

Accepted Manuscript

Molecular phylogeny of the Arctoidea (Carnivora): Effect of missing data on supertree and supermatrix analyses of multiple gene data sets

Tara L. Fulton, Curtis Strobeck

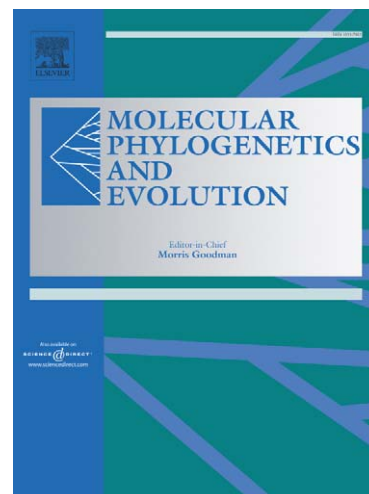
PII: S1055-7903(06)00202-8
DOI: [10.1016/j.ympev.2006.05.025](https://doi.org/10.1016/j.ympev.2006.05.025)
Reference: YMPEV 2251

To appear in: *Molecular Phylogenetics and Evolution*

Received Date: 27 September 2005
Revised Date: 3 May 2006
Accepted Date: 19 May 2006

Cite this article as: Fulton, T.L., Strobeck, C., Molecular phylogeny of the Arctoidea (Carnivora): Effect of missing data on supertree and supermatrix analyses of multiple gene data sets, *Molecular Phylogenetics and Evolution*, doi: [10.1016/j.ympev.2006.05.025](https://doi.org/10.1016/j.ympev.2006.05.025)

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



Molecular phylogeny of the Arctoidea (Carnivora): Effect of missing data on supertree and supermatrix analyses of multiple gene data sets

Tara L. Fulton* and Curtis Strobeck

Department of Biological Sciences, CW405 Biological Sciences Building, University of Alberta,
Edmonton AB Canada T6G 2E9

*Corresponding author. Fax: (780) 492-9234

Phone: (780) 492-8368

E-mail address: tara.fulton@ualberta.ca

Abstract

Phylogenetic relationships of 79 caniform carnivores were addressed based on four nuclear sequence-tagged sites (STS) and one nuclear exon, *IRBP*, using both supertree and supermatrix analyses. We recovered the three major arctoid lineages, Ursidae, Pinnipedia, and Musteloidea, as monophyletic, with Ursidae (bears) strongly supported as the basal arctoid lineage. Within Pinnipedia, Phocidae (true seals) were sister to the Otarioidea [Otariidae (fur seals and sea lions) and Odobenidae (walrus)]. Phocid subfamily and tribal designations were supported, but the otariid subfamily split between fur seals and sea lions was not. All family designations within Musteloidea were strongly supported: Mephitidae (skunks), Ailuridae (monotypic red panda), Mustelidae (weasels, badgers, otters), and Procyonidae (raccoons). A novel hypothesis for the position of the red panda was recovered, placing it as branching after Mephitidae and before Mustelidae + Procyonidae. Within Mustelidae, subfamily taxonomic changes are considered. This study represents the most comprehensive sampling to date of the Caniformia in a molecular study and contains the most complete molecular phylogeny for the Procyonidae. Our data set was also used in an empirical examination of the effect of missing data on both supertree and supermatrix analyses. Sequence for all genes in all taxa could not be obtained, so two variants of the data set with differing amounts of missing data were examined. The amount of missing data did not have a strong effect; instead, phylogenetic resolution was more dependent on the presence of sufficient informative characters. Supertree and supermatrix methods performed equivalently with incomplete data and were highly congruent; conflicts arose only in weakly supported areas, indicating that more informative characters are required to confidently resolve close species relationships.

Keywords

molecular phylogeny; nuclear DNA; Carnivora; Arctoidea; Musteloidea; Procyonidae; supertree; supermatrix; missing data

1. Introduction

Monophyly of the order Carnivora (Mammalia) is well supported through multiple lines of evidence, including morphology (Flynn et al., 1988; Wyss and Flynn, 1993), DNA sequence data (Ledje and Arnason, 1996b, 1996a; Flynn et al., 2000), total evidence (Vrana et al., 1994; Flynn and Nedbal, 1998), and supertree methods (Bininda-Emonds et al., 1999), as is the split between the two major suborders, Feliformia (cat-like carnivores) and Caniformia (dog-like carnivores). Within Caniformia, two infraorders exist: Cynoidea (or Canoidea), comprised of the single family Canidae (dogs and foxes), and Arctoidea, the bear-like carnivores. Arctoidea is further divided into three lineages of uncertain affinities: Ursidae (bears); Pinnipedia [families Phocidae (true seals), Otariidae (fur seals and sea lions), and Odobenidae (walrus)]; and Musteloidea [families Mustelidae (weasels and relatives), Procyonidae (raccoons), Mephitidae (skunks), and Ailuridae (red panda)].

The bear family, Ursidae, has been characterized by rapid radiation events, making phylogenetic inference of the species relationships difficult, and thus, often contentious (Waits et al., 1999; Yu et al., 2004a and references therein). Three subfamilies are generally recognized: Ailuropodinae (monotypic giant panda, *Ailuropoda melanoleuca*), Tremarctinae (monotypic spectacled bear, *Tremarctos ornatus*), and Ursinae (all other extant bears, genera *Ursus*, *Melursus*, and *Helarctos*). Ailuropodinae is now generally accepted as the earliest branching subfamily, though placement of the giant panda within the bear family (vs. its closest relative)

has been a longstanding controversy (see both Waits et al., 1999; Yu et al., 2004a).

Relationships among the ursine bears (sun, sloth, black, brown, and polar bears) remain relatively unresolved. The sloth bear (*Melursus ursinus*) is usually recovered as most basal, but the sun bear (*Helarctos malayans*) has been placed as sister to the brown and polar bears (*Ursus arctos* and *U. maritimus*), the American and Asian black bears (*Ursus americanus* and *U. thibetanus*), or the sloth bear, though most of these placements have been poorly supported by bootstrapping (Talbot and Shields, 1996; Waits et al., 1999; Yu et al., 2004a; Delisle and Strobeck, 2005).

Within Pinnipedia, familial relationships have been contentious, but most current studies support Otariidae and Odobenidae (sometimes superfamily Otaroidea) as sister to Phocidae (Vrana et al., 1994; Ledje and Arnason, 1996b, 1996a; Dragoo and Honeycutt, 1997; Flynn and Nedbal, 1998; Bininda-Emonds et al., 1999; Flynn et al., 2000; Davis et al., 2004; Flynn et al., 2005). The family Otariidae is comprised of two subfamilies: Otariinae (sea lions) and Arctocephalinae (fur seals). These subfamily designations may be inappropriate in light of evidence of various types suggesting that both Arctocephalinae and Otariinae may be paraphyletic (Bininda-Emonds et al., 1999; Brunner, 2000; Wynen et al., 2001; Davis et al., 2004; Delisle and Strobeck, 2005).

Within Musteloidea, familial relationships are more uncertain. Inclusion of the red panda, *Ailurus fulgens*, in Musteloidea is increasingly supported, but with uncertain affinity within the superfamily. Recent molecular studies including representative species from all caniform families place *Ailurus* either with Mephitidae as the sister clade to mustelids and procyonids using primarily mitochondrial DNA (Flynn et al., 2000; Delisle and Strobeck, 2005), or as the most basal musteloid lineage using nuclear and mitochondrial DNA (Flynn et al., 2005).

Support for either hypothesis remains moderate to weak. At the subfamily level, Wozencraft (1993) closely followed Simpson (1945) in dividing Mustelidae into the subfamilies Mustelinae (weasels), Mephitinae (skunks), Melinae (badgers), Taxidiinae (American badger), Mellivorinae (honey badger), and Lutrinae (otters). Recent molecular evidence supports the elevation of skunks to family level (Vrana et al., 1994; Ledje and Arnason, 1996a; Dragoo and Honeycutt, 1997; Flynn et al., 2000; Domingo-Roura et al., 2005; Flynn et al., 2005, but see Sato, 2004), as the Procyonidae are recovered as sister to Mustelidae *sensu stricto* (excluding skunks), making the ‘traditional’ Mustelidae (including skunks) paraphyletic.

Using the most extensive taxon sampling of the Caniformia to date, this study addresses relationships ranging from species-level to superfamily-level, specifically (1) the relationships between Pinnipedia, Ursidae, and Musteloidea, (2) family relationships within Musteloidea, and (3) subfamily-level to species-level relationships within each of the three major arctoid lineages.

As sequencing becomes increasingly efficient via technological advances, molecular data is being quickly obtained on a large scale and is easily accessible through GenBank. This has great advantages in phylogenetics, as more complete taxon sampling within monophyletic lineages can lead to increased accuracy in the estimated phylogeny (Rannala et al., 1998). Conversely, as sequences are often obtained from different studies, sequence is often unavailable for the complete taxon set, leaving entire genes in one or more taxa to be considered as “missing” in multiple gene analysis. Missing data itself may not be problematic for phylogenetic accuracy so long as enough informative characters are present to place the incomplete taxon within the phylogeny (Dragoo and Honeycutt, 1997; Wiens, 2003, 2006). Unfortunately, no *a priori* method exists to determine the number of required informative characters for complete resolution, as this is taxon and tree-dependent. If too few informative characters exist, ‘floating’

taxa may reduce overall support for clades that may otherwise be well supported. However, recent simulations suggest that the addition of taxa with up to 75% missing data may still not have a negative impact on the accuracy of the tree (Wiens, 2006). A more practical problem is the increase in tree search time associated with large amounts of missing data.

In the interest of including all taxa, even highly incomplete ones, in the final phylogeny, a considerable amount of the complete data set is missing. Four type I sequence-tagged sites (STS) and one nuclear intron, *IRBP*, including 190 newly generated sequences, were used in this study to address both the systematic questions for the Arctoidea described above and the effect that missing data may have on the final topologies from different analysis methods. Several methods of phylogenetic estimation were used: maximum parsimony, maximum likelihood, and Bayesian supermatrix analyses, and matrix representation with parsimony (MRP) supertree analysis. The first three methods are employed with incomplete taxa including missing data; the supertree method by definition avoids coding of missing data, as supertrees are constructed from source topologies, not sequence data. The MRP matrix is based on the nodes in each source tree using binary representation for the presence or absence of the taxa derived from each node (Baum and Ragan, 2004). Therefore, the matrix representation of individual topologies will include taxa as missing, but the individual gene source trees themselves will not contain missing data. Each individual gene data set contains between 57 and 71 of the total 85 included taxa, so all partitions are highly overlapping and thus, suitable for supertree methodology. MRP has been shown to be nearly as accurate as analysis of the combined primary data when the MRP input taxon sets are completely overlapping, though this accuracy decreases as the amount of taxon overlap decreases, analogous to an increase in the amount of missing data (Bininda-Emonds and

Sanderson, 2001). Here, we investigate the empirical effect that this missing data may have on the supermatrix and supertree methodologies.

2. Materials and methods

2.1 Samples, amplification, and sequencing strategy

85 samples representing 79 caniforms and 6 feliform outgroup species were included in this study and are listed in Table 1. Sequences were obtained for four type I sequence-tagged sites (STS); *feline sarcoma protooncogene (FES)* intron 14, *cholinergic receptor, nicotinic, alpha polypeptide 1 precursor (CHRNA1)* intron 8, *growth hormone receptor (GHR)* intron 9, *rhodopsin (RHO1)* intron 3; and one exon; *interphotoreceptor retinoid-binding protein (IRBP)* exon 1. Each region was PCR amplified and sequenced using amplification primers and additional internal primers, when necessary. Sources for previously available data from GenBank, including accession numbers, are listed in Table 1. New sequences have been entered into GenBank under the accession numbers **DQ205725** to **DQ205914**.

