Bears in a Forest of Gene Trees: Phylogenetic Inference Is Complicated by Incomplete Lineage Sorting and Gene Flow

Verena E. Kutschera,*[1] Tobias Bidon,[1] Frank Hailer,[1] Julia L. Rodi,[1] Steven R. Fain,[2] and Axel Janke*[1,3]

[1]Biodiversity and Climate Research Centre (BiK-F), Senckenberg Gesellschaft für Naturforschung, Frankfurt am Main, Germany

[2]National Fish and Wildlife Forensic Laboratory, Ashland, OR

[3]Institute for Ecology, Evolution and Diversity, Goethe University Frankfurt, Frankfurt am Main, Germany

***Corresponding author:** E-mail: v.kutschera@gmx.net, axel.janke@senckenberg.de.

Associate editor: David Irwin

Abstract

Ursine bears are a mammalian subfamily that comprises six morphologically and ecologically distinct extant species. Previous phylogenetic analyses of concatenated nuclear genes could not resolve all relationships among bears, and appeared to conflict with the mitochondrial phylogeny. Evolutionary processes such as incomplete lineage sorting and introgression can cause gene tree discordance and complicate phylogenetic inferences, but are not accounted for in phylogenetic analyses of concatenated data. We generated a high-resolution data set of autosomal introns from several individuals per species and of Y-chromosomal markers. Incorporating intraspecific variability in coalescence-based phylogenetic and gene flow estimation approaches, we traced the genealogical history of individual alleles. Considerable heterogeneity among nuclear loci and discordance between nuclear and mitochondrial phylogenies were found. A species tree with divergence time estimates indicated that ursine bears diversified within less than 2 My. Consistent with a complex branching order within a clade of Asian bear species, we identified unidirectional gene flow from Asian black into sloth bears. Moreover, gene flow detected from brown into American black bears can explain the conflicting placement of the American black bear in mitochondrial and nuclear phylogenies. These results highlight that both incomplete lineage sorting and introgression are prominent evolutionary forces even on time scales up to several million years. Complex evolutionary patterns are not adequately captured by strictly bifurcating models, and can only be fully understood when analyzing multiple independently inherited loci in a coalescence framework. Phylogenetic incongruence among gene trees hence needs to be recognized as a biologically meaningful signal.

Key words: **species tree, introgressive hybridization, Ursidae, phylogenetic network, coalescence, multi-locus analyses.**

Introduction

Our understanding of evolutionary processes relies on a backbone of phylogenetic inferences from molecular data, but recombination imposes limits on the resolution that can be obtained from a single autosomal locus. High-resolution phylogenies can be obtained in multilocus analyses. In traditional phylogenetic analyses, several loci are concatenated and analyzed as one "superlocus." However, incomplete lineage sorting (ILS), a process by which ancestral polymorphisms can persist through species divergences up to several million years, and gene flow across species boundaries caused by introgressive hybridization generate gene tree discordance, hampering species tree estimation (Tajima 1983; Pamilo and Nei 1988; Leaché et al. 2014). These evolutionary processes are not considered in phylogenetic analyses of concatenated data and can result in inconsistent phylogenetic estimates and high statistical support for an incorrect species tree topology (Kubatko and Degnan 2007).

Bears (Ursidae) are emerging as a prominent example of a mammalian family with a complex speciation history, showing discrepancies among mitochondrial and nuclear phylogenies (Yu et al. 2007; Krause et al. 2008; Nakagome et al. 2008; Pagès et al. 2008; Hailer et al. 2012, 2013; Miller et al. 2012; Cahill et al. 2013). Within bears, the ursine subfamily comprises the American and Asian black bear (*Ursus americanus, U. thibetanus*), sun bear (*Helarctos malayanus*), sloth bear (*Melursus ursinus*), brown bear (*U. arctos*), polar bear (*U. maritimus*), plus numerous extinct taxa. In addition, bears also include the giant panda (*Ailuropoda melanoleuca*) and spectacled bear (*Tremarctos ornatus*). In phylogenetic analyses of genes from the nuclear genome, the placement of the sun bear, sloth bear, and Asian black bear remained unclear (Yu et al. 2004; Nakagome et al. 2008; Pagès et al. 2008). These analyses were performed using a combination of intron and exon sequences, rendering it difficult to interpret whether nodes with low statistical support resulted from insufficient resolution or from actual conflict in evolutionary signals among loci. Moreover, in these studies only one (consensus) sequence per species was analyzed and data from several markers were concatenated, precluding the identification of paraphyletic relationships among species.

Recently, coalescence-based multilocus species tree approaches have been developed (e.g., Heled and Drummond 2010). These analytical advances make it possible

to specifically model the complexity of lineage sorting and to incorporate intraspecific variation and heterozygosity within individuals. Accuracy of such multilocus species trees can be additionally improved by sampling several individuals per species, especially at shallow phylogenetic depths at which lineages are not completely sorted (Maddison and Knowles 2006). This is especially relevant in ursine bears, because the fossil record and dated phylogenies of mitochondrial genome sequences suggested a rapid radiation (Wayne et al. 1991; Yu et al. 2007; Krause et al. 2008), including time frames in which ILS is expected (Nichols 2001).

Another cause of gene tree discordance can be introgressive hybridization, resulting in gene flow across species boundaries, which can only be estimated when intraspecific variation is considered. Although ILS can be modeled in currently available species tree approaches, they cannot account for gene flow. A recent simulation study showed that gene flow can affect species tree inferences by decreasing posterior clade probabilities, underestimating divergence time estimates, and, in cases of high levels of gene flow, by altering the species tree topology (Leaché et al. 2014). Discordance among loci that differ in ploidy and inheritance mode can be explained by contrasting patterns of female and male gene flow (Chan and Levin 2005). In brown and polar bears, discordance between the mitochondrial gene tree and the nuclear species tree has been found (Hailer et al. 2012, 2013; Miller et al. 2012; Cronin et al. 2013), and explained with introgressive hybridization. Previous studies have also indicated phylogenetic discrepancies between mitochondrial and nuclear genes in American and Asian black bears (Yu et al. 2004; Nakagome et al. 2008; Pagès et al. 2008), suggesting that similar processes may have affected their evolution. To examine whether incongruences among nuclear loci and/or discordance between nuclear and mitochondrial phylogenies can be explained by introgression, coalescence-based multilocus gene flow analyses (e.g., Nielsen and Wakeley 2001; Hey 2010; Yu et al. 2012, 2013) can be used to complement species tree inferences. Thus, to more fully understand the evolutionary history of bears, it is crucial to analyze multiple independently inherited markers with a high resolution in several individuals per species. Such data sets need to be analyzed using coalescence models, tracing the evolutionary histories of individual alleles back in time, from extant individuals to their ancestral populations.

We here study the evolutionary history of bears, using a combination of coalescence-based species tree approaches and gene flow analyses. For this purpose, we generated sequence data of 14 independently inherited autosomal introns in 30 individuals and of 5.9 kb from the Y chromosome in 11 males from all eight extant bear species. We combine this with previous data into data sets comprising 29 kb of nuclear sequence and 10.8 kb of mitochondrial sequence to analyze the complexity of phylogenetic signals in bears through multilocus species tree and network analyses, and in statistical model comparisons. Further, we use coalescent-based gene flow analyses to specifically investigate whether remaining conflicts in phylogenetic signals in bears can be explained by introgressive hybridization.

Results

Basic Variability Statistics and Allele Sharing among Ursinae

We sequenced 14 autosomal introns from two to seven individuals per species yielding 7,991 bp, and nine markers from the Y chromosome yielding 5,907 bp in 11 male individuals, representing all extant bear species (supplementary table S1, Supplementary Material online). For giant panda, spectacled bear, sloth bear, sun bear, and Asian black bear, Y-chromosomal data were obtained from all available male individuals. Because of low intraspecific variability of Y chromosomes in brown, polar, and American black bears (Bidon et al. 2014), we included Y-chromosomal data from only one individual of each of these species.

The number of variable sites was 515 across the 14 sequenced autosomal introns and 325 at Y-chromosomal sequence. The total sequence data generated in this study thus comprised 840 variable sites. In contrast, upon concatenation of the autosomal intron data, collapsing all variation within and among individuals into a 50% majority-rule consensus sequence per species, only 396 variable sites remained. Thus, intraspecific and intraindividual polymorphism contributed more than 30% to the phylogenetic signal in our autosomal data. Accordingly, interspecific *p*-distances of our autosomal introns including all phased individuals were on average 115% of the *p*-distances of the same 14 concatenated autosomal introns, and on average 178% of the *p*-distances of previously published autosomal sequences that did not consider intraspecific variability and that included both exon and intron sequences (Pagès et al. 2008; supplementary table S2, Supplementary Material online). High levels of shared polymorphisms were found between brown and Asian black bears, between American black and Asian black bears, and between brown and American black bears (supplementary table S3, Supplementary Material online). All ursine species pairs had similar mean genetic distances. Haplotype networks revealed various combinations of interspecific haplotype sharing for 12 of 14 autosomal introns (fig. 1, supplementary fig. S1 and table S4, Supplementary Material online). At eight introns, haplotypes were shared between closely related species, and at four introns, haplotypes were shared between more distantly related species. Across pairwise comparisons among species, the ratio of polymorphic sites to fixed differences increased toward shallower divergences (supplementary table S3, Supplementary Material online).

Haplotype networks showed Y chromosomes from different species as clearly distinct from each other (fig. 1). In contrast to autosomal markers, no haplotype sharing was found. At marker *579.3C*, a large insertion in sloth and sun bears (222 and 221 bp, respectively) was 93% identical to a transposable element from the giant panda (SINEC1_Ame). Mean pairwise distances between species were similar for the Y-chromosomal and autosomal data sets, when at least one of the compared species was giant panda or spectacled bear (supplementary table S3, Supplementary Material online). Within Ursinae, however, relatively fewer Y-chromosomal than

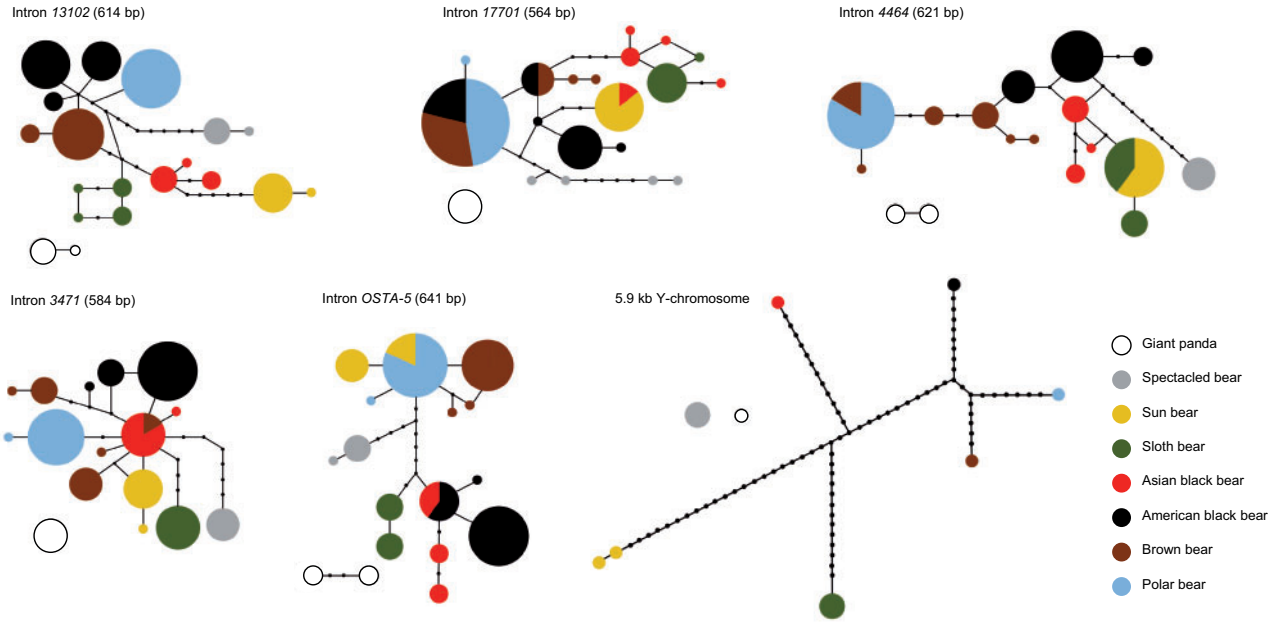


Fig. 1. Statistical parsimony networks for five autosomal intron markers and 5.9 kb of Y-chromosomal sequence in bears. Circle areas are proportional to haplotype frequencies and inferred intermediate states are shown as black dots. For some loci, spectacled bear and giant panda haplotypes were too divergent to be connected at the 95% credibility limit. Likewise, in the Y-chromosomal data set, sun bear haplotypes were connected at the 94% credibility limit. Haplotype networks for nine additional autosomal intron markers are shown in supplementary figure S1, Supplementary Material online.

autosomal substitutions were observed, a pattern also reported by Nakagome et al. (2008). We found a total of three pseudoheterozygous sites on the Y chromosome, all located within 119 bp of marker *403*. The respective columns were removed from the alignment prior to any analysis. Pseudoheterozygous sites on the generally haploid Y chromosome can occur due to segmental duplications (Sachidanandam et al. 2001; Hallast et al. 2013).

Multilocus Species Tree Analyses

*BEAST, a multilocus coalescence approach, jointly estimates gene trees from independently inherited loci, as well as the species tree in which the gene trees are embedded. By including two phased haplotypes per individual and autosomal locus, and data from several individuals per species, variation among and within individuals could be explicitly considered. A multilocus analysis of all nuclear markers from this study yielded a topology placing the American black bear as sister-taxon to a brown/polar bear clade, which was supported by high posterior probability (fig. 2*A*). A clade consisting of Asian black, sun, and sloth bears was recovered with high statistical support. Topological uncertainty within this clade was represented in a cloudogram of species trees sampled from the posterior distribution (Bouckaert 2010) by lines (topologies) connecting the sloth bear with the Asian black bear, and a horizontal line indicating a placement of the sloth bear as sister-taxon to sun and Asian black bear (fig. 2*A*). A topology placing the Asian black bear as sister-taxon to the American black bear, brown bear, and polar bear was represented by faint lines in the cloudogram. Conflicting signals in our nuclear data were further illustrated in a consensus network of the 14 autosomal gene trees from all phased individuals (fig. 3). Although there was a clear separation between an American black, brown, polar bear clade on the one side and an Asian black, sun, sloth bear clade on the other side, the topology deviated from a bifurcating tree. In particular, conflict among Asian black, sun, and sloth bears was depicted by a cuboid, and brown and American black bears were grouped closely together. Using a minimum estimate of 11.6 Ma for the divergence time of the giant panda from the other bear species resulted in a divergence time estimate of the ursine bears from the spectacled bear around the transition from the Miocene to the Pliocene (median: 5.88 Ma; fig. 2*A*, table 1). The divergence between the Asian black, sun, sloth bear clade and the American black, brown, polar bear clade was placed to the early Pleistocene (median: 1.78 Ma). Subsequent divergences within Ursinae occurred during the Pleistocene, within about 1.8 My. The average median posterior estimate of the substitution rate across loci obtained from our calibrated *BEAST analysis was 0.95×10^{-8} substitutions per site per generation, assuming an average generation time for bears of 7.2 years.

