U. malayanus</i> divergence (node 3)	0.2-1 or 5-10	-	5	-	5.67
<i>U. thibetanus</i> - <i>U. americanus</i> split (node 2)	1-3.5	-	5	-	5.19
<i>U. arctos</i> - <i>U. maritimus</i> split (node 1)	0.07-0.1	2-3	1-2	1-1.5	1.32

- no divergence time evaluation in the study

a from Kurten 1968 [9]; Thenius 1979 [27]; Kurten and Anderson 1980 [35]; Savage and Russell 1983 [37]; Wayne et al. 1991 [26]

very useful benchmark that can be tested with future independent evidence.

Our genome analyses provide important insights into not only Ursidae phylogeny, but also the phylogenetic utility of different mt genes. Our data add to the well-studied performance of individual mt genes, mostly protein-coding genes, for estimating phylogeny of deep divergence [16,36], we are interested to see their relative efficiencies, adding mt RNA genes and control region as well, in those of extremely recent split. Our results suggest that combined mt protein-coding genes are more informative than the other subsets of mt genes regarding the lower-level bear relationships resolution. Only by combining all genes is it possible to reach a fully-resolved tree with moderate to strong support from MP, ML, and Bayesian methods of analysis. Ranking single genes by their respective contribution to the total PBS values of the genome tree, as a rough indicator of phylogenetic utility, reveals that some genes, such as ND5, ND4, CYTB, ND2, and 16SrRNA are better indicators of Ursidae evolution than are other genes, such as ND3, ND4L, and ND1 (Figure 5). Our results add to previous findings from Zardoya and Meyer (1996) [16] and Russo, Takezaki, and Nei (1996) [36] that did not include concatenated tRNAs, 2 rRNAs, and control regions in the evaluation of phylogenetic performance, and also agree globally with their conclusions about the rough classification of 12 mt protein-coding genes into good, medium, and poor categories. These conclusions are upheld even though significantly different evolutionary time frames between our studies and theirs (i.e., distantly related vertebrates) (Figure 5) are involved. In this sense, general knowledge of phylogenetic values of the mt genes makes it possible to preselect subsets of mt genes for different-level phylogenetic questions in the case of mt genomes unavailable. In fact, in some previous studies, Ursidae phylogenies based on the combined analysis of a few mt genes have also, to a degree, demonstrated the potential valuable information as those based on complete genome analysis [8,12].

Ursidae has one of the most extensive fossil records of extant Carnivora families [9,26,27,35,37]. Given a good fossil documentation and strongly supported phylogenetic relationships from the largest available data set, it is of interest to draw a comparison between the present mitogenomic dating results and those from previous paleontological and molecular evidence (Table 3). According to estimates based on our genome data, all the separation times within genus *Ursus* appear to have occurred in the Late Miocene or Early Pliocene (5–6 MYA), except that in the Early Pleistocene the most closely related bear species, the brown and polar bears diverged. The rather recent origins and rapid succession of these bear lineages are in line with the observation that most short mtDNA sequences

used in previous studies lack sufficiently strong phylogenetic signals and provide limited resolving power for recovering a strongly supported Ursidae phylogeny. As is seen in Table 3, our estimates of the diversification of the Ursidae family was more in agreement with those obtained with partial mt genes analysis [12] and protein electrophoresis analysis [6] than those obtained with the fossil record [9,26,35] and nuclear sequence analysis [13], which are slightly younger than the present results.

Conclusion

Identification of base composition bias and utilization of the combined data of whole mitochondrial genome sequences has allowed recovery of a strongly supported phylogeny that is upheld when using multiple alternative outgroups for the Ursidae, a mammalian family that underwent a rapid radiation since the mid- to late Pliocene.

The suggestion of Delisle and Strobeck (2002) [38] that application of mitogenomic datasets would be likely to be useful for distinguishing nodes resulting from rapid radiation episodes such as the ursine speciation events is validated by these findings. It remains to be seen if the reliability of mt genome analysis will hold up in studies of other difficult phylogenetic issues. Although the whole mitochondrial DNA sequence based phylogeny is robust, it remains in conflict with phylogenetic relationships suggested by analysis of limited nuclear-encoded data, a situation that will require gathering more nuclear DNA sequence information.