Total genomic DNA was isolated from either tissue or blood using the QIAgen DNeasy Tissue Extraction kit. New sequences were generated using PCR amplification of each region, using ~150ng of template in a 50 μ l reaction containing 160 μ M dNTPs, 2.5mM MgCl₂, 1.5U of *Taq* DNA polymerase, and 1X PCR buffer (10mM Tris-HCl, pH 8.8, 0.1% Triton X-100, 50mM KCl, and 0.16 mg/mL bovine serum albumin). All PCR reactions were performed in a Perkin-Elmer 9600 thermal cycler; PCR primers and cycling conditions for the STS regions were as described in Koepfli and Wayne (2003): 3 minutes at 94°C, 30 cycles of 94°C for 30 sec, 54°C for 30 sec, 72°C for 45 sec, followed by a 5 minute hold at 72°C. Amplification of *IRBP* was performed as in Stanhope et al. (1992): 94°C for 5 min., followed by 35 cycles of 94°C for 1

min., 70°C for 3 min.. PCR products were resolved on a 1% agarose gel and isolated from excised gel fragments with the QIAquick gel extraction kit (Qiagen). Bi-directional sequencing was performed using BigDye v.3.1.1 (Applied Biosystems) following the manufacturer's protocol using the amplification primers. Internal sequencing primers were designed for *GHR* (*GHRU.int* GGAAAATTAGAAGAGGT, *GHRL.int* AAGAGTCATCGTTGTAGAA) and *IRBP* (+785 GGTACAGTGCCGACAAAGATG; -913 GCTTCTGGAGGTCCAGGGC) to ensure complete double-stranded sequencing. Sequences were resolved using an Applied Biosystems 3100 sequencer.

2.2 Phylogenetic analysis

Sequences were analyzed, basecalled, and aligned using 3100 Data Collection Software v.1.0.1 and Sequence Navigator v.1.0.1 (both from Applied Biosystems). Heterozygous sites (equal peak heights in electropherograms observed in both directions of sequence) were coded as polymorphisms. Alignments of newly generated sequences and those obtained from GenBank were performed both by eye and using the default settings in the EMBL-EBI online version of ClustalW (Chenna et al., 2003) to ensure consistency; no significant differences between methods were noted and final alignments were made by combining methods to achieve the minimal number of gaps (in frame, if applicable). Alignments have been entered into TreeBASE under the accession number XXXX. All genes contained informative indels; these were coded as 0/1 (absence/presence of DNA, regardless of gap length) and included in maximum parsimony analyses with gaps in alignment delineated as "missing". Indel information was not included in maximum likelihood or Bayesian analyses, due to the DNA sequence evolution models employed. Sequences were partitioned into and analyzed by (1) individual genes, (2) all STS introns combined (i.e. excluding *IRBP*), and (3) all genes together. The partition containing only

the STS introns was used to examine the effect of missing data on the maximum parsimony, maximum likelihood, and Bayesian methods, as there is considerably less missing data in this partition compared to the complete data set.

Prior to concatenation, partition homogeneity tests (ILD test, Farris et al., 1995) were performed in PAUP* v.4.10b (Swofford, 2003) to ensure congruence between all partitions. Twenty-four tests included: all data, all STS only, each gene to all remaining genes (5 tests), all pair-wise gene comparisons (10 tests), and, for *IRBP* alone: all coding positions, each coding position to the other two positions (3 tests), and all pair-wise coding positions comparisons (3 tests). All tests were non-significant at $\alpha=0.05$.

Maximum parsimony (MP) analyses were hindered by relatively long search times and large numbers of equally parsimonious trees due to the large number of missing characters and lack of sufficient informative characters, thus, limited search replicates were used. Heuristic searching was performed in PAUP* with TBR branch-swapping and as-is addition of taxa. Bootstrap support (BP) was obtained by 100 bootstrap replicates of heuristic searching with TBR branch-swapping and 10 random sequence addition replicates. Due to computer memory constraints, searches for individual genes were limited to hold 100 trees/addition sequence replicate; concatenated data sets were limited to 1000 trees/addition sequence replicate.

Modeltest v.3.0.6 (Posada and Crandall, 2001) was implemented for each gene, all STS genes, and all genes to determine the most appropriate model of DNA evolution based on hierarchical likelihood ratio testing. The selected models for each partition are shown in Table 2. An iterative approach to maximum likelihood (ML) heuristic searching (Sullivan et al., 2005, and references therein) in PAUP* was employed for each partition using the selected DNA evolution model from Modeltest, with the starting parameters estimated from the MP topology.

Two rounds of searching with set parameters followed by estimation of parameters on the optimized topology resulted in stable parameter estimates and log likelihoods for all searches.

Bayesian analyses were performed using MrBayes v.3.1.1 (Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003). For each individual gene partition, the DNA evolution model selected by ModelTest was implemented. For combined data sets, parameters were estimated separately for each gene partition based on the individual gene models. Four chains (using default temperature = 0.2 for the three heated chains) were run for 1 000 000 iterations, sampling the “cold” chain every 100 iterations. The first 10% of samples were discarded as “burn-in” after visualization in the program Tracer v.1.3 (Rambaut and Drummond, 2003). Two runs were performed for each data set to ensure proper convergence.

Three supertree analyses were performed: all individual gene maximum parsimony trees combined, all individual gene Bayesian trees combined, and all maximum likelihood trees combined. Hereon in, these supertrees will be respectively referred to as the MP-input, Bayes-input, and ML-input supertrees. The MP source trees were 50% majority-rule consensus trees; the Bayesian source trees were 90% majority-rule consensus trees of the Markov chain Monte Carlo (MCMC) sampled trees. The program RadCon (Thorley and Page, 2000) was used to create the MRP supertree matrix for each and the resultant matrices were analysed in PAUP* by maximum parsimony heuristic searching with random addition of taxa. All input trees were equally weighted.

To statistically assess the various hypotheses concerning the placement of the red panda (*Ailurus fulgens*) within Musteloidea, the nonparametric Shimodaira-Hasegawa test (SH test, Shimodaira and Hasegawa, 1999) was performed using all data. The SH test was implemented in PAUP* using 10 000 RELL approximation replicates, estimating parameters appropriate to

the TVMef+I+ Γ likelihood model. Three alternate topologies were tested simultaneously: the maximum likelihood topology where Mephitidae (skunks) are the most basal musteloid lineage, the ML topology modified such that the red panda is the most basal musteloid lineage, and the ML topology modified to place the red panda as sister to the Mephitidae. Pairwise Templeton tests (Templeton, 1983) were implemented in PAUP* for the purpose of a parsimony-based hypothesis test, comparing the two alternate topologies, in turn, to the ML topology.

3. Results

3.1 General sequencing and tree estimation results

All ILD tests showed no significant incongruence between any of the partitions. All individual genes contained informative indels; the number of informative indels are shown in Table 2. All indels supported clades with high maximum parsimony bootstrap (MP BP) support and did not contribute support to any weakly supported clades. The aligned length of each gene partition is shown in Table 2. A 204 base pair (bp) insertion in all canid species in *GHR* was removed from the alignment, as was a 219 bp insertion in the red panda in *RHO1*. Sequence for *FES* for domestic cat, tiger, and bobcat was obtained but removed from the alignment due to difficulty in assessing homology (results not shown). In some cases, amplification products could not be obtained due to poor DNA quality or the presence of numerous non-specific amplification products under all PCR conditions tested. In these cases, sequence was not included in the final data matrix (Table 1).

For all partitions, multiple most parsimonious trees were obtained (Table 2). Both the *GHR* and complete data set MP searches were stopped due to computer memory constraints (92,600 MP trees). MP tree length, number of informative sites, consistency index (CI), and

retention index (RI) are reported for each partition in Table 2. Individual gene topologies (MP BP) contained polytomies but did not differ significantly from each other or from the concatenated total data MP BP topology. Differences between trees were only found in areas of weak support; no “hard” incongruencies (opposing topologies supported by $\geq 80\%$ bootstrap support) were observed.

All maximum likelihood searches yielded a single most likely tree (log likelihoods in Table 2), except *FES*, which recovered six equally likely trees. All *FES* trees were identical in topology and all parameters except the transition:transversion (Ti:Tv) ratio, Ti:Tv kappa, and gamma shape. However, the greatest difference in any one parameter estimate was only 2^{-4} .

Multiple most parsimonious trees were found for all supertree analyses. A strict consensus was made for each search and all results are based upon these strict consensus trees.

3.2 Relationships between Pinnipedia, Ursidae, and Musteloidea

All phylogenies constructed from concatenated data sets (all STS and all data) yielded Arctoidea as monophyletic with Ursidae as the most basal arctoid lineage (Fig 1) with high support (98-99% MP BP; BPP=1.0). Individual gene trees were either unresolved between arctoid lineages, or had lower support (MP BP=67-93%). The only phylogenies that did not contain a monophyletic Arctoidea were the ML and Bayesian *IRBP* gene trees. These placed Canidae within the Arctoidea, as sister to Musteloidea, as was also found by Yu et al.'s (2004b) MP analysis (but not by ML or Bayesian analyses) of caniform *IRBP* sequences. However, the Canidae branch was approximately twice as long as the average branch from musteloid root to tip, indicating some discrepancy within the tree (results not shown). This result was not seen using the MP method, suggesting that long-branch attraction is not the cause of the unusual

topology. Imposing a constraint for arctoid monophyly created a polytomy among the three arctoid lineages; all other relationships were unchanged and $-\ln L$ decreased from -6567.5753 to -6572.6696 (results not shown). When constrained out of Arctoidea in this fashion, the canids remained on an unusually long branch. Removal of canids from the *IRBP* data set did not change the remaining topology.

The ML phylogram (Fig 2) illustrates the close evolutionary relationships between lineages and between taxa, as indicated by the many short branches. Fewer than 0.005 substitutions/site occur between the branching of Ursidae before Pinnipedia and Musteloidea. Within lineages, the average branch length from root to tip is approximately 3-4 times greater within Musteloidea than any other major clade. This discrepancy in rate may be due to the reduced body size of musteloids compared to pinnipeds and ursids, and the many factors associated with body size such as metabolic rate or generation time. We are further investigating this discrepancy in evolutionary rate.

3.3 Species relationships within Ursidae

Relationships within Ursidae were highly unresolved by both supermatrix and supertree analyses of the complete data set (Fig 3a). All methods recovered the giant panda, *Ailuropoda melanoleuca*, as the most basal lineage (68% MP BP, 0.98 BPP for all data), followed by the spectacled bear, *Tremarctos ornatus* (MP BP=80%, BPP=1.0). ML and the MP-input supertree recovered the sun bear, *Helarctos malayanus*, as branching next, followed by the sloth bear, *Melursus ursinus*. The MP-input supertree placed the sloth bear in a polytomy with the *Ursus* species. All other methods included the sun bear in this polytomy, only recovering the brown bear, *Ursus arctos*, and polar bear, *Ursus maritimus*, as a clade. The STS-only data set contained

only four ursid species (Fig 3b). The sloth bear was the most basal, followed by the sun bear, then the American black bear, *Ursus americanus*, and polar bear. The clade formed by the latter two was strongly supported (MP BP=95%, BPP=1.0) and consistent with the topology obtained from analyses of all genes combined. The placement of the sloth bear as basal to the other ursines was also well supported (MP BP=87%, BPP=0.94), but is inconsistent with the all-data topology. However, this is not a strong incongruency, as placement of the sun bear as more basal than the sloth bear in the all-data topology is not supported by either Bayesian or parsimony with bootstrapping. Both data sets recovered topologies consistent with present subfamilial classifications. The Ailuropodinae (giant panda, *Ailuropoda melanoleuca*) was most basal, then the Tremarctinae (spectacled bear, *Tremarctos ornatus*) and the Ursinae (all other ursids).