In a *BEAST analysis of the 14 autosomal introns alone (data not shown), and in a BEAST analysis of the Y-chromosomal sequences alone (fig. 2*B*), the same topology was obtained as in the combined species tree analysis (fig. 2*A*), but with lower statistical support for an Asian black, sun, sloth bear clade. Phylogenetic analyses of concatenated nuclear data were conducted for comparison and are described in the supplementary material, Supplementary Material online. A *BEAST analysis of a combined data set including our data and previously published sequences (29 kb from 30 nuclear

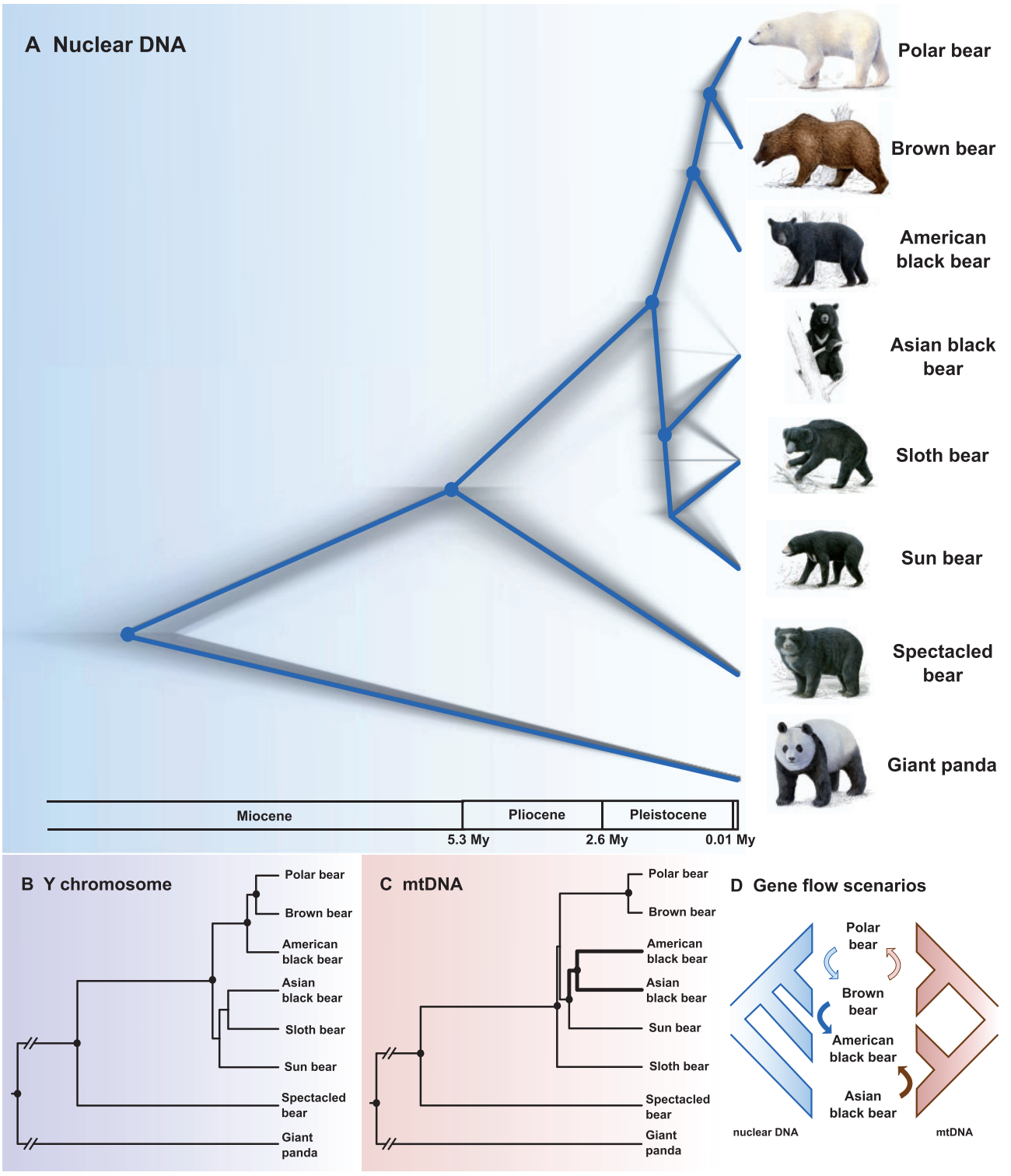


Fig. 2. (*A*) Cloudogram of species trees from *BEAST analysis, based on 14 autosomal introns and 5.9 kb of Y-chromosomal sequence (90,000 species trees). The consensus tree of the most frequently occurring topology in the posterior distribution is superimposed onto the cloudogram in blue. Blue dots at nodes indicate posterior support >0.96 in the maximum-clade-credibility tree. Frequency of different topologies occurring in the posterior distribution is illustrated by width and intensity of grey branches. Variation in density along the *x* axis portrays variation in time estimates of divergences. (*B*) Gene tree of 5.9-kb Y-chromosomal sequence from BEAST. Note that in a *BEAST analysis of the 14 autosomal introns alone, the same topology was obtained, with low statistical support ($P < 0.95$) for a clade of Asian black bears, sun bears, and sloth bears (data not shown). (*C*) Gene tree of mitochondrial genome data (protein-coding regions, excluding *ND6*) from BEAST. Black dots at nodes indicate posterior support >0.95. (*D*) Schematic scenarios for interspecific gene flow that could explain discordance between mitochondrial and nuclear phylogenies. Blue arrows: Nuclear gene flow, brown arrows: Introgression of mtDNA. Light blue and light brown arrows indicate gene flow identified in previous studies (Hailer et al. 2012, 2013; Miller et al. 2012; Cahill et al. 2013; Liu et al. 2014). Note that IMa2 identified additional introgression signals from Asian black into sloth bears (supplementary fig. S3*B*, Supplementary Material online).

Table 1. Divergence Time Estimates Obtained from *BEAST Based on 15 Nuclear Markers (14 autosomal introns and Y-chromosomal sequence).

Prior	Estimated Divergence Time, Ma (95% HPD interval)						
	Giant Panda/Spect. Bear + Ursinae	Spect. Bear/Ursinae	Polar + Brown + Am. Black Bear/Asian Black + Sun + Sloth Bear	Asian Black Bear/Sun + Sloth Bear	Sun/Sloth Bear	Am. Black Bear/Polar + Brown Bear	Polar/Brown Bear
Root height min. 11.6 Ma	12.46	5.88	1.78	1.56	1.42	0.94	0.62
(Abella et al. 2012)	(11.6–14.48)	(4.67–7.18)	(1.42–2.2)	(1.2–1.96)	(1.04–1.81)	(0.67–1.25)	(0.38–0.89)

markers; supplementary table S5, Supplementary Material online, lists all analyzed data sets) did not converge within 2×10^9 generations, likely due to incongruent signals among loci. In a cloudogram of this analysis (results not shown), the three most frequent topologies were the same as obtained from our 15 loci data set (fig. 2*A*), but the third most common topology, which was identical to those previously published by Nakagome et al. (2008) and Pagès et al. (2008), was represented by thick lines placing the Asian black bear as sister-taxon to the American black, brown, polar bear clade. Thus, this third topology occurred more often in the data from previous studies than in our own intron and Y chromosome data, illustrating the heterogeneity of phylogenetic signals in bears.

Contrasting Signals from Nuclear and Mitochondrial DNA

When reanalyzing mitochondrial genomes from all eight extant bear species in BEAST, we obtained a topology with the sloth bear as sister-taxon to all other ursines with limited support and the sun bear as sister-taxon to an American and Asian black bear clade (fig. 2*C*; Yu et al. 2007; Krause et al. 2008). This topology differed from nuclear phylogenies (fig. 2*A* and *B*).

We evaluated the phylogenetic signal from the mitochondrial data set and two Y-chromosomal data sets (supplementary tables S5 and S6, Supplementary Material online) for their fit on 105 different tree topologies that can be built for five operational taxonomical units. In these topologies, polar and brown bears were constrained to be sister taxa, the spectacled bear as sister-taxon to all ursines, and the giant panda as outgroup.

In approximately unbiased (AU) tests of mitochondrial data, all three possible positions of the sloth bear in this phylogeny obtained high probability (*P*) values, with low differences in the log-likelihood values ($\Delta \log L$) relative to the best tree (supplementary table S6*A*, Supplementary Material online). All topologies obtained from analyses of nuclear DNA in this and in previous studies (Nakagome et al. 2008; Pagès et al. 2008) were incompatible with the mitochondrial data set ($P < 0.01$; supplementary table S6*A*, Supplementary Material online). Conversely, all three mitochondrial topologies were incompatible with the Y-chromosomal data, regardless whether our Y-chromosomal sequences were analyzed alone, or when combining Y-chromosomal sequences from this study with Y-linked markers from Nakagome et al. (2008) and Pagès et al. (2008) (supplementary table S6*B* and *C*, Supplementary Material online). For both these Y-chromosomal data sets, the highest *P* value was observed for the topology that was also reconstructed in BEAST using our own Y-chromosomal data set (fig. 2*B*) Additional topologies could not be rejected ($P \geq 0.05$), including the species tree topology (fig. 2*A*). Topologies from previous publications were characterized by large $\Delta \log L$ values, and some were incompatible ($P < 0.05$).

To perform statistical comparisons of the mitochondrial and the nuclear species tree topologies, we conducted analyses of our nuclear data in *BEAST, in which we constrained the species tree topology to either the mitochondrial topology (fig. 2*C*) or the species tree topology (fig. 2*A*), respectively. The latter analysis was carried out to ensure that constraining per se did not affect the analysis. To test the two hypotheses, posterior probabilities were compared using Bayes factors (BF) (Kass and Raftery 1995; Suchard et al. 2005), the Bayesian analog of likelihood ratio (LLR) tests. Considering a $\log_{10}(BF) > 2$ (or $BF > 100$) as "decisive" (Kass and Raftery 1995), the nuclear species tree topology was favored over the mitochondrial gene tree topology with high statistical support ($\log_{10}[BF] = 4.2$, or $BF = 15{,}811$).

Gene Flow and Demographic Analyses

Multilocus coalescence approaches such as *BEAST can efficiently accommodate ILS, but they do not model gene flow, although the latter can significantly impact phylogenetic inferences (Leaché et al. 2014). We therefore used IMa2, which is based on an isolation-with-migration model and jointly estimates six demographic parameters, including population migration rates between populations since their divergence from a common ancestral population. We analyzed species pairs where conflict between mitochondrial and species tree topologies was found (brown bear–American black bear, American black bear–Asian black bear), or based on shared haplotypes between distantly related species (polar bear–sun bear). Pairs of Asian bear species (Asian black bear–sun bear, Asian black bear–sloth bear, sloth bear–sun bear) were selected to investigate whether past introgression may explain the uncertain branching order among these species ($P = 0.67$; fig. 2*A*).

IMa2 analyses indicated significant unidirectional gene flow from the brown bear into the American black bear lineage (table 2 and supplementary fig. S3*A*, Supplementary Material online), irrespective of the upper prior boundaries chosen. This was also evident from haplotype sharing

Table 2. Demographic Parameters (modal values; 95% HPD interval in parentheses) from Analyses of Bear Species Pairs in IMa2, Based on 14 Autosomal Introns.

Species 1	Species 2	N_{e1}	N_{e2}	$2N_1M_1$	$2N_2M_2$
American black bear	Asian black bear	21,432 (8,664–44,233)	44,233 (18,696–94,394)	0 (0–0.16)	0.03 (0–0.38)
American black bear	Brown bear	20,178 (8,550–37,963)	43,435 (24,282–76,267)	0.08[a] (0.01–0.24)	0 (0–0.12)
Polar bear	Sun bear	3,967 (1,231–11,355)	16,279 (6,703–33,517)	0.01 (0–0.06)	0 (0–0.09)
Asian black bear	Sun bear	46,969 (21,432–89,834)	19,608 (7,752–44,233)	0.03 (0–0.23)	0 (0–0.12)
Asian black bear	Sloth bear	46,969 (22,344–88,922)	4,104 (1,368–16,872)	0 (0–0.18)	0.03[a] (0–0.1)
Sloth bear	Sun bear	1,368 (0–10,488)	4,104 (1,368–16,872)	0.01 (0–0.07)	0.04 (0–0.16)

N_{e1} and N_{e2}, effective population sizes for species 1 and 2, respectively; $2N_1M_1$, population migration rate into species 1 from species 2 per generation; $2N_2M_2$, population migration rate into species 2 from species 1 per generation. Posterior probability distributions for parameters are shown in supplementary figure S3, Supplementary Material online.

[a]Migration rates that are significantly different from zero at the $P < 0.05$ level in LLR tests (Nielsen and Wakeley 2001; Hey 2010).

between brown and American black bears, which shared four haplotypes at three introns (fig. 1, supplementary fig. S1 and table S4, Supplementary Material online). Between American and Asian black bears, two haplotypes were shared at two introns, but multilocus analyses in IMa2 revealed no significant gene flow between these two species. The posterior distribution for gene flow from American into Asian black bears showed a peak at 0.03 migrants per generation, but the 95% highest posterior density (HPD) interval included zero (table 2 and supplementary fig. S3*A*, Supplementary Material online). The same applied to sun and polar bears, which also shared two haplotypes at two introns. Although the 95% HPD interval for gene flow from sun into polar bears also included zero, the posterior distribution had a clear peak at 0.01 migrants per generation.

In IMa2 analyses of Asian bear species pairs, significant unidirectional gene flow was detected from the Asian black bear lineage into the sloth bear lineage at a rate of 0.03 migrants per generation (table 2 and supplementary fig. S3*B*, Supplementary Material online), consistent with shared variation between the two species (supplementary table S3, Supplementary Material online). Neither between Asian black and sun bears, nor between sloth and sun bears, significant signals of gene flow were detected (table 2), although in both cases, two haplotypes were shared at two introns (fig. 1, supplementary fig. S1 and table S4, Supplementary Material online). The posterior distributions for gene flow from sun into Asian black bears, from sloth into sun bears, and from sun into sloth bears showed clear peaks at 0.01–0.04 migrants per generation, but the 95% HPD intervals included zero (table 2 and supplementary fig. S3*B*, Supplementary Material online).

IMa2 showed small effective population sizes (N_e) for polar bears and sloth bears, and much larger values for brown and Asian black bears (table 2 and supplementary fig. S3*C*, Supplementary Material online), consistent with current nucleotide diversity levels (supplementary table S7, Supplementary Material online). In all IMa2 runs, the posterior distributions of ancestral population size had a clear peak, but for some species pairs, the upper tails did not approach zero, even in runs based on much wider priors (supplementary fig. S3*D*, Supplementary Material online). The right tails of the posterior distributions of the time since population splitting also did not converge on zero, so this parameter could not be estimated with certainty for any species pair (supplementary fig. S3*E*, Supplementary Material online). However, when restricting the prior for the splitting time to the minimum age of the youngest *Ursavus* fossil (ca. 7.1 My; Fortelius 2003), the genus that is believed to have given rise to the *Ursus* lineage (Kurtén 1968), the highest peaks of the posterior distributions coincided with the geological ages of time estimates inferred in *BEAST (table 1 and supplementary fig. S3*E*, Supplementary Material online). In summary, our gene flow analyses thus indicated that besides ILS, introgression also played a role during the evolutionary history of bears.

Discussion

Introgression and ILS both lead to variation in the phylogenetic signal among loci and individuals from the same species, causing gene tree discordance. Especially in rapidly diverged species such as ursine bears, disentangling the effects of ILS and introgression remains challenging. Because concatenation approaches cannot model or portray either of these processes, we instead used coalescent-based multilocus methods to analyze multiple independently inherited loci sequenced in several individuals from each extant bear species.