Methods

DNA Samples and Sequence Determination

The complete mt sequences of three bear species in genus *Ursus*, the polar bear (*U. maritimus*), the brown bear (*U. arctos*), and the American black bear (*U. americanus*), have been determined in previous studies of genome evolution [38]. Thus, the availability of the other five mt genome sequences from Ursidae was of considerable interest for phylogenetic reconstruction. We extracted total DNA from fresh blood or frozen tissues of the Asiatic black bear (*U. thibetanus*), the sloth bear (*U. ursinus*), the sun bear (*U. malayanus*), the spectacled bear (*Tremarctos ornatus*) and the giant panda (*Ailuropoda melanoleuca*) using standard proteinase K, phenol/chloroform extraction [39].

Mt genome sequences were initially amplified with sets of universal primers (73 in total) described in Delisle and Strobeck's original study (2002) [38]. In the case of poor PCR performance with universal primers, 31 additional species-specific oligonucleotide primers were designed (underlined in Figure 6). Primer sequence information was available upon request. A "touch-down" PCR amplification was carried out using the following parameters:

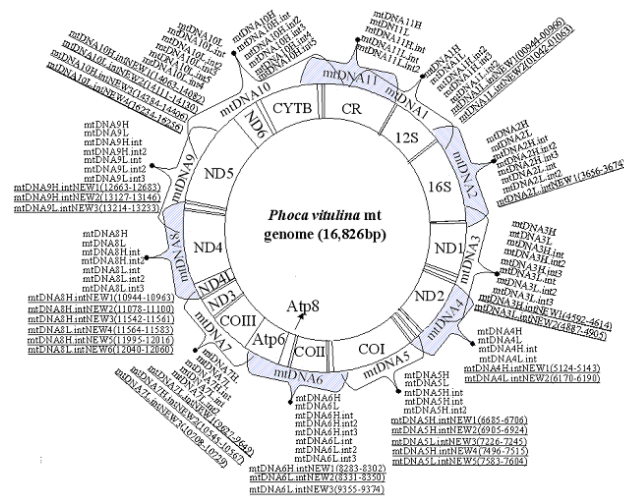


Figure 6
 Universal primers [35] and newly designed species-specific primers (underlined) used for amplifying Ursidae mt genomes. Locations of new primers are indicated in brackets and correspond to nucleotide numbers from the harbor seal sequence (Accession No. X63726). Eleven mtDNA fragments (mtDNA1-11) covering the entire mt genome are labeled as described in Delisle and Strobeck (2002) [38].

95°C hot start (5 min), 10 cycles of 94°C denaturation (1 min), 60–50°C annealing (1 min; °C/cycle), 72°C extension (1 min), and finally 25 cycles of 94°C denaturation (1 min), 50°C annealing (1 min), 72°C extension (1 min). The amplified DNA fragments were purified and sequenced in both directions with an ABI PRISM™ 3700 DNA sequencer following the manufacturer's protocol. Mt sequences obtained were checked carefully to ensure that they did not include nuclear copies of mtDNA-like pseudogenes. The exact length of the control region in the mt genome cannot be determined due to the presence of long tandem repeated sequences. Newly determined genomes have been deposited in GenBank under Accession No. EF19661–EF19665.

Sequence Data Analyses

The complete mt genome sequences of all extant bear species were aligned with program CLUSTAL X [40] and verified by eye. Five data sets comprising (1) all protein-coding genes combined except NADH6 gene, (2) 12S and 16S tRNA genes combined, (3) all 22 rRNA genes combined, (4) control region (CR), and (5) all genes combined were analyzed for phylogenetic reconstruction. The NADH6 gene is excluded from the analyses due to its anomalous nucleotide composition which can confound phylogenetic inferences.

Each of these data sets was subjected to unweighted maximum parsimony (MP) and maximum likelihood (ML) analyses using PAUP *4.0b8 [41]. For MP analyses, we adopted an exhaustive search algorithm with TBR branch swapping. For model-based ML analyses, we introduced hierarchical likelihood ratio tests (hLRT) to compare the goodness of fit of 56 nucleotide substitution models using ModelTest version 3.06 [42]. Once an appropriate model was established, a ML tree was constructed based on this model of sequence evolution. The reliability of phylogenetic relationships was evaluated by bootstrap analysis [43] for MP and ML trees (BS; 1000 replicates for MP and 100 replicates for ML). In addition, we performed a partitioned Bayesian inference (pBI) analysis [44] for phylogenetic reconstruction using MrBayes 3.1 [45], allowing a separate general time reversible (GTR) + I + Γ model and set of parameters for each gene partition, with the assumption that the underlying evolutionary process was potentially different across these partitions. Four Metropolis-coupled Markov chain Monte Carlo (MCMC) analyses were run for 2 million generations, sampling trees every 100 generations. Robustness for branches in pBI analysis was assessed by posterior probability (PP). We also conducted partitioned Bremer support analysis (PBS) [46-48] with TreeRot.v2 [25] to measure the contribution of each data partition to the total Bremer support for the nodes of genome-based tree topology.