3.4 Relationships within Pinnipedia

Both concatenated data sets yielded pinniped monophyly (100% MP BP; BPP=1.0) in all construction methods, with strong support (100% MP BP; BPP=1.0) for the sister relationship of the Odobenidae (walrus) to the Otariidae (Fig 4). Below the family level, relationships are less well-supported. Within the Otariidae, the subfamily division between fur seals, Arctocephalinae, and sea lions, Otariinae, is not supported. Steller's sea lion, *Eumetopias jubatus*, is well supported (MP BS=94% for all-data, 83% for STS-only; BPP=1.0 for both data sets) as the most basal otariid included in this study; all other relationships were often polytomous depending on analysis method employed, though the South American fur seal, *Arctocephalus australis*, and New Zealand fur seal, *A. forsteri*, were moderately supported as sister taxa when all data were included (MP BP=76%; BPP=1.0). The positions of the Antarctic fur seal, *A. gazella*, and South

American sea lion, *Otaria byronia*, differed depending on the construction method used. Both likelihood-based methods placed the two as a clade sister to the remaining fur seals (Fig 4), the ML-input supertree recovered the Antarctic fur seal as basal to the remaining otariids (Fig 4); all other methods did not resolve the relationship.

Relationships within the Phocidae are fairly well supported for higher-level designations, but often unresolved at the species level (Fig 4). Monophyly of the subfamily Monachinae (southern seals and monk seals) is strongly supported by all combined analyses (MP BP >90%; BPP=1.0). Remaining relationships are consistent with traditional tribal designations: the Monachini (monk seals) were the deepest branching lineage, followed by the northern elephant seal (tribe: Miroungini) and the Lobodontini (all other southern seals). Relationships within the Lobodontini were unresolved by all methods except ML, which only recovers a single tree without support and the resolution may therefore not be significant. The exception to this is the sister relationship between the leopard, *Hydrurga leptonyx*, and Weddell, *Leptonychotes weddellii*, seals, found in all trees except the MP-input and ML-input supertrees (polytomies). Bayesian analysis weakly recovered (BPP=0.6 in both all-genes and STS-only data sets) the Ross seal, *Omatophoca rossii*, as the earliest branching lobodontine, followed by the crabeater seal, *Lobodon carcinophagus*. Monophyly of the subfamily Phocinae (northern seals) is moderately supported, with the bearded seal, *Erignathus barbatus*, as the earliest branching lineage. The Bayes-input supertree did not resolve the position of the bearded seal, placing it in a polytomy with the remaining northern seals and the southern seals. ML and Bayesian and the MP-input and Bayes-input supertree analyses of the complete combined data set supported the harp seal, *Phoca groenlandica*, as the next branching taxon (MP BP<50%, BPP=1.0), while combined STS data and ML-input supertree supported the hooded seal, *Cystophora cristata*, in

that position (MP BP= 71%; BPP=1.0). The ML-input supertree alternately placed the harp seal in a clade with the ringed seal, *Phoca hispida*. The ML topology placed the grey, *Halichoerus grypus*, and Caspian, *Phoca caspica*, seals as sister and the harbor, *Phoca vitulina*, and spotted, *Phoca largha*, seals as sister; these relationships were very weakly supported by Bayesian. All other species relationships remain unresolved.

3.5 Relationships within Musteloidea

Monophyly of the Musteloidea (including *Ailurus*) was strongly supported (MP BP=100%; BPP=1.0) by all data sets and all tree construction methods (Fig 1), as were the lineages corresponding to the family designations of Ailuridae (red panda), Mephitidae (skunks), Procyonidae (raccoons), and Mustelidae (weasels, badgers, and otters). Mephitidae is recovered as the earliest branching musteloid family, followed by Ailuridae, then Procyonidae and Mustelidae. Support for Procyonidae and Mustelidae as sister families is strong (MP BP=97% for total data, 94% for STS only; BPP=1.0). Placement of Mephitidae as more basal than Ailuridae is strongly supported by Bayesian (BPP=1.0 for total data, BPP=0.98 for STS only), but only weakly supported by maximum parsimony (MP BP=59% for total data, <50% for STS only). This arrangement was also found in all strict consensus supertrees except the ML-input supertree, which recovered a trichotomy between Ailuridae, Mephitidae, and Procyonidae+Mustelidae. The Shimodaira-Hasegawa (SH) test was implemented to test if the placement of Mephitidae as the most basal musteloid lineage is a statistically better arrangement than either the placement of Ailuridae as the most basal lineage or as sister to Mephitidae. The sister grouping of Ailuridae with Mephitidae was rejected at $\alpha=0.05$ and Ailuridae as the most basal musteloid lineage was rejected at $\alpha=0.10$ ($p=0.092$). The Templeton test, or Wilcoxon

signed-ranks test (Templeton, 1983) was implemented under the parsimony criterion. Neither of the alternate topologies could be rejected at $\alpha=0.05$, consistent with the low parsimony bootstrap support (59%) recovered for the most parsimonious arrangement of Mephitidae as the most basal lineage.

The procyonid phylogeny is highly supported (MP BP=100%, BPP=1.0) at all nodes, using all data sets and construction methods (Fig 2). The kinkajou, *Potos flavus*, is basal to the other procyonids, consistent with the subfamily designations dividing the kinkajou and olingo (subfamily Potosinae) from the Procyoninae [raccoons, *Procyon*; ringtail, *Bassariscus*; and coatis, *Nasua*). However, an olingo (*Bassaricyon*) was not included in this study so the validity of these subfamily designations cannot be confirmed. The Procyoninae is also supported by one deletion event. Coatis are recovered as the most basal of the Procyoninae.

The mephitid phylogeny was also strongly supported (MP BP=100%, BPP=1.0) at every node, using all data sets and analysis methods. All three genera (*s.s.* excluding *Mydaus*) are represented, with the hog-nosed skunk, *Conepatus mesoleucus*, as basal to the western spotted skunk, *Spilogale gracilis*, and the striped skunk, *Mephitis mephitis* (Fig 2).

Within Mustelidae, major clades were well supported, but species relationships range from 100% support (MP BP and BPP) to a lack of resolution; some relationships differed between construction methods (Fig 5). In all combined analyses, the American badger, *Taxidea taxus*, is the earliest branch (all data and STS-only, respectively: MP BS=99%, 85%; BPP=0.99, 0.94), supporting its subfamilial status as Taxidiinae. Parsimony, Bayesian, and MP-input supertree analyses of both combined data sets weakly supported (MP BP=63% all-data, 88% STS-only; BPP=0.65 all-data, 0.99 STS-only) the Melinae (Old world badgers), excluding ferret-badgers (*Melogale*), as sister to the clade containing the genera *Martes*, *Eira*, and *Gulo* (Fig 5).

Maximum likelihood did not place these two clades as sister, instead placing Melinae (excluding *Melogale*) as sister to all mustelines and lutrines. Bayes-input and ML-input supertree analyses could not resolve this relationship. Within the *Martes-Eira-Gulo* clade, the wolverine (*Gulo gulo*), tayra (*Eira barbara*), fisher (*Martes pennanti*), and yellow-throated marten (*Martes flavigula*) were basal to the remaining *Martes* species, but otherwise, relationships were poorly resolved. The Chinese ferret badger, *Melogale moschata*, is strongly supported (MP BP=100%, BPP=1.0) as the basal lineage to the remaining mustelids, making Melinae polyphyletic. The remaining mustelids formed three lineages: the genus *Mustela*, the Lutrinae (otters), and the grison (*Galictis vittata*) + zorilla (*Ictonyx striatus*). All primary data supermatrix analyses placed *Mustela* as sister to a Lutrinae+(grison+zorilla) clade. Conversely, the ML-input supertree place the grison+zorilla clade as sister to *Mustela*, not Lutrinae. In all trees, the American mink, *Mustela vison*, and long-tailed weasel, *M. frenata*, formed a clade on the earliest branch in the *Mustela* clade, followed by the ermine, *M. ermina*, then the least and mountain weasels (*M. nivalis* and *M. altaica*) as a clade, and the European mink, *M. lutreola* as basal to the remaining *Mustela* species, which were found in a polytomy. The giant otter, *Pteronura brasiliensis*, was the most basal lutrine; all other otters were divided into two well-supported (MP BP>90%; BPP=1.0) clades. The *Lontra* clade is comprised of the river otter, *L. canadensis*, basal to the marine otter, *L. felina*, and neotropical otter, *L. longicaudis*. The second clade is composed of the sea otter, *Enhydra lutris*, and spotted-necked otter, *Lutra maculicollis*, on one branch, and the Eurasian otter, *Lutra lutra*, basal to the Oriental small-clawed and African clawless otters (*Amblonyx cinereus* and *Aonyx capensis*). The grouping of these genera is consistent with van Zyll de Jong's (1991) tribe Aonychini. Only the MP supertree method yielded a different lutrine

topology, placing the sea otter and spotted-necked otters as branching sequentially instead of being sister taxa.

3.6 Effect of missing data: all STS vs. total data

To determine the effect of missing data on the different methodologies and their associated support values, tree searches were conducted for both the STS introns only and for the total data set (including *IRBP*). The total (all-data) data set has considerably more missing data than the STS-only data set, as 17 species only have sequence for *IRBP* and 10 species had sequence only for the four STS introns. The STS-only data set only contains 14 taxa that are incomplete to varying degrees (Table 1). The most significant difference between taxon sets is found within Ursidae and Mustelidae, where four and twelve taxa are added (respectively) to the STS-only taxon set to form the combined data set. The all-data analysis of Ursidae recovered the same topology (Fig 3) as the *IRBP* gene tree, though all of the MP BP values decreased as much as 18% (*IRBP* gene tree results not shown). BPP values were essentially unchanged. No MP BP nodal support within Ursidae was higher than 85% in either tree, indicating that *IRBP* does not contain enough information to clearly resolve the ursid topology, a problem only exacerbated by the addition of significant amounts of missing data in the complete data set. The addition of taxa within the Ursidae also had an effect. In the STS-only tree containing the polar, American black, sun, and sloth bears, the sloth bear was basal to the other bears. When all taxa were included in the analysis of all the data, the ML tree and the MP-based supertree recovered the sun bear as more basal. However, these results are not strongly incongruent, as placement of the sloth bear as basal by the STS-data set was strongly supported by both bootstrapping and Bayesian posterior probabilities, the ML all-data placement of the sun bear as more basal had no

associated support. Within Mustelidae, addition of taxa with significant amounts of missing data (i.e. *IRBP* sequence only) did not conflict with the STS-only topology and a significant decrease in support was not observed, but most nodes in the all-data topology were not present in the STS-only topology because there were significantly fewer taxa included.

The converse situation occurred within the Lutrinae, where only two (the sea and river otters) of the nine included species have sequence for *IRBP*. Despite the lack of *IRBP* sequence for nearly all the lutrines, the topology and its Bayesian posterior probabilities remained unchanged between data sets and the MP bootstrap support was very similar. Of the eight nodes found within the Lutrinae, two had identical MP BP values in the STS and the total data trees, three had increased support with the total data set, and three had decreased support. However, the largest discrepancy between MP BP values was only 6%. Therefore, within the Lutrinae, the missing *IRBP* data appears to have had little to no effect, as the STS information was enough to resolve the topology. On the overall topology, the same is true. The two topologies are completely congruent, though the STS-only taxon set was smaller so several nodes cannot be observed compared to the total data tree. All nodes that were highly supported by MP BP in the STS data set were also highly supported in the total data set. Other less supported nodes had similar values between trees and no trend for either an increase or decrease in support between trees was observed. Only the Ursidae experienced a decrease in bootstrap support by the addition of a large amount of missing data.

