The evolutionary history of bears is shaped by gene flow across species

Vikas Kumar[1,2], Fritjof Lammers[1,2], Tobias Bidon[1,2], Markus Pfenninger[1,2], Lydia Kolter[3], Maria A. Nilsson[1], Axel Janke[1,2*]

Affiliations

[1]Senckenberg Biodiversity and Climate Research Centre, Senckenberg Gesellschaft für Naturforschung, Senckenberganlage 25, D-60325 Frankfurt am Main, Germany

[2]Goethe University Frankfurt, Institute for Ecology, Evolution & Diversity, Biologicum, Max-von-Laue-Str.13, D-60439 Frankfurt am Main, Germany

[3] AG Zoologischer Garten Cologne, Riehler Straße 173, 50735 Cologne, Germany

***Corresponding author**: Axel Janke

Senckenberg Biodiversity and Climate Research Centre, Senckenberg Gesellschaft für Naturforschung, Senckenberganlage 25, D-60325 Frankfurt am Main, Germany

Tel: +49 6975421840, email: Axel Janke Axel.Janke@senckenberg.de

Abstract

Bears are iconic mammals with a complex evolutionary history. Natural bear hybrids and studies of few nuclear genes indicate that gene flow among bears may be more common than expected and not limited to the closely related polar and brown bears. Here we present a genome analysis of the bear family with representatives of all living species. Phylogenomic analyses of 869 mega base pairs divided into 18,621 genome fragments yielded a well-resolved coalescent species tree despite signals for extensive gene flow across species. However, genome analyses using three different statistical methods show that gene flow is not limited to closely related species pairs. Strong ancestral gene flow between the Asiatic black bear and the ancestor to polar, brown and American black bear explains numerous uncertainties in reconstructing the bear phylogeny. Gene flow across the bear clade may be mediated by intermediate species such as the geographically wide-spread brown bears leading to massive amounts of phylogenetic conflict. Genome-scale analyses lead to a more complete understanding of complex evolutionary processes. The increasing evidence for extensive inter-specific gene flow, found also in other animal species, necessitates shifting the attention from speciation processes achieving genome-wide reproductive isolation to the selective processes that maintain species divergence in the face of gene flow.

Introduction

Ursine bears are the largest living terrestrial carnivores and have evolved during the last three million years, attaining a wide geographical distribution range (Fig. 1). Bears are a prominent case where conflicting gene trees and an ambiguous fossil record [1] make the interpretation of their evolutionary history difficult [2]. Introgressive gene flow resulting from inter-species mating is believed to be rare among mammals [3]. However, some 600 mammalian hybrids are known [4] and the importance of hybridization has started to gain attention in evolutionary biology [5]. Yet, our knowledge of the extent of post speciation gene flow is limited, because few genomes of closely related species have been sequenced.

In bears, natural mating between grizzlies (brown bears *Ursus arctos*), and polar bears (*Ursus maritimus*) results in hybrid offspring, the grolars [6,7]. Genome scale studies in brown and polar bears find that up to 8.8% of individual brown bear genomes have an origin in the polar bear [8]. Additionally, the brown bear mitochondrial (mt) genome was captured by polar bears during an ancient hybridization event [9,10] and polar bear alleles are distributed across brown bear populations all over the world by male-biased migration and gene flow [8,11].

Polar and brown bears belong to the sub-family Ursinae, which comprises six extant, morphological and ecological distinct species [12], but hybridization among some ursine bears is possible. A natural hybrid has been reported also between the Asiatic black bear (*Ursus thibetanus*) and the sun bear (*Ursus malayanus*) [13]. In captivity more bear hybrids are known (between sun and sloth bear, Asiatic black and brown bear, and between American black and brown bear), some of them have been fertile [4]. Despite limited population sizes for most bears and apparently distinct habitats, morphology and ecology, molecular phylogenetic studies have been unable to unequivocally reconstruct the relationship among the six ursine bear species [2,14]. Especially, the evolution of the American (*Ursus americanus*) and Asiatic black bear is difficult to resolve, despite

75 being geographically well separated (Fig. 1).

76 Evidence from the fossil record, morphology and mitochondrial phylogeny suggested a
77 closer relationship between the Asiatic and the American black bears [14–17]. In contrast, autosomal
78 and Y-chromosomal sequences support a different grouping with the American black bear being
79 sister group to the brown/polar bear clade [2,11,14,18,19]. Another apparent conflict between
80 mitogenomics, morphology and autosomal sequence data is the position of the morphologically
81 distinct sloth bears (*Ursus ursinus*). Mitochondrial DNA (mtDNA) analyses and morphological
82 studies placed sloth bears as sister-group to all other ursine bears, while nuclear gene analyses favor
83 a position close to the sun bears [2,17,20]. A study of nuclear introns with multiple individuals for each
84 ursine species was unable to reconstruct a well-supported species tree and suggested that
85 incomplete lineage sorting (ILS) and/or gene flow caused the complexities in the ursine tree [2].
86 However, previous molecular studies did not have access to genome data from all bear species and
87 were thus limited to small parts of the genome.

88 The genomic era allows for a detailed analyses of how gene flow from hybridization affects
89 genomes, and has revealed much more complex evolutionary histories than previously anticipated
90 for many species, including our own [21–24]. Multiple genomic studies on polar, brown bears and the
91 giant panda [25–28] lead to a wealth of available genomic data for some ursine bears. We investigated
92 all living Ursinae and Tremarctinae bear species based on six newly sequenced bear genomes in
93 addition to published genomes. Methods specifically developed to deal with complex genome data
94 [29,30] and gene flow in particular [21,31] are applied to resolve and understand the processes that have
95 shaped the evolution of bears.

96

97

98

99

Results

The newly sequenced individuals were morphologically typical for the respective species. Mapping Illumina reads against the polar bear genome reference [28] yielded for the Asiatic bear species an average coverage of 11X. Tables in (Supplementary Table 1 and 2) detail the sequencing and assembly data, and provide accession numbers of the included species. As a basis for subsequent evolutionary analyses, non-overlapping 100 kb **G**enome **F**ragments (GFs) were extracted from assemblies against polar bear scaffolds larger than 1 megabase (Mb). These have presumably a higher assembly quality than smaller fragments and still represent >96% of the genome (Supplementary Fig. 1). Ambiguous sites, gaps, repetitive sequences, and transposable element sequences were removed from GF alignments (Supplementary Fig. 2). The newly sequenced individuals were morphologically typical for the respective species. Pedigrees (Supplementary Fig. 3) and heterozygosity plots (Supplementary Fig. 4) show that the sequenced individuals are neither hybrids nor, compared to wild specimens, severely inbred.

Network analysis depicts hidden conflict in the coalescent species tree

GFs larger than 25 kb, representing the majority of the length distribution (Supplementary Fig. 2), contain on average 104 substitutions between the Asiatic black bear and respectively, sun and sloth bear (Supplementary Fig. 5). Phylogenetic topology testing on real and simulated sequence data shows that GFs with this information content significantly reject alternative topologies (Supplementary Fig. 6 and Fig. 7). For subsequent coalescence, consensus, and network analyses only GFs larger than 25 kb were used and are thus based on firmly supported ML-analyses.

