TRACING THE EVOLUTION OF MALE LINEAGES IN BEARS USING GENETIC MARKERS ON THE Y CHROMOSOME

Dissertation
zur Erlangung des Doktorgrades
der Naturwissenschaften

vorgelegt beim Fachbereich 15
der Johann Wolfgang Goethe - Universität
in Frankfurt am Main

von
Tobias Bidon
aus Tuttlingen

Frankfurt am Main, 2015
(D 30)

vom Fachbereich 15 (Biowissenschaften) der

Johann Wolfgang Goethe - Universität als Dissertation angenommen.

Dekanin: Prof. Dr. Meike Piepenbring

Gutachter: Prof. Dr. Axel Janke, Prof. Dr. Markus Pfenninger

Datum der Disputation:

10.09.2015

TABLE OF CONTENTS

Summary **1**

Zusammenfassung **3**

Hintergrund *3*

Durchgeführte Studien *4*

Fazit *7*

General introduction **9**

The quest for genetic variation *9*

Characteristics and structure of the Y chromosome *11*

The male perspective 11

Its special features render the Y chromosome a valuable marker in evolutionary studies 13

Sex chromosomes evolved from autosomes 15

Y chromosomes are rich in repetitive sequences 17

Study organisms *18*

Phylogenetic relationship of bears 18

Brown and polar bears have different levels of intraspecific variation 20

Thesis objectives *21*

General Discussion **23**

The Y chromosome as a resource – developing male-specific markers *24*

Molecular sex determination in conservation and forensics 24

Identifying male-specific markers is complicated by the structure of the Y chromosome 26

The Y chromosome as an evolutionary marker – uncovering the paternal history of bears *28*

Expanding phylogeographic inferences using male-specific markers 28

Lack of patrilineal structuring in brown bears 30

Male-biased gene flow shapes genetic variation in brown bears and other species 31

Divergent Y-chromosomal clades in polar bears are geographically widespread 33

The Y chromosome and ursine speciation 35

Y-chromosomal introgression is unlikely in bears 38

The role of the Y chromosome in speciation might have been underappreciated 39

On the relevance of a single locus in the era of genomics *40*

Conclusion *41*

References **43**

Publications **55**

Publication 1: A sensitive and specific multiplex PCR approach for sex identification of ursine and tremarctine bears suitable for non-invasive samples *55*

Publication 2: Brown and Polar Bear Y Chromosomes Reveal Extensive Male-Biased Gene Flow within Brother Lineages *65*

Publication 3: Bears in a Forest of Gene Trees: Phylogenetic Inference Is Complicated by Incomplete Lineage Sorting and Gene Flow *93*

Appendix **123**

Genome-wide search identifies 1.9 megabases from the polar bear Y chromosome for evolutionary analyses *123*

Acknowledgements **157**

Curriculum Vitae **159**

SUMMARY

The mammalian family of bears (Ursidae) comprises eight extant species, occurring on four different continents. Among them are the iconic and well-known brown and polar bears, both widely distributed across the Northern hemisphere. Their intraspecific genetic structuring has been extensively investigated, albeit with a focus on genetic markers from maternally inherited parts of their genomes (mitochondrial DNA). The evolutionary relationship and divergence time between brown and polar bears have recently triggered an extensive debate, while less focus has been put on to other parts of the ursid phylogeny, particularly to a clade of three Asian bear species. To date, whole genomes of more than 100 bear individuals from four different species have been sequenced. Yet, one fundamental part of the genome has been largely omitted from specific analyses, in bears as well as in most other mammals: the Y chromosome.

The mammalian Y chromosome provides a unique perspective on the evolutionary history of organisms due to its distinct features, and specifically reflects the patriline because of its male-specific inheritance. The characteristics of this chromosome make it well suited to complement and contrast evolutionary inferences based on other genetic markers, and to uncover processes like sex-biased gene flow and hybridization. The unique insights that can be gained from analyses of Y-linked genetic variation made me utilize this part of the genome to investigate the evolution of male lineages in bears. Studying the patriline is particularly promising in this taxonomic group because of male-biased dispersal and a complex and fast radiation of bears. The analysis of Y-chromosomal genetic markers is thus the common theme of this dissertation: I present the identification of large amounts of Y-chromosomal sequence, the development of male-specific markers from such sequences, and the application of these markers to trace the evolution of male lineages of different bear species.

Specifically, I developed a molecular sex determination system based on the detection of two Y-linked fragments that allows to reliably discriminate between females and males from seven different bear species (Bidon *et al.* 2013). The approach is highly sensitive, bear-specific, and can be applied in standard molecular

laboratories. This makes it valuable in conservation genetics and forensic applications, e.g. to analyze non-invasively collected samples.

Furthermore, I used Y-linked markers in a comprehensive and range-wide sample of brown and polar bears, and show that male-biased gene flow plays an important role in distributing genetic material throughout the ranges of both species (Bidon *et al.* 2014). In brown bears, I detected a lack of paternal population structuring which is in strong contrast to the detailed structuring of the matriline.

Analyzing Y-chromosomal sequences from all eight bear species, I present a phylogeny of the patriline that largely resembles the topology from other nuclear markers but is different from the topology of the mitochondrial gene tree (Kutschera *et al.* 2014). This discordance among loci generates interesting hypotheses about inter-species gene flow, particularly among American and Asiatic black bears.

With the identification of almost two million basepairs of Y-chromosomal sequence and the analysis of an unprecedented large male-specific dataset in polar bears, a high-resolution view on the distribution of their intraspecific variation was obtained (Bidon *et al*. 2015). In particular, two clades that are divergent but do not show pronounced phylogeographic structure were detected, confirming the great dispersal capacity of males of this high arctic species.

This dissertation thus represents a comprehensive investigation of Y-linked genetic variation on the intra- and interspecific level in a non-model organism. With my research, I contribute to an increased understanding of the complex evolutionary history of bears. In particular, I show that male-biased gene flow strongly influences the distribution of nuclear genetic variation, and that the contrast between phylogenies of differentially inherited markers can help to understand interspecific hybridization between closely related species. Moreover, my findings demonstrate the potential of Y-chromosomal markers to uncover unknown evolutionary patterns and processes. This applies not only to bears but to many species, even such that are generally well known and well described.

Zusammenfassung

Untersuchung der Evolution von männlichen Erblinien in Bären mithilfe genetischer Marker auf dem Y Chromosom

Hintergrund

Das Y Chromosom ist ein wesentlicher Bestandteil des Säugetiergenoms, und kann als männchen-spezifischer genetischer Marker einzigartige und interessante Einblicke in die Evolutionsgeschichte von Organismen liefern. Es hat einige besondere Eigenschaften, die es von anderen Teilen des Genoms unterscheiden. Dazu gehören unter anderem die Vererbung von Vater zu Sohn (paternale Vererbung), sowie das weitgehende Fehlen von interchromosomaler Rekombination (Jobling *et al.* 2014). Evolutive Analysen beruhen meist auf der Analyse von mütterlicherseits (maternal) vererbter DNA der Mitochondrien oder auf Erbinformation der Autosomen des Zellgenoms. Letztere werden von beiden Geschlechtern an die Nachkommen weitergeben. Da unterschiedlich vererbte Teile des Genoms unterschiedliche Muster und Prozesse abbilden können, bleibt etwa der durch die Wanderung von Männchen vermittelte Genfluss durch die Analyse von mitochondrieller DNA unerkannt. Die kombinierte Analyse unterschiedlicher Marker ermöglicht es nun, potentielle Kontraste zwischen ihnen, z. B. die Abbildung unterschiedlicher Verwandtschaftsbeziehungen (Topologien der Stammbäume), darzustellen. Es sind diese Kontraste, die uns z. B. auf Hybridisierung zwischen Arten schließen lassen, und damit unser Verständnis für evolutive Muster und Prozesse erweitern können (Fahey *et al.* 2014).

Das Y Chromosom ist bisher nur in sehr wenigen Arten, hauptsächlich in Modellorganismen wie Mensch und Maus, ausgiebig sequenziert und charakterisiert worden. Auch liegen nur für wenige Arten evolutive Analysen basierend auf diesem genetischen Lokus vor. Die Ursachen dafür sind in den spezifischen Eigenschaften des Y Chromosoms zu finden, z.B. in der großen Zahl sich wiederholender (repetitiver) Sequenzabschnitte, oder der geringen Dichte an Protein-kodierenden Genen. Deshalb wurde das Y Chromosom auch in vielen Studien, die genomweite Daten erheben, entweder gar nicht sequenziert (indem Weibchen untersucht wurden) oder zumindest nicht spezifisch analysiert (Greminger *et al.* 2010).

Eine Gruppe innerhalb der Säugetiere, für die das Y Chromosom als genetischer Marker vielversprechend ist, ist die Familie der Bären (Ursidae). Es gibt heute acht Bärenarten, die auf vier verschiedenen Kontinenten vorkommen (McLellan & Reiner 1994). Unter ihnen sind die gut untersuchten und bekannten Braun- und Eisbären und der Große Panda. Die Evolutiongeschichte und der Zeitpunkt der Artaufspaltung von Braun- und Eisbären hat in den letzten Jahren eine intensive Debatte erlebt (z. B. Hailer *et al.* 2012, Liu *et al*. 2014). Es gibt aber auch Arten, deren Evolutionsgeschichte weniger gut untersucht ist, etwa drei nahverwandte Bärenarten Asiens. Die evolutive Entwicklung der Bären fand innerhalb einer relativ kurzen Zeit von wenigen Millionen Jahren statt, weshalb Hybridisierung und andere evolutive Prozesse die Entschlüsselung ihrer Verwandtschaftsverhältnisse entscheidend beeinflussen können. Die Analyse unterschiedlich vererbter genetischer Marker kann hier wertvolle Einblicke liefern. Geschlechtsspezifische Unterschiede im Ausbreitungsverhalten, wie sie für Braunbären gut dokumentiert sind (McLellan & Hovey 2001), lassen außerdem Unterschiede in der Populationsstruktur basierend auf maternal bzw. paternal vererbten Markern erwarten. Trotzdem wurden Marker auf dem Y Chromosom bisher kaum verwendet um phylogenetische bzw. phylogeographische Muster und Prozesse in Bären zu untersuchen.