We first reconstructed phylogenetic trees based on nuclear data. Next, we specifically investigated whether gene flow could explain observed incongruences among nuclear loci, and the conflict between the nuclear species tree and the mitochondrial phylogeny. This approach provided a more comprehensive understanding of the evolutionary process than by simply aiming at a fully resolved bifurcating tree. By explicitly considering intraspecific and intraindividual variation, we demonstrate that both ILS and introgression have shaped the evolutionary history of ursine bears.

Species Tree Inferences in the Presence of ILS and Introgression

The multilocus species tree of autosomal introns and Y-chromosomal sequence from this study (fig. 2*A*) is similar, but not identical, to phylogenetic trees reconstructed in previous studies based on concatenated nuclear data. In contrast to the concatenation approach, however, ILS is specifically considered and modeled in our species tree estimation. We obtained high posterior support for a placement of the giant

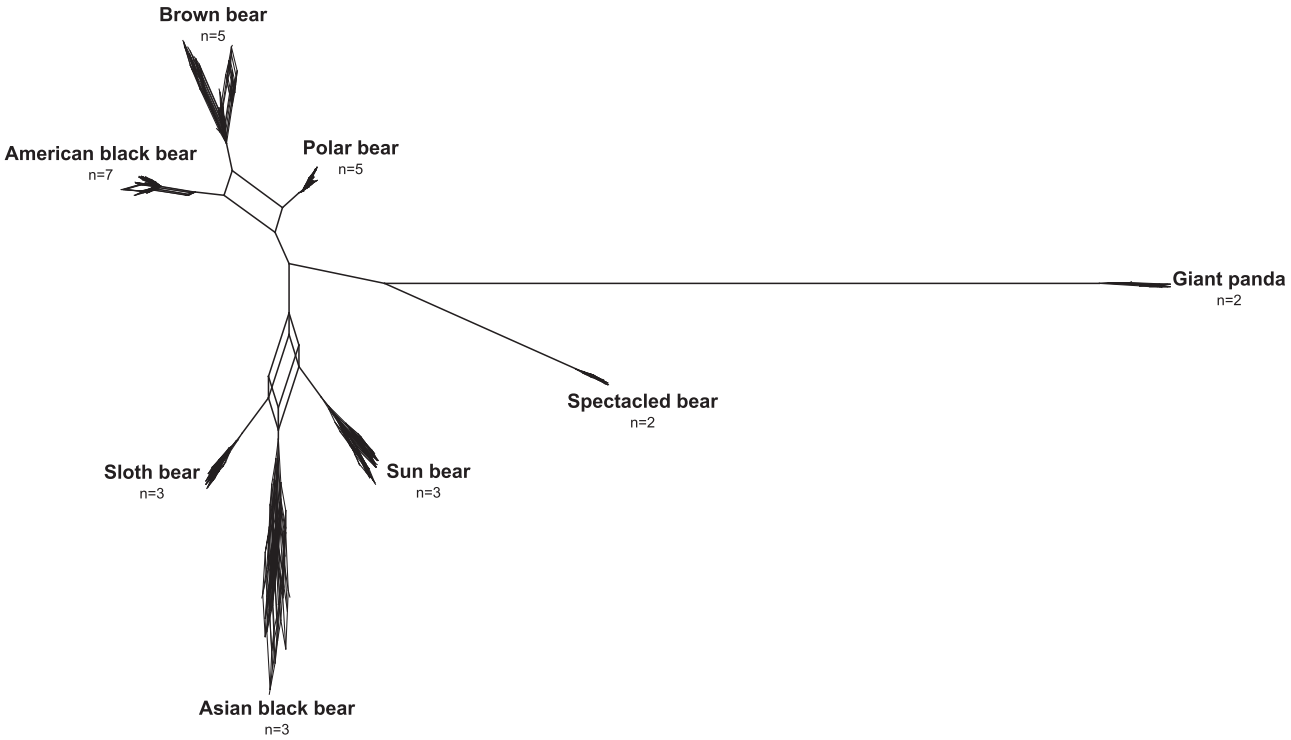


Fig. 3. Consensus network of 14 autosomal gene trees obtained from a *BEAST analysis of 14 nuclear introns. All splits found in at least two gene trees (2/14, threshold = 0.14) are shown. *n*, number of individuals analyzed per species.

panda and the spectacled bear outside the variation of all Ursinae, and for a brown, polar, and American black bear clade. A previous study placed the Asian black bear as sister-taxon to the brown, polar, and American black bear with high statistical support (Pagès et al. 2008). In our species tree, however, sun, sloth, and Asian black bear, the three species whose current distributions are limited to Asia, form a highly supported clade. The sun, sloth, and Asian black bear clade is distinct from the brown, polar, and American black bear clade also in our consensus network of autosomal gene trees (fig. 3). Because the sun and sloth bear are currently not included in the *Ursus* genus, our findings render *Ursus*, as it is currently defined, paraphyletic.

The exact branching order within the clade of Asian bear species is complex, however, as illustrated by a cuboid connecting Asian black bears, sun bears, and sloth bears in the consensus network. Some support for a sister relationship between the sun bear and the sloth bear comes from a sequence insertion in sun and sloth bears in the Y chromosome, which is 93% identical to a transposable element from the giant panda (SINEC1_Ame). We note, however, that more insertions are required to obtain statistical significance (Waddell et al. 2001). Low statistical support for a sister relationship of sun and sloth bears in the species tree (fig. 2*A*) can result from introgression, as *BEAST does not model gene flow. A recent simulation study showed that even low levels of gene flow between nonsister species reduce statistical support for the true sister species clade in species tree inferences using *BEAST (Leaché et al. 2014). Indeed, we detect weak, but significant unidirectional gene flow from the Asian black bear lineage into the sloth bear lineage (table 2 and supplementary fig. S3*B*, Supplementary Material online). This is consistent with low statistical support for a sun and sloth bear clade, and with alternative topologies in the cloudogram of species trees showing Asian black and sloth bears as sister species. Thus, a combination of phylogenetic and gene flow estimation approaches suggests that sun and sloth bears may be sister species that have been impacted by introgression from a bear lineage related to extant Asian black bears.

Due to their haploid nature and uniparental inheritance, mitochondrial and Y-chromosomal loci are expected to sort more rapidly than biparentally inherited autosomal loci. In contrast to mtDNA, intraspecific variation on the Y chromosome is low in many mammals (Hellborg and Ellegren 2004), but differences are predicted to accumulate quickly among lineages (Petit et al. 2002). Furthermore, the Y chromosome lacks recombination over most of its length. Therefore, it constitutes a high-resolution record of evolutionary history. Accordingly, the Y chromosome shows haplotypes from different species as clearly distinct (fig. 1). Despite differences in the pattern of haplotype sharing and in the mean distances between pairs of ursine species, the Y-chromosomal gene tree and the autosomal species tree show congruent phylogenetic signals, and both marker systems contrast with the phylogenetic signal of mtDNA with high statistical confidence (fig. 2 and supplementary table S6, Supplementary Material online).

Rapid Speciation and ILS in Ursine Bears

Several lines of evidence suggest extensive ILS for autosomal loci in ursine bears. A large number of polymorphic sites

within species compared with the number of fixed differences between ursine species pairs confirm that intraspecific polymorphism makes a major contribution to the overall phylogenetic signal on autosomal loci—a signal that needs to be considered. However, this is not possible in noncoalescence-based phylogenetic analyses of concatenated data. We show that haplotype sharing in bears occurs most frequently between closely related species. Neither haplotypes nor polymorphic sites are shared between giant pandas, spectacled bears, and ursine bears. Our divergence time estimates indicate that speciation events in ursine bears occurred within only about 1.8 My. Assuming an average N_e of 28,000 individuals for brown and polar bears (Miller et al. 2012; Hailer et al. 2013; Nakagome et al. 2013) and a generation time of 10 years (Tallmon et al. 2004; Cronin et al. 2009), lineage sorting for most autosomal loci in bears requires 1.1–2.0 My, based on coalescence theory (corresponding to 4–7 N_e generations; Nichols 2001). Considering the rapid radiation of ursine bears, ILS is thus expected to be common in the autosomal part of their genome.

Ursine bears descended directly from *U. minimus* (Kurtén 1968), a species known from the fossil record. Thus, modern ursine bears most likely radiated after the last occurrence of this species in the fossil record. Indeed, our time estimate for the onset of the ursine radiation is younger than the youngest *U. minimus* fossil, which was dated to 2.6–3.4 Ma (Fortelius 2003). Our estimation places the onset of the radiation of Ursinae to the early Pleistocene, and the most recent speciation event, the polar/brown bear divergence, to the mid Pleistocene. In contrast to divergence time estimates based on mitochondrial genomes (Yu et al. 2007; Krause et al. 2008), our estimated time frame excludes the Miocene. Our polar/brown bear divergence time estimate is similar to other recent estimates from nuclear data (Edwards et al. 2011; Hailer et al. 2012; Cahill et al. 2013; Liu et al. 2014), but younger than the 4–5 Ma proposed by Miller et al. (2012). We note that our estimates may underestimate the actual divergence times, and that the incorporation of sequence data from ancient bear specimens as fossil tip calibration points will likely allow for more refined divergence time estimates. The average substitution rate across all loci obtained from our calibrated *BEAST analyses of 0.95×10^{-8} substitutions per site per generation is lower than a rate estimated for primates (2.5×10^{-8} substitutions per site per generation; Nachman and Crowell 2000). Applying the faster rate from primates would lead to even younger divergence time estimates for bears. Regardless of the exact timing, the Plio-/Pleistocene epoch was characterized by climatic fluctuations, dramatic changes in habitat characteristics and habitat fragmentation, promoting population differentiation and speciation but also allowing for secondary contact.

Our study shows that the rapid radiation of bears did not allow for complete lineage sorting on their autosomes. This is reflected in the high degree of shared polymorphic sites and haplotypes between ursine species, in our network analyses, and in the short internal branches found in the present and in previous phylogenetic analyses of ursines (Yu et al. 2007; Krause et al. 2008; Nakagome et al. 2008; Pagès et al. 2008). These findings highlight that the extent of ILS on the autosomes of species with similar population sizes and speed of speciation as ursine bears is not to be underestimated.

Accounting for ILS was only possible because we consider intraspecific variability within a coalescence framework. In contrast, previous phylogenetic studies of the bear family analyzed concatenated sequences of only one (consensus) individual per species, without being able to specifically model the genealogical history of intraspecific variation, which was made possible by recent methodological developments. A recent simulation study demonstrated that sampling effort in terms of number of individuals and markers had a large effect on species tree accuracy, especially when lineage sorting was incomplete (Lanier and Knowles 2012). In that study, accurate species tree estimates were obtained by sampling three individuals per species and nine independent loci, suggesting that our sampling scheme should yield reliable results. Thus, by extending the available data on bears with sequences of high resolution from several individuals per species, and by using an advanced coalescence multilocus approach that specifically models ILS, complemented by multilocus gene flow analyses, our data set allows for the estimation of a statistically robust species tree of bears, including divergence time estimates.

Haplotype networks of autosomal introns further illustrate the effect of sampling several individuals per species. For example, depending on which Asian black bear individual is chosen for phylogenetic analysis, the signal would be altered, as each Asian black bear individual shares different haplotypes with different other bear species. Moreover, data sets analyzed in previous studies contained less than half of the number of variable sites of our data set, highlighting that a considerable amount of genealogical information resides within species, including the variation found among individuals, as well as intraindividual variability (heterozygous sites).

Discordance between Mitochondrial and Nuclear Phylogenies of Bears

We find evidence for ILS among ursine bear species and gene flow from Asian black bears into sloth bears, causing incongruences among genealogical histories of nuclear loci. Similarly, discordances between mitochondrial and nuclear phylogenies in bears have been reported previously, but without explicitly testing alternative hypotheses considering ILS or introgression. We show that the nuclear species tree of ursine bears conflicts with the mitochondrial gene tree topology using statistical model comparisons in a coalescence framework, and that the Y-chromosomal and the mitochondrial gene tree are mutually exclusive using likelihood-based statistical tests, both with high statistical significance. Such discordance can be explained by differences in ploidy and inheritance mode of the maternally inherited mtDNA, the paternally inherited Y chromosome, and the biparentally inherited autosomal loci, which capture different aspects of evolutionary history. Therefore, comparing differentially inherited loci allows for the identification of possibly

contrasting patterns of female and male gene flow, and of introgression events.

Discordance between the mitochondrial gene tree on the one side and the autosomal species tree and the Y-chromosomal gene tree on the other side has already been documented for brown bears and polar bears (Hailer et al. 2012, 2013; Miller et al. 2012; Cahill et al. 2013; Bidon et al. 2014). This pattern was explained with introgressive hybridization between the two species and the replacement of the polar bear mitochondrial genome (mitochondrial capture; fig. 2*D*). Hybridization between different bear species has been observed in zoos and in the wild (Gray 1972; Kelly et al. 2010). The discordant placement of the American black bear in the nuclear species tree and in the mitochondrial gene tree (fig. 2*A*–*D*), and the detection of unidirectional gene flow from the brown bear into the American black bear lineage suggest a similar process for American black, Asian black, and brown bears.

Two hybridization scenarios could explain the incongruent placement of the American black bear in the nuclear species compared with the mitochondrial gene tree (fig. 2*D*): A) The replacement of the original American black bear mtDNA by an Asian black bear-like lineage through introgressive hybridization (mitochondrial capture), leading to a matrilineal sister-relationship of the two species. Alternatively B), nuclear swamping of the American black bear genome by genetic material from the brown bear through male-mediated introgressive hybridization, causing the placement of the American black bear with the brown/polar bear clade in the nuclear species tree (see Leaché et al. 2014).

Mitochondrial capture (scenario A) would require hybridization between Asian and American black bears (fig. 2*D*). The current distribution of Asian and American black bears is allopatric. However, the Bering land bridge connected eastern Asia and North America several times for long time periods during the Pleistocene (Hoffecker and Elias 2007). Today, populations from both species occur proximal to this region: Asian black bears in eastern Russia, the Korean Peninsula and Japan, and American black bears in Alaska and Yukon, Canada (Servheen et al. 1990). The Bering land bridge may thus have provided opportunity for sympatry of American and Asian black bears in former times. Asian and American black bears share two haplotypes at two intron loci, and are polymorphic for the same variants at four sites (fig. 1, supplementary fig. S1 and tables S3 and S4, Supplementary Material online), but we find no significant multilocus signal of gene flow between the two species under the isolation-with-migration model. mtDNA was shown to introgress more easily than paternally or biparentally inherited genetic material (Chan and Levin 2005). Numerous cases of mitochondrial introgression across species boundaries have been documented, often with lower levels or without introgression of nuclear DNA, for example in polar and brown bears (Hailer et al. 2012), elephants (Roca et al. 2005), chipmunks (Good et al. 2008), colobine monkeys (Roos et al. 2011), hares (Melo-Ferreira et al. 2012), and in black rats (Pagès et al. 2013). Thus, mitochondrial capture can explain our observations.