A coalescent species tree utilizing 18,621 GFs >25 kb (869,313,834 bp) resolved the relationships among bears with significant support for all branches (Fig. 2A, Supplementary Fig. 8). In the coalescent-based species tree, sun and sloth bears form the sister group to the Asiatic black bear, and the American black bear groups with polar and brown bears. The spectacled bear (*Tremarctos*

125 *ornatus*) is, consistent with previous results [2,18], placed as sister taxon to Ursinae. The well-resolved
126 coalescent species tree appears to be without conflict from the genomic data.
127 However, a network analysis [32] gained from the same 18,621 GFs identifies conflicting
128 phylogenetic signal in the data (Fig. 2B). The square and cuboid-like structures are indicative of
129 alternative phylogenetic signals, particularly among brown and polar bears, but also among the
130 Asiatic bear species. The brown bear from the Admiralty, Baranof, and Chichagof (ABC) islands
131 groups in different arrangements with the other brown and polar bears, consistent with gene flow
132 between the two species [8,10,28]. When the threshold level for depicting conflicting branches is
133 reduced in the network analysis, the signal becomes increasingly complex, illustrating the
134 conflicting signals from the 18,621 ML-trees (Supplementary Fig. 9). Still, the network analysis
135 agrees with the species tree when the spectacled bear is viewed as outgroup relative to the major
136 cluster of American black, brown and polar bear, and Asiatic black bear with, sun and sloth bear.
137 The phylogenetic conflict can be caused by incomplete lineage sorting (ILS) or gene flow, but less
138 likely from lack of resolution due to the strong phylogenetic signal of each GF (Supplementary Fig.
139 6 and Fig. 7). Analyses of 8,050 protein coding sequences (10,303,323 bp) and GFs from scaffolds
140 previously identified as X chromosomal (total 74Mb) [27], also conform to the species tree and
141 network (Supplementary Fig. 10). Finally, the paternal side of the bear evolution based on Y
142 chromosome sequences [33] cannot be analyzed for all species, but the resulting tree (Supplementary
143 Fig. 11) is consistent with the inferred species tree.
144 The Bayesian mtDNA tree (Fig. 3, Supplementary Fig. 12) from 11,529 bp of sequence information
145 that remained after alignment and filtering conforms to previous studies [2,17,34], making this the
146 hitherto largest taxonomic sampling of 38 complete bear mt genomes. However, several nodes of
147 the mtDNA tree differ notably from the coalescent species tree (Fig. 2A). In the mtDNA tree, the
148 brown bears are paraphyletic, because the brown bear mt genome introgressed into the polar bear

149 population [10,35]. The extinct cave bear (*Ursus spelaeus*) is the sister group to polar and brown bears.
150 The American black bear is the sister group to the Asiatic black bear, and the sloth bear is the sister
151 group to all ursine bears. The topological agreement of the mtDNA tree to previous mtDNA studies
152 and expected placement of the new individuals corroborates the matrilineal lineages and that the
153 studied individuals are representative for their species and likely not hybrids.
154 Finally, a consensus analysis based on ML-trees from GF (Supplementary Fig. 13) produces
155 a tree that is identical to the coalescent species tree, but highlights that numerous individual GF
156 trees support alternative topologies (Supplementary Table 3). Closer inspection of the individual
157 18,621 GF ML topologies shows that 38.1% (7,086) support a topology where Asiatic black bear is
158 the sister group to the American black/brown/polar bear clade. The Asiatic black bear groups in
159 different arrangements with the two other Asiatic bears: 18.7% (3,474) of the branches support a
160 grouping with the sun bear, and 7.5% (1,394) with the sloth bear in other topologies.
161
162 **Gene flow among bears is common**
163 Seemingly conflicting phylogenetic signals in evolutionary analyses can be explained either by
164 incomplete lineage sorting (ILS) [36] or gene flow among species. In contrast to the largely random
165 process of ILS, gene flow produces a bias in the phylogenetic signal, because it is a directed process
166 between certain species. The *D*-statistic measures the excess of shared polymorphisms of two
167 closely related lineages with respect to a third lineage [21] and can thus discriminate between gene
168 flow and ILS. The test assumes that the ancestral population of the three in-group taxa was
169 randomly mating and recently diverging [37]. These assumptions might be compromised in wide-
170 spread, structured species like bears. However, speciation is rarely instantaneous, but is rather
171 preceded by a period of population divergence. Such a period should not compromise the test as
172 long as there was a panmictic population ancestral to the progenitor populations of the eventual

daughter species at some point in time, which is a reasonable assumption.

The *D*-statistics analyses find evidence of gene flow between most sister bear species (Fig. 4, Supplementary Table 4-5 and Supplementary Fig. 14). Regardless if spectacled bear or giant panda is used as outgroup, the direction and relative signal strengths of gene flow in the tested topologies remain the same (Supplementary Table 6). The *D*-statistic is limited to four-taxon topologies and therefore gene flow signals are difficult to interpret when they occur between distant species, as it cannot determine if it is a direct, indirect, or ancestral signal. For taking more complex phylogenetic signals and gene flow patterns into account, and to determine the direction of gene flow, we applied the recently introduced D_{FOIL}-statistics. This method uses a symmetric five-taxon topology and has specifically been developed to detect and differentiate gene flow signal among ancestral lineages [31].

In agreement with the phylogenetic conflict and *D*-statistics, the D_{FOIL}- statistic finds strong gene flow between the ancestor of the American black bear/brown/polar bear clade, possibly the extinct Etruscan bear (*Ursus etruscus)*, and the Asiatic black bear (Fig. 4, Table 1). The Etruscan bear was geographically overlapping with other bear species and was, like the Asiatic black bear, widely distributed [38]. It has been identified in fossil layers of Europe between 2.5 Ma to 1.0 Ma [1,39,40]. The wide geographical distribution would explain the nearly equally strong gene flow from Asiatic black bear into brown bear also observed in the *D*-statistics analyses (Supplementary Fig. 14). Finally, there is a strong signal for gene flow between the American and Asiatic black bears. The gene flow could have taken place either on the American or Asiatic side of the Bering Strait and is consistent with the suggested mitochondrial capture between the species [2] and the deviating mitochondrial tree (Fig. 3). It should be noted that most of the weaker gene flow signals in Fig. 4 (dashed-lines) need to be corroborated and do not necessarily reflect direct species hybridization. The signals are possibly remnants of ancestral gene flow that is not detected because of allelic loss

or signal of indirect gene flow by ghost lineages or intermediate species. Permutations of species for the D_{FOIL} analysis including other polar, sloth and brown bear individuals show that the estimates are taxon independent (Table 1).

PhyloNet [41] has been specifically developed to detect hybridization events in genomic data while accounting for ILS. We applied the maximum likelihood approach implemented in PhyloNet [41,42] to detect signs of hybridization among bear species. We sampled 4,000 ML trees from putatively independent GFs with one individual representing each species. The ABC island brown bear was chosen as another representative for brown bears and positive control, because for the ABC island brown bear population hybridization with polar bears has been documented [8,10,33]. The outgroup, the spectacled bears were removed to reduce the computational complexity and, because previous analyses could not detect gene flow between tremarctine and ursine bears. The complex phylogeny requires exceptional computational time so we analyzed only networks with up to two reticulations. The resulting PhyloNet network with the highest likelihood (Supplementary Fig. 15) shows reticulations between ABC island brown bear and polar bears, and also between the Asiatic black bear and the ancestral branch to American black, brown and polar bears. It is noteworthy, that the second reticulation has a high inheritance probability (41.8%), which agrees with the strongest gene flow signal identified by D_{FOIL} analyses (Fig. 4, Table 1). Due to computational limits so far only two reticulations that represent the strongest hybridization signals could be identified. For three and more reticulations the network-space becomes extremely large.

Additional analysis using CoalHMM [43] supports the findings of gene flow from D-, D_{FOIL}, and PhyloNet analyses (Supplementary Fig. 16). It shows that a migration model fits most pair wise comparisons significantly better than ILS. A sensitivity analysis shows that this finding is robust under a broad range of parameters (Supplementary Fig. 17 and 18). Thus, gene flow among bears throughout most of their history is the major factor for generating conflicting evolutionary signals.

221

222 **Estimation of divergence times and population splits**

223 The phylogenomic based divergence time estimates (Fig. 5) are older than recent estimates based on
224 nuclear gene data [2], but are consistent with that on mtDNA data [17,44] (Supplementary Table 7). The
225 amount of heterozygous sites differs among species and individuals, and is highest in the Asiatic
226 black bear genome and, as expected [2] lowest in the polar bears and spectacled bears (Supplementary
227 Fig. 4). It is noteworthy that the average numbers of heterozygous sites differ among the two sun
228 bears, which may reflect different population histories.

229 Estimates for past changes in effective population size (N_e) using the pairwise sequentially
230 Markovian coalescent (PSMC) [45] for all newly sequenced bears are shown in Fig. 6 (unscaled
231 PSMC see Supplementary Fig. 19). While PSMC plots from low coverage genomes may vary and
232 not be ultimately accurate, the plots inferred for the brown, polar and American black bear are very
233 similar to previous published on higher coverage genomes (Supplementary Fig. 20) [26]. The
234 demographic histories of the Asian bear individuals vary widely, but do not overlap in bootstrap
235 analyses since 100 ka (Supplementary Fig. 21).