3.7 Supertree vs. supermatrix

The three supertree topologies based on ML, MP, and Bayesian input trees were compared to the ML, MP BP, and Bayesian supermatrix topologies. In general, very few discrepancies were found between the supermatrix topologies and supertrees and most

discrepancies occurred in areas that were not well resolved by any method. In general, the differences between supertrees and supermatrix topologies were not greater than the differences found between the ML, Bayesian, and MP supermatrix methodologies. All trees resulted in identical family-level topologies, although the ML-input supertree did not resolve the relative position of the red panda within the Musteloidea. Five different species positions were obtained within the strict consensus supertrees that were not observed in the supermatrix trees. The MP-input supertree placed the sea otter and spotted-necked otters as branching sequentially (in that order), instead of forming a clade sister to the remaining species in the larger clade (Fig 5). The ML-input supertree exchanged the positions of the wolverine and yellow-throated marten within the *Martes-Eira-Gulo* clade and placed the grison+zorilla clade as sister to *Mustela* within Mustelidae (Fig 5). The ML-input supertree additionally placed the harp and ringed seals as sister within the Phocinae and did not group the Antarctic fur seal and South American sea lion as sister, instead, they branched sequentially (Fig 4). The Bayes-input supertree did not recover any clades that were not present in at least one of the supermatrix analyses. Although all of the supertrees resulted in many equally parsimonious trees, the strict consenses of these remain relatively well resolved, except in areas that remain poorly supported or unresolved by all methods.

4. Discussion

4.1 Higher-level systematics of the Arctoidea

Both data sets; all data combined and STS-only; and all analysis methods yielded Arctoidea as monophyletic and strongly supported Ursidae (MP BP=98%, BPP=1.0) as the most basal arctoid lineage (Fig 1). This arrangement for the Arctoidea was also strongly supported by

the nuclear analysis of Flynn et al. (2005) and weakly supported by supertree analysis (Bininda-Emonds et al., 1999), but is not supported by mtDNA alone (Delisle and Strobeck, 2005), and is only sometimes supported by morphology (Wolsan, 1993; Wyss and Flynn, 1993). Within Arctoidea, monophyly of each family and major lineage (Ursidae, Pinnipedia, Musteloidea) is strongly supported (MP BP \geq 93%, BPP=1.0).

4.2 Relationships within Ursidae

All analyses of the total data set for the Ursidae supported the subfamilial designations of Ailuropodinae (giant panda, *Ailuropoda melanoleuca*), Tremarctinae (spectacled bear, *Tremarctos ornatus*), and Ursinae (all other ursids), with the latter two as sister (Fig 3a). Resolution within Ursinae was poor in the all-data analyses (Fig 3a), and though the sun bear, *Helarctos malayanus*, was recovered as more basal than the sloth bear, *Melursus ursinus*, this was a very short branch and was unsupported. Little confidence is placed in this resolution. The STS-only analyses containing only the sloth, sun, polar (*Ursus maritimus*), and American black bears (*Ursus americanus*) were strongly supported by both Bayesian and MP BP recovered the sloth bear, as the most basal ursine, and the sun bear, and the genus *Ursus* as a clade (Fig 3b). This placement of the sloth bear is congruent with previous molecular studies (Waits et al., 1999; Yu et al., 2004a), but strong support for the sun bear as more basal than both *Ursus* species is unusual, and contradictory to the mitochondrial DNA-based hypothesis placing the sun bear as sister to the brown and polar bears to the exclusion of the black bears (Waits et al., 1999; Delisle and Strobeck, 2005). This suggests that inclusion of the sun bear within the genus *Ursus* (vs. retaining *Helarctos*) as suggested by Nowak (1991), should not be implemented until a more well-supported consensus is reached.

4.3 Relationships within Pinnipedia

Within Pinnipedia, the Otariidae (fur seals and sea lions) and Odobenidae (walrus) were recovered as a sister clade to Phocidae (true seals), supporting the superfamily association of the former two as Otaroidea (Flynn and Wesley-Hunt, 2005). Otariidae is strongly supported as monophyletic, but the subfamilies Arctocephalinae (fur seals) and Otariinae (sea lions) are not. Steller's sea lion, *Eumetopias jubatus*, is strongly supported as the most basal otariid included, but other species relationships are not supported by all methods and are often unresolved (Fig 4). The placement of Steller's sea lion as more basal than the South American sea lion, *Otaria byronia*, conflicts with the results of Wynen et al. (2001), although their placement of *Otaria* as more basal was very weakly supported (MP BP<50%). Though the topologies differ, our results concur with that of the more comprehensive study of Wynen et al. (2001, and references therein), in that the present subfamily designations within Otariidae are misleading.

Phocidae is divided into two subfamilies, Phocinae (northern seals) and Monachinae (southern, elephant, and monk seals). This division is well supported by all methods (Fig 4), as are tribal distinctions to a lesser extent. However, at the species level, very little resolution exists and any resolved species relationships are generally associated with low support. Within the Phocinae, the bearded seal, *Erignathus barbatus*, is the only member of the tribe Erignathini and is well supported as the basal member of the subfamily. Relationships between and within the remaining two tribes, Cystophorini (monotypic hooded seal, *Cystophora cristata*) and Phocini (all other northern seals) are highly unresolved and differ based on the analysis method employed. Likelihood-based methods (ML and Bayesian) for the all-data set strongly supported the basal placement of the harp seal, *Phoca groenlandica*, to *Cystophora*, making the Phocini paraphyletic. Within the remaining Phocini, the harbor and spotted seals (*Phoca vitulina* and *P.*

largha) grouped as sister, as did the grey seal, *Halichoerus grypus*, and Caspian seal, *Phoca caspica*, but with poor support. In contrast, analyses of the STS-only data set, and the ML-input supertree strongly supported the hooded seal as branching after the bearded seal, consistent with mtDNA results (Davis et al., 2004) and supporting the tribal status of the hooded seal as Cystophorini and monophyly of the Phocini. Within Monachinae, the tribal relationships of Monachini (monk seals) as most basal, then Miroungini (elephant seals) and Lobodontini (southern seals) were supported (Fig 4). Within the Lobodontini, all methods except the ML-input supertree resolved the leopard seal, *Hydrurga leptonyx*, and Weddell seal, *Leptonychotes weddellii*, as sister. No other lobodontine species relationships were resolved, as is frequently the case within this tribe (Davis et al., 2004; Fyler et al., 2005).

4.4 Musteloid family-level systematics: phylogenetic affinity of the red panda

Within Musteloidea, monophyly is strongly supported (MP BP=100%, BPP=1.0) for each family: Mephitidae (skunks), Mustelidae (weasels, badgers, otters), and Procyonidae (raccoons). The family designation Ailuridae (monotypic red panda) is also supported. We propose a novel placement of the Ailuridae as sister to Mustelidae + Procyonidae, with Mephitidae as the most basal musteloid family (Fig 1). This topology was recovered by all analysis methods, except by the ML-input supertree, which could not resolve the relative position of the red panda. The basal placement of the Mephitidae has strong Bayesian support (1.0 for all genes, 0.98 for STS-only), but poor MP BP support (59% for all genes, <50% for STS-only). Flynn et al. (2005) obtained nearly identical support values (MP BP<50%, BPP=0.99-1.0) for their placement of the red panda as the most basal musteloid, as did Delisle and Strobeck (2005) for their placement of the red panda as sister (MP BP=73-77%, BPP=1.0) to the Mephitidae (as proposed by Flynn et al.,

2000). Likelihood-based statistical testing using the nonparametric Shimodaira-Hasegawa test of these alternate placements compared to the proposed placement rejected the first alternate hypothesis with $p < 0.1$ and the second with $p < 0.05$. The SH test is usually conservative, testing whether or not alternate hypotheses are equally good explanations of the data (Goldman et al., 2000). Because the test itself is so conservative, rejection at $\alpha = 0.1$ is still a strong result. Parsimony-based hypothesis testing (Templeton test) could not reject either of the alternate topologies, consistent with the relative inability of parsimony to resolve this relationship. Because Ailuridae is monotypic and Mephitidae is on a relatively long branch, it is possible that long-branch attraction may be a factor in parsimony analysis. At present, we propose this novel placement of the red panda as a third hypothesis, but conservative interpretation would retain a polytomous relationship between Ailuridae, Mephitidae, and Mustelidae + Procyonidae. Based on available data, it appears that mitochondrial DNA analyses recover Mephitidae and Ailuridae as sister (Flynn et al., 2000; Delisle and Strobeck, 2005), nuclear DNA supports Ailuridae as branching after Mephitidae (this study), and a combination of the two yields a ‘compromised’ position of Ailuridae as the most basal musteloid lineage (Flynn et al., 2005). Future work combining morphology, nuclear DNA, mtDNA, and other molecular techniques will hopefully aid in the resolution of the phylogenetic affinity of the ‘enigmatic’ red panda.

4.5 Species relationships within Mephitidae and Procyonidae

The Mephitidae are comprised of four genera: *Spilogale*, spotted skunks; *Mephitis*, striped skunks; *Conepatus*, hog-nosed skunks; and *Mydaus*, the recently included stink badgers (Bryant et al., 1993; Dragoo and Honeycutt, 1997; Flynn et al., 2000). *Mydaus* was not represented in this study, but the remaining three genera were strongly supported as a

monophyletic group, with *Conepatus* as most basal. This arrangement is congruent with the findings based on nuclear and mitochondrial DNA (Flynn et al., 2005), morphology (Bryant et al., 1993), and total evidence (Dragoo and Honeycutt, 1997).

Monophyly of Procyonidae is also strongly supported, as are all nodes within the family topology, using both data sets and all methods. We present the most comprehensive molecular phylogeny of the Procyonidae to date, including four of the five widely recognized genera. These are grouped into two subfamilies: Potosinae, represented by the kinkajou (*Potos flavus*) in this study, and Procyoninae, which include the genera *Procyon* (raccoons), *Bassariscus* (ringtail), and *Nasua* (coatis). The fifth genus *Bassaricyon* (olingos), assigns to Potosinae, but it is not included in this study, precluding any definitive conclusions regarding subfamilial designations. The recovered topology (Fig 2) is consistent with the above subfamilial designations suggested by morphology (Decker and Wozencraft, 1991). The Procyoninae were recovered as monophyletic, with the clade supported by a single deletion event. Within Procyoninae, coatis are strongly supported as the most basal lineage, in contrast to morphology, which places the ringtail as more basal (Decker and Wozencraft, 1991). As more molecular data is obtained for *Bassaricyon*, both subfamilial designations and species relationships can be properly assessed.