Several other observations argue for nuclear swamping (scenario B). Such a forceful process could result from male-biased gene flow from brown into American black bears, with physically larger male brown bears mating with female black bears, without mtDNA passing the species boundary. Such gene flow must have stopped at some time in the past to explain the level of differentiation observed between brown bear and American black bear Y chromosomes. Indeed, we find significant, but weak signals of gene flow from the brown bear lineage into the American black bear lineage (table 2 and supplementary fig. S3*A*, Supplementary Material online), consistent with three haplotypes and three polymorphic sites shared between brown and American black bears (fig. 1, supplementary fig. S1 and tables S3 and S4, Supplementary Material online). Similarly, Miller et al. (2012) observed gene flow between brown and American black bears since their speciation, lasting until the late Pleistocene. Scenario B postulates that the mitochondrial gene tree reflects the speciation history of American and Asian black bears. Indeed, there is paleontological evidence for a sister-species relationship between American and Asian black bears (Kurtén and Anderson 1980). Remains of the ancestral nuclear genome, from times prior to introgression of brown bear genes into the American black bear lineage should still be detectable in American black bears. These ancestral remains may be represented by two haplotypes and four polymorphisms shared between American and Asian black bears. There is evidence for nuclear swamping affecting the genomes of brown and polar bears (fig. 2*D*): At the mitochondrial genome, polar bears were found to be closely related to brown bears from the Alaskan ABC (Admiralty, Baranof, and Chichagof) islands and from Ireland (now extinct) (Cronin et al. 1991; Edwards et al. 2011). At the nuclear genome, unidirectional gene flow has been detected from polar bears into North American brown bears, including ABC island brown bears (Cahill et al. 2013; Liu et al. 2014). Based on these findings, ABC island brown bears have been suggested to carry a mitochondrial haplotype that derives from an initial polar bear ancestry, whereas extensive male-biased gene flow from mainland brown bears has replaced much of the original polar bear-like genome with genetic material from immigrant brown bears (Cahill et al. 2013; Bidon et al. 2014). Considering these observations from different bear species, nuclear swamping is a reasonable explanation for the different placement of the American black bear lineage in nuclear and mitochondrial phylogenies.

Both hypotheses regarding American black bears appear rather drastic. Another source of conflict between nuclear and mitochondrial phylogenies can be the faster lineage sorting of the mitochondrial genome compared with autosomal DNA, due to the smaller effective population size of mtDNA (Funk and Omland 2003; McKay and Zink 2010). However, ILS was accounted for in our statistical comparisons of mitochondrial and nuclear topologies in a coalescence framework, rendering differences in lineage sorting an unlikely cause for the observed discrepancies between mitochondrial and nuclear phylogenies. Nonetheless, a scenario including several hybridization events during the evolutionary history of

ursine bears is conceivable, involving ancient hybridization of American and Asian black bears, gene flow from Asian black bears into sloth bears, and/or male-biased gene flow from brown bears into American black bears. Extended population-level and/or genome-wide studies and analytical approaches that incorporate both ILS and introgression into species tree estimation will be required to fully understand the evolutionary processes leading to the observed discrepancies between nuclear and mitochondrial phylogenies in these species.

Capturing the Complexity of Evolutionary Processes

Charles Darwin pointed out that many closely related species are not completely reproductively isolated (Darwin 1859), and in recent decades, molecular studies have identified introgressive hybridization as a pervasive evolutionary process (Schwenk et al. 2008). At least 10% of animal species hybridize with closely related species in well-studied taxa (Gray 1972; Mallet 2005). In addition, based on predictions from coalescence theory, lineage sorting of autosomal genes should be completed within about four to seven N_e generations (Nichols 2001). Thus, ILS spans time scales of up to several million years, often covering longer time frames than required for speciation in mammals. ILS has been shown to affect a large proportion of the genomes of humans and their closest relatives (Hobolth et al. 2011; Prüfer et al. 2012; Scally et al. 2012), but only few studies have specifically examined both ILS and gene flow in vertebrates that diverged several million years ago. Notably, many species have a larger population size than bears and great apes, so their genomes will be even more affected by ILS.

Initially, when technological advances made it feasible to sequence multiple loci, phylogenetic methods developed for single loci were used to analyze a concatenated superlocus. This approach ignored the heterogeneity of the phylogenetic signal among loci, and disregarded the vast amount of phylogenetic information that resides within individuals and species by including only one individual per species. Indeed, simulation studies have shown that the concatenation procedure can provide high statistical support for an incorrect species tree, because lineage sorting processes are not modeled (Kubatko and Degnan 2007). Finally, branch length estimates are affected when heterozygous sites are excluded from phylogenetic analyses (Lischer et al. 2014), which was common practice in phylogenetic analyses of concatenated autosomal data. Conceptual advances and recently developed coalescence-based multilocus species tree approaches now provide a means to infer overall phylogenetic relationships (species trees), against which individual gene trees can be contrasted to identify the underlying evolutionary processes. Although species tree approaches such as *BEAST (Heled and Drummond 2010) do not model gene flow, coalescence-based gene flow analyses can be used to complement phylogenetic inferences of evolutionary history: For example, in orioles (Jacobsen and Omland 2012), hares (Melo-Ferreira et al. 2012), and gibbons (Chan et al. 2013). By comparing marker systems with different inheritance modes and ploidy, sex-biased mechanisms and introgression events can be identified. To depict the complexity of evolutionary processes, networks of individual loci and multilocus networks (Holland et al. 2004; Bapteste et al. 2013) are better suited than bifurcating trees, because the latter may obscure evolutionary signals (Morrison 2005; Hallström and Janke 2010; Bapteste et al. 2013). In summary, advanced phylogenetic studies that aim to capture the full complexity of the evolutionary process need to consider "phylogenetic incongruence [as] a signal, rather than a problem" (Nakhleh 2013).

Materials and Methods

Samples and DNA Extraction

Samples were obtained from one giant panda, two spectacled bears, three sloth bears, three sun bears, three Asian black bears, one American black bear, two brown bears, and three polar bears (supplementary table S1, Supplementary Material online). All samples originated from zoo individuals or from animals legally hunted for purposes other than this study. Total DNA was extracted from muscle, skin, and blood samples using a standard salt extraction protocol (Crouse and Amorese 1987), or a standard phenol–chloroform extraction protocol (Sambrook and Russell 2000).

Amplification and Sequencing

We used primer pairs for 14 independently inherited autosomal markers (Hailer et al. 2012) to amplify intron sequences with flanking exon sequences in 15 individuals. We amplified nine Y-chromosomal markers in 11 male individuals (supplementary table S8, Supplementary Material online), using primers that were either described in Bidon et al. (2014), or newly designed (*322*, *389*, *403*) based on the polar bear genome (Liu et al. 2014), or based on male giant panda reads (Zhao et al. 2013) mapped against the polar bear genome. Polymerase chain reactions (PCRs) were performed using 5–15 ng of genomic DNA, and each PCR setup contained no-template controls. For amplification of Y-chromosomal markers, female DNA controls were included to ensure male-specificity throughout all experiments. PCR conditions and primers are listed in supplementary table S8, Supplementary Material online. PCR products were detected using standard agarose gel electrophoresis, and cycle sequenced with BigDye 3.1 chemistry (Applied Biosystems, Foster City, CA) in both directions according to the manufacturer's recommendation, and detected on an ABI 3100 instrument (Applied Biosystems). Electropherograms were checked manually. For autosomal introns, sequence data were included from Hailer et al. (2012) and from the giant panda genome assembly (Li et al. 2010), the final data set comprised 30 individuals. The Y-chromosomal data set included sequence data from Bidon et al. (2014). Therefore, American black bear and polar bear individuals differed between this and the autosomal intron data set. Accession numbers are listed in supplementary table S1, Supplementary Material online. Sequences were aligned using ClustalW implemented in Geneious 5.6.6 and 6.1.6 (Biomatters, Auckland, New Zealand; Drummond et al.

2012). We compared Y-chromosomal sequences from our single male giant panda individual with the mapped panda reads. Although this genome's Y-chromosomal sequence could not be included in our analyses because of some missing data, we found that all panda-specific divergent sites that were covered by both individuals were identical.

Data Analyses

We resolved heterozygous indels at autosomal markers using Champuru (Flot 2007) and Indelligent (Dmitriev and Rakitov 2008). Haplotypes were deduced using PHASE implemented in the software DnaSP v5.0 (Librado and Rozas 2009), based on alignments containing all available unphased sequences from the present and from a previous study (Hailer et al. 2012), allowing for recombination within haplotypes and using a cutoff value of 0.6 (Harrigan et al. 2008; Garrick et al. 2010). Twelve heterozygous sites could not be resolved and respective alignment columns were discarded from analyses. Sites containing floating indels, gaps, or missing data (*N*) were deleted from the alignments. In the Y-chromosomal alignment, three pseudoheterozygous sites were removed. Sequence diversity and differentiation statistics were calculated in Arlequin 3.5 (Excoffier and Lischer 2010), MEGA 5.2.2 (Tamura et al. 2011), and DnaSP v5.0 (Librado and Rozas 2009). To investigate the heterogeneity among different loci, statistical parsimony networks were reconstructed using TCS 1.21 (Clement et al. 2000). For this analysis, indels were treated as single mutational events, and gaps as a fifth character state. Longer gaps were treated as single mutational changes. The connection probability limit was set to 0.95 (autosomal loci) or 0.94 (Y-chromosomal sequence).

We reconstructed multilocus species trees from different data sets (supplementary table S5, Supplementary Material online), using *BEAST 1.7.5 (Drummond et al. 2012). Recombination is not modeled in *BEAST, but sampling effort (number of loci, number of individuals) has a much larger effect on species tree accuracy than the error introduced by recombination (Lanier and Knowles 2012). Hence, by reducing an alignment to its largest nonrecombining section, abundant phylogenetic information is discarded. We therefore used the total sequence length of the 14 autosomal introns (8 kb) in all *BEAST analyses. *BEAST was run applying a Yule prior on the species tree and a normal prior of 0.001 ± 0.001 (mean $\pm$ SD) on the substitution rates. We used a strict clock, because a relaxed, uncorrelated lognormal clock approach (Drummond et al. 2006) showed no significant departure from the strict clock model for our data. Models of sequence evolution were used as indicated by jModeltest (Posada 2008) and *BEAST was run for 2×10^9 generations, sampling every 10,000th iteration. Convergence was checked in Tracer with effective sampling sizes (ESS) >200. Two runs with identical settings were combined in LogCombiner v1.7.5 using a burnin of 10%, and a maximum clade credibility tree was constructed using TreeAnnotator.

For divergence time estimates, we assumed a minimum age of 11.6 My for the divergence of the giant panda from other bears, based on the oldest described fossil from the subfamily Ailuropodinae (Abella et al. 2012). Generation time for American black bears has been estimated at 6.27 years (Onorato et al. 2004) and 10 years for brown and polar bears (Tallmon et al. 2004; Cronin et al. 2009). For spectacled, sloth, sun, Asian black bears, and giant pandas, no adequate data were available, but as generation time is correlated with body size in mammals (Bonner 1965), we used the estimate of 6.27 years for American black bears also for these species. Based on the arithmetic mean of these generation time estimates, we assumed an overall generation time of 7.2 years to transform per-year estimates of ursid mutation rates from *BEAST into per-generation values. For statistical comparisons of the mitochondrial and the species tree topologies, we performed *BEAST analyses of autosomal introns and Y-chromosomal data combined. The species tree topology was either constrained to the mitochondrial topology (monophyly of American black bear and Asian black bear, and monophyly of American black bear, Asian black bear, and sun bear), or to the species tree topology (monophyly of polar bear, brown bear, and American black bear). BF were estimated in Tracer based on likelihood traces of the two constrained analyses (Suchard et al. 2005), using 1,000 bootstrap replicates.

To illustrate the extent of phylogenetic conflict in the nuclear signal, DensiTree (Bouckaert 2010) was used to generate a cloudogram of the posterior distribution of species trees from *BEAST, and a consensus network (Holland et al. 2004) was generated using SplitsTree4 (Huson and Bryant 2006). For the latter, *BEAST maximum clade credibility gene trees from the 14 autosomal introns were used as input gene trees, displaying splits that occurred in at least 2 of the 14 gene trees (edge threshold: 0.14).

For phylogenetic analyses of concatenated mitochondrial and Y-chromosomal data, we reconstructed different data sets (supplementary table S5, Supplementary Material online) from sequence data generated in the present and in previous studies (Jameson et al. 2003; Nakagome et al. 2008; Pagès et al. 2008). Pagès et al. (2008) published a consensus sequence of several individuals per species, with intraspecific polymorphisms coded by ambiguity codes. Alignment columns with these sites were disregarded in all analyses. Protein-coding regions from the mitochondrial genomes of all eight bear species (excluding ND6) were obtained from OGRe (Jameson et al. 2003) (for accession numbers, see supplementary table S1, Supplementary Material online), and aligned and concatenated in Geneious 5.6.6. For each data set, the optimal model of sequence evolution was determined using jModeltest (Posada 2008). Concatenated Y-chromosomal data (present study) and mitochondrial sequences were analyzed in BEAST 1.7.5 (Drummond et al. 2012) using a Yule prior on the species tree and a normal prior of 0.001 ± 0.001 on the substitution rates. BEAST was run for 1×10^9 generations, sampling every 10,000th iteration. Convergence was checked in Tracer (ESS > 200) and maximum-clade credibility trees were reconstructed in TreeAnnotator using a burnin of 10%. The AU test (Shimodaira 2002) was performed in Treefinder (Jobb et al. 2004) with 50,000 bootstrap replicates each, using

mitochondrial and two Y-chromosomal data sets. Likelihoods and tree statistics were calculated in Treefinder in an exhaustive search among all 105 topologies that are possible for five operational taxonomical units. The giant panda served as outgroup, with the spectacled bear as sister-taxon to all ursines, and the polar and brown bear were constricted to be sister lineages.

We used IMa2 (Hey 2010) on the 14 autosomal introns to assess the level of gene flow among species. This software is based on an isolation-with-migration model and estimates effective population sizes (present and ancestral), splitting times, and population migration rates using Markov chain Monte Carlo (MCMC) simulations. As the isolation-with-migration model assumes no recombination within and free recombination between markers (Hey and Nielsen 2004), the nonrecombining sections of the 14 autosomal introns (in total 5.1 kb) were used as reconstructed in IMgc (Woerner et al. 2007). Substitution rates per marker per year were estimated from the average divergence ($D_{XY} = 2T\mu$) between the giant panda and polar bear, assuming a divergence time (T) of 12 Ma (Abella et al. 2012), and the Hasegawa–Kishino–Yano model of sequence evolution. We assumed a generation time of 8 years for the pairwise comparisons brown bear—American black bear and polar bear—sun bear, and a generation time of 6 years for the other pairwise comparisons. Generation times were based on estimates of 6.27 years for American black bears (Onorato et al. 2004) and of 10 years for brown and polar bears (Tallmon et al. 2004; Cronin et al. 2009). Preliminary runs were performed to evaluate various prior settings, heated chain conditions, and the necessary MCMC lengths. To set an upper bound for the splitting time, we assumed that time since divergence could not be older than the minimum age of the youngest *Ursavus* fossil (ca. 7.1 My; Fortelius 2003), the genus from which the *Ursus* lineage is thought to have descended (Kurtén 1968). For effective population sizes, we defined an upper bound for the prior by multiplying the arithmetic mean of θ_π (Tajima 1983) of each species pair by approximately nine, allowing for larger population sizes in the past (Miller et al. 2012). Four independent runs, each with different starting seeds, were performed with optimized priors and heating schemes, using 40 Markov chains. After a burnin period with stationary already reached, 25,000 genealogies were saved. Convergence was assessed based on ESS >50, stable parameter trend plots, and similar parameter estimates from the first and the second half of the runs. Marginal posterior probability density estimates and LLR tests to assess whether migration rates were significantly different from zero were calculated in "L mode" of IMa2, using 100,000 sampled genealogies from each of the four independent runs.

Supplementary Material

Supplementary material is available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

This work was supported by LOEWE Landes-Offensive zur Entwicklung Wissenschaftlich-ökonomischer Exzellenz, the Arthur und Aenne Feindt-Stiftung, Hamburg, and the RISE Research Internships in Science and Engineering (RISE) program of the German Academic Exchange Service (DAAD). The findings and conclusions in this article are those of the authors and do not necessarily represent the views of the US Fish and Wildlife Service. The authors thank U. Arnason, S.B. Hagen, H.-G. Eiken N. Lecomte, M. Onucsán, and B. Steck for providing samples, J.B. Hlíðberg (www.fauna.is) for the artwork, and the editor and reviewers for helpful comments on a previous version of the manuscript.