Durchgeführte Studien

In der vorliegenden Dissertation präsentiere ich die Identifizierung von fast zwei Millionen Basenpaaren Y-chromosomaler DNA-Sequenz im Eisbären, die Verwendung solcher Sequenzen um Männchen-spezifische genetische Marker für verschiedene Bärenarten zu entwickeln, und die Anwendung dieser Marker, um die männlichen Erblinien von Bären zu untersuchen. Darüber hinaus zeige ich, dass das Y Chromosom ein wichtiges Werkzeug in der Forensik und Naturschutzgenetik ist.

So beschreibe ich die Entwicklung eines zuverlässigen Tests für die molekulare Geschlechtsbestimmung bei sieben Bärenarten (Bidon *et al*. 2014). Der Test basiert auf der PCR-Amplifikation von zwei DNA-Fragmenten des Y Chromosoms, und einem DNA-Fragment des X Chromosoms und deren Nachweis mittels Gelelektrophorese. Dadurch können Weibchen und Männchen zuverlässig voneinander unterschieden werden. Durch die geringe Länge der zu amplifizierenden

Fragmente und der hohen Sensitivität eignet sich die Methode auch für nicht-invasiv gesammeltes Probenmaterial, wie z. B. Kot und Haare, das sich oft durch geringe DNA Quantität und Qualität auszeichnet. Der Test kann in molekularbiologischen Standardlaboren leicht angewendet werden und ist daher vielfältig einsetzbar. Darüber hinaus ist er spezifisch für Bären, sodass bei Proben, die mit DNA einer anderen Art kontaminiert sind, das Ergebnis nicht verfälscht werden kann.

Des Weiteren habe ich in meiner Arbeit die innerartliche Populationsstruktur der männlichen Erblinien von Braun- und Eisbären über deren gesamte Verbreitungsgebiete hinweg untersucht und mit dem bekannten Muster der weiblichen Erblinien der mitochondriellen DNA verglichen (Bidon *et al*. 2013). Hierzu wurden mit Hilfe des Eisbärgenoms Sequenz- und Mikrosatellitenmarker auf dem Y Chromosom entwickelt. Bevor die Marker in 130 Individuen angewendet wurden, wurde deren Y-chromosomaler Ursprung durch den Nachweis Männchen-spezifischer Amplifikation überprüft. Das ausgeprägte Ausbreitungsverhalten männlicher Braunbären kann durch Y-spezifische Marker abgebildet werden. Es war deshalb zu erwarten, dass sich die paternale Populationsstruktur von dem Muster mütterlicherseits vererbter, mitochondrieller DNA, unterscheidet. Tatsächlich ist dieser Unterschied so stark ausgeprägt, dass identische Y Chromosomen in geographisch weit voneinander entfernten Gebieten vorkamen. Dies stellt einen deutlichen Kontrast zur maternalen Populationsstruktur von Braunbären dar, die durch geographisch klar getrennte und nicht überlappende Gruppen charakterisiert ist (Davison *et al*. 2011). Eine Sequenzvariante (Haplotyp) des Y Chromosoms war besonders häufig und kam über das gesamte Verbreitungsgebiet hinweg vor. Analysen basierend auf Y-chromosomalen Mikrosatellitendaten zeigten eine größere Variabilität, die jedoch keine geographische Struktur widerspiegelte. Auch in Eisbären war ein Y-chromosomaler Haplotyp besonders häufig und über große geographische Distanzen verbreitet. Der Unterschied zur mitochondriellen Struktur war jedoch weniger stark ausgeprägt, was das beträchtliche Ausbreitungspotential beider Geschlechter dieser Art veranschaulicht. Zusammenfassend konnte gezeigt werden, dass Männchen beider Bärenarten ihr genetisches Material über große räumliche Distanzen verbreiten, und dass dieser wichtige Prozess in der Betrachtung der Evolutionsgeschichte berücksichtigt werden muss.

In meiner Dissertation habe ich neben der innerartlichen Variation auch die zwischenartliche Variation des Y Chromosoms untersucht. Nicht alle

Verwandtschaftsbeziehungen der acht heute lebenden Bärenarten sind eindeutig geklärt (Pagès *et al*. 2008, Nakagome *et al*. 2008). Auf Grund von Diskrepanzen zwischen mitochondriellen und nukleären Stammbaumtopologien, verspricht die Analyse des Y Chromosoms hier aufschlussreiche Einblicke in potentielle Hybridisierung zwischen verschiedenen Bärenarten. Ich konnte zeigen, dass die Topologie des auf Y-chromosomalen Daten basierenden Stammbaums weitgehend der Topologie des auf mehreren autosomalen Genorten basierenden Artenbaums entspricht (Kutschera *et al*. 2014). Besonders interessant ist hierbei die Position des Amerikanischen Schwarzbären. Während dieser in der Y-chromosomalen Topologie ein Geschwistertaxon zu Braun- und Eisbären bildet, bildet er in der mitochondriellen Topologie ein Geschwistertaxon zu Asiatischen Schwarzbären (Yu *et al*. 2007, Krause *et al*. 2008). Genfluss und Introgression von mitochondrieller DNA von Asiatischen in Amerikanische Schwarzbären könnte diese Diskrepanz zwischen maternaler und paternaler Topologie erklären. Jedoch bleibt zu klären wo diese Hybridisierung stattgefunden haben könnte, da Fossilien, die eine geographische Überlappung dieser beider Arten belegen könnten, bisher fehlen (McLellan & Reiner 1994).

Um die Analysen des Y Chromosoms auf eine größere Datengrundlage stellen zu können, ist es wichtig weitere Sequenzen dieses Chromosoms im Genom zu identifizieren. Zwar gibt es ein Referenzgenom eines männlichen Eisbären (Li *et al*. 2011), jedoch wurden die Sequenzfragmente (Scaffolds) bisher keinen bestimmten Chromosomen zugeordnet. Mit der Suche nach bekannten Y-chromosomalen Genen aus anderen Säugetieren konnten Scaffolds mit Y-chromsomalem Ursprung identifiziert werden (Bidon *et al*. 2013, 2014, 2015, Kutschera et al. 2014). Mit Hilfe eines Vergleiches der relativen Abdeckung (Coverage) von Sequenzdaten eines weiteren Männchens und eines Weibchens (Miller *et al*. 2012) im Referenzgenom mit erwarteten Coverage-Werten, konnte ein Großteil der Scaffolds einer bestimmten chromosomalen Kategorie (autosomal, X-chromosomal, Y-chromosomal) zugeordnet werden (Bidon *et al*. 2015). So gelang es, fast zwei Millionen Basenpaare DNA-Sequenz auf dem Y Chromosom zu identifizieren. Damit konnten die intraspezifische Populationsstruktur des Eisbären mit sehr hoher Auflösung untersucht, und jedem Individuum spezifische Substitutionen zugeordnet werden. Es zeigten sich außerdem zwei divergente Y-chromosomale Gruppen, die jeweils Individuen aus unterschiedlichen Regionen

(Alaska und Spitzbergen) umfassen und damit keine deutliche geographische Struktur abbilden.

FAZIT

Die Verwendung des Y Chromosoms für die Untersuchung evolutiver Fragestellungen kann Signale liefern, die sich stark vom bekannten Muster anderer Marker unterscheiden. Diese Kontraste zwischen unterschiedlich vererbten Genorten liefern uns wertvolle Einblicke in die Evolutionsgeschichte von Organismen. Diskrepanzen in der Populationsstruktur etwa können auf geschlechtsspezifischen innerartlichen Genfluss hindeuten. Mütterlicherseits vererbte mitochondrielle Marker, wie sie in der Phylogeographie traditionell häufig verwendet wurden (Garrick *et al*. 2015), können die Populationsstruktur einer Art überschätzen, wenn, wie bei vielen Säugetieren der Fall, sich Männchen weiter ausbreiten als Weibchen (Greenwood 1980). Die Unterschiede zwischen Topologien von Genbäumen, die auf verschiedenen Markern basieren, erlauben uns außerdem Rückschlüsse über Hybridisierung zwischen Arten zu ziehen.

Auf Bären bezogen bedeuten die Ergebnisse meiner Dissertation, dass männlicher Genfluss die räumliche Verbreitung von genetischem Material stark beeinflusst. Es konnte sowohl für Braun- als auch für Eisbären gezeigt werden, dass Y Chromosomen über große geographische Distanzen verbreitet wurden. Des Weiteren zeigen der Vergleich der Stammbaumtopologien unterschiedlich vererbter Genorte und die dabei deutlich werdenden Diskrepanzen, dass Hybridisierung zwischen Bärenarten einen bedeutsamen Prozess in deren Evolutionsgeschichte darstellt.

Damit ist diese Arbeit eine umfassende Untersuchung der Variation des Y Chromosoms in einem Nicht-Modell-Organismus auf inner- und zwischenartlicher Ebene. Auch machen meine Studien deutlich, dass das Y Chromosom als genetischer Marker einen besonderen Blickwinkel auf evolutive Muster und Prozesse ermöglicht und für die molekulare Geschlechtsbestimmung unerlässlich ist. Die Untersuchung dieses Teil des Genoms trägt damit entscheidend dazu bei, unser Verständnis der Evolution zu erweitern. Die Generierung immer längerer Y-chromosomaler Sequenzen wird evolutive Analysen der männlichen Erblinie mit immer höherer Auflösung ermöglichen, in Bären und in anderen Säugetieren.

General introduction

The Y chromosome has distinct properties that give a unique and important view on the evolutionary history of organisms. The male perspective can complement and contrast inferences based on biparentally (autosomal) or maternally inherited loci (mitochondrial DNA). Moreover, it can reveal processes that cannot be uncovered by using established marker systems, for example male-biased gene flow and introgression. However, analyses of the Y chromosome have been neglected in most mammals, including the mammalian family of bears (Ursidae). Bears are especially interesting study organisms for Y-chromosomal markers because sex-biased dispersal and a complex and fast radiation suggest the presence of differences between differently inherited loci. The interspecific relationships as well as the intraspecific genetic structuring of different bear species are well studied based on maternally inherited markers. Therefore, the investigation of the patriline is especially promising to increase our understanding of the evolutionary history of this taxonomic group.

In my thesis, I developed a male perspective on bear evolution by identifying large-scale Y-chromosomal sequences from genome-wide sequencing data, by establishing a new bear-specific molecular sex determination approach based on the amplification of Y-specific fragments, by investigating the population structuring and the evolutionary history of paternal lineages of brown and polar bears, and by studying the phylogenetic relationship of patrilines from all extant bear species.

In the following, I will briefly outline different marker systems, explain the special features of the Y chromosome, introduce the family of bears, and discuss the development and application of male-specific markers to investigate the evolutionary history of bears.