To avoid potential tree estimation bias introduced by nucleotide composition [49,50] or saturation, two additional weighting strategies were applied in the analysis of combined 12 protein-coding genes: (1) excluding the 3rd codon positions (P12), and (2) recoding the 3rd codon position nucleotides to two-state categories, R (purine) and Y (pyrimidine), (RY-coding). The RY-coding was used here based on the growing observation that it can greatly improve consistency in phylogenetic resolution by reducing bias from differences in nucleotide composition [51-54]. In the combined analysis, portions of the 12S rRNA and 16S rRNA genes were also partitioned into two separate subsets according to their secondary structures: single-strand stems and base-paired loops [55,56].

The giant panda was used as an outgroup for estimating phylogenetic relationships within genus *Ursus*. To examine if the resulting tree topologies were sensitive to outgroup alteration, we also carried out phylogenetic analysis with Pinnipedia, a non-ursine superfamily in Carnivora, for the rooting.

The molecular clock hypothesis was examined using the likelihood-ratio test [57] with PUZZLE [58,59]. When clock-like behavior was not rejected by the test, the divergence times among them were calculated and compared to previous molecular results and fossil records.

Competing interests

The author(s) declares that there are no competing interests.

Authors' contributions

LY and YPZ designed the study. YWL carried out the experiment work. LY performed the analyses. LY, OAR and YPZ prepared the manuscript. All authors read and approved the final manuscript.

Acknowledgements

This work was supported by grants from the State Key Basic Research and Development Plan (2007CB411600), National Natural Science Foundation of China (30600067, 30460026, 30621092), and Bureau of Science and Technology of Yunnan Province, and by a start-up fund from the Yunnan University. We thank Andrea Johnson at CRES for helpful comments on the manuscript.