References

Abella J, Alba DM, Robles JM, Valenciano A, Rotgers C, Carmona R, Montoya P, Morales J. 2012. *Kretzoiarctos* gen. nov., the oldest member of the giant panda clade. *PLoS One* 7:e48985.

Bapteste E, van Iersel L, Janke A, Kelchner S, Kelk S, McInerney JO, Morrison DA, Nakhleh L, Steel M, Stougie L, et al. 2013. Networks: expanding evolutionary thinking. *Trends Genet.* 29:439–441.

Bidon T, Janke A, Fain SR, Eiken HG, Hagen SB, Saarma U, Hallström BM, Lecomte N, Hailer F. 2014. Brown and polar bear Y chromosomes reveal extensive male-biased gene flow within brother lineages. *Mol Biol Evol.* 31:1353–1363.

Bonner JT. 1965. Size and cycles: an essay on the structure of biology. Princeton (NY): Princeton University Press.

Bouckaert RR. 2010. DensiTree: making sense of sets of phylogenetic trees. *Bioinformatics* 26:1372–1373.

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

Chan KMA, Levin SA. 2005. Leaky prezygotic isolation and porous genomes: rapid introgression of maternally inherited DNA. *Evolution* 59:720–729.

Chan Y-C, Roos C, Inoue-Murayama M, Inoue E, Shih C-C, Pei KJ-C, Vigilant L. 2013. Inferring the evolutionary histories of divergences in *Hylobates* and *Nomascus* gibbons through multilocus sequence data. *BMC Evol Biol.* 13:82.

Clement M, Posada D, Crandall KA. 2000. TCS: a computer program to estimate gene genealogies. *Mol Ecol.* 9:1657–1659.

Cronin MA, Amstrup SC, Garner GW, Vyse ER. 1991. Interspecific and intraspecific mitochondrial DNA variation in North American bears (*Ursus*). *Can J Zool.* 69:2985–2992.

Cronin MA, Amstrup SC, Talbot SL, Sage GK, Amstrup KS. 2009. Genetic variation, relatedness, and effective population size of polar bears (*Ursus maritimus*) in the southern Beaufort Sea, Alaska. *J Hered.* 100: 681–690.

Cronin MA, McDonough MM, Huynh HM, Baker RJ. 2013. Genetic relationships of North American bears (*Ursus*) inferred from amplified fragment length polymorphisms and mitochondrial DNA sequences. *Can J Zool.* 91:626–634.

Crouse J, Amorese D. 1987. Ethanol precipitation: ammonium acetate as an alternative to sodium acetate. *Focus* 19:13–16.

Darwin CR. 1859. On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life. London: John Murray.

Dmitriev DA, Rakitov RA. 2008. Decoding of superimposed traces produced by direct sequencing of heterozygous indels. *PLoS Comput Biol.* 4:e1000113.

Drummond AJ, Ashton B, Buxton S, Cheung M, Cooper A, Duran C, Field M. 2012. Geneious v5.6. [Internet]. [2012 Mar]. Available from: http://www.geneious.com.

Drummond AJ, Ho SYW, Phillips MJ, Rambaut A. 2006. Relaxed phylogenetics and dating with confidence. *PLoS Biol.* 4:e88.

Drummond AJ, Suchard MA, Xie D, Rambaut A. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol.* 29: 1969–1973.

Edwards CJ, Suchard MA, Lemey P, Welch JJ, Barnes I, Fulton TL, Barnett R, O'Connell TC, Coxon P, Monaghan N, et al. 2011. Ancient hybridization and an Irish origin for the modern polar bear matriline. *Curr Biol.* 21:1251–1258.

Excoffier L, Lischer HEL. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour.* 10:564–567.

Flot J-F. 2007. champuru 1.0: a computer software for unraveling mixtures of two DNA sequences of unequal lengths. *Mol Ecol Notes.* 7: 974–977.

Fortelius M (coordinator). 2003. New and old worlds database of fossil mammals (NOW). University Helsinki. [cited 2014 Jan]. Available from: http://www.helsinki.fi/science/now.

Funk DJ, Omland KE. 2003. Species-level paraphyly and polyphyly: frequency, causes, and consequences, with insights from animal mitochondrial DNA. *Annu Rev Ecol Evol Syst.* 34:397–423.

Garrick RC, Sunnucks P, Dyer RJ. 2010. Nuclear gene phylogeography using PHASE: dealing with unresolved genotypes, lost alleles, and systematic bias in parameter estimation. *BMC Evol Biol.* 10:118.

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

Gray A. 1972. Mammalian hybrids. A check-list with bibliography. Slough (United Kingdom): Commonwealth Agricultural Bureaux.

Hailer F, Kutschera VE, Hallström BM, Fain SR, Leonard JA, Arnason U, Janke A. 2013. Response to comment on "Nuclear genomic sequences reveal that polar bears are an old and distinct bear lineage". *Science* 339:1522.

Hailer F, Kutschera VE, Hallström BM, Klassert D, Fain SR, Leonard JA, Arnason U, Janke A. 2012. Nuclear genomic sequences reveal that polar bears are an old and distinct bear lineage. *Science* 336: 344–347.

Hallast P, Balaresque P, Bowden GR, Ballereau S, Jobling MA. 2013. Recombination dynamics of a human Y-chromosomal palindrome: rapid GC-biased gene conversion, multi-kilobase conversion tracts, and rare inversions. *PLoS Genet.* 9:e1003666.

Hallström BM, Janke A. 2010. Mammalian evolution may not be strictly bifurcating. *Mol Biol Evol.* 27:2804–2816.

Harrigan RJ, Mazza ME, Sorenson MD. 2008. Computation vs. cloning: evaluation of two methods for haplotype determination. *Mol Ecol Resour.* 8:1239–1248.

Heled J, Drummond AJ. 2010. Bayesian inference of species trees from multilocus data. *Mol Biol Evol.* 27:570–580.

Hellborg L, Ellegren H. 2004. Low levels of nucleotide diversity in mammalian Y chromosomes. *Mol Biol Evol.* 21:158–163.

Hey J. 2010. Isolation with migration models for more than two populations. *Mol Biol Evol.* 27:905–920.

Hey J, Nielsen R. 2004. Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. *Genetics* 167:747–760.

Hobolth A, Dutheil JY, Hawks J, Schierup MH, Mailund T. 2011. Incomplete lineage sorting patterns among human, chimpanzee, and orangutan suggest recent orangutan speciation and widespread selection. *Genome Res.* 21:349–356.

Hoffecker JF, Elias SA. 2007. The human ecology of Beringia. New York: Columbia University Press.

Holland BR, Huber KT, Moulton V, Lockhart PJ. 2004. Using consensus networks to visualize contradictory evidence for species phylogeny. *Mol Biol Evol.* 21:1459–1461.

Huson DH, Bryant D. 2006. Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol.* 23:254–267.

Jacobsen F, Omland KE. 2012. Extensive introgressive hybridization within the northern oriole group (Genus *Icterus*) revealed by three-species isolation with migration analysis. *Ecol Evol.* 2: 2413–2429.

Jameson D, Gibson AP, Hudelot C, Higgs PG. 2003. OGRe: a relational database for comparative analysis of mitochondrial genomes. *Nucleic Acids Res.* 31:202–206.

Jobb G, von Haeseler A, Strimmer K. 2004. TREEFINDER: a powerful graphical analysis environment for molecular phylogenetics. *BMC Evol Biol.* 4:18.

Kass RE, Raftery AE. 1995. Bayes factors. *J Am Stat Assoc.* 90:773–795.

Kelly BP, Whiteley A, Tallmon D. 2010. The Arctic melting pot. *Nature* 468:891.

Krause J, Unger T, Nocon A, Malaspinas A-S, Kolokotronis S-O, Stiller M, Soibelzon L, Spriggs H, Dear PH, Briggs AW, et al. 2008. Mitochondrial genomes reveal an explosive radiation of extinct and extant bears near the Miocene-Pliocene boundary. *BMC Evol Biol.* 8:220.

Kubatko LS, Degnan JH. 2007. Inconsistency of phylogenetic estimates from concatenated data under coalescence. *Syst Biol.* 56: 17–24.

Kurtén B. 1968. Pleistocene mammals of Europe. Chicago (IL): Aldine.

Kurtén B, Anderson E. 1980. Pleistocene mammals of North America. New York: Columbia University Press.

Lanier HC, Knowles LL. 2012. Is recombination a problem for species-tree analyses? *Syst Biol.* 61:691–701.

Leaché AD, Harris RB, Rannala B, Yang Z. 2014. The influence of gene flow on species tree estimation: a simulation study. *Syst Biol.* 63: 17–30.

Li R, Fan W, Tian G, Zhu H, He L, Cai J, Huang Q, Cai Q, Li B, Bai Y, et al. 2010. The sequence and de novo assembly of the giant panda genome. *Nature* 463:311–317.

Librado P, Rozas J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25:1451–1452.

Lischer HEL, Excoffier L, Heckel G. 2014. Ignoring heterozygous sites biases phylogenomic estimates of divergence times: implications for the evolutionary history of *Microtus* voles. *Mol Biol Evol.* 31: 817–831.

Liu S, Lorenzen ED, Fumagalli M, Li B, Harris K, Xiong Z, Zhou L, Korneliussen TS, Somel M, Babbitt C, et al. 2014. Population genomics reveal recent speciation and rapid evolutionary adaptation in polar bears. *Cell* 157:785–794.

Maddison WP, Knowles LL. 2006. Inferring phylogeny despite incomplete lineage sorting. *Syst Biol.* 55:21–30.

Mallet J. 2005. Hybridization as an invasion of the genome. *Trends Ecol Evol.* 20:229–237.

McKay BD, Zink RM. 2010. The causes of mitochondrial DNA gene tree paraphyly in birds. *Mol Phylogenet Evol.* 54:647–650.

Melo-Ferreira J, Boursot P, Carneiro M, Esteves PJ, Farelo L, Alves PC. 2012. Recurrent introgression of mitochondrial DNA among hares (*Lepus* spp.) revealed by species-tree inference and coalescent simulations. *Syst Biol.* 61:367–381.

Miller W, Schuster SC, Welch AJ, Ratan A, Bedoya-Reina OC, Zhao F, Lim Kim H, Burhans RC, Drautz DI, Wittekindt NE, et al. 2012. Polar and brown bear genomes reveal ancient admixture and demographic footprints of past climate change. *Proc Natl Acad Sci U S A.* 109: E2382–E2390.

Morrison DA. 2005. Networks in phylogenetic analysis: new tools for population biology. *Int J Parasitol.* 35:567–582.

Nachman MW, Crowell SL. 2000. Estimate of the mutation rate per nucleotide in humans. *Genetics* 156:297–304.

Nakagome S, Mano S, Hasegawa M. 2013. Comment on "Nuclear genomic sequences reveal that polar bears are an old and distinct bear lineage". *Science* 339:1522.

Nakagome S, Pecon-Slattery J, Masuda R. 2008. Unequal rates of Y chromosome gene divergence during speciation of the family Ursidae. *Mol Biol Evol.* 25:1344–1356.

Nakhleh L. 2013. Computational approaches to species phylogeny inference and gene tree reconciliation. *Trends Ecol Evol.* 28: 719–728.

Nichols R. 2001. Gene trees and species trees are not the same. *Trends Ecol Evol.* 16:358–364.

Nielsen R, Wakeley J. 2001. Distinguishing migration from isolation: a Markov chain Monte Carlo approach. *Genetics* 158:885–896.

Onorato DP, Hellgren EC, Bussche RA, van D, Doan-Crider DL. 2004. Phylogeographic patterns within a metapopulation of black bears

(*Ursus americanus*) in the American southwest. *J Mammal.* 85: 140–147.
Pagès M, Bazin E, Galan M, Chaval Y, Claude J, Herbreteau V, Michaux J, Piry S, Morand S, Cosson J-F. 2013. Cytonuclear discordance among Southeast Asian black rats (*Rattus rattus* complex). *Mol Ecol.* 22: 1019–1034.
Pagès M, Calvignac S, Klein C, Paris M, Hughes S, Hänni C. 2008. Combined analysis of fourteen nuclear genes refines the Ursidae phylogeny. *Mol Phylogenet Evol.* 47:73–83.
Pamilo P, Nei M. 1988. Relationships between gene trees and species trees. *Mol Biol Evol.* 5:568–583.
Petit E, Balloux F, Excoffier L. 2002. Mammalian population genetics: why not Y? *Trends Ecol Evol.* 17:28–33.
Posada D. 2008. jModelTest: phylogenetic model averaging. *Mol Biol Evol.* 25:1253–1256.
Prüfer K, Munch K, Hellmann I, Akagi K, Miller JR, Walenz B, Koren S, Sutton G, Kodira C, Winer R, et al. 2012. The bonobo genome compared with the chimpanzee and human genomes. *Nature* 486:527–531.
Roca AL, Georgiadis N, O'Brien SJ. 2005. Cytonuclear genomic dissociation in African elephant species. *Nat Genet.* 37:96–100.
Roos C, Zinner D, Kubatko LS, Schwarz C, Yang M, Meyer D, Nash SD, Xing J, Batzer MA, Brameier M, et al. 2011. Nuclear versus mitochondrial DNA: evidence for hybridization in colobine monkeys. *BMC Evol Biol.* 11:77.
Sachidanandam R, Weissman D, Schmidt SC, Kakol JM, Stein LD, Marth G, Sherry S, Mullikin JC, Mortimore BJ, Willey DL, et al. 2001. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* 409:928–933.
Sambrook J, Russell DW. 2000. Molecular cloning: a laboratory manual. New York: Cold Spring Harbor Laboratory Press.
Scally A, Dutheil JY, Hillier LW, Jordan GE, Goodhead I, Herrero J, Hobolth A, Lappalainen T, Mailund T, Marques-Bonet T, et al. 2012. Insights into hominid evolution from the gorilla genome sequence. *Nature* 483:169–175.
Schwenk K, Brede N, Streit B. 2008. Introduction. Extent, processes and evolutionary impact of interspecific hybridization in animals. *Philos Trans R Soc B Biol Sci.* 363:2805–2811.
Servheen C. 1990. The status and conservation of the bears of the world. *Int Conf Bear Res and Manage Monogr Series* 2:1–32.
Shimodaira H. 2002. An approximately unbiased test of phylogenetic tree selection. *Syst Biol.* 51:492–508.
Suchard MA, Weiss RE, Sinsheimer JS. 2005. Models for estimating Bayes factors with applications to phylogeny and tests of monophyly. *Biometrics* 61:665–673.
Tajima F. 1983. Evolutionary relationship of Dna sequences in finite populations. *Genetics* 105:437–460.
Tallmon DA, Bellemain E, Swenson JE, Taberlet P. 2004. Genetic monitoring of Scandinavian brown bear effective population size and immigration. *J Wildl Manag.* 68:960–965.
Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol.* 28:2731–2739.
Waddell PJ, Kishino H, Ota R. 2001. A phylogenetic foundation for comparative mammalian genomics. *Genome Inform.* 12:141–154.
Wayne RK, Van Valkenburgh B, O'Brien SJ. 1991. Molecular distance and divergence time in carnivores and primates. *Mol Biol Evol.* 8: 297–319.
Woerner AE, Cox MP, Hammer MF. 2007. Recombination-filtered genomic data sets by information maximization. *Bioinformatics* 23: 1851–1853.
Yu L, Li Q-W, Ryder OA, Zhang Y-P. 2004. Phylogeny of the bears (Ursidae) based on nuclear and mitochondrial genes. *Mol Phylogenet Evol.* 32:480–494.
Yu L, Li Y-W, Ryder OA, Zhang Y-P. 2007. Analysis of complete mitochondrial genome sequences increases phylogenetic resolution of bears (Ursidae), a mammalian family that experienced rapid speciation. *BMC Evol Biol.* 7:198.
Yu Y, Barnett RM, Nakhleh L. 2013. Parsimonious inference of hybridization in the presence of incomplete lineage sorting. *Syst Biol.* 62: 738–751.
Yu Y, Degnan JH, Nakhleh L. 2012. The probability of a gene tree topology within a phylogenetic network with applications to hybridization detection. *PLoS Genet.* 8:e1002660.
Zhao S, Zheng P, Dong S, Zhan X, Wu Q, Guo X, Hu Y, He W, Zhang S, Fan W, et al. 2013. Whole-genome sequencing of giant pandas provides insights into demographic history and local adaptation. *Nat Genet.* 45:67–71.