4.6 Subfamily and species relationships within Mustelidae

Relationships within Mustelidae have been subject to considerable study both recently and in the past, with most recent studies in agreement upon the status (monophyletic or paraphyletic) of the subfamilies, though not necessarily on the species relationships within them (Bryant et al., 1993; Dragoo and Honeycutt, 1997; Koepfli and Wayne, 1998, 2003; Sato et al.,

2003; Sato et al., 2004; Flynn et al., 2005). At the subfamily level for badgers, Taxiidinae (monotypic American badger, *Taxidea taxus*) is the most basal lineage and the traditional Melinae (Old World badgers) is polyphyletic if *Mydaus* (stink badgers) and *Melogale* (ferret-badgers) are included. If *Mydaus* and *Melogale* are excluded in accordance with recent studies, monophyly of the Melinae (*s.s.*) comprising of *Meles* (Old World badgers) and *Arctonyx* (hog badgers) is strongly supported. However, the placement of Melinae (*s.s.*) is uncertain. It either branches after *Taxidea* and before the remaining mustelids as found by ML, or as sister to the *Martes-Eira-Gulo* clade, as recovered and moderately supported by all other analysis methods (Fig 5). The Chinese ferret-badger, *Melogale moschata*, is strongly supported by all methods as the next branching lineage, further supporting its exclusion from Melinae. The remaining mustelids grouped in three strongly supported clades: *Mustela*, Lutrinae (otters), and the African zorilla (*Ictonyx striatus*) + the Central and South American-ranging grison (*Galictis vittata*). Most analysis methods placed the grison+zorilla clade as sister to Lutrinae, but the ML-input supertree instead recovered the clade as sister to *Mustela*. This sister relationship with *Mustela* was recovered by the individual gene trees for *CHRNA1*; all other individual gene trees did not resolve the relationship, indicating that the *CHRNA1* gene tree may have led to this relationship in the ML-input supertree. Monophyly of the Lutrinae was strongly supported (Fig 5). The three major clades recovered within the Lutrinae correspond to the tribe Aonychini, the New World otters (*Lontra*), and the giant otter (*Pteroneura brasiliensis*). Most lutrine species relationships were well supported with one exception. The sea otter, *Enhydra lutris*, is recovered as either the most basal lineage in the Aonychini (*Enhydra-Lutra-Amblyonyx-Aonyx*) by the MP-input supertree or as sister to the spotted-necked otter, *Lutra maculicollis* by all other methods. The former arrangement was recovered by Koepfli and Wayne (2003) in their combined analysis of

five nuclear STS loci (four of which are used in this study), but when only the STS loci were analyzed, the latter was found. However, though this sister relationship was recovered most often, it is essentially unsupported (MP BP=52% for both all-data and STS-only; BPP=0.69 for all-data and 0.64 for STS-only) and should be interpreted with caution. More in-depth systematics of the Lutrinae are discussed by Koepfli and Wayne (1998). Results within *Mustela* were the same as recovered by the MP analysis of Sato et al. (2003), though with somewhat less resolution as some relationships within *Mustela* are solely based on the *IRBP* gene tree. Species relationships are poorly resolved within the *Martes-Eira-Gulo* clade (Fig 5), though the clade itself is well supported. Four taxa, the fisher (*Martes pennanti*), tayra (*Eira barbara*), wolverine (*Gulo gulo*), and yellow-throated marten (*Martes flavigula*), are more basal on the tree than the remaining *Martes* species, but the precise relationships of these taxa is uncertain. However, the well-supported inclusion of the tayra and wolverine in this clade make the genus *Martes* paraphyletic. As a result, it is perhaps misleading for the fisher and/or the yellow-throated marten to be included in *Martes*, or for *Eira* and *Gulo* to be retained as distinct genera. Though genus-level classifications remain contentious, given the strongly supported monophyly of this group and the subfamilial status of the Melinae (*s.s.*) as the potential sister clade to the *Martes-Eira-Gulo* clade, we propose that this clade also be afforded subfamilial status as Martinae. Accordingly, if Mustelinae are restricted to the genus *Mustela*; the subfamily Galictinae (see Anderson, 1989) is reinstated to include *Galictis* and *Ictonyx* but not *Eira*; and the ferret-badgers (*Melogale*) become the subfamily Helictidinae as proposed by Sato et al. (2004) (originally as tribe Helictidina by Gray in 1825, elevated to subfamily by Gill in 1872), the subfamilies Taxidiinae, Melinae (*s.s.*), and Lutrinae would remain valid. The mustelid genera *Vormela*, *Lyncodon*, *Poecilictis*, *Poecilogale*, and *Mellivora* are not included in this study and further work

is necessary to confirm the phylogenetic affinities of these genera. *Mellivora* (honey badger) is generally considered as the monotypic subfamily Mellivorinae. *Poecilogle* has been closely allied with *Ictonyx* (Flynn et al., 2005) and *Poecilictis* may not even be a distinct genus from *Ictonyx*, thus, both will likely be included in the redefined Galictinae. Very little molecular work has been performed on *Vormela* and *Lyncodon*, but *Lyncodon* was originally included in Galictinae (Anderson, 1989) and *Vormela* has also been allied with *Ictonyx* (Bininda-Emonds et al., 1999). However, definitive classification of these relatively unstudied species should be delayed until more information is acquired.

4.7 Effect of missing data: all STS vs. total data

In addition to systematic questions within the Arctoidea, we were also interested in the effect of missing data on the final topologies and their associated measures of support. It is sometimes the case that taxa missing data for entire genes are excluded from analysis to avoid the potential problems of reduced support and resolution (Flynn et al., 2005). However, increased taxon sampling within monophyletic lineages, especially those taxa whose placement will break up long branches (thereby reducing long-branch attraction), has been shown to increase phylogenetic accuracy, more so than the addition of characters (Graybeal, 1998; Rannala et al., 1998). It has also been proposed that missing data itself, or even the proportion of missing data may not be problematic, so long as enough informative characters are present to place the taxon on the tree (Wiens, 2003). Inclusion of even highly incomplete taxa (10-25% complete) using model-based analyses such as ML or Bayesian or moderately incomplete taxa (50%) with MP analyses has been shown to have a strong positive effect on accuracy in simulation studies (Wiens, 2006). Based on this, it was our intention to include as many taxa as

possible within the Arctoidea for which information for at least one of the included genes was available. To address the effects of missing data, two data sets were analyzed: one containing all five genes and 79 taxa, and one containing only the four STS loci for 62 taxa. With the addition of the *IRBP* gene to the STS data set, 17 taxa containing only sequence for *IRBP* were added, introducing a considerable number of missing data cells into the data set, especially within Ursidae and Mustelidae (Table 1). The result of the addition of these taxa (discussed below) and the associated missing data supported the idea (Wiens, 2003, 2006) that additional taxa have a greater impact on topology than additional characters and that the amount of missing data is less important than the number of informative characters for those taxa. Within Ursidae, four of the eight taxa were included with only *IRBP* sequence, resulting in a topology that matched that of the *IRBP* gene tree. The addition of taxa (and thus, missing data for all STS introns in those species) led to a general reduction in MP bootstrap support from the *IRBP* gene tree to the combined data set. However, all MP BP values were moderate to low and the largest decrease was 18%. BPP values were essentially unchanged between the *IRBP* gene tree and the all-data tree, indicating that Bayesian analyses are more robust when faced with substantial missing data (~60% missing), likely due to lower susceptibility to long-branch attraction than MP (Huelsenbeck, 1995; Alfaro et al., 2003). Within Mustelidae, taxa with only *IRBP* sequence (~40% missing) were mainly added within *Martes* and *Mustela*. Within both genera, the STS-only and combined data topologies were completely congruent and support was not substantially reduced. Species relationships that could not be resolved in the *IRBP* gene tree remained unresolved in the combined tree, as no additional informative characters were added to successfully resolve relationships. Conversely, within Lutrinae (otters), the addition of *IRBP* added sequence for only two of the nine species, introducing missing data without much benefit

of additional information. In this case, the topology was completely resolved by the STS-only sequence information and the addition of missing data had no clear effect. Equal numbers of nodes experienced increased MP BP support to those that decreased in support and the changes were small. In all cases where support was 100% MP BP or BPP=1.0, it remained unchanged, indicating that enough informative characters were present to nullify any effect of missing data. Overall, topologies that were well resolved prior to the incorporation of missing data remained well-resolved, while poorly resolved topologies experienced some decreased MP BP support. Lack of resolution in the supertrees appears to be due to the absence of enough informative characters to fully resolve input trees. Although not conclusive, these findings support the idea that if enough informative characters exist to place a taxon within a topology, the addition of missing data (to an extent) will have little to no effect.

4.8 Supertree vs. supermatrix

A second approach used to deal with the amount of missing data in the complete data set was the application of the matrix representation with parsimony (MRP) supertree method. Due to non-identical taxon sets between genes, a traditional consensus of the gene trees could not be made. Supertree methodologies are used to combine topologies, but do not require identical taxon sets, merely overlapping ones, as its construction algorithm is node-based, not tree-based (Baum and Ragan, 2004). The supertree method provided two major advantages to the analysis in this study. First, compared to a supermatrix approach, it is a considerably faster method for creating a ML-based tree. The creation of the supermatrix ML tree took approximately twice as long as ML searches for individual gene trees and subsequent construction of the ML-input supertree. Second, it allowed the gene information interpreted under MP, ML, and Bayesian

criteria to be combined in a different way for comparison with the supermatrix-derived topologies from these methods to examine the potential effect that missing data was having on these topologies.

The MRP method and its applications have been heavily criticized since it was proposed independently by Baum in 1990 and Ragan in 1991 (see Gatesy and Springer, 2004). We do not address any methodological shortcomings that have been raised against the MRP method, but do believe that the method employed here avoids many of the shortcomings in previous applications of the MRP method. Combination of newly constructed gene trees via supertree analysis has been proposed in the past (Bininda-Emonds, 2004), avoiding problems such as non-independence (i.e. pseudoreplication) of source-tree data and questionable quality of source trees.

In this study, supermatrix and supertree topologies were highly congruent. Five branching patterns were recovered by only supertree methods and not by supermatrix analyses, and one of these “novel” arrangements, the sequential branching of the sea otter and spotted-necked otter (vs. sister), is actually the preferred topology by another molecular study (Koepfli and Wayne, 2003). The other four novel groupings were found in the ML-input supertree, likely reflecting the more resolved, but unsupported input topologies rather than truly novel clades. All supertree and supermatrix methods recovered the same subfamily and higher topologies, excepting the different placement of Melinae (*s.s.*) found only by ML supermatrix analysis (Fig 5). Conflicts between analytical methods and novel supertree clades were confined to areas of weak topological support, generally species relationships within subfamilies. In this study, differences appear to result from the need for increased data, not from specific shortcomings of any method (Bininda-Emonds, 2004).

Differences between ML and Bayesian methods are thought to have arisen from the ability to apply and estimate parameters for distinct data partitions (genes) within MrBayes for Bayesian analysis, and not within PAUP* for ML, as well as the fact that ML nearly always returns a single resolved topology as it does not rely on any type of consensus. If it were feasible to perform bootstrapping on our ML data set, we would have been able to identify areas that, while resolved, were weakly supported and were perhaps over-contributing to the ML-input supertree. All of the MP searches of the MRP-constructed matrices recovered thousands of equally parsimonious trees. The 50% majority-rule consensus of each of these searches returned a nearly completely resolved topology (results not shown). However, almost all of the nodes that were not found in 100% of the MP trees were found in less than 75%. We therefore used only the strict consensus of these trees (vs. majority-rule), in favour of obtaining a single supertree at the expense of some (potentially misleading) resolution (Steel et al., 2000). Overall, when the resultant supertrees were compared to their supermatrix counterparts, very few topological differences existed, and both methods appeared to suffer equally from inadequacies in the primary data (too few informative characters). Based on this, the ML-input supertree analysis is advantageous to a ML supermatrix analysis, based on the shorter tree construction time and its more consensus-like basis when support cannot be obtained for the ML topology. However, the Bayesian supermatrix analysis of the combined data set was faster than the creation of individual Bayesian gene trees and their subsequent supertree construction. Bayesian analysis in MrBayes also has the benefit of applying gene-specific models in combined gene analysis and has an associated support measure, making it the best choice for a fast likelihood-based analysis. However, based on the results of this study, we advocate the use of several analysis methods for

any data set, particularly those lacking enough informative characters to fully resolve relationships.

Acknowledgements

We thank the following individuals and institutions for their generous donations of samples: R.A. Mead, D.W. Coltman, P. Lemelin, M. Culver, «GreetingLine», K. Kloecker, B. McClymont, W. Amos, S. Atkinson, A. Blix, I. Boyd, T. Gelatt, F. Gulland, L. Harwood, E. Nordoy, B. Robson, B. Stewart, B. Sjare, M. Wainstein, N. Gemmell, I. Stirling, Folsom City Zoo, USGS Alaska Science Center, Alberta Fish and Wildlife, Calgary Zoo, Cleveland Metroparks Zoo, and the Parks Canada tissue repository. Thanks also to I. Delisle and two anonymous reviewers for their helpful comments. This work was supported by an NSERC research grant to C.S..