References

- Wayne RK, Benveniste RE, Janczewski DN, O'Brien SJ: **Molecular and biochemical evolution of the Carnivora.** In *Carnivore Behavior, Ecology, and Evolution* Edited by: Gittleman JL. Cornell Univ. Press, Ithaca; 1989:465-494.
- Nowak RM: **Walker's Mammals of the World.** Johns Hopkins Press, Baltimore; 1991.
- Kitchener AC: **A review of the evolution, systematics, functional morphology, distribution and status of the Ursidae.** *Int Zoo News* 1994, **242**:4-24.
- O'Brien SJ, Nash WG, Wildt DE, Benveniste RE: **A molecular solution to the riddle of the giant panda's phylogeny.** *Nature* 1985, **317**:140-144.
- Nash WG, O'Brien SJ: **A comparative chromosome banding analysis of the ursidae and their relationship to other carnivores.** *Cytogene Cell Genet* 1987, **45**:206-212.
- Goldman D, Giri PR, O'Brien SJ: **Molecular genetic-distance estimates among the ursidae as indicated by one- and two-dimensional protein electrophoresis.** *Evolution* 1989, **43**:282-295.
- Hashimoto T, Otaka E, Adachi J, Mizuta K, Hasegawa M: **The giant panda is closer to a bear judged by alpha- and beta-hemoglobin sequences.** *J Mol Evol* 1993, **36**:282-289.
- Waits LP, Sullivan J, O'Brien SJ, Ward RH: **Rapid radiation events in the family ursidae indicated by likelihood phylogenetic estimation from multiple fragments of mtDNA.** *Mol Phylogenet Evol* 1999, **13**:82-92.
- Kurten B: **The Pleistocene Mammals of Europe.** Aldine, Chicago. Ill; 1968.
- Zhang YP, Ryder OA: **Mitochondrial DNA sequence evolution in the Arctoidea.** *Proc Natl Acad Sci USA* 1993, **90**:9557-9561.
- Zhang YP, Ryder OA: **Phylogenetic relationships of bears (the ursidae) inferred from mitochondrial DNA sequences.** *Mol Phylogenet Evol* 1994, **3**:351-359.
- Talbot SL, Shields GF: **A phylogeny of the bears (ursidae) inferred from complete sequences of three mitochondrial genes.** *Mol Phylogenet Evol* 1996, **5**:567-575.
- Yu L, Li QW, Ryder OA, Zhang YP: **Phylogeny of the bears (Ursidae) based on nuclear and mitochondrial genes.** *Mol Phylogenet Evol* 2004, **32**:480-494.
- Brown WM, George M, Wilson AC: **Rapid evolution of animal mitochondrial DNA.** *Proc Natl Acad Sci USA* 1979, **76**:1967-1971.
- Moore WS: **Inferring phylogenies from mtDNA variation: Mitochondrial gene trees versus nuclear gene trees.** *Evolution* 1995, **49**:718-726.
- Zardoya R, Meyer A: **Phylogenetic performance of mitochondrial protein-coding genes in resolving relationships among vertebrates.** *Mol Biol Evol* 1996, **13**:933-942.
- Springer MS, Debry RW, Douady C, Amrine HM, Madsen O, de Jong WW, Stanhope MJ: **Mitochondrial versus nuclear gene sequences in deep-level mammalian phylogeny reconstruction.** *Mol Biol Evol* 2001, **18**:132-143.
- Cummings MP, Otto SP, Wakeley J: **Sampling properties of DNA sequence data in phylogenetic analysis.** *Mol Biol Evol* 1995, **12**:814-822.
- Cao Y, Adachi J, Janke A, Paabo S, Hasegawa M: **Phylogenetic relationships among eutherian orders estimated from inferred sequences of mitochondrial proteins: instability of a tree based on a single gene.** *J Mol Evol* 1994, **39**:519-527.
- Cao Y, Fujiwara M, Nikaido M, Okada N, Hasegawa M: **Interordinal relationships and timescale of eutherian evolution as inferred from mitochondrial genome data.** *Gene* 2000, **259**:149-158.
- Miya M, Kawaguchi A, Nishida M: **Mitogenomic exploration of higher teleostean phylogenies: a case study for moderate-scale evolutionary genomics with 38 newly determined complete mitochondrial DNA sequences.** *Mol Biol Evol* 2001, **18**:1993-2009.
- Ingman M, Kaessmann H, Paabo S, Gyllenstein U: **Mitochondrial genome variation and the origin of modern humans.** *Nature* 2001, **408**:708-713.
- Kimura M: **A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide.** *J Mol Evol* 1980, **16**:111-120.
- Kumar S, Tamura K, Nei M: **MEGA3: Integrated Software for Molecular Evolutionary Genetics Analysis and Sequence Alignment.** *Briefings in Bioinformatics* 2004, **5**:150-163.
- Sorenson MD: **TreeRot, version 2.** Boston Univ., Boston, MA; 1999.
- Wayne RK, VanValkenburgh B, O'Brien SJ: **Molecular distance and divergence time in carnivores and primates.** *Mol Biol Evol* 1991, **8**:297-319.
- Thenius E: **Zur systematischen und phylogenetischen Stellung des Bambusbaren Ailuropoda melanoleuca David (Carnivora:Mammalia).** *Z Sauge tierk* 1979, **44**:286-305.
- Rokas A, Carroll S: **Bushes in the tree of life.** *Plos Biol* 2006, **4**:1899-1904.
- Shields GF, Kocher TD: **Phylogenetic relationships of North American ursids based on analysis of mitochondrial DNA.** *Evolution* 1991, **45**:218-221.
- Talbot SL, Shields GF: **Phylogeography of brown bears (Ursus arctos) of Alaska and parapatry within the Ursidae.** *Mol Phylogenet Evol* 1996, **5**:477-494.
- Shields GF, Adams D, Garner G, Labelle M, Pietsch J, Ramsay M, Schwartz C, Titus K, Williamson S: **Phylogeography of mitochondrial DNA variation in brown bears and polar bears.** *Mol Phylogenet Evol* 15:319-326.
- Arnason U, Johnsson E: **The complete mitochondrial DNA sequence of the harbor seal, Phoca vitulina.** *J Mol Evol* 1992, **34**:493-505.
- Arnason U, Adegoke JA, Bodin K, Born EW, Esa YB, Gullberg A, Nilsson M, Short RV, Xu X, Janke A: **Mammalian mitogenomic relationships and the root of the eutherian tree.** *Proc Natl Acad Sci USA* 2002, **99**:8151-8156.
- Allen GM: **"The mammals of China and Mongolia." National History of Central Asia.** *Am Must Nat Hist, New York* 1938:328-336.
- Kurten B, Anderson E: **Pleistocene Mammals of North America.** Cornell Univ Press Ithaca; 1980.
- Russo CAM, Takezaki N, Masatoshi N: **Efficiencies of different genes and different tree-building methods in recovering a known vertebrate phylogeny.** *Mol Biol Evol* 1996, **13**:525-536.
- Savage DE, Russell DE: **Mammalian paleofaunas of the world.** Addison-Wesley, London; 1983.
- Delisle I, Strobeck C: **Conserved primers for rapid sequencing of the complete mitochondrial genome from carnivores, applied to three species of bears.** *Mol Biol Evol* 2002, **19**:357-361.
- Sambrook E, Fritsch F, Maniatis T: **Molecular Cloning.** Cold Spring Harbor Press, Cold Spring Harbor, NY; 1989.
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG: **The clustalx windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools.** *Nucleic Acids Research* 1997, **24**:4876-4882.
- Swofford DL: **PAUP*: phylogenetic analysis using parsimony (* and other methods). Version 4.0b8.** Sinauer Associates, Sunderland, MA; 2001.
- Posada D, Crandall KA: **Modeltest: testing the model of DNA substitution.** *Bioinformatics* 1998, **14**:817-818.