Supplementary Information

Bears in a forest of gene trees: Phylogenetic inference is complicated by incomplete lineage sorting and gene flow

Verena E. Kutschera, Tobias Bidon, Frank Hailer, Julia L. Rodi, Steven R. Fain, Axel Janke

Corresponding authors:

Verena E. Kutschera, Biodiversity and Climate Research Centre (BiK-F), Senckenberg Gesellschaft für Naturforschung, Senckenberganlage 25, 60325 Frankfurt am Main, Germany. Email: v.kutschera@gmx.net

Axel Janke, Biodiversity and Climate Research Centre (BiK-F), Senckenberg Gesellschaft für Naturforschung, Senckenberganlage 25, 60325 Frankfurt am Main, Germany and Goethe University Frankfurt, Institute for Ecology, Evolution and Diversity, 60438 Frankfurt am Main, Germany. Email: axel.janke@senckenberg.de

Phylogenetic analyses of concatenated data

We conducted phylogenetic analyses of concatenated autosomal data to highlight the extent of reduction in variation resulting from the concatenation procedure, and for comparison with the multi-locus species tree (fig. 2*A*) and phylogenetic trees from previous studies. Intraspecific and intra-individual polymorphisms were disregarded, because for concatenation, for each species all variation within and among individuals had to be collapsed into one single 50% majority-rule-consensus sequence. Unresolved sites with each variant occurring 50% were deleted from the alignments. Phylogenetic trees from concatenated nuclear data were calculated in MrBayes 3.2 (Ronquist et al. 2012) and in Treefinder version 2008 (Jobb et al. 2004). For Bayesian inferences in MrBayes we used one cold and three heated chains and ran the analyses for 10,000,000 Markov chain Monte Carlo generations sampling every 2,000th generation, with a burnin of 25%. We confirmed convergence in Tracer v1.5 (effective sampling size >200). Maximum likelihood analyses were performed in Treefinder with 10,000 bootstrap replicates. In Bayesian and maximum likelihood analyses of concatenated (1) Y-chromosomal, (2) autosomal, and (3) autosomal/Y-chromosomal markers combined, the American black bear was placed as sister-taxon to the brown and polar bear lineage with high statistical support, and the sun bear was sister-taxon to a clade including the sloth and Asian black bear (supplementary figure S2, Supplementary Material online). When analyzing the autosomal and Y-chromosomal data separately, support for the sun/sloth/Asian black bear clade was limited, but it was high in the combined analyses. Statistical support for Ursinae forming a monophyletic group and for the spectacled bear as sister-taxon to all ursines was high for all three datasets (Y-chromosomal, autosomal, autosomal/Y-chromosomal combined); the giant panda was the outgroup.

Supplementary tables

Supplementary table S1: Details of samples and sequences used in the study.

	Species name	Scientific name	Lab ID	Geographic origin	Sex	Accession numbers and/or source study		
						Autosomal markers	Y-chromosomal markers	Mitochondrial genomes
14 autosomal introns and 5.9 kb Y-chromosomal sequence (present study)								
1	Giant panda	*Ailuropoda melanoleuca*	AmeC85	unknown	male	HG974607-HG974634	HG975027-HG975031	--
2	Giant panda	*Ailuropoda melanoleuca*	Amegenom	unknown	female	Giant panda genome (Li et al. 2010)	--	--
3	Spectacled bear	*Tremarctos ornatus*	TorCha	Zoo Basel, Switzerland; ISO Fdx 250229600006729	male	HG974803-HG974830	HG975052-HG975056; HG423284-HG423285 (Bidon et al. 2014)	--
4	Spectacled bear	*Tremarctos ornatus*	TorNob	Zoo Basel, Switzerland; ISO Fdx 968000002054943	male	HG974831-HG974858	HG975057-HG975061	--
5	Sloth bear	*Melursus ursinus*	MURL42	Sunset Zoo Manhattan, KS, USA; Studbook# 460	male	HG974719-HG974746	HG975042-HG975046	--
6	Sloth bear	*Melursus ursinus*	MURL43	Philadelphia Zoo, PA, USA	male	HG974747-HG974774	HG975047-HG975051	--
7	Sloth bear	*Melursus ursinus*	MURL44	India; Studbook# 442	female	HG974775-HG974802	--	--
8	Sun bear	*Helarctos malayanus*	HMAL45	Miami Metro Zoo, FL, USA; Studbook# 635	male	HG974635-HG974662	HG975032-HG975036	--
9	Sun bear	*Helarctos malayanus*	HMAL46	San Diego Zoo, CA, USA; Studbook# 617	male	HG974663-HG974690	HG975037-HG975041	--
10	Sun bear	*Helarctos malayanus*	HMAL47	St. Louis Zoo, MO, USA; Studbook# 644	female	HG974691-HG974718	--	--
11	Asian black bear	*Ursus thibethanus*	UTHL48	John Ball Zoo, MI, USA; Studbook# 401	male	HG974943-HG974970	HG975077-HG975081	--
12	Asian black bear	*Ursus thibethanus*	UTHL49	Southwick's Zoo, MA, USA	female	HG974971-HG974998	--	--
13	Asian black bear	*Ursus thibethanus*	UTHL50	Denver Zoo, CO, USA; Studbook# 585	female	HG974999-HG975026	--	--
14	American black bear	*Ursus americanus*	Uam1203	Yosemite NP, Mariposa, CA, USA	male	see Hailer et al. (2012)	--	--
15	American black bear	*Ursus americanus*	Uam13724	Wesley, Washington, ME, USA	female	see Hailer et al. (2012)	--	--
16	American black bear	*Ursus americanus*	Uam16103	Tanana Flats, AK, USA	female	see Hailer et al. (2012)	--	--

	Species name	Scientific name	Lab ID	Geographic origin	Sex	Accession numbers and/or source study		
						Autosomal markers	Y-chromosomal markers	Mitochondrial genomes
17	American black bear	*Ursus americanus*	Uam24064	Sixes, Curry, OR, USA	female	see Hailer et al. (2012)	--	--
18	American black bear	*Ursus americanus*	Uam6586	Garfield, CO, USA	female	see Hailer et al. (2012)	--	--
19	American black bear	*Ursus americanus*	Uam6616	Humboldt, CA, USA	female	see Hailer et al. (2012)	--	--
20	American black bear	*Ursus americanus*	UamC122	unknown	male	see Hailer et al. (2012)	--	--
21	American black bear	*Ursus americanus*	UamMTM33	MT, USA	male	--	HG975062-HG975066; HG423286-HG423287 (Bidon et al. 2014)	--
22	Brown bear	*Ursus arctos*	Uar001	Rumania	female	see Hailer et al. (2012)	--	--
23	Brown bear	*Ursus arctos*	UarKamK05	Kamchatka, Russia	male	HG974859-HG974886	--	--
24	Brown bear	*Ursus arctos*	Uar1254	Shoshone NF Park, WY, USA	female	see Hailer et al. (2012)	--	--
25	Brown bear	*Ursus arctos*	UarA9106	Admiralty Island, AK, USA	male	see Hailer et al. (2012)	--	--
26	Brown bear	*Ursus arctos*	UarBT1-8	Norway	male	see Hailer et al. (2012)	HG975067-HG975071; HG423290-HG423291 (Bidon et al. 2014)	--
27	Polar bear	*Ursus maritimus*	UmaB26	Turner Island, eastern Greenland	female	see Hailer et al. (2012)	--	--
28	Polar bear	*Ursus maritimus*	UmaB38	Savissivik, western Greenland	male	see Hailer et al. (2012)	--	--
29	Polar bear	*Ursus maritimus*	UmaAKL29	Chukchi Sea population, AK, USA	male	HG974887-HG974914	--	--
30	Polar bear	*Ursus maritimus*	UmaDSL57	Davis Strait population, Canada	male	HG974915-HG974942	--	--
31	Polar bear	*Ursus maritimus*	Uma009	Point Lay, AK, USA	male	see Hailer et al. (2012)	--	--
32	Polar bear	*Ursus maritimus*	UmaDSL51	Davis Strait population, Canada	male	--	HG975072-HG975076; HG423302-HG423303 (Bidon et al. 2014)	--

	Species name	Scientific name	Lab ID	Geographic origin	Sex	Accession numbers and/or source study		
						Autosomal markers	Y-chromosomal markers	Mitochondrial genomes
Autosomal and Y-chromosomal markers (previous studies)								
33	Giant panda	*Ailuropoda melanoleuca*	--	unknown	unknown	see Pagès et al. (2008)	see Pagès et al. (2008)	--
34	Giant panda	*Ailuropoda melanoleuca*	--	unknown	unknown	--	see Nakagome et al. 2008	--
35	Spectacled bear	*Tremarctos ornatus*	--	unknown	unknown	see Pagès et al. (2008)	see Pagès et al. (2008)	--
36	Spectacled bear	*Tremarctos ornatus*	--	unknown	unknown	--	see Nakagome et al. 2008	--
37	Sloth bear	*Melursus ursinus*	--	unknown	unknown	see Pagès et al. (2008)	see Pagès et al. (2008)	--
38	Sloth bear	*Melursus ursinus*	--	unknown	unknown	--	see Nakagome et al. 2008	--
39	Sun bear	*Helarctos malayanus*	--	unknown	unknown	see Pagès et al. (2008)	see Pagès et al. (2008)	--
40	Sun bear	*Helarctos malayanus*	--	unknown	unknown	--	see Nakagome et al. 2008	--
41	Asian black bear	*Ursus thibethanus*	--	unknown	unknown	see Pagès et al. (2008)	see Pagès et al. (2008)	--
42	Asian black bear	*Ursus thibethanus*	--	unknown	unknown	--	see Nakagome et al. 2008	--
43	American black bear	*Ursus americanus*	--	unknown	unknown	see Pagès et al. (2008)	see Pagès et al. (2008)	--
44	American black bear	*Ursus americanus*	--	unknown	unknown	--	see Nakagome et al. 2008	--
45	Brown bear	*Ursus arctos*	--	unknown	unknown	see Pagès et al. (2008)	see Pagès et al. (2008)	--
46	Brown bear	*Ursus arctos*	--	unknown	unknown	--	see Nakagome et al. 2008	--
47	Polar bear	*Ursus maritimus*	--	unknown	unknown	see Pagès et al. (2008)	see Pagès et al. (2008)	--
48	Polar bear	*Ursus maritimus*	--	unknown	unknown	--	see Nakagome et al. 2008	--

	Species name	Scientific name	Lab ID	Geographic origin	Sex	Accession numbers and/or source study		
						Autosomal markers	Y-chromosomal markers	Mitochondrial genomes
Mitochondrial genomes (previous studies)								
49	Giant panda	*Ailuropoda melanoleuca*	--	unknown	unknown	--	--	NC_009492 (Peng et al. 2007)
50	Spectacled bear	*Tremarctos ornatus*	--	unknown	unknown	--	--	NC_009969 (Yu et al. 2007)
51	Sloth bear	*Melursus ursinus*	--	unknown	unknown	--	--	NC_009970 (Yu et al. 2007)
52	Sun bear	*Helarctos malayanus*	--	unknown	unknown	--	--	NC_009968 (Yu et al. 2007)
53	Asian black bear	*Ursus thibethanus*	--	unknown	unknown	--	--	NC_009971 (Yu et al. 2007)
54	American black bear	*Ursus americanus*	--	unknown	unknown	--	--	NC_003426 (Delisle and Strobeck 2002)
55	Brown bear	*Ursus arctos*	--	unknown	unknown	--	--	NC_003427 (Delisle and Strobeck 2002)
56	Polar bear	*Ursus maritimus*	--	unknown	unknown	--	--	NC_003428 (Delisle and Strobeck 2002)

Supplementary table S2: Mean p-distances between species (number of differences/total length).

Species pairs	14 autosomal introns (present study)	14 autosomal introns – consensus (present study)	11 autosomal exon and intron markers (Pagès et al. 2008)
Giant Panda – Spectacled bear	0.033	0.032	0.022
Giant Panda – Sun bear	0.033	0.032	0.021
Giant Panda – Sloth bear	0.032	0.031	0.021
Giant Panda – Asian black bear	0.033	0.032	0.020
Giant Panda – American black bear	0.031	0.030	0.020
Giant Panda – Brown bear	0.032	0.031	0.021
Giant Panda – Polar bear	0.033	0.032	0.021
Spectacled – Sun bear	0.017	0.016	0.011
Spectacled – Sloth bear	0.017	0.016	0.012
Spectacled – Asian black bear	0.017	0.016	0.012
Spectacled – American black bear	0.016	0.015	0.011
Spectacled – Brown bear	0.017	0.015	0.013
Spectacled – Polar bear	0.017	0.016	0.013
Sun – Sloth bear	0.008	0.007	0.004
Sun – Asian black bear	0.009	0.007	0.004
Sun – American black bear	0.009	0.008	0.005
Sun – Brown bear	0.009	0.007	0.005
Sun – Polar bear	0.009	0.008	0.005
Sloth – Asian black bear	0.007	0.005	0.004
Sloth – American black bear	0.008	0.007	0.005
Sloth – Brown bear	0.009	0.007	0.005
Sloth – Polar bear	0.009	0.008	0.005
Asian black – American black bear	0.007	0.006	0.003
Asian black – Brown bear	0.008	0.006	0.003
Asian black – Polar bear	0.008	0.007	0.003
American black – Brown bear	0.007	0.006	0.003
American black – Polar bear	0.007	0.007	0.003
Brown – Polar bear	0.005	0.003	0.003

Calculations are based on (1) 14 autosomal introns (present study; 30 phased individuals), (2) 14 autosomal introns (present study; eight 50% majority-rule consensus individuals), and (3) 11 autosomal exon and intron markers from Pagès et al. (2008) (eight consensus individuals).

Supplementary table S3: Pairwise divergence statistics for 5.9 kb from the Y chromosome and at 14 autosomal introns.