The quest for genetic variation

The study of genetic variation and its distribution among and within species allows investigating the evolution and natural history of organisms. The first genetic polymorphisms to be identified at the beginning of the 20th century were those among human individuals with different blood groups (Landsteiner 1900). In the following

decades, immunological and sequence comparisons of proteins as well as protein electrophoresis were used to investigate genetic diversity (Avise 1994). High but variable amounts of genetic variation were found among human (Harris 1966) and natural populations (Lewontin & Hubby 1966). The first detection of variation at the DNA level became possible with restriction enzymes and the analysis of the resulting fragments (restriction fragment length polymorphism - RFLP) (Potter *et al.* 1975). With the development of DNA sequencing by Sanger (Sanger *et al.* 1977) and the polymerase chain reaction (PCR) in 1986 (Mullis *et al.* 1986), direct comparisons of the number of nucleotide substitutions among different DNA sequences became feasible for the first time.

One molecular marker has since been used intensively in evolutionary genetic studies: maternally inherited mitochondrial (mt)DNA, organelle DNA of the mitochondria. It occurs in high copy number per cell, facilitating analyses from low quality or low quantity samples. The mutation rate is higher compared to the rest of the genome, allowing for the efficient detection of variation. MtDNA is haploid and has a smaller effective population size compared to autosomes, leading to a higher rate of genetic drift and a faster sorting of lineages (Wilson *et al.* 1985; Avise 1986). The first studies that investigated mtDNA variation used RFLP (Avise *et al.* 1979), and a complete human mitochondrial genome sequence was already published in 1981 (Anderson *et al.* 1981). Its high mutation rate and the development of conserved primers for PCR amplification enabled evolutionary analyses of a wide range of different taxa early on (Kocher *et al.* 1989; Janke *et al.* 1994; Ballard & Whitlock 2004).

Only with the possibility to analyze DNA variation from biparentally inherited autosomes several independent estimates of phylogenetic relationships among taxa (Pamilo & Nei 1988), and a more complete understanding of past population processes, e.g. population growth (Brumfield *et al.* 2003), could be obtained. Moreover, multi-locus analyses can provide genome-wide estimates of the average species tree and of divergence times (Rannala & Yang 2003). Different types of autosomal markers can provide different levels of resolution to infer evolutionary relationships at different timescales, e.g. fast evolving microsatellites (Tautz & Schlötterer 1994, Jarne & Lagoda 1996), putatively neutrally evolving intron sequences (Matthee *et al.* 2007), or conserved protein-coding genes (Hallström & Janke 2010). Advances in sequencing technology allow us to generate sequences of

an increasing number of loci across entire genomes with ever decreasing cost and time (Pareek *et al.* 2011).

Evidence for the existence of a male-specific chromosome appeared in the beginning of the 20th century (Painter 1921). In the following decades, the debate circled around specifically Y-linked traits (Stern 1957), how sex chromosomes evolved (Ohno 1969), and how many genes actually remain on the Y chromosome (see below). Also, one important Y-linked gene, the sex-determining factor, *SRY*, was identified (Sinclair *et al.* 1990). The Y chromosome was then also used as a source for genetic variation to study the evolution of human (Casanova *et al.* 1985; Malaspina *et al.* 1990) and natural populations (Sundqvist *et al.* 2001; Greminger *et al.* 2010). Indeed, the Y chromosome continues to be vitally important for human population genetics, evidenced by multiple recent publications (e.g., Wei *et al.* 2013; Scozzari *et al.* 2014; Hallast *et al.* 2014; Karmin et al. 2015). Genetic variation of many members of the bear family has been extensively investigated above and below the species level (e.g., Yu *et al.* 2004; Davison *et al.* 2011; Zhao *et al.* 2013; Cronin *et al.* 2014). However, the particular male perspective is largely missing in analyses of ursids.

Characteristics and structure of the Y chromosome

The male perspective

Not all parts of the genome are inherited in an analogous manner. While mitochondrial DNA occurs in the cells of both males and females, it is typically maternally inherited, i.e. only females pass on their mtDNA to the next generation (Jobling *et al.* 2014). In contrast, both parents pass on one copy of each of their autosomes. The Y chromosome, one of the two sex chromosomes of a typical mammalian genome, is unique among other parts of the genome because it is uniparentally inherited from father to son and never occurs in females (Jobling *et al.* 2014). Successive generations of male individuals can thus be related by an easily interpretable phylogeny of the patriline (see **Figure 1** and below).

Their differential inheritance renders markers from different genomic locations informative for distinct scientific questions and inferences. MtDNA specifically represents one part of the evolutionary history of a species, the female lineage. Evolutionary patterns inferred from mtDNA can be affected by demographic

asymmetries, like sex-biased gene flow and hybridization (Toews & Brelsford 2012; Fahey *et al.* 2014). As sex-specific migration differentially influences variation and geographic patterns on mtDNA and the Y chromosome, the comparison of these differentially inherited markers can give unique insights into such processes.

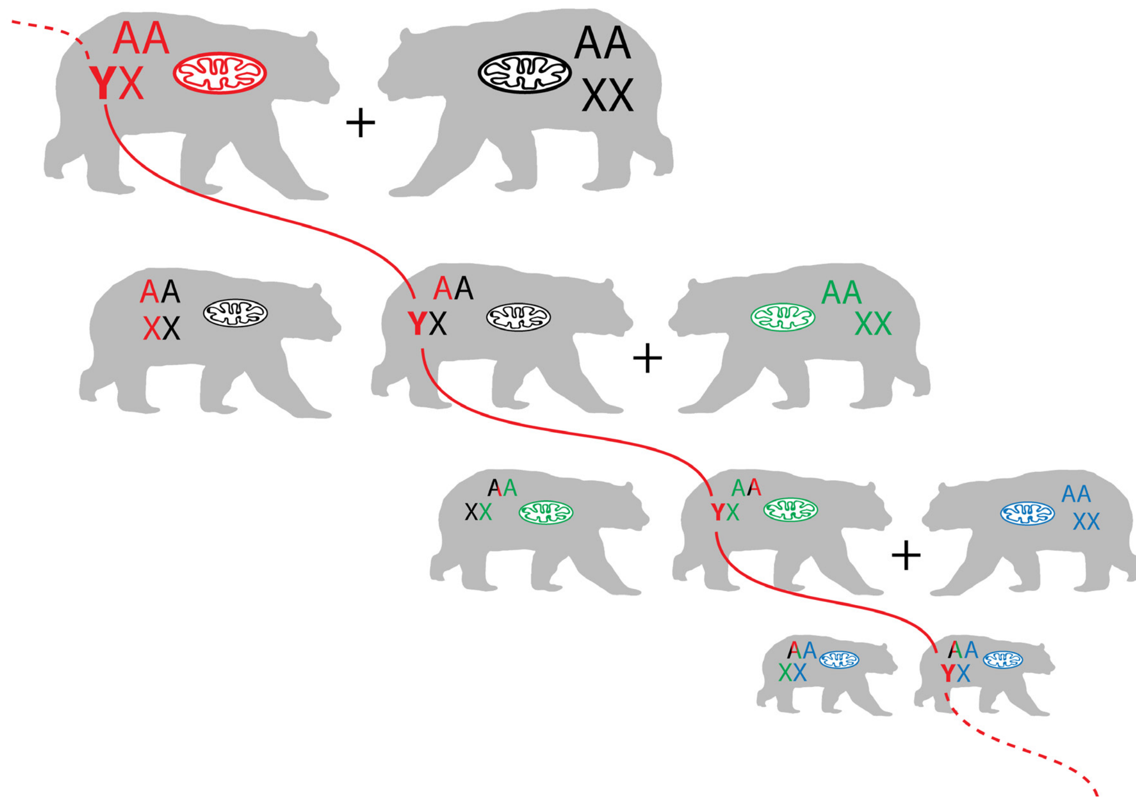


Figure 1. Schematic representation of the differential inheritance of different parts of the genome from one generation to the next. The red line indicates the patriline that relates Y chromosomes of successive generations. Colors represent genetic differences between individuals. Y: Y chromosome, X: X chromosome, A: autosome. mtDNA is illustrated by the drawing of a mitochondrium. Note that due to reshuffling by recombination the inheritance of autosomes and X chromosomes is more complex than shown.

The focus on maternally inherited mtDNA alone can thus result in biased conclusions if drawn for the species as a whole. Although the Y chromosome also constitutes one locus that, in this case, specifically represents the paternal lineage, its male-specific inheritance makes it well suited to contrast and complement markers that are biparentally or maternally inherited. Therefore, the Y chromosome provides a unique view on organismal evolution, essential to identify contrasting patterns of male and female gene flow and hybridization.

Additionally, the male perspective provided by the Y chromosome and its specific occurrence in male individuals play a major role in forensic work, e.g. for

paternity testing (Kayser 2007) and sex determination (Cadamuro *et al.* 2015), and in medical research, e.g. in studies of male infertility (Repping *et al.* 2002).

ITS SPECIAL FEATURES RENDER THE Y CHROMOSOME A VALUABLE MARKER IN EVOLUTIONARY STUDIES

The male-specific Y chromosome is only present in males, and only in one copy per cell, its homologous chromosome being the X chromosome. The largest part of the human Y chromosome, roughly 95% of its entire length of ~60 Mb, came to be known as the non-recombining region (NRY), and is nowadays recognized as the male-specific region of the Y chromosome (MSY; **Figure 2**; Skaletsky *et al.* 2003). This region does not undergo recombination with the X chromosome, is clonally passed on from father to son, and is truly haploid. The remaining ~3 Mb are homologous to and can pair with the X chromosome during meiosis. They are called pseudo-autosomal regions (PAR; Freije *et al.* 1992).