236 According to the PSMC analyses the Asiatic black bear maintained a stable and a relatively
237 high long-term *Ne* since 500 ka (Fig. 6). This is consistent with its wide geographic distribution and
238 its high degree of heterozygous sites in the genome [2]. The effective population size of the Asiatic
239 black bear declined some 20 ka, correlating with the end of the last ice age. By contrast, the
240 spectacled bear maintained a relatively low long-term effective population size, consistent with
241 their lower population diversity [2,46]. The demography of two sun bear individuals is strikingly
242 different from each other since 100 ka. As the bootstrap replicates do not overlap, the different
243 curves support a hypothesis of separate population dynamics (Supplementary Fig. 21). Their
244 distinct mitochondrial lineages (Fig. 3) might indicate that the two sun bear individuals belong to

the described subspecies *U. m. malayanus* (Sumatra and Asian mainland) and *U. m. euryspilus* (Borneo) [47] respectively. The ancestor of extant sun bears might have settled in the Malay Archipelago during the marine isotope stage (MIS) 6. In the following Eemian interglacial, Borneo got isolated, thereby giving rise to different environmental conditions and to a distinct sun bear subspecies, but without samples from multiple individuals from known locations and high coverage genomes, this remains speculative.

Discussion

Previously, nuclear gene trees and mitochondrial trees have been in conflict [17,18,34], and a forest of gene trees made it difficult to conclusively reconstruct the relationships among bears [2]. Now, phylogenomic analyses resolve a solid coalescent species tree and provide a temporal frame of the evolutionary history of the charismatic ursine and tremarctine bears. This exemplifies the power of multi-species-coalescent methods that are becoming increasingly important in genomic analyses [48]. Phylogenetic networks show that evolutionary histories of numerous GFs, i.e. various regions of their genome, are significantly different. This is also evident from large-scale evolutionary analysis of insertion patterns of transposable elements into the bear genomes, which yield a similarly complex history bears [49]. However, phylogenetic networks show that evolutionary histories of numerous GFs, i.e. various regions of their genome, are significantly different. It is important to realize that bifurcating species trees, even coalescence based, can only convey a fraction of the evolutionary information contained in entire genomes and that network analyses are needed to identify underlying conflict in the data [29,50]. The analyses of the ursine phylogeny suggest that gene flow and not incomplete lineage sorting are the major cause for the reticulations in the evolutionary tree. These two processes can be distinguished from each other by methods and programs like *D*-statistics, D_{FOIL} and Phylo-Net [21,31,41] that are specifically developed for this task.

Some of the inferred gene flow between bear species appears weak or episodic and thus requires further corroboration by additional sampling of individuals. Population analyses show that American black bears are divided into two distinct clades that diverged long before the last glacial maximum, indicating a long and isolated evolutionary history on the North American continent [51]. Thus, it is unlikely that American black bears came into contact with the Asiatic sun and sloth bears [15,51]. Likewise, introgressive gene flow between south-east Asiatic bear species and polar bears requires an explanation, because they have been evolving in geographically and climatically distinct areas, from the time when polar bears diverged from brown bears and began parapatric speciation in the Arctic. It is therefore possible that some gene flow events occurred through an intermediate species. The brown bear has been shown to distribute polar bear alleles across its range [8] and may therefore be a plausible vector species for the exchange of genetic material between Asiatic bears and the polar, or American black bear. The brown bear is a likely extant candidate, because it has been and is geographically wide-spread [52]. Furthermore, the geographical range of brown bears overlaps with all other ursine bear species (Fig. 1), they have reportedly migrated several times across continents and islands [52,53], and numerous brown bear hybrids with other bears in either direction are known [4]. While also the Asiatic black bear was widely distributed across Asia and had, like the brown bear [26], a large effective population size (Fig. 6), a migration of the Asiatic black bear into North America has not been shown. Likewise, migration of the American black bear in the opposite direction, from the American to the Asian continent, is not evident from fossil data. The D_{FOIL} and PhyloNet analyses [31,41] are powerful tools to detect ancestral gene flow, such as the prominent signal between the Asiatic black bear and the ancestor to the American black, brown and polar bears (Fig. 4, Table 1). In fact, gene flow during early ursine radiation from extinct bear species, such as the Etruscan bear or the cave bear is to be expected to leave signatures in genomes of their descendants and thus causing conflict in a bifurcating model of evolution.

293

294 **Speciation as a selective rather than an isolation process**

295 There is no question that bears are morphologically, geographically and ecologically distinct and
296 they are unequivocally accepted as species even by different species concepts [54]. Yet, our genome-
297 wide analyses identify gene flow among most ursines, making their genome a complex mosaic of
298 evolutionary histories. Increasing evidence for post-speciation gene flow among primates, canines,
299 and equids [22,23,55] suggests that interspecific gene flow is a common biological phenomenon. The
300 occurrences of gene flow and to a lesser extent ILS, of which a fraction in the phylogenetic signal
301 cannot be excluded, suggest that the expectation of a fully resolved bifurcating tree for most species
302 might be defied by the complex reality of genome evolution. Recent genome-scale analyses of basal
303 divergences of the avian [56], insect [57] and metazoan [58] tree share the same difficulties to resolve
304 certain branches as observed for mammals [59]. Detecting gene flow for these deep divergences is
305 difficult and therefore most of the reticulations and inconsistent trees have been so far attributed to
306 ILS [60].

307 The recent discoveries of gene flow by introgressive hybridization in several mammalian
308 species [22,23,55,61] and in bears over extended periods of their evolutionary history have a profound
309 impact of our understanding of speciation. If, in fact gene flow across several species is frequent,
310 and can last for several hundred-thousand years after divergence, evolutionary histories of genomes
311 will be inherently complex and phylogenetic incongruence will depict this complexity. Therefore,
312 speciation should not only be viewed as achieving genome-wide reproductive isolation but rather as
313 selective processes that maintain species divergence even under gene flow [62].

314

315

316

317

Materials and Methods

Genome sequencing, mapping and creation of consensus sequences

Prior to sampling and DNA extraction and evolutionary analyses, pedigrees from zoo studbooks and appearance of the individuals confirmed that these individuals are not hybrids (Supplementary Fig. 3). DNA extraction from opportunistically obtained tissue and blood samples was done in a pre-PCR environment on different occasions to avoid confusion by standard phenol/chloroform protocols and yielded between 1 to 6 µg DNA for each of the six bear individuals (S1 Data 1.1). Paired end libraries (500 bp) were made by Beijing Genome Institute (BGI) using Illumina TrueSeq and sequencing was done on Illumina HiSeq2000 resulting in 100 bp reads. Raw reads were quality-trimmed by Trimmomatic [63] with a sliding window option, minimum base quality of 20 and minimum read length of 25 bp. The assembled polar bear genome [28] was used for reference mapping using BWA version 0.7.5a [64] with the BWA-MEM algorithm on scaffolds larger than 1Mb. Scaffolds shorter than 1 Mb in length were not involved in the mapping and analyses, due to potential assembly artefacts [65] and for reducing the computational time in downstream analyses. Duplicate Illumina reads were marked by Picard tools version 1.106 (http://picard.sourceforge.net/) and the genome coverage was estimated from Samtools version: 0.1.18 [66].

Freebayes version 0.9.14-17 [67] called Single Nucleotide Variants (SNVs) using the option of reporting the monomorphic sites with additional parameters as -min-mapping-quality 20, -min-alternate-count 4, -min-alternate-fraction 0.3 and -min-coverage 4. A custom Perl script created consensus sequences for each of the mapped bear individuals from the Variant Call Format (VCF) files, keeping the heterozygous sites and removing indels. In order to complete the taxon sampling of the ursine bears, reads from six previously published genomes (S1 Data 1.1) were downloaded on the basis of geographic distribution, availability and sequence depth and SNVs were called as

342 described above. For the two high coverage (>30X) genomes, SNVs calling parameters were set as
343 one-half of the average read depth after marking duplicates. Genome error rates [68] were calculated
344 on the largest scaffold (67 Mb) for all bear genomes, confirming a high quality of the consensus
345 sequences. For additional details see Supplementary Methods and Supplementary Fig. 22.