Species pairs	Y chromosome	Autosomal introns					
	Mean distance	Mean distance	Fixed differences	Shared polymorphisms	Polymorphic in species 1, fixed in 2	Polymorphic in species 2, fixed in 1	Sum of polymorphic sites
Giant Panda – Spectacled bear	212	259.8	253	0	12	7	19
Giant Panda – Sun bear	215.5	266.5	253	0	12	24	36
Giant Panda – Sloth bear	207	258.5	251	0	12	10	22
Giant Panda – Asian black bear	205	262.8	235	0	12	56	68
Giant Panda – American black bear	208	250.4	232	0	12	34	46
Giant Panda – Brown bear	208	258	232	0	12	64	76
Giant Panda – Polar bear	209	262.7	255	0	12	13	25
Spectacled – Sun bear	115.5	136.6	124	0	7	24	31
Spectacled – Sloth bear	107	133.3	127	0	7	10	17
Spectacled – Asian black bear	108	134.1	108	0	7	56	63
Spectacled – American black bear	111	126	110	0	7	33	40
Spectacled – Brown bear	111	133.2	107	0	7	63	70
Spectacled – Polar bear	112	136.8	130	0	7	13	20
Sun – Sloth bear	34.5	62.3	49	0	24	10	34
Sun – Asian black bear	34.5	68.2	34	0	24	56	80
Sun – American black bear	40.5	73.6	52	0	24	34	58
Sun – Brown bear	40.5	69.6	36	1	23	63	87
Sun – Polar bear	41.5	70.1	57	1	23	12	36
Sloth – Asian black bear	24	55.1	32	1	9	55	65
Sloth – American black bear	32	63.3	47	0	10	34	44
Sloth – Brown bear	32	74.3	48	0	10	64	74
Sloth – Polar bear	33	73.5	66	0	10	13	23
Asian black – American black bear	30	58.1	23	4	52	30	86
Asian black – Brown bear	30	61.6	21	10	46	54	110
Asian black – Polar bear	31	65.7	38	0	56	13	69
American black – Brown bear	16	56.6	20	3	31	61	95
American black – Polar bear	17	59.6	42	0	34	13	47
Brown – Polar bear	13	38.9	14	1	63	12	76

Supplementary table S4: Haplotype sharing among bear species (fig. 1, supplementary figure S1).

	Giant panda	Spectacled bear	Sun bear	Sloth bear	Asian black bear	American black bear	Brown bear	Polar bear
Giant panda	---							
Spectacled bear	-	---						
Sun bear	-	-	---					
Sloth bear	-	-	2 loci, 2 haplotypes	---				
Asian black bear	-	-	2 loci, 2 haplotypes	-	---			
American black bear	-	-	-	-	2 loci, 2 haplotypes	---		
Brown bear	-	-	-	-	2 loci, 2 haplotypes	3 loci, 4 haplotypes	---	
Polar bear	-	-	2 loci, 2 haplotypes	-	-	1 locus, 1 haplotype	5 loci, 5 haplotypes	---

Supplementary table S5: Datasets analyzed in the present study.

Dataset	Sequence alignments	Alignment length [bp]
A. Concatenated alignments reconstructed for traditional phylogenetic analyses		
14 autosomal introns	14 autosomal introns (present study)	7,991
5.9 kb Y-chromosomal sequence	9 Y-chromosomal markers (present study)	5,907
15 nuclear markers	14 autosomal introns (present study) + 9 Y-chromosomal markers (present study)	13,898
9.7 kb Y-chromosomal sequence (total evidence)	9 Y-chromosomal markers (present study) + 2 Y-chromosomal markers from Nakagome et al. (2008) + 3 Y-chromosomal markers from Pagès et al. (2008)	9,794
Mitochondrial genomes	Protein-coding regions from the mitochondrial genomes (excluding ND6) (Jameson et al. 2003)	10,807
B. Alignments included in multi-locus species tree analyses (*BEAST) and population genetic analyses under the isolation-with-migration model (IMa2)		
14 autosomal introns	14 autosomal introns (present study)	7,991
14 autosomal introns (non-recombining)	Largest non-recombining sections from 14 autosomal introns (present study) as reconstructed in IMgc (Woerner et al. 2007)	5,127
15 nuclear markers	14 autosomal introns (present study) + 5.9 kb Y-chromosomal sequence (present study)	13,898
30 nuclear markers (total evidence)	14 autosomal introns (present study) + 9.7 kb Y-chromosomal sequence (total evidence) + 4 X chromosomal markers from Nakagome et al. (2008) + 11 autosomal markers from Pagès et al. (2008)	28,681

A. Concatenated alignments reconstructed for traditional phylogenetic analyses and for topology tests. For concatenation, one 50%-majority-rule-consensus individual was reconstructed per species from sequence data generated in the present study. B. Alignments included in multi-locus species tree analyses (*BEAST) and population genetic analyses under the isolation-with-migration model (IMa2).

Supplementary table S6: Results from approximately unbiased (AU) topology tests.

A. Mitochondrial genomes (protein-coding regions, excl. ND6).

Topologies	p-value AU test	LogL	ΔLogL
Top 5 topologies, ranked according to p-value			
((((Uma,Uar),Mur),((Uam,Uth),Hma)),Tor,Ame);	0.75	-36097.13	0.00
(((Uma,Uar),(((Uam,Uth),Hma),Mur)),Tor,Ame);	0.65	-36097.02	-0.11
((((Uma,Uar),((Uam,Uth),Hma)),Mur),Tor,Ame); (Krause et al. 2008)	0.44	-36096.99	-0.14
((((Uma,Uar),((Uam,Hma),Uth)),Mur),Tor,Ame);	0.26	-36101.45	4.32
((((Uma,Uar),Mur),((Uam,Hma),Uth)),Tor,Ame);	0.21	-36101.57	4.44
Nuclear DNA topologies			
((((Uma,Uar),Uam),(Uth,(Hma,Mur))),Tor,Ame); (fig. 2*A*)	0.00	-36131.84	34.71
((((Uma,Uar),Uam),((Uth,Mur),Hma)),Tor,Ame); (fig. 2*B*, suppl. fig. S2)	0.00	-36131.86	34.73
(((((Uma,Uar),Uam),Uth),(Hma,Mur)),Tor,Ame); (Nakagome et al. 2008; Pagès et al. 2008)	0.00	-36131.79	34.66
((((((Uma,Uar),Uam),Uth),Mur),Hma),Tor,Ame); (Pagès et al. 2008)	0.00	-36131.53	34.40
((((((Uma,Uar),Uam),Uth),Hma),Mur),Tor,Ame); (Pagès et al. 2008)	0.00	-36128.41	31.28

B. 5.9 kb Y-chromosomal sequence (present study).

Topologies	p-value AU test	LogL	ΔLogL
Top 5 topologies, ranked according to p-value			
((((Uma,Uar),Uam),((Uth,Mur),Hma)),Tor,Ame); (fig. 2*B*, suppl. fig. S2)	0.89	-10141.62	0.00
((((Uma,Uar),Uam),(Uth,(Hma,Mur))),Tor,Ame); (fig. 2*A*)	0.71	-10142.80	1.18
((((((Uma,Uar),Uam),Uth),Hma),Mur),Tor,Ame); (Pagès et al. 2008)	0.44	-10144.61	2.99
(((((Uma,Uar),Uam),Uth),(Hma,Mur)),Tor,Ame); (Nakagome et al. 2008; Pagès et al. 2008)	0.34	-10144.77	3.15
(((((Uma,Uar),Uam),Hma),(Uth,Mur)),Tor,Ame);	0.30	-10143.67	2.05
Additional nuclear topologies			
((((((Uma,Uar),Uam),Uth),Mur),Hma),Tor,Ame); (Pagès et al. 2008)	0.00	-10144.6	2.98
mtDNA topologies			
((((Uma,Uar),Mur),((Uam,Uth),Hma)),Tor,Ame);	0.00	-10214.81	73.19
(((Uma,Uar),(((Uam,Uth),Hma),Mur)),Tor,Ame);	0.00	-10213.06	71.44
((((Uma,Uar),((Uam,Uth),Hma)),Mur),Tor,Ame); (Krause et al. 2008)	0.00	-10214.85	73.23

C. 9.7 kb of Y-chromosomal sequence (total evidence): 5.9 kb Y-chromosomal sequence (present study) concatenated with five Y-linked markers from Pagès et al. (2008) and Nakagome et al. (2008).

Topologies	p-value AU test	LogL	ΔLogL
Top 5 topologies, ranked according to p-value			
((((Uma,Uar),Uam),((Uth,Mur),Hma)),Tor,Ame); (fig. 2*B*, suppl. fig. S2)	0.92	-16672.99	0.00
(((((Uma,Uar),Uam),(Uth,Mur)),Hma),Tor,Ame);	0.75	-16674.60	1.61
((((Uma,Uar),Uam),(Uth,(Hma,Mur))),Tor,Ame); (fig. 2*A*)	0.62	-16674.33	1.34
((((((Uma,Uar),Uam),Uth),Mur),Hma),Tor,Ame); (Pagès et al. 2008)	0.32	-16675.57	2.58
((((((Uma,Uar),Uam),Uth),Hma),Mur),Tor,Ame); (Pagès et al. 2008)	0.19	-16675.57	2.58
Additional nuclear topologies			
(((((Uma,Uar),Uam),Uth),(Hma,Mur)),Tor,Ame); (Nakagome et al. 2008; Pagès et al. 2008)	0.14	-16676.03	3.04
mtDNA topologies			
((((Uma,Uar),Mur),((Uam,Uth),Hma)),Tor,Ame);	0.00	-16795.51	122.52
(((Uma,Uar),(((Uam,Uth),Hma),Mur)),Tor,Ame);	0.00	-16794.17	121.18
((((Uma,Uar),((Uam,Uth),Hma)),Mur),Tor,Ame); (Krause et al. 2008)	0.00	-16795.70	122.71

Ame: giant panda, Tor: spectacled bear, Hma: sun bear, Mur: sloth bear, Uth: Asian black bear, Uam: American black bear, Uar: brown bear, Uma: polar bear.

Supplementary table S7: Genetic diversity within bear species based on 14 autosomal introns and 5.9 kb from the Y chromosome.

Species	*n* (total)	Autosomes			Y chromosome		
		n	**S**	**π (x 10^{-3})**	***n***	**S**	**π (x 10^{-3})**
Giant panda	2	2	12	1.4 ± 1.0	1	-	-
Spectacled bear	2	2	7	0.6 ± 0.5	2	0	0.0 ± 0.0
Sun bear	3	3	24	2.0 ± 1.2	2	1	0.2 ± 0.2
Sloth bear	3	3	10	1.1 ± 0.7	2	0	0.0 ± 0.0
Asian black bear	3	3	55	5.1 ± 3.0	1	-	-
American black bear	8	7	34	2.1 ± 1.1	1	-	-
Brown bear	5	5	64	4.4 ± 2.4	1	-	-
Polar bear	6	5	13	0.8 ± 0.5	1	-	-

n = number of analyzed individuals; S = number of segregating sites; π = Tamura-Nei-corrected nucleotide diversity (π ± S.D.). Note that for American black bears and polar bears, different individuals were sequenced for autosomal and Y-chromosomal markers (see supplementary table S1).

Supplementary table S8: Primers (in 5' to 3' orientation) and amplification conditions of 14 autosomal introns and nine Y-linked markers.

Marker	Forward primer	Reverse primer	gene (intron#)	size [bp]	T [°C]
1247	TATTGGTGGAGGCTTCACAG	AGACATCCAACAAAGGGCTG	*AZIN1*(6)	805	65-58[d]
2331	CCAGGATATTTTGAYGCAATC	CTCAGCTTTYGGTAGGCAAC	*LRGUK*(14)	674	63-56[d]
3471	AKACTGAGTCCCAGCAGCAG	CCRTTCTGGGAAACTTGCTC	*SPTBN1*(31)	712	67-60[d]
4464	TCCTTTCCCAGAGCAARAAG	TGGTCCTTGGCGAAGTTTAC	*ABCA1*(49)	738	58
4779	TTTGCAAATCTRAGAGCAGAG	CAGTTGCTGCTTAACTTGTTCC	*CCDC90B*(4)	708	63-56[d]
7545	GGGAAAGTTCCGGTTTTTG	TTTCTCAGACACCCTGTCCC	*GGA3*(3)	726	58
9072	TTTCATCGGTGTCATCATCG	TGTCATGAAGATGTCCTGGC	*SCN5A*(24)	694	65-58[d]
9205	CYAAATGTCGGAGTGCRGGG	CTTGGCAATAGCTTTGGCTG	*ATP12A*(12)	817	65-58[d]
11080	AAGGGCAAGCTGTTGTAAGR	TCAGCTTTRGTTCCATTTCC	*PREX2*(29)	753	67-60[d]
13102	ACACYGTGGGKTTATGGAGC	TCCACACAGATAGCCCAAGG	*TRAPPC10*(8)	690	65-58[d]
15923 [a]	CTGAGCCCAAGTTCGAGAAG [a]	GTGTAGTCCTCCAGGGAGATATAG [a]	*SPTA1*(51)	739	68-58[d]
17701 [b]	CTCAGTGGTAGCCAAGGACC	GCTGGAGTTGGAGGAATCAG	*IGSF22*(15)	692	67-60[d]
22245	TTCYTGGAAATTGACCCAAC	GGCTGAAGGACTCCTCRCTG	*SEL1L3*(20)	714	58
OSTA-5	TGMWGGYCATGGTGGAAGGCTTTG	AGATGCCRTCRGGGAYGAGRAACA	*OSTA*(5)	724	67-60[d]
318.2C_Ame [c]	AAGAACTGTATTCCATCTRTCCC	AGKAAATGTGAAAGTACTGGTTTAC	Y chr.[g]	979	69-62[d]
318.3C	CGACCTTGACCAACAAGAGG	GAGATGGTCTCTGCAAGATGG	Y chr.	1216	66-61[e]
318.3C_Ame [c]	CCCTRTGCCATCATAAATCCC	TTAAGGCTGTGTTTGAGTGCC	Y chr.	724	69-62[d]
318.7C	TCTTCGTCTTCATGCTGTGG	CCAGCTCCTTATATGCTGAACC	Y chr.	1095	68-58[f]
318.7C_Ame [c]	TTGGAGGAGTCAGCTGATGAG	TGTTGGTGTTTCAGTTGTATGTTG	Y chr.	766	69-62[d]
579.3C	TTAACTGCTCTGACCTTCATCG	GTGCACAGGCAAGTGTTAGG	Y chr.	1157	68-58[f]
579.3C_Ame [c]	AATGAACTGCTTGACCTTCG	TGATGGAGGAAATTGAGTGC	Y chr.	1174	68-61[d]
318.10B	TGCACAGTTCAATGGCTACAG	TCAGCAGACATTTTCTTGGAAC	Y chr.	529	66-61[b]
318.11C	GATGATGCATAAGCAATCCTTG	TGCAACCATAACTTGTTTACTTCC	Y chr.	1012	69-64[e]
389	ACCCACTGCTGTTCTGTATCC	CCAACAGTGTAGTGGTTGTGC	Y chr.	679	68-61[d]
322	GAGTAGAGCTGGTGCTTGTGAG	GAAGCAGAGCTCAAGTCTGAAG	Y chr.	821	70-63[d]
403	CACTCAGGAGAGCACAGGTC	TGTGTGTCGTAAGCAGAGGTC	Y chr.	796	69-62[d]

Markers are named consecutively based on our list of aligned giant panda and dog sequences, with the numbering reflecting their relative position along the dog chromosomes. Gene and intron numbers in the giant panda are given; size for nuclear markers denotes expected amplicon size in base pairs in the giant panda; and T is the annealing temperature used in PCR. Y-specific markers are indicated in the column "gene" and named based on the scaffold they were obtained from.

[a] The primers for this locus were newly designed compared to Hailer et al. (2012), to improve specificity.

[b] These sequences correct an error in the primer sequences given in Hailer et al. (2012).

[c] Panda-specific primers were designed, in case no PCR product was obtained in any of giant panda, spectacled bear, Asian black bear, sun bear or sloth bear.