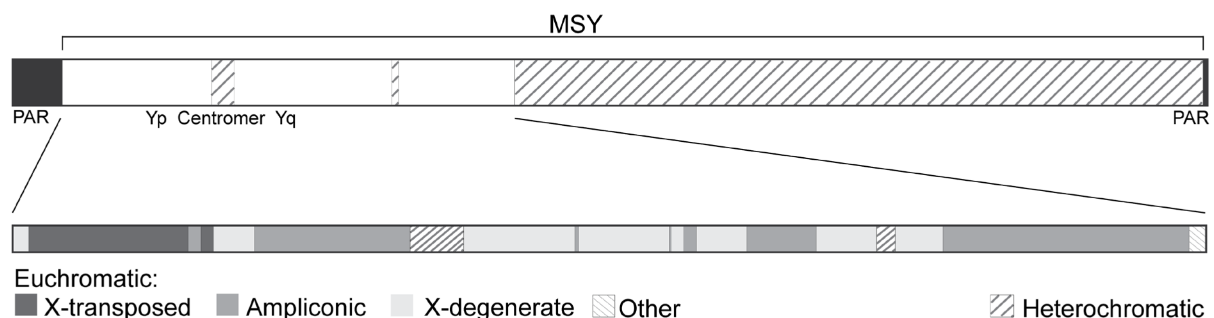


Figure 2. Schematic representation of the human Y chromosome. The euchromatic part is shown enlarged, with different sequence classes illustrated. MSY: male-specific region of the Y, PAR: pseudo-autosomal regions (recombining with the X chromosome), Yp: short arm, Yq: long arm of the chromosome. Modified from Skaletsky *et al.* 2003.

Due to the lack of recombination over most of its length, mutations sequentially accumulate on the Y chromosome, and historical patterns are maintained in successive generations. Haplotypes, i.e. the particular combination of Y-linked alleles in one individual, are passed on from generation to generation and can be related by a single phylogeny – the patriline (**Figure 1**). This is contrary to biparentally inherited autosomal markers. Here, tracing back single maternal or paternal lineages is

complicated by the concealing effect of recombination, i.e. the reshuffling of chromosomes in every generation (Posada *et al.* 2002).

Without recombination, changes on the Y chromosome are introduced only by mutation, which then is the only evolutionary force increasing Y-chromosomal DNA variation (Greminger *et al.* 2010). Although the Y-chromosomal substitution rate is higher compared to autosomes (more germ cell divisions per generation) (Li *et al.* 2002), the intraspecific sequence diversity on the Y chromosome is relatively low (Jobling & Tyler-Smith 1995; Charlesworth & Charlesworth 2000; Hellborg & Ellegren 2004). This is due to the smaller effective population size (Ne) compared to autosomes (one-quarter, because the Y chromosome only occurs in males, and only in one copy), which makes the Y chromosome more susceptible to genetic drift, the random sampling of chromosomes from one generation to the next. This leads to a more recent common ancestor of Y lineages and thus there is less time to accumulate diversity (Jobling *et al.* 2014). Moreover, selection acting on any part of this single locus removes variability on the entire chromosome because all sites are linked. This can happen either through background selection or through selective sweeps (Charlesworth & Charlesworth 2000; Petit *et al.* 2002): Purifying selection is strong on the Y chromosome due to its role in male fitness (Hurles & Jobling 2001; Wilson Sayres *et al.* 2014), and diversity can thus be removed by background selection. Moreover, a strongly selected beneficial mutation with a net benefit over all the linked deleterious mutations could reach fixation, dragging with it all linked mutations in a selective sweep. Thereby diversity would be reduced at any linked neutral site (Charlesworth 1996).

While the low effective population size implies low diversity *within* a lineage, it leads to an increased level of differentiation *among* lineages. Drift randomly changes the frequencies of different Y chromosomes in each lineage, thereby establishing differences among the lineages. Geographic structure is thus quickly generated, rendering the Y chromosome well suited for phylogeographic inferences. Very long haplotypes can be inferred without having to consider the complex reshuffling processes during meiosis because the Y chromosome is the longest non-recombining fraction of the entire genome, three orders of magnitude larger than mitochondrial DNA (~60 megabases (mb) versus ~16 kilobases (kb) in humans; Jobling *et al.* 2014). This can translate into a very detailed phylogeny with many

branch-specific substitutions. In that way, the Y chromosome can be a rich source of polymorphisms and can provide high resolution for evolutionary inferences.

The Y chromosome and its comparison to other marker systems are indispensible strategies to recover signatures of male-specific gene flow and hybridization. Both processes are relevant to understand the evolutionary history of bears. While brown and polar bears have been important study organisms in phylogeography due to their wide distribution across the Northern hemisphere, the availability of samples and the general interest in understanding population structuring in large-bodied mammals, no phylogeographic study using male-specific markers had been conducted. Indeed, all range-wide phylogeographic investigations of brown bears have so far relied on mtDNA (Davison *et al.* 2011). Moreover, the fast radiation of several bear species implies that interspecific hybridization has shaped their genomes and influences phylogenetic reconstruction. Analysis of the Y chromosome gene tree topology therefore promises interesting insights into the evolutionary history of bears.

SEX CHROMOSOMES EVOLVED FROM AUTOSOMES

The specific characteristics of the Y chromosome are a consequence of its evolutionary history. Sex chromosomes of many species have originated from a pair of homologous ancestral chromosomes, and have done so independently in many evolutionary lineages. Today, sex chromosomes occur in mammals, birds, reptiles, amphibians, fish and many invertebrate taxa (Bachtrog *et al.* 2014), and also in plants (Ming *et al.* 2011). Male heterogamety and female homogamety is the standard condition in mammals, with males having one X and one Y, and females two X chromosomes (Charlesworth 1991; Ellegren 2011), but other systems exist (e.g., X0-system, Bachtrog *et al*. 2014).

Starting from a pair of autosomes, three steps were involved in the evolution of distinct (heteromorphic) sex chromosomes in humans. First, one of the two ancestral chromosomes acquired a mutation that established a sex-determining factor, such as a gene that controls development of male sex (Foster & Graves 1994). Second, recombination was suppressed between these two chromosomes, probably enhanced by mutations that were beneficial for one sex but disadvantageous for the other (Rice 1987). Linkage between those mutations and the

sex-determining gene was selected for, and recombination that disrupted this linkage was suppressed (Charlesworth *et al.* 1991). Furthermore, multiple successive inversion events in the Y chromosome led to a cessation in recombination with the X chromosome of a discrete region at a particular point in time. Five such regions, called evolutionary strata, have been identified in humans (Lahn & Page 1999; Ross *et al.* 2005). Once a region on the Y chromosome was isolated from recombination with the X chromosome, it was subject to degeneration – the third step involved in sex chromosome evolution.

Y chromosomes undergo a process of genetic erosion, losing activity of most genes that were present in their ancestors. On the human Y chromosome only 78 protein-coding genes remain (Skaletsky *et al.* 2003), compared to more than 800 protein-coding genes on the X chromosome (Ross *et al.* 2005). Thus, only few X-linked genes have an active homolog on the Y chromosome. This degeneration is one of the most striking features common to sex chromosomes among different taxa, although other modes of sex chromosome evolution exist (Carvalho *et al.* 2009).

Several mechanisms have been proposed to explain Y-chromosomal degeneration (Charlesworth & Charlesworth 2000). In the absence of recombination, those Y chromosomes that carry the least amount of deleterious mutations cannot be replaced once lost. This process is called Muller´s ratchet (Muller 1964), and leads to the successive accumulation of deleterious mutations on the Y chromosome. Selection is another force possibly contributing to degeneration and gene loss, because it would always affect the entire chromosome due to the linkage of all sites. However, it has been difficult to gather evidence for the action of positive selection on the human and *Drosophila* Y chromosomes (Hurles & Jobling 2001; Wilson Sayres *et al.* 2014; Singh *et al.* 2014), likely because it is difficult to disentangle it from purifying selection and from demography (Fahey *et al.* 2014).

The degeneration of Y-linked genes has led to a rather dreary picture of this chromosome, with just a few genes and one function left: sex determination. Thus, the Y chromosome was assumed to face its complete disappearance (Marshall Graves 2000). We now know, however, that it contains distinct sequence classes and conserved genes with important biological functions (Bellott *et al.* 2014). Moreover, the Y chromosome has acquired new sequences during its evolutionary history and is thus more than merely a degenerate version of the X chromosome (Hawley 2003; Skaletsky *et al.* 2003).

Y CHROMOSOMES ARE RICH IN REPETITIVE SEQUENCES

Most of what we know about the architecture of the Y chromosome comes from a few well-studied organisms. The most completely sequenced, assembled, and annotated Y chromosomes are those of three primate species, human (Skaletsky *et al.* 2003), chimpanzee (Hughes *et al.* 2010), and rhesus macaque (Hughes *et al.* 2012), as well as the mouse (Soh *et al*. 2014). Detailed analyses are also available from the Y chromosomes of cat and dog (Li *et al*. 2013), from marmoset, rat, bull and opossum (Bellott *et al*. 2014), from the largely heterochromatic Y chromosome of drosophila (Carvalho *et al*. 2009), and from the male-specific region of the very young medaka fish Y chromosome (Kondo *et al.* 2006). One of the reasons for the paucity of Y-chromosomal sequence information in many species is the repetitive nature of the Y chromosome that impedes the correct assembly of short reads.

The human Y chromosome consists of transcriptionally active euchromatic (roughly 23 Mb) and highly condensed and repetitive heterochromatic DNA sequences (roughly 40 Mb). While the Y chromosomes of some species contain less heterochromatin than the human Y chromosome (e.g. mouse, chimpanzee), even the euchromatic parts are highly enriched in repeated (ampliconic) sequence elements (Hughes *et al.* 2010; Soh *et al.* 2014). The euchromatic portion of the human Y chromosome contains all 156 Y-linked transcription units that have been identified, including 78 protein-coding genes. These genes fall into two categories: one group is expressed ubiquitously, in many tissues throughout the body, the other is expressed specifically in testes (Lahn & Page 1997).

The euchromatic parts can further be divided into three discrete sequence classes: X-transposed, X-degenerate, and ampliconic (**Figure 2**; Skaletsky *et al.* 2003). The X-transposed sequences contain two genes and are 99% identical to their X-linked homologs, but are not involved in crossing-over during meiosis. The X-degenerate sequences are remains of the ancient autosomes from which X and Y chromosomes originate. This sequence class encodes 16 different proteins, among them all 12 ubiquitously expressed Y-linked genes. The ampliconic sequences consist of long repeat units that are extremely similar to each other. This sequence class contains the highest density of genes, often occurring in several copies, and those genes are expressed predominantly in testes. Eight gene-rich large inverted repeats (palindromes) exhibit directional gene conversion activity, thereby

maintaining >99.9% similarity between the two arms of a palindrome. This protects the genes in these regions, many of them important for spermatogenesis, from decay. Thus, intra-chromosomal recombination between arms of palindromes, and also between other Y-linked paralogous sequences, excludes some regions from degeneration (Bosch *et al.* 2004; Hallast *et al.* 2013). Although degeneration is the general mode of Y chromosome evolution, gene acquisition, and gene conservation play important roles, too.