43. Felsenstein J: **Confidence limits on phylogenies: An approach using the bootstrap.** *Evolution* 1985, **39**:783-791.
44. Larget B, Simon DL: **Markov chain Monte Carlo algorithms for the Bayesian analysis of phylogenetic trees.** *Mol Biol Evol* 1999, **16**:750-759.
45. Ronquist F, Huelsenbeck JP: **MRBAYES 3: Bayesian phylogenetic inference under mixed models.** *Bioinformatics* 2003, **19**:1572-1574.
46. Bremer K: **The limits of amino acid sequence data in angiosperm phylogenetic reconstruction.** *Evolution* 1988, **42**:795-803.
47. Bremer K: **Branch support and tree stability.** *Cladistics* 1994, **10**:295-304.
48. Baker RH, DeSalle R: **Multiple sources of character information and the phylogeny of Hawaiian drosophilids.** *Syst Biol* 1997, **46**:674-698.
49. Saccone C, Lanave C, Pesole G, Preparata G: **Influence of base composition on quantitative estimates of gene evolution.** *Methods Enzymol* 1990, **183**:570-583.
50. Saccone C, Gissi C, Lanave C, Larizza A, Pesole G, Reyes A: **Evolution of the mitochondrial genetic system: an overview.** *Gene* 2000, **261**:153-159.
51. Delsuc F, Phillips MJ, Penny D: **Comment on "Hexapod origins: monophyletic or paraphyletic?".** *Science* 2003, **301**:1482.
52. Phillips MJ, Penny D: **The root of the mammalian tree inferred from whole mitochondrial genomes.** *Mol Phylogenet Evol* 2003, **28**:171-185.
53. Harrison GL, McLenachan PA, Phillips MJ, Slack KE, Cooper A, Penny D: **Four new avian mitochondrial genomes help get to basic evolutionary questions in the late cretaceous.** *Mol Biol Evol* 2004, **21**:974-983.
54. Gibson AV, Gowri-Shankar P, Higgs G, Rattray M: **A comprehensive analysis of mammalian mitochondrial genome base composition and improved phylogenetic methods.** *Mol Biol Evol* 2005, **22**:251-264.
55. Springer MS, Douzery E: **Secondary structure and patterns of evolution among mammalian mitochondrial 12S rRNA molecules.** *J Mol Evol* 1996, **43**:357-373.
56. Mears JA, Sharma MR, Gutell RR, McCook AS, Richardson PE, Caulfield TR, Agrawal RK, Harvey SC: **A structural model for the large subunit of the mammalian mitochondrial ribosome.** *J Mol Biol* 2006, **358**:193-212.
57. Felsenstein J: **Phylogenies from molecular sequences: inference and reliability.** *Annu Rev Genet* 1988, **22**:521-565.
58. Strimmer K, von Haeseler A: **Quartet puzzling: A quartet maximum likelihood method for reconstructing tree topologies.** *Mol Biol Evol* 1996, **13**:964-969.
59. Schmidt HA, Strimmer K, Vingron M, von Haeseler A: **TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing.** *Bioinformatics* 2002, **18**:502-504.
60. Fulton TL, Strobeck C: **Molecular phylogeny of the Arctoidea (Carnivora): Effect of missing data on supertree and supermatrix analyses of multiple gene data sets.** *Mol Phylogenet Evol* 2006, **41**:165-181.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