346 **Data filtration, simulation of sequence length and topology testing**

347 The next step was to create multi-species alignments that could be used for further phylogenetic
348 analysis from all 13 bear individuals included in this study. In order to create a data set with
349 reduced assembly and mapping artefacts, genome data was masked for TEs and simple repeats [22]
350 using the RepeatMasker [69] output file of the reference genome (polar bear) available from
351 http://gigadb.org/ [28]. Since the polar bear reference genome RepeatMasker output file did not contain
352 the simple repeat annotation, we repeatmasked the polar bear reference genome with the option (-
353 int) to mask simple repeats. Then all the bear genomes were masked with the help of bedtools
354 version 2.17.0 [70] and custom Perl scripts. Non-overlapping, sliding window fragments of 100 kb
355 were extracted using custom perl scripts together with the program splitter from the Emboss
356 package [71] from scaffolds larger than 1 Mb (Supplementary Fig. 1), creating a dataset of 22,269 GFs
357 from 13 bear individuals. In addition, heterozygous sites, and repeat elements and ambiguous sites
358 were all marked "N" and removed using custom Perl scripts. An evaluation of the minimum
359 sequence length of GFs needed for phylogenetic analysis was done by estimating how much
360 sequence data is needed to reject a phylogenetic tree topology with using the approximate unbiased,
361 AU test [72]. This is important, because only sufficiently long sequences can differentiate between
362 alternative trees with statistical significance. The evaluation was done in two separate analyses: (a)
363 with a simulated data set and (b) on a data set of 500 random GFs. For additional details see
364 Supplementary Methods.

365 **Phylogenetic tree analysis using Genomic Fragment (GF) data**

For phylogenetic analysis, all GFs with length <25 kb were removed from the initial 22,269 GFs resulting in a data set consisting of 18,621 GFs (mean sequence length of 46,685 bp and standard deviation of 9,490 bp). The dataset was then used to create a coalescent phylogenetic species tree. First the selected GFs were used to create individual maximum likelihood (ML) trees using RAxML version 8.2.4 [73]. The best fitting substitution model was selected on 10 Mb of genomic data using jModelTest 2.1.1 [74] available in RAxML version 8.2.4 [73] and applied to all ML analyses. From 18,621 ML trees, Astral program [30] constructed a coalescent species tree. For bootstrap support of the coalescent species tree, GF were bootstrapped 100 times, generating a total of 1,862,100 ML trees. The bootstrapped ML-trees and the coalescent species tree were used as input in Astral [30] using default parameters to generate bootstrap support. The consense program in Phylip version 3.69 [75] built from 18,621 ML trees a majority rule consensus tree. SplitsTree version 4 [76] created a consensus network from the 18,621 GF ML-trees with various threshold settings (5%, 7%, 10% and 30%), to explore the phylogenetic conflict among the bear species.

Phylogenetic analysis of nuclear protein-coding genes and mitochondrial genomes

The annotation of the polar bear genome [28] was used to extract the protein coding sequences (CDS) from the genome that could be used for phylogenetic analysis. The species alignments were complemented by giant panda (*Ailuropoda melanoleuca*, ailMel1) sequences from Ensembl (http://www.ensembl.org/) to provide an outgroup to the analyses. In order to determine orthologous CDS between the polar bear reference genome and the giant panda Proteinortho version 5.06 [77] was used. The CDS of the new bear genomes that corresponded to the polar bear CDS were aligned with MAFFT version 7.154b [78]. Gaps were removed using Gblocks version 0.91b [79] and a custom perl script removed ambiguous sites. CDS <300 bp were not used for phylogenetic analyses. The best evolutionary model, GTR+G+I [80], was estimated using jModelTest 2.1.1[74]. A coalescent species tree was constructed with bootstrap support with Astral [30] from individual CDS using the GTR+G+I

model of sequence evolution. In addition, a concatenated analysis of the coding sequence was also done to estimate the concatenated CDS tree. The CDSs were concatenated and the substitution model GTR+G+I was used to create an ML tree with RaxML [73]. A AU topology test was made on the CDS topology using the CONSEL version 1.20 [81] and the species tree (Fig. 2A).

In order to extract the complete mt genomes from Illumina sequence data, the reads for different bear species were mapped to their respective published complete mt genome sequences using BWA version 0.7.5a [64]. Consensus sequences were created using Samtools version: 0.1.18 [66], aligned by MAFFT version 7.154b [78] to 32 published sequences (Accession numbers see Supplementary Fig. 12), and MrBayes version 3.2.2 [82] was used to create the Bayesian phylogenetic tree using the best fitting GTR+G+I model of sequence evolution. The analysis was run for 4,000,000 generations with a sample frequency of 4,000 with default priors and an arbitrary burn in of 25% of the samples. Convergence was assessed using the average standard deviation of split frequency which reached < 0.01 and potential scale reduction factor close to 1.00.

Gene flow analysis using *D*-statistics and the D_{FOIL}-method

The program ANGSD [83] was used for admixture analysis (*D*-statistics) among the ursine bears using the spectacled bear-Chappari as outgroup. The reads of other bears were mapped to the consensus sequence of the spectacled bear as previously described. In addition, insertion/deletion (indel) realignment was done using GATK version 3.1-1 [84]. All possible four-taxon topologies of the bear species including sun bear-Anabell, brown bear-Finland, Brown bear-ABC, Polar bear-2, American black bear, Asiatic black bear, Sloth bear were involved for gene flow analysis using *D*-statistics. A block jackknife procedure (with 10 Mb blocks) with parameters: -minQ 30 and -minMapQ30, was used to assess the significance of the deviation from zero. We also mapped the sun bear-Anabell, the Asiatic black bear and the sloth bear against the giant panda genome (ailMel1) http://hgdownload.soe.ucsc.edu/goldenPath/ailMel1/bigZips/ and repeated the analyses described above on to

414 investigate, if the choice of the outgroup affected our conclusions. In addition, we analyzed the data
415 using D_{FOIL}-statistics [31], to detect the signatures of introgression. For this analysis we assumed the
416 coalescent species tree (Fig. 2A) and selected a window size of 100 kb with --mode dfoil as
417 suggested by the authors [31]. Other parameters were left at default.

418 **Hybridization inference using PhyloNet**

419 A data set of 4,000 random (every fourth) GFs, that are putatively in linkage equilibrium, was
420 created to calculate rooted ML trees with RAxML as described earlier. The trees were pruned to
421 contain one individual of each ursine species plus the ABC- brown bear to reduce computational
422 complexity of the ML analyses. Maximum likelihood networks in a coalescent framework, thus
423 incorporating ILS and gene flow, were inferred using PhyloNet [41,42] allowing 0, 1 and 2 reticulations
424 in 50 runs and returning the five best networks.

425 **Estimation of heterozygosity, past effective population size and divergence times**

426 In order to calculate the amount of heterozygous sites as well as their distribution in all the bear
427 genomes, their genomes were fragmented into 10 Mb regions using custom Perl scripts. The
428 number of heterozygous sites was counted using a custom Perl script and plotted as distributions
429 using R. The pairwise sequentially Markovian coalescent (PSMC) [45] analysis was done to assess
430 past changes in effective population size over time. We used default parameters and 100 bootstrap
431 replicates assuming a generation time for brown and polar bears of ten years, and six years for the
432 other bear species for the PSMC analysis. We selected a mutation rate of 1×10^{-8}
433 changes/site/generation for all species. These parameters were used in previous brown and polar
434 bear [26] and enable comparability between the studies. A generation time of six years has been shown
435 for the American black bear [85] and was deemed realistic for the other relatively small-bodied bears.
436 The mutation rate is close to a pedigree-based mutation rate of 1.1×10^{-8} changes/site/generation in
437 humans [86] that is considered to be typical for mammals.