STUDY ORGANISMS

PHYLOGENETIC RELATIONSHIP OF BEARS

Bears are a mammalian family (Ursidae, order: Carnivora), comprising eight extant species: giant panda (*Ailuropoda melanoleuca*, David 1869), spectacled bear (*Tremarctos ornatus*, Cuvier 1825), sun bear (*Helarctos malayanus*, Raffles 1821), sloth bear (*Melursus ursinus*, Shaw 1791), Asiatic black bear (*Ursus thibetanus*, Cuvier 1823), American black bear (*U. americanus*, Pallas 1780), brown bear (*U. arctos*, Linnaeus 1758), and polar bear (*U. maritimus*, Phipps 1774), the latter six belonging to the subfamily Ursinae (McLellan & Reiner 1994). Bears occur in North America, South America, Europe and Asia, occupying many different habitats, from the high arctic to lowland tropical forests (**Figure 3**). They are iconic, large-bodied mammals, with symbolic importance in many human cultures. Where food is not available through some parts of the year, bears hibernate. Their feeding ecology ranges from entirely carnivorous to omnivorous to almost completely herbivorous (Sacco & Van Valkenburgh 2004). Six of eight ursids are considered vulnerable or endangered by the International Union for the Conservation of Nature (IUCN).

Genetic variation among different bear species was first investigated in the 1970s and 1980s, employing immunological approaches and protein electrophoresis (Sarich 1973; O´Brien *et al.* 1985; Goldman *et al.* 1989). These studies focused on the riddle in which carnivore family to place the giant panda, which is now known to be the first-diverging lineage of extant ursids (Waits *et al.* 1999), its divergence followed by that of the spectacled bear (Wayne *et al.* 1991; Zhang & Ryder 1994). With restriction fragment analyses of mtDNA and sequencing of mtDNA control region and the *Cytb* gene, a consensus was also reached on the close relationship of the two most recently evolved ursine bears, brown and polar bears (Cronin *et al.*

1991; Shields & Kocher 1991). It also became clear that polar bears are nested within brown bear variation on the mt locus (Talbot & Shields 1996b). However, the exact branching pattern of the remaining four ursines remained ambiguous, despite a number of studies analyzing mtDNA sequence variation in the 1990s (Zhang & Ryder 1994; Talbot & Shields 1996a; Waits *et al.* 1999), and nuclear exon and intron sequences later on (Yu *et al.* 2004).

Figure 3. Approximate extent of the distribution ranges of eight bear species. Note that the current ranges of sun bear, sloth bear, and Asiatic black bear are fragmented and only spots remain within the area indicated here (Fredriksson *et al.* 2008; Garshelis & Steinmetz 2008; Garshelis *et al.* 2008a; b; Goldstein *et al.* 2008; Lü *et al.* 2008; McLellan *et al.* 2008; Schliebe *et al.* 2008).

With the analysis of the complete mitochondrial genomes from all eight bear species, strong support was finally achieved for the matrilineal relationships among most ursine species (Yu *et al.* 2007; Krause *et al.* 2008): An early divergence of the sloth bear, and a placement of the sun bear as sister taxon to a clade comprising American and Asiatic black bear was suggested. In contrast, signals from nuclear autosomal, X-chromosomal and Y-chromosomal genes placed the American black bear as sister taxon to brown and polar bears (and not as sister to the Asiatic black bear), but could not resolve the exact placement of sun, sloth and Asiatic black bear in the phylogeny (Yu *et al.* 2004; Pagès *et al.* 2008; Nakagome *et al.* 2008). The four Y-linked genes sequenced by Pagès *et al.* (2008) and Nakagome *et al.* (2008) had been the only Y-sequences analyzed in bears.

Recent studies using nuclear intron sequences and genomic data (Hailer *et al.* 2012; Miller *et al.* 2012; Cahill *et al.* 2013, 2015; Cronin *et al.* 2014; Liu *et al.* 2014)

have focused on the evolution of brown and polar bears. With this data it is now clear that they are distinct sister lineages, with brown bear paraphyly for mtDNA being likely due to mitochondrial introgression (Hailer *et al.* 2012, Miller *et al.* 2012), or genome conversion (Cahill et al. 2013, 2015), or both (Hailer 2015). The divergence time of brown and polar bears is subject of an extensive, ongoing debate: estimates based on nuclear data range from 343 thousand years ago (kya) to ca. five million years ago (mya) (Hailer *et al.* 2012, Miller *et al*. 2012, Liu *et al*. 2014, Cahill *et al*. 2013, Cronin *et al.* 2014).

This discordance between nuclear and mitochondrial markers and the ambiguity surrounding some aspects of ursine phylogeny render the male-specific part of the genome important and promise interesting insights into bear evolution.

BROWN AND POLAR BEARS HAVE DIFFERENT LEVELS OF INTRASPECIFIC VARIATION

Analyses of genetic variation shed light on phylogenetic relationships among ursid species, and allow studying the spatial structuring of the phylogenetic history of intraspecific lineages, i.e. to study phylogeographic patterns (Avise *et al.* 1987). Indeed, both brown and polar bears have attained considerable attention in this respect for more than 20 years (Cronin *et al.* 1991; Taberlet & Bouvet 1994; Paetkau *et al.* 1999; Hewitt 2000; Davison *et al.* 2011). Despite their close relatedness, they differ remarkably in many of their biological characteristics, and also in the degree to which their populations are genetically structured.

Polar bears are marine top predators, distributed circumpolar, and highly specialized to their arctic sea-ice environment (Stirling 2012). Nineteen populations are currently recognized (Schliebe *et al.* 2008), but genetic differences among them are generally small based on mtDNA (Cronin *et al.* 1991), nuclear microsatellites (Paetkau *et al.* 1999; Cronin & MacNeil 2012; Peacock *et al.* 2015), and genome-wide data (Miller *et al*. 2012, Cahill *et al.* 2013, Cronin *et al.* 14).

The brown bear is one of the world´s most widely distributed terrestrial mammals, a generalist that occupies a great diversity of habitats (McLellan *et al.* 2008). It has pronounced population structuring based on mtDNA variation, with two main maternal lineages comprising several well-separated clades on three continents (Waits *et al.* 1998; Leonard *et al.* 2000; Davison *et al.* 2011; Edwards *et al.* 2011; Hirata *et al.* 2013; Keis *et al.* 2013). Autosomal microsatellites show structuring of

brown bear populations on regional scales (Paetkau *et al.* 1997; Waits *et al.* 2000; Tammeleht *et al.* 2010; Kopatz *et al.* 2012, 2014), however, this structuring does not necessarily reflect the boundaries between mtDNA clades (Paetkau *et al.* 1997, Waits *et al.* 2000). Genome-wide data reflect the elevated structuring of brown bears, showing more intraspecific variation than polar bears (Cahill *et al.* 2013, Cronin *et al.* 2014). However, no range-wide phylogeographic analyses based on nuclear data have been conducted to date, and phylogeographic studies based on Y-specific markers are entirely missing.

THESIS OBJECTIVES

The common thread of this dissertation is the utilization of genetic markers on the Y chromosome, their identification, development and application to trace the paternal lineages of bears. The main research goal was to develop a male-specific perspective on the evolutionary history of bears that likely differs in some respects from the information provided by other markers. In particular, I expected the Y chromosome to contrast other markers due to sex-biased gene flow and hybridization among bear species. The Y chromosome is understudied in many species, including bears. This renders the approaches that were used in my thesis as unique in this taxonomic group.

Y-linked sequences were identified from whole-genome sequence data, and used as a base to develop male-specific markers. These markers were then tested for their male-specificity, i.e. they were required to only amplify in males but not in females.

Specifically, I describe the development of a new molecular sex determination method for ursine and tremarctine bears (Bidon *et al*. 2013, see Publications, Publication 1). The Y chromosome plays an essential part in such an application due to its specificity for the male lineage. The method is characterized by a dual-Y approach, where two markers from the Y chromosome and one positive control from the X chromosome are being detected to discriminate between the sexes. Reliable molecular sex determination, and thus knowledge of the sex of DNA samples, is also a prerequisite to undertake further investigations using Y-linked markers.

Moreover, paternal lineages of brown and polar bears from across their ranges were investigated using sequence information and allele sizes of

microsatellite markers from the Y chromosome (Bidon *et al*. 2014, see Publications, Publication 2). The geographic distribution of male-specific versus female-specific genetic variation was used to infer the impact of male-biased dispersal, which is a well-described process in brown bears. Moreover, the divergence time and the clear separation of the patrilines of brown and polar bears promise to add valuable information to the debate about the evolution of the polar bear lineage.

Furthermore, Y-specific sequence markers were developed and applied in all eight extant bear species in order to detect potentially contrasting topologies inferred from biparentally, maternally, and paternally inherited markers (Kutschera *et al*. 2014, see Publications, Publication 3). Discrepancies among markers can be informative about evolutionary processes, such as hybridization, which has likely shaped the genomes of closely related bear species.

In total, almost two million basepairs of Y-linked sequences were identified in a genome-wide search in polar bears (Bidon *et al*. 2015, see Appendix). The available polar bear reference assembly in combination with short read data from an additional female and male polar bear was used to reliably identify these large amounts of previously unknown Y-linked sequences. This strategy subsequently enabled the utilization of large-scale Y-chromosomal datasets that helped to reconstruct the phylogeny and phylogeography of polar bear patrilines with high resolution.

GENERAL DISCUSSION

The Y chromosome provides a unique view on the evolutionary history of organisms and can reveal processes that cannot be uncovered by the sole analysis of maternally or bi-parentally inherited parts of the genome. Nevertheless, surprisingly little is known about patterns of Y-chromosomal variation in most species, including the mammalian family of bears.

The discovery of polymorphic Y-linked markers that can specifically be amplified in males as a single locus often requires substantial effort, especially in species were genomic information is largely absent. This is aggravated as most genome sequencing projects have excluded the Y chromosome (Willard 2003; Murphy *et al.* 2006; Greminger *et al.* 2010; Li *et al.* 2013). This relates to difficulties in sequencing and assembling of highly repetitive Y-chromosomal sequences, to the low gene content of the Y chromosome (the assumption that not much is missed), or because equal coverage of X-linked and autosomal sequences is desired (Hughes & Rozen 2012; Bachtrog 2013).