References

- Alfaro, M., Zoller, S., Lutzoni, F., 2003. Bayes or bootstrap? A simulation study comparing the performance of Bayesian Markov chain Monte Carlo sampling and bootstrapping in assessing phylogenetic confidence. *Mol. Biol. Evol.* 20, 255-266.
- Anderson, E., 1989. The phylogeny of mustelids and the systematics of ferrets. In: Seal, U.S., Thorne, E.T., Bogan, M.A., Anderson, S.H., (Eds.), *Conservation biology and the black-footed ferret*. Yale University Press, New Haven, Connecticut, pp. 10-20.
- Baum, B. R., Ragan, M. A., 2004. The MRP method. In: Bininda-Emonds, O.R.P., ed. *Phylogenetic supertrees: combining information to reveal the tree of life*. Kluwer Academic Publishers, Dordrecht, The Netherlands, pp. 17-34.

- Bininda-Emonds, O. R. P., 2004. The evolution of supertrees. *Trends Ecol. Evol.* 19, 315-322.
- 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.
- Bininda-Emonds, O. R. P., Sanderson, M. J., 2001. Assessment of the accuracy of matrix representation with parsimony analysis supertree construction. *Syst. Biol.* 50, 565-579.
- Brunner, S., 2000. Cranial morphometrics of fur seals and sea lions (Family: Otariidae) - systematics, geographic variation and growth. *Veterinary Anatomy and Pathology*. University of Sydney, NSW 2006, Sydney, 254 pp.
- Bryant, H. N., Russell Fls, A. P., Fitch, W. D., 1993. Phylogenetic relationships within the extant Mustelidae (Carnivora): appraisal of the cladistic status of the Simpsonian subfamilies. *Zoological Journal of the Linnean Society* 108, 301-334.
- Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T. J., Higgins, D. G., Thompson, J. D., 2003. Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Research* 31, 3497-3500.
- Davis, C. S., Delisle, I., Stirling, I., Siniff, D. B., Strobeck, C., 2004. A phylogeny of the extant Phocidae inferred from complete mitochondrial DNA coding regions. *Mol. Phylogenet. Evol.* 33, 363-377.
- Decker, D. M., Wozencraft, W. C., 1991. Phylogenetic analysis of recent procyonid genera. *J. Mammal.* 72, 42-55.
- Delisle, I., Strobeck, C., 2005. A phylogeny of the Caniformia (order Carnivora) based on 12 complete protein-coding mitochondrial genes. *Mol. Phylogenet. Evol.* 37, 192-201.

- Domingo-Roura, X., Lopez-Giraldez, F., Saeki, M., Marmi, J., 2005. Phylogenetic inference and comparative evolution of a complex microsatellite and its flanking regions in carnivores. *Genetical Research Cambridge* 85, 223-233.
- Dragoo, J. W., Honeycutt, R. L., 1997. Systematics of Mustelid-like carnivores. *J. Mammal.* 78, 426-443.
- Farris, J. S., Källersjö, M., Kluge, A. G., Bult, C., 1995. Constructing a significance test for incongruence. *Syst. Biol.* 44, 570-572.
- Flynn, J. J., Finarelli, J. A., Zehr, S., Hsu, J., Nedbal, M. A., 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.
- Flynn, J. J., Nedbal, M. A., Dragoo, J. W., Honeycutt, R. L., 2000. Whence the red panda? *Mol. Phylogenet. Evol.* 17, 190-199.
- Flynn, J. J., Neff, N. A., Tedford, R. H., 1988. Phylogeny of the Carnivora. In: Benton, M.J., ed. *The Phylogeny and Classification of the Tetrapods*. Clarendon Press, Oxford, pp. 73-116.
- Flynn, J. J., Wesley-Hunt, G. D., 2005. Carnivora. In: Rose, K.D., Archibald, J.D., (Eds.), *The rise of placental mammals*. The John Hopkins University Press, Baltimore, Maryland, pp. 175-198.
- Fyler, C. A., Reeder, T. W., Berta, A., Antonelis, G., Aguilar, A., Androukaki, E., 2005. Historical biogeography and phylogeny of monachine seals (Pinnipedia: Phocidae) based on mitochondrial and nuclear DNA data. *J. Biogeography* 32, 1267-1279.

- Gatesy, J., Springer, M. S., 2004. A critique of matrix representation with parsimony supertrees. In: Bininda-Emonds, O.R.P., ed. Phylogenetic supertrees: combining information to reveal the tree of life. Kluwer Academic Publishers, Dordrecht, The Netherlands, pp. 369-388.
- Goldman, N., Anderson, J. P., Rodrigo, A. G., 2000. Likelihood-based tests of topologies in phylogenetics. *Syst. Biol.* 49, 652-670.
- Graybeal, A., 1998. Is it better to add taxa or characters to a difficult phylogenetic problem? *Syst. Biol.* 47, 9-17.
- Huelsenbeck, J. P., 1995. The performance of phylogenetic methods in simulation. *Syst. Biol.* 44, 17-48.
- Huelsenbeck, J. P., Ronquist, F., 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17, 754-755.
- Koepfli, K.-P., Wayne, R. K., 1998. Phylogenetic relationships of otters (Carnivora: Mustelidae) based on mitochondrial cytochrome *b* sequences. *J. Zool. Lond.* 246, 401-416.
- Koepfli, K.-P., Wayne, R. K., 2003. Type I STS markers are more informative than cytochrome *b* in phylogenetic reconstruction of the Mustelidae (Mammalia: Carnivora). *Syst. Biol.* 52, 571-593.
- Ledje, C., Arnason, U., 1996a. Phylogenetic analyses of complete cytochrome *b* genes of the order Carnivora with particular emphasis on the Caniformia. *J. Mol. Evol.* 42, 135-144.
- Ledje, C., Arnason, U., 1996b. Phylogenetic relationships within caniform carnivores based on analyses of the mitochondrial 12S rRNA gene. *J. Mol. Evol.* 43, 641-649.
- Nowak, R. M., 1991. Walker's Mammals of the World. Fifth Edition, 5 edn The John Hopkins University Press, Baltimore.

- Posada, D., Crandall, K. A., 2001. Selecting the best-fit model of nucleotide substitution. *Syst. Biol.* 50, 580-601.
- Rambaut, A., Drummond, A., 2003. Tracer, version v1.0.1.
- Rannala, B., Huelsenbeck, J. P., Yang, Z., Nielson, R., 1998. Taxon sampling and the accuracy of large phylogenies. *Syst. Biol.* 47, 702-710.
- Ronquist, F., Huelsenbeck, J. P., 2003. MrBayes3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19, 1572-1574.
- Sato, J. J., Hosoda, T., Wolsan, M., Suzuki, H., 2004. Molecular phylogeny of Arctoids (Mammalia: Carnivora) with emphasis on phylogenetic and taxonomic positions of the ferret-badgers and skunks. *Zool. Sci.* 21, 111-118.
- Sato, J. J., Hosoda, T., Wolsan, M., Tsuchiya, K., Yamamoto, M., Suzuki, H., 2003. Phylogenetic relationships and divergence times among Mustelids (Mammalia: Carnivora) based on nucleotide sequences of the Nuclear Interphotoreceptor Retinoid Binding Protein and mitochondrial cytochrome *b* genes. *Zool. Sci.* 20, 243-264.
- Shimodaira, H., Hasegawa, M., 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Mol. Biol. Evol.* 16, 1114-1116.
- Simpson, G. G., 1945. The principles of classification and a classification of mammals. *Bulletin of the American Museum of Natural History* 85, 1-350.
- Stanhope, M. J., Czelusniak, J., Si, J.-S., Nickerson, J., Goodman, M., 1992. A molecular perspective on mammalian evolution from the gene encoding Interphotoreceptor Binding Protein, with convincing evidence for bat monophyly. *Mol. Phylogenet. Evol.* 1, 148-160.
- Steel, M. A., Dress, A., Böcker, S., 2000. Simple but fundamental limitations on supertree and consensus tree methods. *Syst. Biol.* 49, 363-368.

- Sullivan, J., Abdo, Z., Joyce, P., Swofford, D. L., 2005. Evaluating the performance of a successive-approximations approach to parameter optimization in maximum-likelihood phylogeny estimation. *Mol. Biol. Evol.* 22, 1386-1392.
- Swofford, D. L. (2003). PAUP*. Phylogenetic Analysis Using Parsimony (*and other methods). (Sunderland, Massachusetts, Sinauer Associates).
- Talbot, S. L., Shields, G. F., 1996. A phylogeny of the bears (Ursidae) inferred from complete sequences of three mitochondrial genes. *Mol. Phylogenet. Evol.* 5, 567-575.
- Templeton, A. R., 1983. Phylogenetic inference from restriction endonuclease cleavage site maps with particular reference to the evolution of humans and the apes. *Evolution* 37, 221-244.
- Thorley, J. L., Page, R. D. M., 2000. RadCon: Phylogenetic tree comparison and consensus. *Bioinformatics* 16, 486-487.
- van Zyll de Jong, C. G., 1991. A brief review of the systematics and a classification of the Lutrinae. *Habitat* 6, 79-83.
- Vrana, P. B., Milinkovitch, M. C., Powell, J. R., Wheeler, W. C., 1994. Higher level relationships of the arctoid Carnivora based on sequence data and "total evidence". *Mol. Phylogenet. Evol.* 3, 47-58.
- Waits, L., Sullivan, J., O'Brien, S. J., Ward, R. H., 1999. Rapid radiation events in the family Ursidae indicated by likelihood phylogenetic estimation from multiple fragments of mtDNA. *Mol. Phylogenet. Evol.* 13, 82-92.
- Wiens, J. J., 2003. Missing data, incomplete taxa, and phylogenetic accuracy. *Syst. Biol.* 52, 528-538.

- Wiens, J. J., 2006. Missing data and the design of phylogenetic analyses. *Journal of Biomedical Informatics* 39, 34-42.
- Wolsan, M., 1993. Phylogeny and classification of early European Mustelida (Mammalia, Carnivora). *Acta Theriol* 38, 345-384.
- Wozencraft, W. C., 1993. Order Carnivora. In: Wilson, D.E., Reeder, D.A., (Eds.), *Mammal species of the world: a taxonomic and geographic reference*. Smithsonian Institution Press, Washington, DC, pp. 279-348.
- Wynen, L. P., Goldsworthy, S. D., Insley, S. J., Adams, M., Bickham, J. W., Francis, J., Gallo, J. P., Hoetzel, A. R., Majluf, P., White, R. W. G., Slade, R. W., 2001. Phylogenetic relationships within the eared seals (Otariidae: Carnivora): Implications for the historical biogeography of the family. *Mol. Phylogenet. Evol.* 21, 270-284.
- Wyss, A. R., Flynn, J. J., 1993. A phylogenetic analysis and definition of the Carnivora. In: Szalay, F., Novacek, M., McKenna, M., (Eds.), *Mammal Phylogeny: Placentals*. Springer-Verlag, New York, pp. 32-52.
- Yu, L., Li, Q.-w., Ryder, O. A., Zhang, Y.-p., 2004a. Phylogeny of the bears (Ursidae) based on nuclear and mitochondrial genes. *Mol. Phylogenet. Evol.* 32, 480-494.
- Yu, L., Li, Q.-w., Ryder, O. A., Zhang, Y. P., 2004b. Phylogenetic relationships within mammalian order Carnivora indicated by sequences of two nuclear DNA genes. *Mol. Phylogenet. Evol.* 33, 694-705.

Figure legends

Fig 1. Family level phylogeny of the Caniformia. Maximum parsimony bootstrap / Bayesian posterior probability values are listed above the branch for the data set containing all genes and below the branch for the data set containing only the STS introns. All supermatrix and supertree methods recovered the same family-level topology, except the ML-input supertree, which could not resolve the position of the red panda (*Ailuridae*) within the musteloids.