438 A well-documented fossil from the giant panda lineage at 12 million years ago (Ma) [20] with
439 a maximum calibration point of 20 Ma based on mitochondrial estimates [87] was used to provide the
440 calibration point needed in PAML MCMCtree [88] to estimate divergence times on 5,151,660 bp
441 coding sequence data. In addition, the divergence time of Tremarctinae was set to 7-13 Ma [89].
442 Divergence time estimates of Ursinae was based on the occurrence of *U. minimus* at 4.3-6 Ma [90],
443 and a polar/brown bear divergence was given a range of 0.48-1.1 Ma [8,10,28]. The calibration points
444 were used to estimate divergence times in MCMC tree in PAML [88] with a sample size of 2,000,000,
445 burn-in of 200,000, and tree sampling every second iteration. Convergence was checked by
446 repeating the analysis again.
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461

462 **References**

1. Wagner, J. Pliocene to early Middle Pleistocene ursine bears in Europe: a taxonomic overview. *J. Natl. Mus. Prague Nat. Hist. Ser.* **179,** 197–215 (2010).
2. Kutschera, V. E. *et al.* Bears in a Forest of Gene Trees: Phylogenetic Inference Is Complicated by Incomplete Lineage Sorting and Gene Flow. *Mol. Biol. Evol.* **31,** 2004–2017 (2014).
3. Coyne, J. A. & Orr, H. A. *Speciation*. **37,** (Sunderland, MA: Sinauer Associates, 2004).
4. Gray, A. *Mammalian hybrids. A check-list with bibliography*. (Commonwealth Agricultural Bureaux, 1972).
5. Mallet, J. Hybridization as an invasion of the genome. *Trends Ecol. Evol.* **20,** 229–237 (2005).
6. Preuß, A., Ganslosser, U., Purschke, G. & Magiera, U. Bear-hybrids: behaviour and phenotype. *Zool. Gart.* **78,** 204–220 (2009).
7. Smol, J. P. Climate Change: A planet in flux. *Nature* **483,** S12–S15 (2012).
8. Cahill, J. A. *et al.* Genomic evidence of geographically widespread effect of gene flow from polar bears into brown bears. *Mol. Ecol.* **24,** 1205–1217 (2015).
9. Edwards, C. J. *et al.* Ancient hybridization and an Irish origin for the modern polar bear matriline. *Curr. Biol. CB* **21,** 1251–1258 (2011).
10. Hailer, F. *et al.* Nuclear genomic sequences reveal that polar bears are an old and distinct bear lineage. *Science* **336,** 344–347 (2012).
11. Bidon, T. *et al.* Brown and polar bear Y chromosomes reveal extensive male-biased gene flow within brother lineages. *Mol. Biol. Evol.* **31,** 1353–1363 (2014).
12. Nowak, R. *Walker's Mammals of the World*. (Johns Hopkins Press, 1991).
13. Galbreath, G. J., Hunt, M., Clements, T. & Waits, L. P. An apparent hybrid wild bear from Cambodia. *Ursus* **19,** 85–86 (2008).
14. Yu, L., Li, Y.-W., Ryder, O. A. & Zhang, Y.-P. Analysis of complete mitochondrial genome sequences increases phylogenetic resolution of bears (Ursidae), a mammalian family that experienced rapid speciation. *BMC Evol. Biol.* **7,** 198 (2007).
15. Kurten, B. & Anderson, E. Pleistocene mammals of North America. *Columbia Univ. Press* (1980).
16. McLellan, B. & Reiner, D. A review of bear evolution. *Bears Their Biol. Manag.* 85–96 (1994).

17. Krause, J. *et al.* Mitochondrial genomes reveal an explosive radiation of extinct and extant bears near the Miocene-Pliocene boundary. *BMC Evol. Biol.* **8,** 220 (2008).

18. Pagès, M. *et al.* Combined analysis of fourteen nuclear genes refines the Ursidae phylogeny. *Mol. Phylogenet. Evol.* **47,** 73–83 (2008).

19. Nakagome, S., Mano, S. & Hasegawa, M. Comment on 'Nuclear Genomic Sequences Reveal that Polar Bears Are an Old and Distinct Bear Lineage'. *Science* **339,** 1522–1522 (2013).

20. Abella, J. *et al.* Kretzoiarctos gen. nov., the Oldest Member of the Giant Panda Clade. *PLoS ONE* **7,** e48985 (2012).

21. Green, R. E. *et al.* A Draft Sequence of the Neandertal Genome. *Science* **328,** 710–722 (2010).

22. Carbone, L. *et al.* Gibbon genome and the fast karyotype evolution of small apes. *Nature* **513,** 195–201 (2014).

23. Jónsson, H. *et al.* Speciation with gene flow in equids despite extensive chromosomal plasticity. *Proc. Natl. Acad. Sci. U. S. A.* **111,** 18655–18660 (2014).

24. Poelstra, J. W. *et al.* The genomic landscape underlying phenotypic integrity in the face of gene flow in crows. *Science* **344,** 1410–1414 (2014).

25. Li, R. *et al.* The sequence and de novo assembly of the giant panda genome. *Nature* **463,** 311–317 (2010).

26. Miller, W. *et al.* Polar and brown bear genomes reveal ancient admixture and demographic footprints of past climate change. *Proc. Natl. Acad. Sci.* **109,** E2382–E2390 (2012).

27. Cahill, J. A. *et al.* Genomic Evidence for Island Population Conversion Resolves Conflicting Theories of Polar Bear Evolution. *PLoS Genet.* **9,** e1003345 (2013).

28. Liu, S. *et al.* Population genomics reveal recent speciation and rapid evolutionary adaptation in polar bears. *Cell* **157,** 785–794 (2014).

29. Bapteste, E. *et al.* Networks: expanding evolutionary thinking. *Trends Genet.* **29,** 439–441 (2013).

30. Mirarab, S. *et al.* ASTRAL: genome-scale coalescent-based species tree estimation. *Bioinformatics* **30,** (2014).

31. Pease, J. B. & Hahn, M. W. Detection and Polarization of Introgression in a Five-Taxon Phylogeny. *Syst. Biol.* **64,** 651–662 (2015).

32. Huson, D. H. H., Regula, R. & Scornavacca, C. *Phylogenetic Networks*. (Cambridge University Press, 2010).

33. Bidon, T., Schreck, N., Hailer, F., Nilsson, M. & Janke, A. Genome-wide search identifies 1.9 megabases from the polar bear Y chromosome for evolutionary analyses. *Genome Biol. Evol.* evv103 (2015). doi:10.1093/gbe/evv103

34. Yu, L., Li, Q., Ryder, O. A. & Zhang, Y. Phylogeny of the bears (Ursidae) based on nuclear and mitochondrial genes. *Mol. Phylogenet. Evol.* **32,** 480–494 (2004).

35. Cronin, M. A., Amstrup, S. C., Garner, G. W. & Vyse, E. R. Interspecific and intraspecific mitochondrial DNA variation in North American bears (*Ursus*). *Can. J. Zool.* **69,** 2985–2992 (1991).

36. Nei, M. *Molecular Evolutionary Genetics*. (Columbia University Press, 1987).

37. Durand, E. Y., Patterson, N., Reich, D. & Slatkin, M. Testing for ancient admixture between closely related populations. *Mol. Biol. Evol.* **28,** 2239-2252 (2011).

38. Baryshnikov, G. & Zakharov, D. Early pliocene bear *Ursus thibetanus* (Mammalia, carnivora) from Priozernoe locality in the Dniester basin (Molodova republic). *Proc. Zool. Inst. RAS* **317,** 3–10 (2013).

39. Pérez-Hidalgo, T. & José, T. The European descendants of *Ursus etruscus* C. Cuvier (Mammalia, Carnivora, Ursidae). (1992).

40. Croitor, R. & Brugal, J.-P. Ecological and evolutionary dynamics of the carnivore community in Europe during the last 3 million years. *Quat. Int.* **212,** 98–108 (2010).

41. Than, C., Ruths, D. & Nakhleh, L. PhyloNet: a software package for analyzing and reconstructing reticulate evolutionary relationships. *BMC Bioinformatics* **9,** 322 (2008).

42. Yu, Y., Dong, J., Liu, K. J. & Nakhleh, L. Maximum likelihood inference of reticulate evolutionary histories. *Proc. Natl. Acad. Sci.* **111,** 16448–16453 (2014).