The availability of an assembled genome of a male polar bear (Li *et al.* 2011), as well as genome-wide sequences from a male brown bear enabled the initial identification of scaffolds presumably originating from the Y chromosome in bears. This was achieved by searching for homologous sequences to genes known to be Y-linked in the dog (*ZFY*, *SMCY*, *RBMY*, *SRY*, *UBEY*, *EIF1AY*). Based on these genes, it was possible to extract the first five putative Y-linked scaffolds from brown and polar bears. They were validated to be male-specific and thus of Y-chromosomal origin through PCR amplification of several regions on these scaffolds in male and female bears. This was the groundwork for the development of Y-linked sequence and microsatellites markers in different bear species (Bidon *et al*. 2013, 2014, Kutschera *et al*. 2014). In total, almost two megabases of sequence data from the Y chromosome were identified and this resource was used for evolutionary analyses (Bidon *et al*. 2015).

The Y chromosome as a resource – developing male-specific markers

Molecular sex determination in conservation and forensics

The Y chromosome holds a unique role in forensic genetics, because of its specificity for the male lineage. It is essential in genotyping male individuals in criminal investigations, e.g. in cases of sexual assault. Moreover, Y-specific markers are indispensable in molecular sex identification (Bidon *et al*. 2013) and in paternity testing (Kayser 2007). For example, Y-chromosomal markers have been used in a prominent case of paternity testing, where it was shown that Thomas Jefferson, former US president, was the father of the oldest son of one of his slaves, Sally Hemings (Taylor *et al.* 1998).

A common approach in molecular sex determination in human forensics is the detection of size differences between the Y- and X-copy of the amelogenin gene. However, a rare, Y-specific deletion exists, and males can wrongly appear as females in this test (Steinlechner *et al.* 2002). Although failures are rare, legal consequences in the field of criminal investigation can be substantial. That wrong assignment is an issue even in human forensic applications exemplifies the importance of accurate and reliable molecular tools, such as the dual-Y approach (Bidon *et al*. 2013).

Knowledge about the sex of individual samples is a prerequisite to develop and apply Y-chromosomal markers for evolutionary inferences – a reliable molecular sex determination method was thus a very useful first step for all further studies of this dissertation. While different molecular sex determination methods had been developed for bears (e.g., Amstrup et al. 1993; Yamamoto et al. 2002; Pagès et al. 2009), their application proved to be difficult in our laboratory and in others (Frosch 2014). A dual-Y approach had previously been successfully developed for Florida manatee (Tringali *et al.* 2008) as well as for elephants (Ahlering *et al.* 2011), showing promise for the development of a new bear-specific approach (Bidon *et al*. 2013).

The essence of the dual-Y approach is that two fragments from the Y chromosome are PCR amplified and detected by gel electrophoresis in addition to an X-linked control (**Figure 4**). This increases robustness and reliability, because even if one of the two Y-specific amplicons fails, the detection of the second amplicon still signals male sex. The control fragment that amplifies in both sexes is critically, since it prevents PCR failure from being interpreted as female sex. Without

such a control, the absence of Y-specific amplicons could mean both female sex of the sample or complete failure of any amplification. Beyond the benefits of the co-amplification of two Y-specific markers, the new approach is both sensitive and employs small amplicon sizes. This facilitates detection in samples with low amounts of potentially degraded DNA, which can be expected from non-invasively collected material. The bear specificity of the primers ensures that if a sample contains material from multiple sources, the sex cannot wrongly be inferred from another species. Importantly, also human contamination during laboratory procedures cannot tamper the results.

Molecular sex determination is a powerful technique when the sex cannot be determined by direct observation or when sample information is lacking. Such molecular investigative techniques are also needed in wildlife law enforcement, e.g. for the protection of endangered species suffering from illegal harvesting and trade (Ogden *et al.* 2009). Moreover, for the application of informed wildlife management strategies, sex ratios are an important parameter. For instance, a reintroduction program for the endangered Pyrenean brown bear population was based on the sex ratio determined by molecular means (Taberlet *et al.* 1993, 1997).

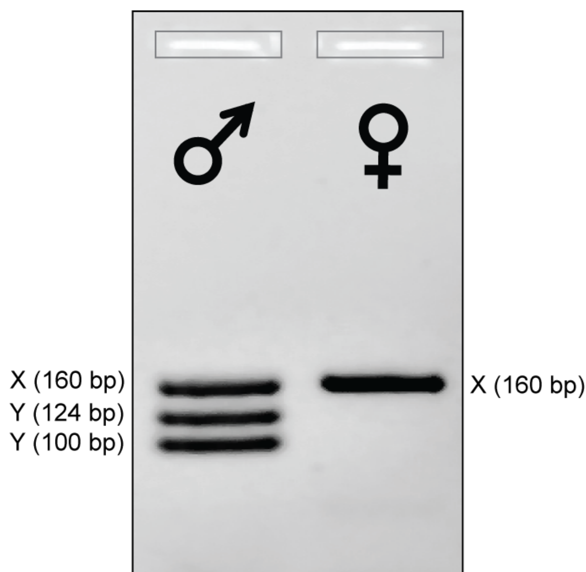


Figure 4. Representative gel picture from the dual-Y approach for sex determination in bears. The test is based on the detection of one X- and two Y-linked amplicons in males (left), versus one X-linked amplicon in females (right). It can be applied to all ursine bears and the spectacled bear. Fragment sizes are shown in brackets; bp: basepairs.

Our bear-specific dual-Y approach has already been proven to be of practical value, since it is now applied in different laboratories (C. Frosch (Senckenberg Gelnhausen), S. Shaarma (Clemson University, USA), J. Schregel (BioForsk,

Norway), pers. comm.). Studies investigating historic translocations of brown bears in the Balkan region (Nowak *et al.* 2014) and brown bear population structure in Bulgaria (Frosch *et al*. 2014) further demonstrate the feasibility of the dual-Y approach. We use the method as a standard analysis of any new samples, for ursine as well as tremarctine bears. Beyond verifying the sex of individuals, the high sensitivity of the primers also allows for a prior assessment of DNA quality and quantity: it became apparent that if amplification with the sexing primers fails, other primers (e.g., autosomal introns) are unlikely to work on these samples. Also, our publication has triggered the development of a similar sexing approach for the common hamster (Fuchs 2014). Recently, a dual-Y approach for increased reliability has been developed for human and non-human primates, illustrating the usefulness of such a technique even in well-studied model species (Cadamuro *et al.* 2015).

IDENTIFYING MALE-SPECIFIC MARKERS IS COMPLICATED BY THE STRUCTURE OF THE Y CHROMOSOME

The repetitive nature of the Y chromosome has hampered the specific identification of sequences from the Y chromosome in most species. Moreover, inferring Y-linked information from other lineages, e.g. for primer design, might be difficult because Y chromosomes can differ considerably morphologically and genetically among lineages (Murphy et al. 2006; Waters et al. 2007; Li et al. 2013). Even relatively closely related species can have highly divergent Y chromosomes, e.g. human and chimpanzees (Hughes *et al*. 2010). The availability of genome-wide data from sequenced male individuals can be a rich source to develop Y-linked markers if Y-chromosomal sequences can be identified from these data, e.g. for primer design, or to find single nucleotide polymorphisms and microsatellites. As genomes from more and more species are becoming available at an increasing speed, so are sequence data from the Y chromosome (as long as male individuals are sequenced). However, Y-linked sequences are hidden among anonymous scaffolds that have not been assigned a chromosomal origin. In bears, a male reference assembly with thousands of scaffolds is available (Li *et al*. 2011), and so are several genomes of male individuals (e.g., Miller *et al*. 2012). In this context, the possibility to identify Y-chromosomal scaffolds from genome-wide sequencing data promises to be an important tool for the development of Y-chromosomal markers in bears.

We used the available polar bear assembly as well as sequence data from other published bear genomes and showed that the different coverage expectations of short reads from males and females on Y-linked, X-linked and autosomal reference scaffolds can be used to reliably infer the chromosomal origin of scaffolds and identify large amounts of Y-chromosomal sequences (Bidon *et al*. 2015). This approach is called average depth (AD)-ratio and works by separately mapping short sequence reads from a male and a female polar bear genome to the reference scaffolds of the male polar bear. By comparing the average read depth (coverage) for each scaffold between the male and the female individual, a value can be retrieved that is indicative of the chromosomal origin of the respective scaffold.

A classical approach to identify Y-linked sequences is to search for similarity to known Y-linked gene sequences (e.g., Carvalho et al. 2000; Krzywinski et al. 2004). This was also successfully conducted to identify the first five Y-linked scaffolds used for the development of male-specific markers (Bidon *et al*. 2013, 2014, Kutschera *et al*. 2015). However, in combination with the AD-ratio approach that was much more efficient in identifying Y-chromosomal scaffolds, limitations of a search strategy became apparent (Bidon *et al*. 2015). For example, scaffolds without Y-linked genes cannot be discovered, and due to sequence similarity between genes on the Y and other chromosomes, the rate of false positives can be relatively high. The combination of different *in-silico* approaches plus the *in-vitro* validation of male-specificity thus appears to be a favorable strategy to efficiently and reliably identify large amounts of Y-chromosomal sequence from whole-genome sequencing data (Bidon *et al*. 2015).

Almost two megabases of identified Y-linked sequences is an amount unprecedented in any non-model organism for which the Y chromosome has not been specifically sequenced. Nevertheless, it must be made clear that we could only identify a small part of the presumed total size of the polar bear Y chromosome, likely due to sequence and assembly problems of highly repetitive regions that are abundant on Y chromosomes of many species (Li *et al.* 2013; Soh *et al.* 2014). It would be interesting to specifically sequence larger parts of the bear Y chromosome, and to use these sequences in a comparative analysis with the other two sequenced carnivore Y chromosomes, those of dog and cat (Li *et al*. 2013). This might give interesting insights into Y chromosome evolution from the perspective of other

organisms than primates, in which considerable structural differences have been observed between closely related species (Hughes *et al*. 2010).

The identification of Y-linked scaffolds in the polar bear assembly enabled the development of primers to resequence smaller fragments of the Y chromosome and to analyze allele sizes of Y-linked microsatellites in many individuals of different bear species (Bidon *et al*. 2014, Kutschera *et al*. 2014). Moreover, the scaffold sequences extracted after mapping of published genomes of male individuals could be directly used to infer the patrilineal evolutionary history of polar bears with high resolution (Bidon *et al*. 2014, 2015).