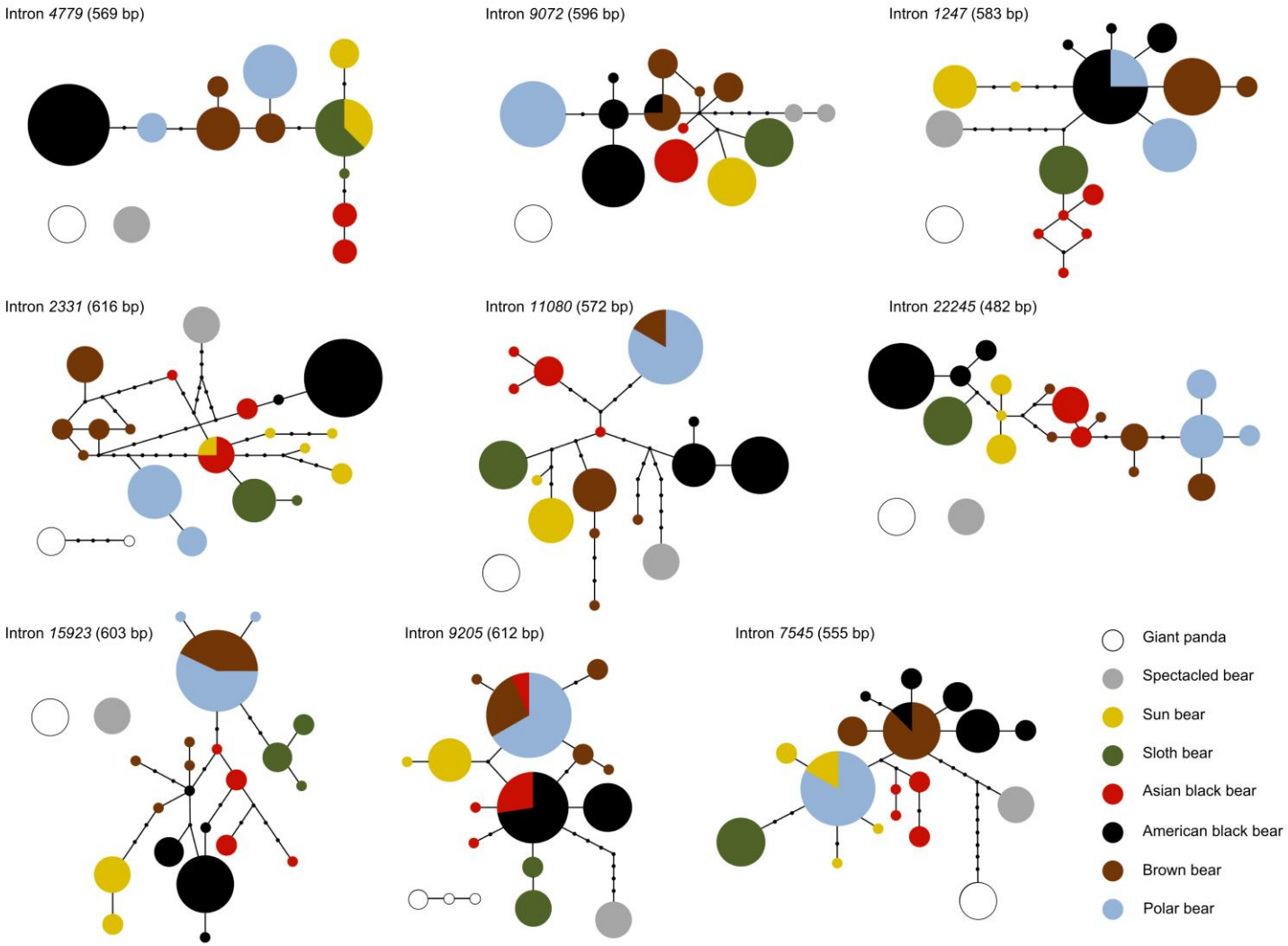
[d] Touchdown PCR, during which the annealing temperature was lowered by 0.5°C in each of 14 cycles, followed by 26 normal cycles.

[e] Touchdown PCR, during which the annealing temperature was lowered by 0.5°C in each of 10 cycles, followed by 30 normal cycles.

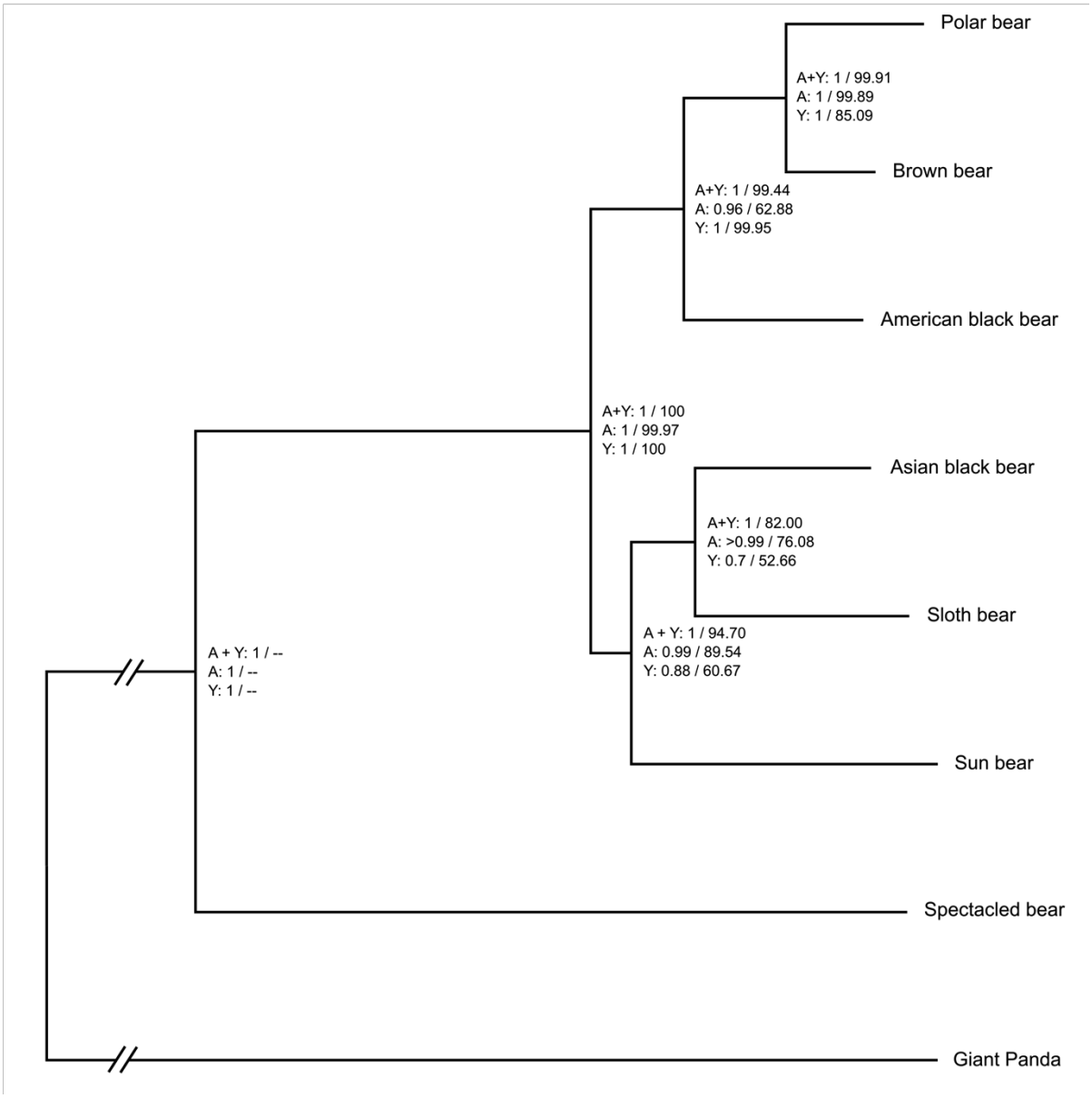
[f] Touchdown PCR, during which the annealing temperature was lowered by 1.0°C in each of 10 cycles, followed by 30 normal cycles.

[g] This marker includes exon 4 of *Usp9Y*.

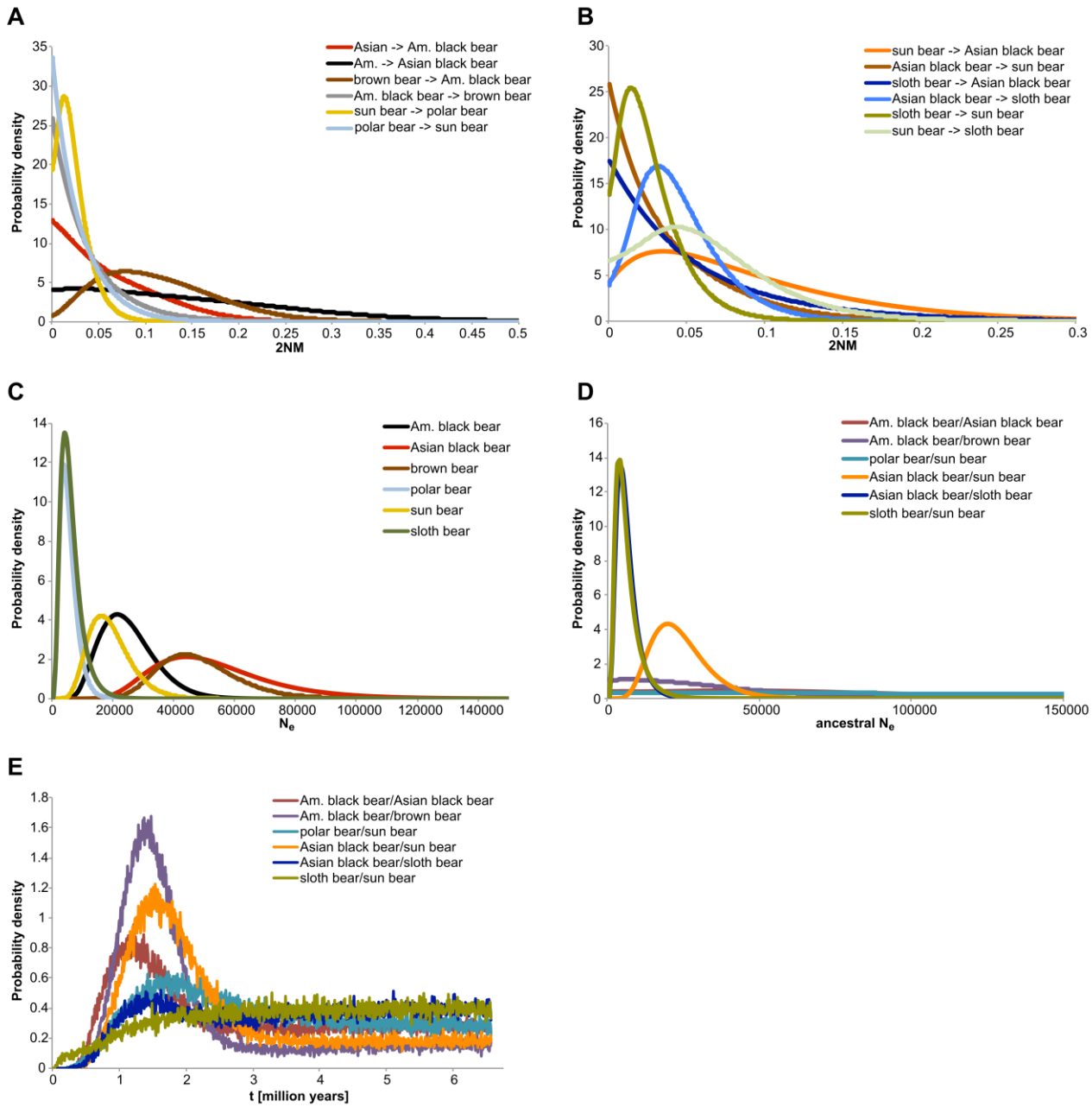
Supplementary figures



Supplementary figure S1: Statistical parsimony networks for nine autosomal intron markers in bears. Circle areas are proportional to haplotype frequencies and inferred intermediate states are shown as black dots. For some loci, spectacled bear and giant panda haplotypes were too divergent to be connected at the 95% credibility limit.



Supplementary figure S2: Phylogenetic tree of 14 concatenated autosomal introns and 5.9 kb Y-chromosomal sequence obtained from MrBayes. Numbers next to nodes denote branching support (first number: posterior probability values from MrBayes, second number: bootstrap values from maximum likelihood analyses in Treefinder), for three different datasets: (1) A+Y: 14 concatenated autosomal introns concatenated with nine Y-chromosomal markers, (2) A: 14 concatenated autosomal introns, and (3) Y: 5.9 kb Y-chromosomal sequence.



Supplementary figure S3: Posterior probability distributions for parameters in IMa2 pairwise comparison analyses. Curves are shown for (*A*) and (*B*) estimated population migration rates (2NM) between species; (*C*) effective population sizes (N_e) of the analyzed species; (*D*) effective population sizes of ancestral populations of the analyzed species pairs; (*E*) splitting time estimates (in million years).

References

Bidon T, Janke A, Fain SR, Eiken HG, Hagen SB, Saarma U, Hallström BM, Lecomte N, Hailer F. 2014. Brown and Polar Bear Y Chromosomes Reveal Extensive Male-Biased Gene Flow within Brother Lineages. Mol. Biol. Evol. 31:1353–1363.

Delisle I, Strobeck C. 2002. Conserved Primers for Rapid Sequencing of the Complete Mitochondrial Genome from Carnivores, Applied to Three Species of Bears. Mol. Biol. Evol. 19:357–361.

Hailer F, Kutschera VE, Hallström BM, Klassert D, Fain SR, Leonard JA, Arnason U, Janke A. 2012. Nuclear Genomic Sequences Reveal that Polar Bears Are an Old and Distinct Bear Lineage. Science 336:344–347.

Jameson D, Gibson AP, Hudelot C, Higgs PG. 2003. OGRe: a relational database for comparative analysis of mitochondrial genomes. Nucleic Acids Res. 31:202–206.

Jobb G, von Haeseler A, Strimmer K. 2004. TREEFINDER: a powerful graphical analysis environment for molecular phylogenetics. BMC Evol. Biol. 4:18.

Krause J, Unger T, Nocon A, Malaspinas A-S, Kolokotronis S-O, Stiller M, Soibelzon L, Spriggs H, Dear PH, Briggs AW, et al. 2008. Mitochondrial genomes reveal an explosive radiation of extinct and extant bears near the Miocene-Pliocene boundary. BMC Evol. Biol. 8:220.

Li R, Fan W, Tian G, Zhu H, He L, Cai J, Huang Q, Cai Q, Li B, Bai Y, et al. 2010. The sequence and de novo assembly of the giant panda genome. Nature 463:311–317.

Nakagome S, Pecon-Slattery J, Masuda R. 2008. Unequal Rates of Y Chromosome Gene Divergence During Speciation of the Family Ursidae. Mol. Biol. Evol. 25:1344–1356.

Pagès M, Calvignac S, Klein C, Paris M, Hughes S, Hänni C. 2008. Combined analysis of fourteen nuclear genes refines the Ursidae phylogeny. Mol. Phylogenet. Evol. 47:73–83.

Peng R, Zeng B, Meng X, Yue B, Zhang Z, Zou F. 2007. The complete mitochondrial genome and phylogenetic analysis of the giant panda (*Ailuropoda melanoleuca*). Gene 397:76–83.

Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. 2012. MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. Syst. Biol. 61:539–542.

Woerner AE, Cox MP, Hammer MF. 2007. Recombination-Filtered Genomic Datasets by Information Maximization. Bioinformatics 23:1851–1853.

Yu L, Li Y-W, Ryder O, Zhang Y-P. 2007. Analysis of complete mitochondrial genome sequences increases phylogenetic resolution of bears (Ursidae), a mammalian family that experienced rapid speciation. BMC Evol. Biol. 7:198.