43. Mailund, T. *et al.* A new isolation with migration model along complete genomes infers very different divergence processes among closely related great ape species. *PLoS Genet* **8,** e1003125 (2012).

44. Heath, T. A., Huelsenbeck, J. P. & Stadler, T. The fossilized birth–death process for coherent calibration of divergence-time estimates. *Proc. Natl. Acad. Sci.* **111,** E2957–E2966 (2014).

45. Li, H. & Durbin, R. Inference of human population history from individual whole-genome sequences. *Nature* **475,** 493–496 (2011).

46. García-Rangel, S. Andean bear *Tremarctos ornatus* natural history and conservation. *Mammal Rev.* **42,** 85–119 (2012).

47. Meijaard, E. Craniometric differences among Malayan sun bears (Ursus malayanus); evolutionary and taxonomic implications. *Raffles Bull. Zool.* **52,** 665–672 (2004).

48. Edwards, S. V. *et al.* Implementing and testing the multispecies coalescent model: A valuable paradigm for

phylogenomics. *Mol. Phylogenet. Evol.* **94,** 447-462 (2016).

49. Lammers, F., Gallus, S., Janke, A. & Nilsson, M. Massive phylogenetic conflict in bears identified by genome-wide discovery of transposable element insertions. Submitted (2016).
50. Nakhleh, L. Computational approaches to species phylogeny inference and gene tree reconciliation. *Trends Ecol. Evol.* **28,** 719–728 (2013).
51. Puckett, E. E., Etter, P. D., Johnson, E. A. & Eggert, L. S. Phylogeographic Analyses of American Black Bears (*Ursus americanus*) Suggest Four Glacial Refugia and Complex Patterns of Postglacial Admixture. *Mol. Biol. Evol.* **32,** 2338–2350 (2015).
52. Davison, J. *et al.* Late-Quaternary biogeographic scenarios for the brown bear (Ursus arctos), a wild mammal model species. *Quat. Sci. Rev.* **30,** 418–430 (2011).
53. Hirata, D. *et al.* Molecular phylogeography of the brown bear (*Ursus arctos*) in Northeastern Asia based on analyses of complete mitochondrial DNA sequences. *Mol. Biol. Evol.* **30,** 1644–1652 (2013).
54. Harrison, R. G. & Larson, E. L. Hybridization, Introgression, and the Nature of Species Boundaries. *J. Hered.* **105,** 795–809 (2014).
55. Freedman, A. H. *et al.* Genome sequencing highlights the dynamic early history of dogs. *PLoS Genet.* **10,** e1004016 (2014).
56. Jarvis, E. D. *et al.* Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science* **346,** 1320–1331 (2014).
57. Misof, B. *et al.* Phylogenomics resolves the timing and pattern of insect evolution. *Science* **346,** 763–767 (2014).
58. Nosenko, T. *et al.* Deep metazoan phylogeny: when different genes tell different stories. *Mol. Phylogenet. Evol.* **67,** 223–233 (2013).
59. Hallström, B. M. & Janke, A. Mammalian Evolution May not Be Strictly Bifurcating. *Mol. Biol. Evol.* **27,** 2804–2816 (2010).
60. Suh, A., Smeds, L. & Ellegren, H. The Dynamics of Incomplete Lineage Sorting across the Ancient Adaptive Radiation of Neoavian Birds. *PLoS Biol.* **13,** e1002224 (2015).
61. Li, G., Davis, B. W., Eizirik, E. & Murphy, W. J. Phylogenomic evidence for ancient hybridization in the genomes of living cats (Felidae). *Genome Res.* (2015). doi:10.1101/gr.186668.114
62. Wu, C.-I. The genic view of the process of speciation. *J. Evol. Biol.* **14,** 851–865 (2001).

63. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30,** 2114–2120 (2014).

64. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinforma. Oxf. Engl.* **25,** 1754–1760 (2009).

65. Baker, M. De novo genome assembly: what every biologist should know. *Nat. Methods* **9,** 333–337 (2012).

66. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinforma. Oxf. Engl.* **25,** 2078–2079 (2009).

67. Garrison, E. & Marth, G. Haplotype-based variant detection from short-read sequencing. arXiv preprint arXiv:1207.3907v2. (2012).

68. Orlando, L. *et al.* Recalibrating equus evolution using the genome sequence of an early Middle Pleistocene horse. *Nature* **499,** 74–78 (2013).

69. Smit, A., Hubley, R. & Green, P. RepeatMakser Open-4.0.2013-2015. (2015).

70. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinforma. Oxf. Engl.* **26,** 841–842 (2010).

71. Rice, P., Longden, I. & Bleasby, A. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet. TIG* **16,** 276–277 (2000).

72. Shimodaira, H. An approximately unbiased test of phylogenetic tree selection. *Syst. Biol.* **51,** 492–508 (2002).

73. Stamatakis, A. RAxML Version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30,** 1312-3. (2014).

74. Darriba, D., Taboada, G. L., Doallo, R. & Posada, D. jModelTest 2: more models, new heuristics and parallel computing. *Nat. Methods* **9,** 772 (2012).

75. Felsenstein, J. PHYLIP (Phylogeny Inference Package) version 3.6. A*vailable from :Author Department of genome sciences, University of Washington Seattle.* (2005).

76. Huson, D. H. & Bryant, D. Application of Phylogenetic Networks in Evolutionary Studies. *Mol. Biol. Evol.* **23,** 254–267 (2006).

77. Lechner, M. *et al.* Proteinortho: detection of (co-)orthologs in large-scale analysis. *BMC Bioinformatics* **12,** 124 (2011).

78. Katoh, K., Misawa, K., Kuma, K. & Miyata, T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30,** 3059–3066 (2002).

79. Talavera, G. & Castresana, J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst. Biol.* **56,** 564–577 (2007).

80. Lanave, C., Preparata, G., Sacone, C. & Serio, G. A new method for calculating evolutionary substitution rates. *J. Mol. Evol.* **20,** 86–93 (1984).

81. Shimodaira, H. & Hasegawa, M. CONSEL: for assessing the confidence of phylogenetic tree selection. *Bioinformatics* **17,** 1246–1247 (2001).

82. Ronquist, F. & Huelsenbeck, J. P. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinforma. Oxf. Engl.* **19,** 1572–1574 (2003).

83. Korneliussen, T. S., Albrechtsen, A. & Nielsen, R. ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics* **15,** 356 (2014).

84. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20,** 1297–1303 (2010).

85. Onorato, D. P., Hellgren, E. C., van Den Bussche, R. A. & Doan-Crider, D. L. Phylogeographic patterns within a metapopulation of black Bears (*Ursus americanus*) in the American Southwest. *J. Mammal.* **85,** 140–147 (2004).

86. Veeramah, K. R. & Hammer, M. F. The impact of whole-genome sequencing on the reconstruction of human population history. *Nat. Rev. Genet.* **15,** 149–162 (2014).

87. Wu, J. *et al.* Phylogeographic and Demographic Analysis of the Asian Black Bear (Ursus thibetanus) Based on Mitochondrial DNA. *PloS One* **10,** e0136398 (2015).

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

89. Tedford, R. H. & Martin, J. Plionarctos, a Tremarctine Bear (Ursidae: Carnivora) from Western North America. *J. Vertebr. Paleontol.* **21,** 311–321 (2001).

90. Gustafson, P. The vertebrate faunas of the Pliocene Ringold Formation, south-central Washington. University of of Oregon, Mus. *Nat Hist Bull* **23,** 1–62 (1978).

463

464

465

466

Acknowledgements

We are grateful to Luay Nakhleh (Rice University) for expert help with Phylo-Net analyses, Yichen Zheng for valuable comments on the manuscript and to Jon Baldur Hlidberg (www.fauna.is), and Aidin Niamir for artwork. Blood samples were kindly provided by Ludwig Carsten (Allwetter Zoo Münster), Tim Schikora (Zoo Schwerin), Christian Wenker (Basel Zoo) and Eva Martinez Nevado (Zoo Madrid). This study was supported by Hesse's funding program LOEWE (Landes-Offensive zur Entwicklung Wissenschaftlich-ökonomischer Exzellenz) and the Leibniz Society.