THE Y CHROMOSOME AS AN EVOLUTIONARY MARKER – UNCOVERING THE PATERNAL HISTORY OF BEARS

EXPANDING PHYLOGEOGRAPHIC INFERENCES USING MALE-SPECIFIC MARKERS

The investigation of variation among lineages within a species, and the geographic distribution of this variation is the realm of phylogeography. MtDNA has been an important tool in this field, because, like the Y chromosome, it can quickly generate phylogeographic structure. This makes both markers well suited for the analyses of relatively recent events and for the detection of phylogeographic patterns (**Figure 5)**.

The higher effective population size of an average autosomal locus implies that incomplete lineage sorting affects phylogeographic inferences at timescales where uniparentally loci are already sorted and reciprocally monophyletic. Uniparentally and biparentally inherited markers thus address different timescales due to their different coalescent times, with mtDNA and the Y chromosome being more informative in the recent past (between Ne and 4Ne), and autosomal markers in the distant past (before 4Ne) (Zink & Barrowclough 2008; Fahey *et al.* 2014).

Sex-specific migration differentially influences the distribution of genetic variation and geographic patterns at differentially inherited markers. This is one of the reasons why mtDNA is often included, even if extensive nuclear datasets are analyzed (Prugnolle & de Meeus 2002; Toews & Brelsford 2012; Garrick *et al.* 2015). Some of the many examples that compare patterns of differentiation at matrilineal and biparental markers are studies of gazelles (Lerp *et al.* 2013), South American foxes (Tchaicka *et al.* 2007), tiger quolls (Firestone *et al.* 1999), Australian robins (Pavlova *et al.* 2013), and Galápagos giant tortoises (Garrick *et al.* 2014). This is also

the reason why the Y chromosome would be ideal to complement and contrast other loci, as has been noted before (Davison *et al.* 2011, Prugnolle & de Meeus 2002). However, male-specific markers have been included in only few studies of natural populations (see below), and have not been used in phylogeographic analyses of brown and polar bears.

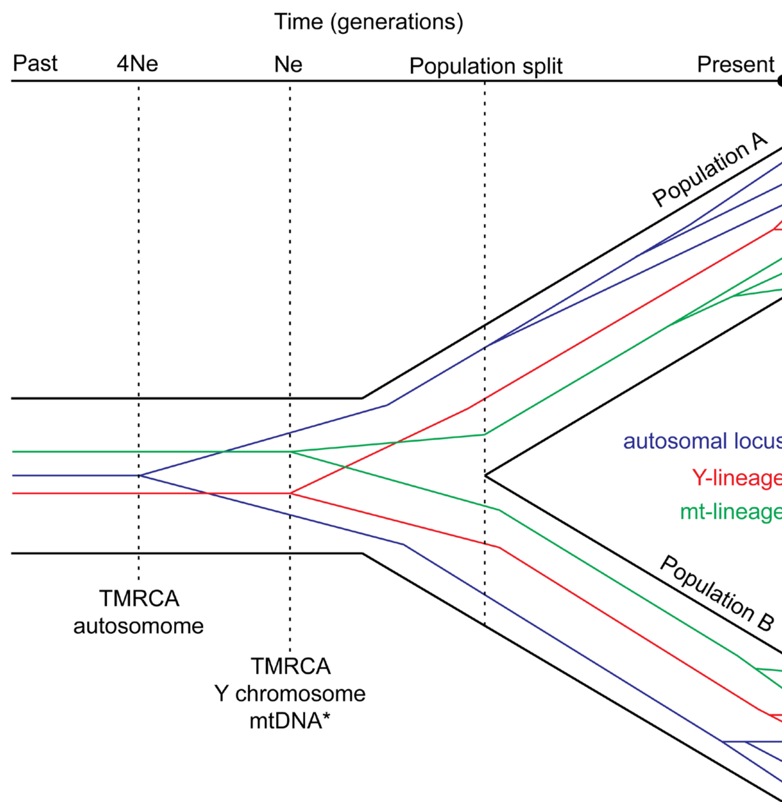


Figure 5. Coalescent times of different markers sampled in two populations. The Y chromosome and mtDNA achieve monophyly faster, or, conversely, complete lineage sorting sooner, due to their lower effective population size (Ne) compared to an (average) autosomal locus. In the recent past (between Ne and 4Ne generation), Y-chromosomal and mtDNA markers can thus be used for phylogeographic inferences, whereas autosomal markers are more informative in the distant past (Zink & Barrowclough 2008). The split between ancestral populations always occurs after the split of genetic lineages (Edwards & Beerli 2000). Note that the time axis is not to scale and that this schematic illustration is based on the assumption of neutrality (no selection) and random mating, i.e. a reproductive skew would lead to unequal effective population sizes of mtDNA and the Y chromosome. TMRCA: time to the most recent common ancestor.

Due to the well-documented male-biased dispersal in brown bears (see below), investigation of the patriline in comparison to the established structure of the matriline promised interesting insights into the species´ phylogeography. While mtDNA should reflect female philopatry, the Y chromosome is expected to contrast

this pattern due to migrating males. In polar bears, both sexes are highly mobile and a less pronounced difference between the matri- and patriline should be observed.

LACK OF PATRILINEAL STRUCTURING IN BROWN BEARS

Using sequence data and allele sizes of microsatellites from the Y chromosome, a striking contrast to the established, mtDNA-based phylogeographic pattern in brown bears was found (Bidon *et al*. 2014). Instead of geographically restricted maternal clades, a shallow divergence among Y-chromosomal haplotypes, and the occurrence of one haplotype across the entire sampled range, from Norway, throughout Asia to northern Canada, was detected (**Figure 6**). Additionally, a small number of haplotypes closely related to the common haplotype were found at different locations of the distribution range, but in very few individuals. This means that almost every sampled brown bear has an identical Y chromosomal sequence for the analyzed markers. Utilizing faster evolving microsatellites, a considerable amount of variation in allele sizes among brown bear Y chromosomes was detected. However this variation did not translate into geographic structure: a mix of different Y-chromosomal haplotypes occurred in any given region, and closely related haplotypes occurred in regions far apart.

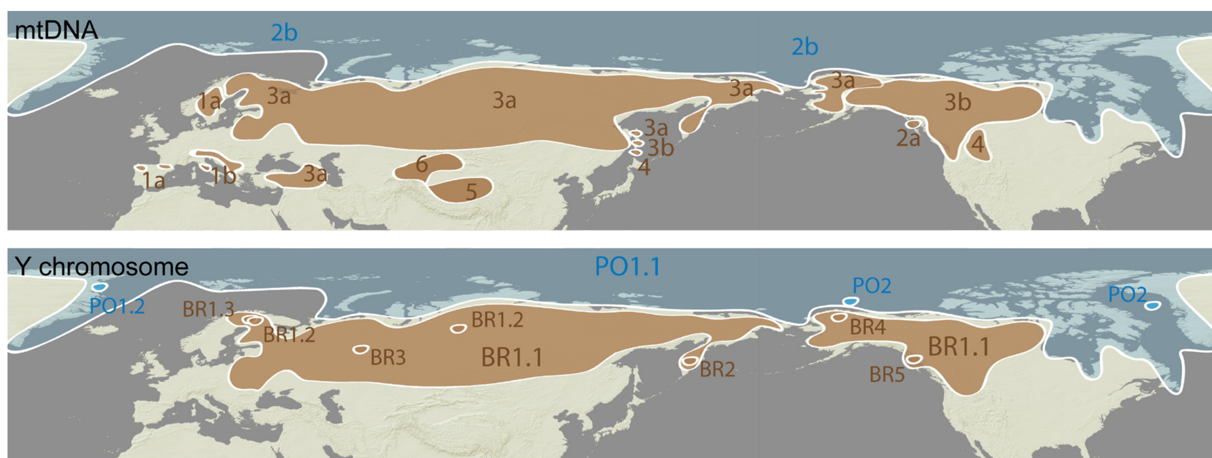


Figure 6. Range-wide population structuring in brown and polar bears, and the contrast of structuring based on maternally inherited mtDNA versus the male-specific Y chromosome. For mtDNA, geographically distinct and non-overlapping brown bear clades exist (1a, 1b, 2a, etc.), while one Y-chromosomal haplotype is spread over the entire sampled range (BR1.1). In polar bears, both markers show one clade that is spread across the entire range. Distribution of mt-clades is based on Davison *et al.* (2011).

These results indicate that males are spreading genetic variation over vast geographic areas, whereas females remain and reproduce closer to their place of birth. It is therefore important to note that matrilineal markers alone can overestimate population structuring if male-biased gene flow exists. This has also been suggested in a study of brown bears at a contact zone between matrilineal clades 1 and 3 in Bulgaria (Frosch *et al.* 2014). The authors present evidence of genetic admixture between two subpopulations based on autosomal markers, along with the detection of two male long-distance dispersers.

The lack of deep splits and the low variability of the Y chromosome are consistent with positive selection that may have acted on a specific Y chromosome variant (Bidon *et al*. 2014). Such a variant could have reached high frequencies, and could have been spread across the range by dispersing males. However, it was not possible to distinguish between selection and demographic scenarios (population expansion) based on tests for deviations from neutral expectations. Similarly, it has been remarkably difficult to detect signatures of positive selection on the Y chromosome in other species, including humans (Lawson Handley et al. 2006; Wilson Sayres et al. 2014, Singh et al. 2014).

MALE-BIASED GENE FLOW SHAPES GENETIC VARIATION IN BROWN BEARS AND OTHER SPECIES

My findings of widespread Y-chromosomal haplotypes likely reflect male brown bears spreading their Y chromosomes over large geographic areas (Bidon *et al*. 2014, **Figure 6**). The pronounced structuring of brown bear mtDNA haplotypes, in contrast, reflects the tendency of females to live in matrilineal assemblages close to their natal site. Consistent with these considerations is the well-documented male-bias in dispersal in brown bears.

Dispersal from the natal area in order to breed is an important life history trait in mammals. The majority of mammalian species exhibit male-biased dispersal, where males show a stronger tendency of leaving their natal area, or to disperse farther from that area (Greenwood 1980; Pusey 1987). Sex-biased differences in dispersal lead to a sex-bias in gene flow, if dispersal is effective, i.e. if it is followed by successful reproduction (Lawson Handley & Perrin 2007). Gene flow resulting from dispersal ultimately influences the geographic distribution of genetic variation.