Fig 2. Maximum likelihood tree recovered from the combined data set containing five nuclear genes ($-\ln L=18759.18170$). Branch lengths are optimized and correspond to the number of substitutions/site indicated by the scale bar. The TVMef + I + Γ substitution model (transversional model, equal frequencies) was implemented, as selected by ModelTest 3.06. Base frequencies were therefore set to be equal, the proportion of invariant sites = 0.229823, gamma shape = 0.932748, and substitution rates were: A-C= 1.07832, A-G=C-T= 5.24403, A-T= 0.58459, C-G= 0.97498, G-T=1.0. Bars along the right-hand side of the tree indicate family and higher-level taxonomic groups.

Figure 3. Maximum likelihood cladogram for Ursidae, recovered (a) from all genes and (b) from STS introns only. Support shown is MP BP/BPP for each data set. Nodes that do not have support values indicated were polytomous in both the MP BP and Bayesian topologies. The MP BP tree for all genes additionally resolved the American black bear as closer to the brown and polar bears (59% MP BP). The Bayes-input and ML-input supertrees were identical to the Bayesian topology found using all genes (i.e. resolving only three nodes), while the MP-input supertree additionally resolved the sun bear in the same position as in (a).

Figure 4. Phylogeny of the Pinnipedia. The maximum likelihood cladogram is shown with the MP BP/BPP support from the data set with all genes (above the branch) and from STS only (below the branch). Values of <50% MP BP or <0.5 BPP indicate polytomies in the respective MP BP or Bayesian trees. The MP-input and Bayes-input supertrees recovered consistent topologies but were less resolved. The dotted line indicates the topology (with support) recovered using the STS-only data set, placing the hooded seal as more basal than the harp seal. This pattern was also observed in the ML-input supertree, which also placed the ringed and harp seals as sister and the Antarctic fur seal as more basal than the South American sea lion (indicated by dotted branches). Subfamily designations are indicated on the right.

Fig 5. Phylogeny of the Mustelidae adapted from the maximum likelihood topology (Fig 1). MP BP / BPP values are listed above the branch for the all-data set and below the branch for the STS-only data set. Nodes that do not have support values indicated were polytomous in both the MP BP and Bayesian topologies. The above topology was recovered by all supertree and supermatrix methodologies with the following exceptions. The double-dotted arrow indicates (with associated support) the placement of the *Martes-Eira-Gulo* clade as sister to Melinae, as recovered by Bayesian, MP BP, and the MP-input supertree (ML-input and Bayes-input supertrees were unresolved). The solid arrow indicates the sister relationship between the grison+zorilla clade and the *Mustela* clade, while the double-headed arrow indicates the reciprocal placement of the wolverine and the yellow-throated marten; both as recovered by the ML-input supertree. The dotted arrow indicates the sequential branching of the sea otter and

spotted-necked otters (vs. sister) recovered in the MP-input supertree. Bars along the right-hand side indicate subfamily designations.

ACCEPTED MANUSCRIPT

Tables

Table 1 Taxa included in study and accession numbers for sequences used for phylogenetic analysis. Blank cells = sequence not available.‡

Family	Common Name	Scientific Name	IRBP Exon I	FES	GHR	CHRNA1	RHO1
Ursidae	Brown bear	<i>Ursus arctos</i>	<u>AY303842^a</u>				
	Polar bear	<i>Ursus maritimus</i>	<u>AY303843^a</u>	<u>DQ205765**</u>	<u>DQ205798**</u>	<u>DQ205725**</u>	<u>DQ205839**</u>
	American black bear	<i>Ursus americanus</i>	<u>AY303837^a</u>	<u>DQ205766**</u>	<u>DQ205799**</u>	<u>DQ205726**</u>	<u>DQ205840**</u>
	Asian black bear	<i>Ursus thibetanus</i>	<u>AY303841^a</u>				
	Sun bear	<i>Helarctos malayanus</i>	<u>AY303839^a</u>	<u>DQ205767**</u>	<u>DQ205800**</u>	<u>DQ205727**</u>	<u>DQ205841**</u>
	Sloth bear	<i>Melursus ursinus</i>	<u>AY303838^a</u>	<u>DQ205768**</u>	<u>DQ205801**</u>	<u>DQ205728**</u>	<u>DQ205842**</u>
	Giant panda	<i>Ailuropoda melanoleuca</i>	<u>AY303836^a</u>				
	Spectacled bear	<i>Tremarctos ornatus</i>	<u>AY303840^a</u>				
Procyonidae	Raccoon	<i>Procyon lotor</i>	<u>AB082981^d</u>	<u>AF498183^c</u>	<u>AF498207^c</u>	<u>AF498152^c</u>	<u>AF498231^c</u>
	Crab-eating raccoon	<i>Procyon cancrivorus</i>	<u>AB109332^d</u>				
	White-nosed coati	<i>Nasua narica</i>	<u>DQ205878**</u>	<u>DQ205769**</u>	<u>DQ205802**</u>	<u>DQ205729**</u>	<u>DQ205843**</u>
	Coati	<i>Nasua nasua</i>	<u>AY525031^a</u>				
	Ringtail	<i>Bassariscus astutus</i>	<u>DQ205879**</u>	<u>AF498182^c</u>	<u>AF498206^c</u>	<u>DQ205730**</u>	<u>DQ205844**</u>
	Kinkajou	<i>Potos flavus</i>	<u>DQ205880**</u>	<u>DQ205770**</u>	<u>DQ205803**</u>	<u>DQ205731**</u>	<u>DQ205845**</u>
Ailuridae	Red panda	<i>Ailurus fulgens</i>	<u>DQ205881**</u>	<u>DQ205771**</u>	<u>DQ205804**</u>	<u>DQ205732**</u>	<u>DQ205877**</u>
Mephitidae	Striped skunk	<i>Mephitis mephitis</i>	<u>AB109331^d</u>		<u>DQ205805**</u>	<u>DQ205733**</u>	<u>DQ205846**</u>
	Western spotted skunk	<i>Spilogale gracilis latifrons</i>	<u>DQ205882**</u>	<u>DQ205772**</u>	<u>DQ205806**</u>	<u>DQ205734**</u>	<u>DQ205847**</u>
	Hog-nosed skunk	<i>Conepatus mesoleucus</i>	<u>DQ205883**</u>	<u>DQ205773**</u>	<u>DQ205807**</u>	<u>DQ205735**</u>	<u>DQ205848**</u>
Mustelidae	Sea otter	<i>Enhydra lutris</i>	<u>AB082978^b</u>	<u>AF498162^c</u>	<u>AF498186^c</u>	<u>AF498131^c</u>	<u>AF498210^c</u>
	African clawless otter	<i>Aonyx capensis</i>		<u>AF498160^c</u>	<u>AF498184^c</u>	<u>AF498129^c</u>	<u>AF498208^c</u>
	Oriental small-clawed otter	<i>Amblonyx cinereus</i>		<u>AF498161^c</u>	<u>AF498185^c</u>	<u>AF498130^c</u>	<u>AF498209^c</u>
	River otter	<i>Lontra canadensis</i>	<u>DQ205884**</u>	<u>AF498163^c</u>	<u>AF498187^c</u>	<u>AF498132^c</u>	<u>AF498211^c</u>
	Marine otter	<i>Lontra felina</i>		<u>AF498164^c</u>	<u>AF498188^c</u>	<u>AF498133^c</u>	<u>AF498212^c</u>
	Neotropical otter	<i>Lontra longicaudis</i>		<u>AF498165^c</u>	<u>AF498189^c</u>	<u>AF498134^c</u>	<u>AF498213^c</u>
	Eurasian otter	<i>Lutra lutra</i>		<u>AF498166^c</u>	<u>AF498190^c</u>	<u>AF498135^c</u>	<u>AF498214^c</u>
	Spotted-necked otter	<i>Lutra maculicollis</i>		<u>AF498167^c</u>	<u>AF498191^c</u>	<u>AF498136^c</u>	<u>AF498215^c</u>
	Giant otter	<i>Pteronura brasiliensis</i>		<u>AF498168^c</u>	<u>AF498192^c</u>	<u>AF498137^c</u>	<u>AF498216^c</u>
	American badger	<i>Taxidea taxus</i>	<u>DQ205885**</u>	<u>AF498179^c</u>	<u>AF498203^c</u>	<u>AF498148^c</u>	<u>AF498227^c</u>
	European badger	<i>Meles meles</i>	<u>AB082980^b</u>	<u>AF498178^c</u>	<u>AF498202^c</u>	<u>AF498147^c</u>	<u>AF498226^c</u>
	Japanese badger	<i>Meles meles anakuma</i>	<u>AB082979^b</u>				
	Hog badger	<i>Arctonyx collaris</i>	<u>AY525049^a</u>	<u>AF498180^c</u>	<u>AF498204^c</u>	<u>AF498149^c</u>	<u>AF498228^c</u>
	Chinese ferret badger	<i>Melogale moschata</i>	<u>AB109330^d</u>	<u>AF498181^c</u>	<u>AF498205^c</u>	<u>AF498150^c</u>	<u>AF498229^c</u>
	Fisher	<i>Martes pennanti</i>		<u>AF498173^c</u>	<u>AF498197^c</u>	<u>AF498142^c</u>	<u>AF498221^c</u>
	American marten	<i>Martes americana</i>	<u>AB082963^b</u>	<u>AF498172^c</u>	<u>AF498196^c</u>	<u>AF498141^c</u>	<u>AF498220^c</u>
	Yellow-throated marten	<i>Martes flavigula</i>	<u>AB082964^b</u>				
	Sable	<i>Martes zibellina</i>	<u>AB109329^d</u>				
	Beech marten	<i>Martes foina</i>	<u>AB082965^b</u>				
	Japanese marten	<i>Martes melampus</i>	<u>AB082967^b</u>				
	Pine marten	<i>Martes martes</i>	<u>AB082966^b</u>				
	Wolverine	<i>Gulo gulo</i>	<u>AB082962^b</u>	<u>AF498174^c</u>	<u>AF498198^c</u>	<u>AF498143^c</u>	<u>AF498222^c</u>
	American mink	<i>Mustela vison</i>	<u>AB082977^b</u>	<u>AF498171^c</u>	<u>AF498195^c</u>	<u>AF498140^c</u>	<u>AF498219^c</u>
	Domestic ferret	<i>Mustela putorius furo</i>	<u>AB082974^b</u>	<u>DQ205774**</u>	<u>DQ205808**</u>	<u>DQ205736**</u>	<u>DQ205849**</u>
	Ermine	<i>Mustela ermina</i>	<u>AB082969^b</u>	<u>AF498169^c</u>	<u>AF498193^c</u>	<u>AF498138^c</u>	<u>AF498217^c</u>
	Long-tailed weasel	<i>Mustela frenata</i>	<u>DQ205886**</u>	<u>AF498170^c</u>	<u>AF498194^c</u>	<u>AF498139^c</u>	<u>AF498218^c</u>
	Least weasel	<i>Mustela nivalis</i>	<u>AB082973^b</u>	<u>DQ205775**</u>	<u>DQ205809**</u>	<u>DQ205737**</u>	<u>DQ205850**</u>
	European mink	<i>Mustela lutreola</i>	<u>AB082972^b</u>				
	European polecat	<i>Mustela putorius</i>	<u>AB082975^b</u>				
	Siberian weasel	<i>Mustela sibirica</i>	<u>AB082976^b</u>				
	Steppe polecat	<i>Mustela eversmannii</i>	<u>AB082970^b</u>				
	Japanese weasel	<i>Mustela itatsi</i>	<u>AB082971^b</u>				
	Mountain weasel	<i>Mustela altaica</i>	<u>AB082968^b</u>				
Tayra	<i>Eira barbara</i>		<u>AF498175^c</u>	<u>AF498199^c</u>	<u>AF498144^c</u>	<u>AF498223^c</u>	
Greater grison	<i>Galictis vittata</i>		<u>AF498176^c</u>	<u>AF498200^c</u>	<u>AF498145^c</u>	<u>AF498224^c</u>	
Zorilla	<i>Ictonyx striatus</i>		<u>AF498177^c</u>	<u>AF498201^c</u>	<u>AF498146^c</u>	<u>AF498225^c</u>	