Author contributions

A.J. designed the research and obtained funding. A.J. and T.B. collected the data; V.K. and F.L. conducted the analyses; L.K. provided pedigrees and located samples; A.J., V.K., M.P., M.N., F.L., and T.B. interpreted the results; A.J. and V.K. wrote the paper with the help of all authors.

Completing Financial Interests

The authors declare no competing interests.

Data submission

The raw reads of the genome sequences have been deposited in the European Nucleotide Archive under the BioProject accession code PRJEB9724.

Figures

Figure 1: Approximate geographic distribution of extant bears according to IUCN data. Figure has been created using ArcGIS 10 (http://desktop.arcgis.com/en/arcmap/). The base map is obtained from the Global Administrative Areas (2012), GADM database of Global Administrative Areas, version 2.0 (http://www.gadm.org). Species range maps are from IUCN 2015 (http://www.iucnredlist.org).

Figure 2: A coalescent species tree and a split network analysis from 18,621 GF ML trees. (A) In the coalescent species tree all branches receive 100% bootstrap support. The position of the root and the depicted branch lengths were calculated from 10 Mb of GF data. (B) A split network analysis with a 7% threshold level for conflicting branches depicts the complex phylogenetic signal found in the bear genome. The ABC-island brown bear (asterisk) shares alleles with polar bears; among the Asiatic bears allele sharing is more complex. The bear paintings were made by Jon Baldur Hlidberg (www.fauna.is).

Figure 3: Phylogenetic relationship among the bears using mtDNA genomes A Bayesian mitochondrial tree from 37 complete mt genomes (colored circles). Support values for p<1.0 are shown in (Supplementary Fig. 12) lists accession numbers. Stars indicate the mt genes that were sequenced in this study. Asterisks indicate new mt genomes and from the scale bar indicates 0.04 substitutions per site.

Figure 4: Graphical summary of gene flow analyses using *D* and D_{FOIL} statistics on a cladogram. The percentage of GFs rejecting the species tree and indicating gene flow estimated by D_{FOIL} analyses is shown by percentages and blue arrows for values >1%, and dashed lavender for <0.1% (Table 1). These percentages do not indicate the total amount of introgressed genetic material, which can be a fraction of the GF sequence. Green arrows depict *D*-statistics results for

521 significant gene flow signal between recent species. Some gene flow cannot have occurred directly
522 between species, because the involved species exist in different habitats. Therefore, it is suggested
523 that such signals are likely remnants of ancestral gene flow or gene flow through a vector species.
524 The bear painting was made by Jon Baldur Hlidberg (www.fauna.is).
525
526 **Figure 5: Phylogenomic estimates of divergence times.**
527 The light blue scale bar below the tree shows divergence times in million years and 95% confidence
528 intervals for divergence times are shown at the nodes as shadings. Exact times and intervals for all
529 divergences are listed in Supplementary Table 7. The bear painting was made by Jon Baldur
530 Hlidberg (www.fauna.is).
531
532 **Figure 6: Historical effective population sizes (N_e) using the pairwise Markovian coalescent**
533 **(PSMC) analyses for the newly sequenced bear genomes.** The x-axis shows the time and the y-
534 axis shows the effective population size (N_e). The two sun bears have radically different population
535 histories that are not overlapping in bootstrap analyses since 100 ka (Supplementary Fig. 21) and
536 the Asiatic black bear had a constant large Ne since 500 ka similar to that of the brown bear which
537 is consistent with the wide geographic distribution and high heterozygosity. Other bears maintained
538 a relatively low long-term Ne, consistent with a lower population diversity and low heterozygosity
539 (Supplementary Fig. 4). The bear paintings were made by Jon Baldur Hlidberg (www.fauna.is).
540
541

Table 1. Gene flow detected by the D_{FOIL} analyses that is based on a five taxon analysis.

	AmB, BrB, SuB, AsB	**AmB, BrB, SlB, AsB**	**AmB, PoB, SuB, AsB**	**AmB, PoB, SlB, AsB**	**Average % (As/AmB, SuB/SlB, PoB/BrB)**	**Average % (BrB, AmB, SuB/SlB, AsB)**	**Average % (PoB, AmB, SuB/SlB, AsB)**
AsB => AmB	0.07% (15)	0.08% (18)	0.14% (31)	0.15% (34)	0.11%		
AmB => AsB	1.22% (270)	1.42%(313)	2.06% (454)	2.48% (547)	1.80%		
Su/SlB => AmB	0.02% (4)	0.01% (1)	0.02% (5)	0.02% (5)		0.01%	0.02%
AmB => SuB/SlB	0.02% (4)	0.01%(1)	0.02% (4)	0.01% (2)		0.01%	0.01%
SuB/SlB => BrB/PoB	0.07% (16)	0.02%(4)	0.02% (5)	0.02% (5)		0.05%	0.02%
BrB/PoB => SuB/SlB	0.03% (6)	0.01%(1)	0.02% (5)	0.01% (3)		0.02%	0.02%
AsB => BrB/PoB	1.25% (276)	1.02%(225)	0.46% (101)	0.29 (64)		1.14%	0.37%
Br/Po => As	9.70% (2159)	11.0% (2415)	8.37% (1846)	9.02 (1989)		10.4%	8.69%
BrB/PoB, AmB <=> SuB/SlB	0.10% (23)	0.06 (14)	0.20% (44)	0.11% (25)	0.12%		
BrB/PoB, AmB <=> AsB	32.2% (7098)	32.0% (7060)	46.3% (10214)	45.8% (10108)	39.1%		

Note – The table shows the percentage of 100 kb fragments that have a signal of gene flow, and in brackets the absolute number is shown. The rows show these values for different combinations of four bear species with the spectacled bear as an outgroup. The last three columns summarize amount of gene flow. The arrows in the table (=>) indicate the direction of the gene flow, between the respective species for each of the combinations analyzed. For example: between Asiatic black and American black bear the D_{FOIL} finds 15 – 34 GF that support gene flow (first row). There is much more gene flow in the other direction (second row). Abbreviations: SuB (Sun bear), SlB (Sloth bear), AsB (Asiatic black bear), AmB (American black bear), BrB (Brown bear, Finland), and PoB (Polar bear, Svalbard).