Otariidae	South American fur seal	<i>Arctocephalus australis</i>	<u>DQ205887</u> **	<u>DQ205776</u> **	<u>DQ205810</u> **	<u>DQ205738</u> **	<u>DQ205851</u> **	
	New Zealand fur seal	<i>Arctocephalus forsteri</i>	<u>DQ205888</u> **	<u>DQ205777</u> **	<u>DQ205811</u> **			
	Antarctic fur seal	<i>Arctocephalus gazella</i>	<u>DQ205891</u> **	<u>DQ205780</u> **	<u>DQ205814</u> **			
	South American sea lion	<i>Otaria byronia</i>	<u>DQ205889</u> **	<u>DQ205778</u> **	<u>DQ205812</u> **	<u>DQ205739</u> **	<u>DQ205852</u> **	
	Steller's sea lion	<i>Eumetopias jubatus</i>	<u>DQ205890</u> **	<u>DQ205779</u> **	<u>DQ205813</u> **	<u>DQ205740</u> **	<u>DQ205853</u> **	
Odobenidae	Walrus	<i>Odobenus rosmarus</i>	<u>DQ205892</u> **	<u>DQ205781</u> **	<u>DQ205815</u> **	<u>DQ205741</u> **	<u>DQ205854</u> **	
	Hooded seal	<i>Cystophora cristata</i>	<u>DQ205893</u> **	<u>DQ205782</u> **	<u>DQ205816</u> **	<u>DQ205742</u> **	<u>DQ205855</u> **	
Phocidae	Bearded seal	<i>Erigonathus barbatus</i>	<u>DQ205894</u> **	<u>DQ205783</u> **	<u>DQ205817</u> **	<u>DQ205743</u> **	<u>DQ205856</u> **	
	Harp seal	<i>Phoca groenlandica</i>	<u>DQ205901</u> **	<u>DQ205790</u> **	<u>DQ205825</u> **	<u>DQ205750</u> **	<u>DQ205863</u> **	
	Grey seal	<i>Halichoerus grypus</i>	<u>DQ205902</u> **	<u>DQ205791</u> **	<u>DQ205826</u> **	<u>DQ205751</u> **	<u>DQ205864</u> **	
	Ringed seal	<i>Phoca hispida</i>	<u>DQ205899</u> **	<u>DQ205788</u> **	<u>DQ205823</u> **		<u>DQ205862</u> **	
	Harbor seal	<i>Phoca vitulina</i>				<u>DQ205752</u> **	<u>DQ205865</u> **	
	Spotted seal	<i>Phoca largha</i>	<u>DQ205904</u> **	<u>DQ205793</u> **	<u>DQ205827</u> **	<u>DQ205754</u> **	<u>DQ205867</u> **	
	Caspian seal	<i>Phoca caspica</i>	<u>DQ205905</u> **		<u>DQ205828</u> **	<u>DQ205755</u> **		
	Ross seal	<i>Ommatophoca rossii</i>	<u>DQ205900</u> **	<u>DQ205789</u> **	<u>DQ205824</u> **	<u>DQ205749</u> **		
	Leopard seal	<i>Hydrurga leptonyx</i>	<u>DQ205895</u> **		<u>DQ205818</u> **	<u>DQ205744</u> **	<u>DQ205857</u> **	
	Crabeater seal	<i>Lobodon carcinophagus</i>	<u>DQ205896</u> **	<u>DQ205784</u> **	<u>DQ205819</u> **	<u>DQ205745</u> **	<u>DQ205858</u> **	
	Weddel seal	<i>Leptonychotes weddelli</i>	<u>DQ205903</u> **	<u>DQ205792</u> **		<u>DQ205753</u> **	<u>DQ205866</u> **	
	Northern elephant seal	<i>Mirounga angustirostris</i>	<u>DQ205897</u> **	<u>DQ205785</u> **	<u>DQ205820</u> **	<u>DQ205746</u> **	<u>DQ205859</u> **	
	Hawaiian monk seal	<i>Monachus schauinslandi</i>	<u>DQ205898</u> **	<u>DQ205786</u> **	<u>DQ205821</u> **	<u>DQ205747</u> **	<u>DQ205860</u> **	
	Mediterranean monk seal	<i>Monachus monachus</i>		<u>DQ205787</u> **	<u>DQ205822</u> **	<u>DQ205748</u> **	<u>DQ205861</u> **	
	Canidae	Dog	<i>Canis familiaris</i>	<u>DQ205906</u> **	<u>DQ205794</u> **	<u>DQ205835</u> **	<u>DQ205756</u> **	<u>DQ205868</u> **
		Wolf	<i>Canis lupus</i>	<u>DQ205907</u> **	<u>DQ205795</u> **	<u>DQ205836</u> **	<u>DQ205757</u> **	<u>DQ205869</u> **
Arctic fox		<i>Alopex lagopus</i>	<u>DQ205908</u> **	<u>DQ205796</u> **	<u>DQ205837</u> **	<u>DQ205758</u> **	<u>DQ205870</u> **	
Swift fox		<i>Vulpes velox</i>	<u>DQ205909</u> **		<u>DQ205838</u> **			
Domestic cat		<i>Felis domesticus</i>	<u>Z11811</u> ^c		<u>DQ205829</u> **	<u>DQ205759</u> **	<u>DQ205871</u> **	
Felidae	Canada lynx	<i>Lynx canadensis</i>	<u>DQ205910</u> **	<u>DQ205797</u> **	<u>DQ205830</u> **	<u>DQ205760</u> **	<u>DQ205872</u> **	
	Cougar	<i>Puma concolor</i>	<u>DQ205911</u> **		<u>DQ205831</u> **	<u>DQ205761</u> **	<u>DQ205873</u> **	
	Siberian tiger	<i>Panthera tigris altaica</i>	<u>DQ205912</u> **		<u>DQ205832</u> **	<u>DQ205762</u> **	<u>DQ205874</u> **	
	Bobcat	<i>Lynx rufus</i>	<u>DQ205913</u> **		<u>DQ205833</u> **	<u>DQ205763</u> **	<u>DQ205875</u> **	
	African lion	<i>Panthera leo</i>	<u>DQ205914</u> **		<u>DQ205834</u> **	<u>DQ205764</u> **	<u>DQ205876</u> **	

^aYu et al. (2004) ^bSato et al. (2003) ^cKoepfli and Wayne (2003) ^dSato et al. (2004) ^eStanhope et al. (1992)

** Sequences generated by this study.

‡ Sequence unavailability is either due to the absence of sequence in GenBank, or to unsuccessful amplification or sequencing.

Table 2 Maximum Parsimony and Likelihood Results

Gene partition	# Taxa	Length (bp)	Informative sites*	Informative Gaps	MP Tree Length	# MP Trees	Consistency Index	Retention Index	ML Score (-lnL)	Model
<i>FES</i>	57	454	171	10	489	59328	0.675	0.903	3105.76606	HKY+ Γ
<i>GHR</i>	65	652	220	13	578	92600	0.73	0.917	4004.29335	K80+I+ Γ
<i>RHO1</i>	62	280	83	3	258	1583	0.593	0.863	1666.39816	K80+I+ Γ
<i>CHRNA1</i>	63	394	156	12	446	7184	0.704	0.907	2798.98509	K80+ Γ
<i>IRBP</i>	71	1194	314	1	905	48	0.627	0.895	6567.5753	HKY+I+ Γ
All STS	67	1780	630	38	1792	15676	0.68	0.9	11958.2738	K81 Γ
All data	84	2974	949	39	2708	92600	0.66	0.898	18759.18170	TVMef+I+ Γ **

*Excludes informative gaps **Transversional model equal frequencies (D. Posada, ModelTest).

Figure 2



Figure 3

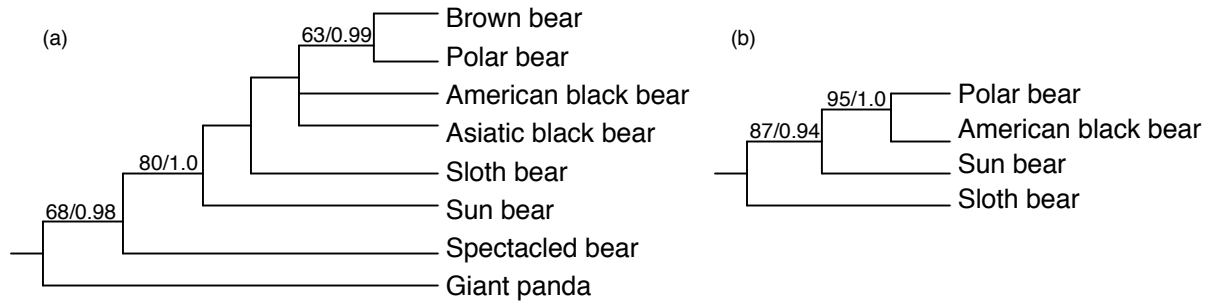


Figure 4

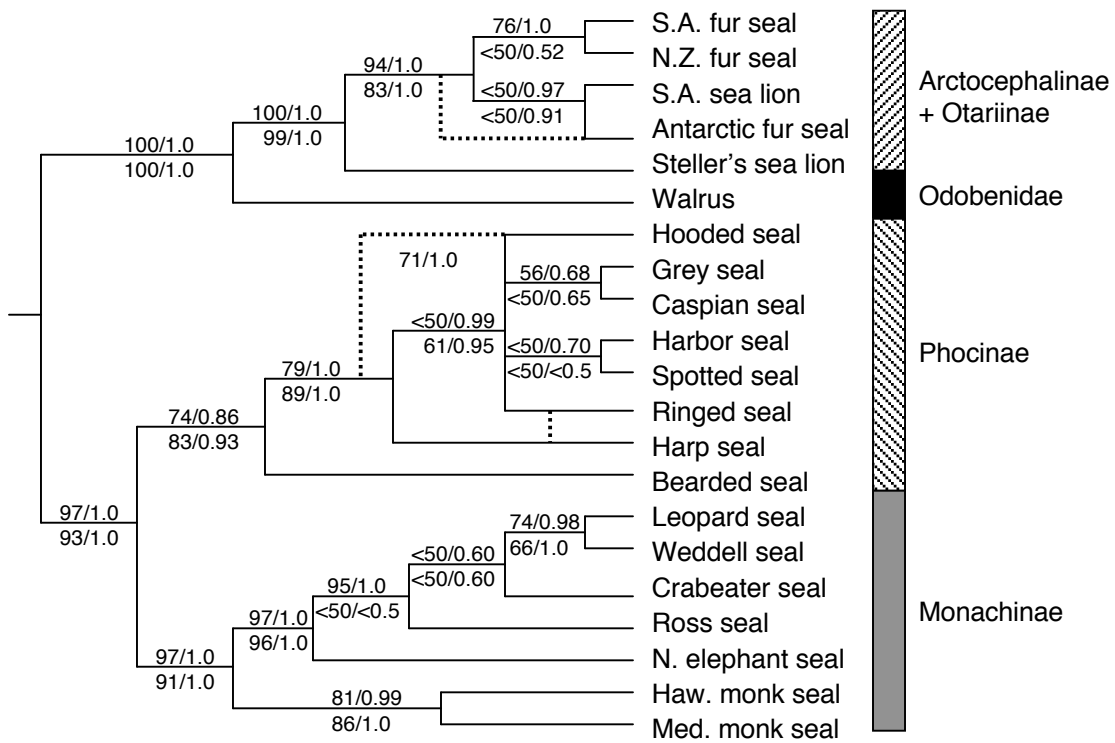


Figure 5