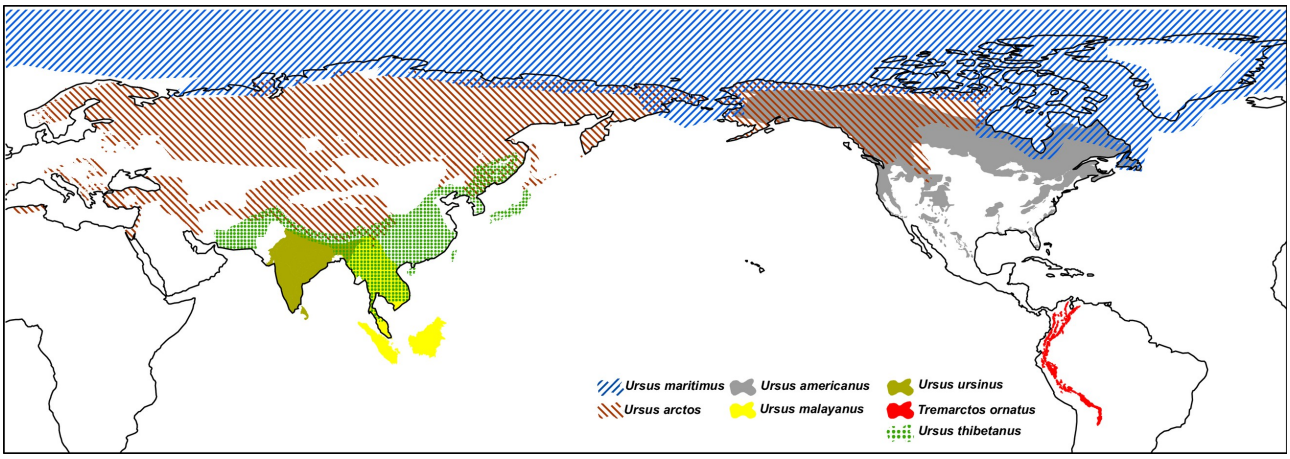
562

563 **Figures**

564

565 **Figure 1**

566



568

569

570

571

572

573

574

575

576

577

578

579

Figure 2

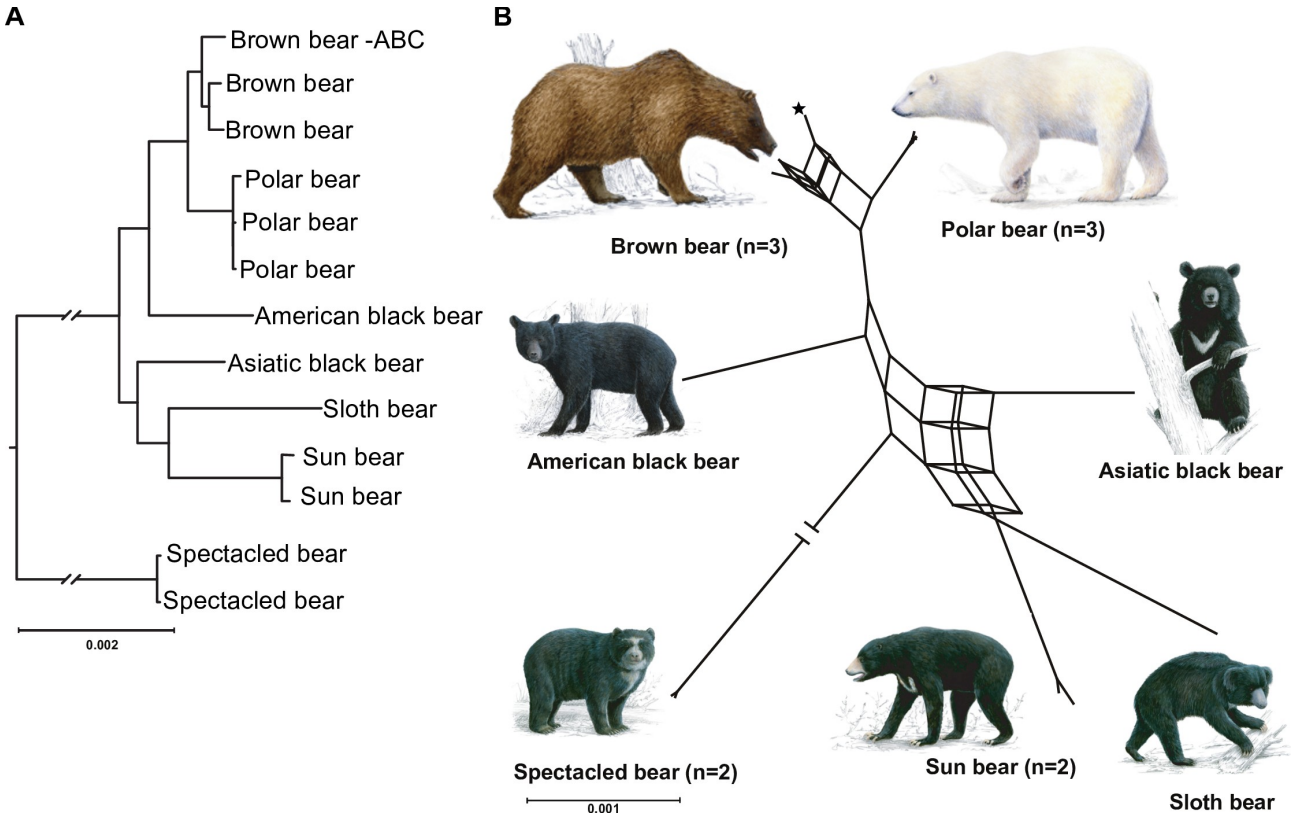


Figure 3

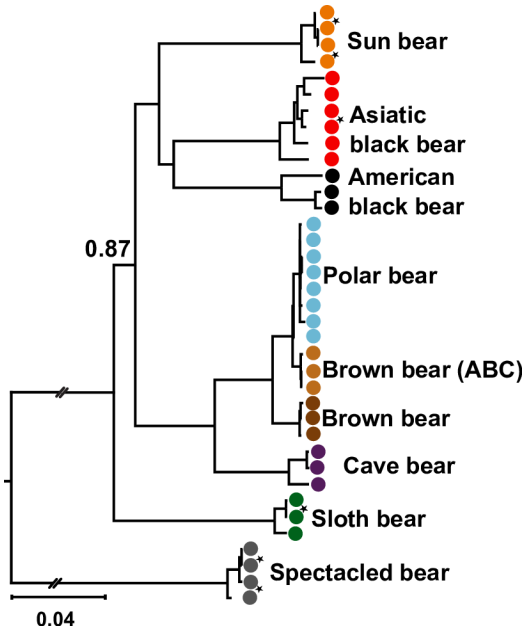
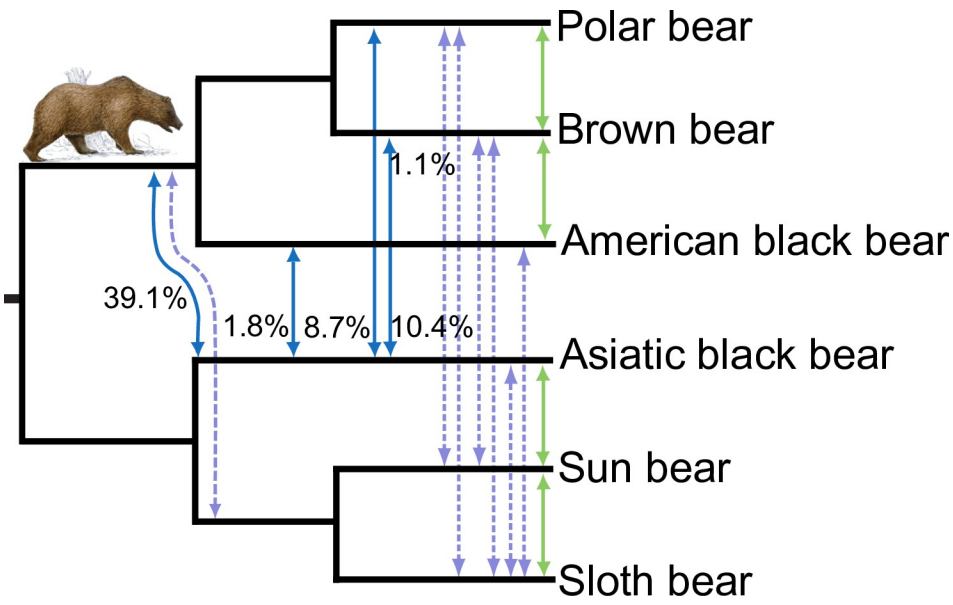


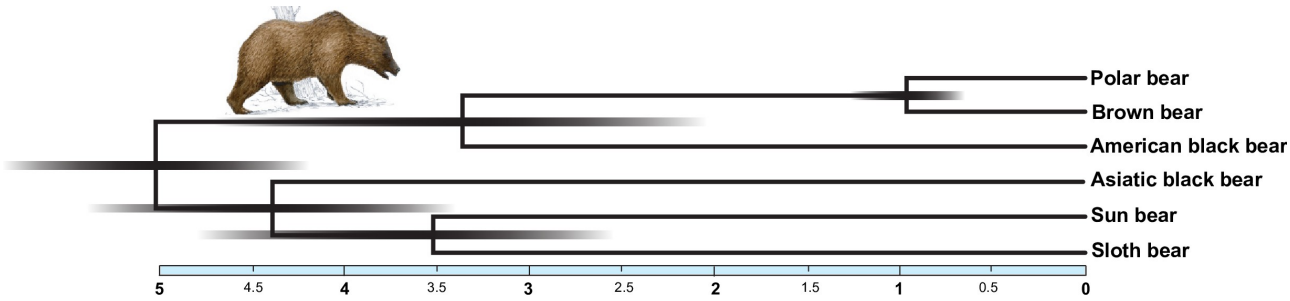
Figure 4



642

643 **Figure 5**

644



646

647

648

649

650

651

652

653

654

655

656

657

658

659

660

661

662

663 **Figure 6**

664