Brown bear males disperse farther (McLellan & Hovey 2001; Proctor *et al.* 2004; Støen *et al.* 2006) and at a higher rate (Zedrosser *et al*. 2007), leave their natal areas earlier in life (Støen *et al*. 2006), and have larger home range sizes (Servheen 1983; Mcloughlin *et al.* 1999) than females. Further, males also show greater daily movements and the longest travelled distances were recorded for males (Craighead Jr. 1976). One famous example of a male bear travelling long distances was "Bruno" from the Italian Adamello-Brenta nature park. It was the first confirmed brown bear in Germany since 1835. It moved several hundred kilometers in about two months before it was shot in Bavaria in June 2006 (Fohrmann 2006). On the other hand, the home ranges of females are more likely to overlap with their mothers´ home range (McLellan & Hovey 2001).

Prominent examples of other carnivores covering distances of up to several hundreds of kilometers are grey wolves (Fritts 1983), red foxes (Trewhella *et al.* 1988), and Eurasian lynx (Zimmermann *et al.* 2005). Sequencing parts of their Y chromosomes would allow for a unique perspective on male gene flow in these vagile species. Indeed, both red foxes and grey wolves are two of the rare examples where studies have directly compared the matri- and patriline. In the red fox, a pattern of pronounced structuring of mtDNA (with several locally restricted clades) but reduced structuring at Y-linked markers (continental scale) was found (Statham *et al.* 2014), but has not been discussed in the context of sex-biased dispersal. Similarly, there is evidence of some spatial genetic structure based on mtDNA in the grey wolf (Pilot *et al.* 2006), while intraspecific Y-chromosomal sequence variability was found to be very low (Hellborg & Ellegren 2004). Y-linked microsatellites have only been applied on a regional scale in wolves: in one of the first studies that used Y-chromosomal markers in a natural population, Sundqvist and colleagues found distinct Y-chromosomal haplotypes in Scandinavian wolves as compared to the Baltic states and Russia, suggesting a contribution of at least two male individuals to the founding of the current Scandinavian population (Sundqvist *et al.* 2001). Other examples of studies where contrasting patterns of mtDNA and Y-chromosomal structuring have been detected include gorillas (Douadi *et al.* 2007), chamois (Pérez *et al.* 2011), and African elephants (Roca *et al.* 2005). For both gorilla and elephant male-biased dispersal has been suggested to have caused the differences between the markers.

The field of human phylogeography is a prominent exception from the general lack of male-specific markers in analyses of evolutionary history. Interestingly, and quite opposite to results from brown bears, it is the patriline that shows more structuring than the matriline in humans. This has been explained with higher female migration rates due to post-marriage movements in human cultures (Seielstad *et al.* 1998; Oota *et al.* 2001). Notably, for large-scale events, e.g. colonization of new continents, congruent patterns between mtDNA and Y chromosome phylogenies have been found (Underhill & Kivisild 2007). The high number of recent publications in this field illustrates that sequences from the Y chromosome continue to be an indispensable part of research of human phylogeography, despite the wealth of genome-wide data available in primates (Underhill & Kivisild 2007; Hughes & Rozen 2012; Wei *et al.* 2013; Scozzari *et al.* 2014; Hallast *et al.* 2014).

DIVERGENT Y-CHROMOSOMAL CLADES IN POLAR BEARS ARE GEOGRAPHICALLY WIDESPREAD

Polar bears spend most of their time on arctic sea ice, where they occur in relatively low density throughout their range (Stirling 2012). They are highly mobile, but have core regions that they preferably travel, their movements being affected by season and availability of prey (Stirling & Oritsland 1995; Amstrup 2003; Laidre *et al.* 2013). Very long movements of single bears have been recorded walking (Durner & Amstrup 1995; Amstrup 2003) and swimming (Amstrup *et al.* 2001). Potential differences between movements of males and females remain difficult to investigate because radio collaring does not work for male polar bears – their neck is thicker than the head. Most recordings of long distance travelling are therefore from adult females. A limited number of telemetry observations from surgically implanted or ear-attached transmitters suggests that overall, the movements and sizes of areas occupied by female and male polar bears do not differ greatly, although differences exist in how linear and directed these movements are (Amstrup et al. 2001; Lone et al. 2012; Laidre et al. 2013). A joint analysis of mtDNA and autosomal microsatellites suggests that gene flow in polar bears is slightly male-biased (Peacock *et al*. 2015). Concordant with these results is that four genetic clusters encompassing 19 populations can be discerned in polar bears at maternally as well as biparentally

inherited markers (Paetkau *et al.* 1999; Campagna *et al.* 2013; Peacock *et al.* 2015), but that the level of structuring is slightly greater for mtDNA.

My results from the analysis of patterns of Y-chromosomal variation are largely consistent with the presumed patterns of gene flow in polar bears from observational and genetic studies (Bidon *et al*. 2014). As expected, the difference between paternal and maternal markers is not as striking as observed for brown bears. Based on sequence data, I found one polar bear haplotype (PO1.1) to be wide spread over the entire sampled range, and no signs of geographic structuring were obtained from analyses of microsatellites.

The Y-chromosomal sequence data revealed two rare haplotypes in addition to one widespread and common haplotype (Bidon *et al*. 2014). One of these rare haplotypes (PO2) was shared by four individuals from different regions (Alaska and Western Greenland). This haplotype thus encompasses 10% of the polar bear individuals analyzed, whereas the most common of the rare haplotypes in brown bears includes only 3% of the individuals analyzed. By identifying almost two megabases of Y-linked scaffolds from the previously published polar bear assembly and polar bear genomes (Li *et al*. 2011, Miller *et al*. 2012), it was possible to analyze a considerably extended dataset in this species (Bidon *et al*. 2015). It became apparent that the two haplotypes in polar bears (PO1.1 and PO2) correspond to two divergent clades: the very same polymorphic site that separates the two haplotypes in the resequencing dataset (Bidon *et al*. 2014) was also present in the respective Y-linked scaffold of additional individuals (published polar bear genomes, Miller *et al.* 2012; Bidon *et al*. 2015). Analyzing more Y-linked scaffolds in these published genome individuals, the same two clades, separating the same two individuals (from Alaska and Svalbard) from other polar bears, were found (**Figure 7**).

Most interestingly, the two divergent clades within polar bears do not show a strong geographic signal, because both contain individuals from different regions. This finding confirms the lack of pronounced population structuring in polar bears across different marker systems, and emphasizes the great dispersal capacity of both sexes. Nevertheless, the differentiation between the two patrilineal clades might indicate an ancient structuring that has not been described for any genetic marker system within this species. According to our divergence time estimate, the two paternal lineages have separated in the middle Pleistocene and have thus apparently been maintained over a considerable amount of time (Bidon *et al*. 2015). They have

now, facilitated by the large dispersal capacities of male polar bears, been spread over large parts of the Arctic. Additional sampling and sequencing will be necessary to reveal the geographic extent of this.

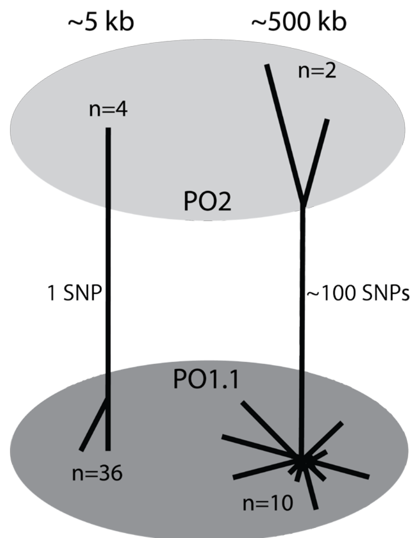


Figure 7. Schematic representation of two patrilineal polar bear clades, PO1.1 and PO2. A small dataset based on resequencing of ~5 kb in 40 individuals (Bidon *et al*. 2014) recovered the same two groups as a larger dataset based on ~500 kb scaffold data from 12 published polar bear genomes (Bidon *et al*. 2015). In the ~5 kb dataset, there is no variation within each clade, with the exception of one diverging individual in PO1.1. The higher amount of variation in the ~500 kb dataset is illustrated by the drawing of individual lineages. Note that the branches are not drawn to scale. Only one substitution differentiates between the two clades in the ~5 kb dataset, whereas ~100 substitutions in the ~500 kb dataset.

The Y chromosome and ursine speciation

Different regions of the genome can show different levels of differentiation among species (Qvarnström & Bailey 2009), resulting in topological discrepancies in their phylogenetic trees. By comparing phylogenetic trees from autosomes and sex-linked loci one can learn about past hybridization events. On the one hand, genomic regions with reduced levels of interchromosomal recombination, like the Y chromosome, can remain highly differentiated despite extensive interspecific gene flow (Geraldes *et al.* 2008; Ellegren *et al.* 2012), reflecting the previous topology (i.e. before gene flow started). On the other hand, lineage sorting is achieved sooner and monophyletic groups arise faster on the Y chromosome. This means that while alleles at autosomal loci have not yet reached reciprocal monophyly between species, markers at the Y chromosome are already completely sorted (**Figure 5**).

The comparison of mtDNA and nuclear topologies has already quarried interesting patterns in bears: the mtDNA topology differs from the nuclear topology, i.e. American and Asiatic black bears are matrilineal sister species (Talbot & Shields 1996a; Krause *et al.* 2008). This is consistent with fossil evidence that suggests that Asiatic and American black bears are closely related and derive from *Ursus minimus* (Kurtén & Anderson 1980; McLellan & Reiner 1994). In contrast, autosomal markers

group American black bears as sister species to brown and polar bears, and the three Asian bears (Asiatic black bear, sun bear, sloth bear) into one clade (Kutschera *et al*. 2015). A fast radiation accompanied by interspecific gene flow between different bear species, and incomplete lineage sorting have likely been important forces shaping the genomes of ursine bears. Indeed, the exact branching pattern between Asiatic black bear, sloth bear and sun bear is not yet resolved, and the actual pattern of organism divergence among ursine bears thus remains unknown. In this context, analysis of the patriline promises another interesting and important perspective on the complex speciation of these species.

As expected, Y-chromosomal haplotypes from different species were clearly separated from each other, with many unique, lineage specific substitutions (Bidon *et al*. 2014, Kutschera *et al*. 2014). This contrasts with haplotype networks from single, autosomal loci that showed extensive allele sharing among different combinations of species (Kutschera *et al*. 2014; Hailer *et al*. 2012). However, there are only very few shared substitutions between any two of the three Asian bears on the sequenced fragments from the Y chromosome, reflecting the relatively low support for the grouping of these species. The giant panda is the most divergent of all ursids, and the spectacled bear is a sister taxon to all ursines. These results mirror findings from the three previously analyzed Y-linked genes (Nakagome *et al*. 2008). Except for the lower statistical support and the exact branching order within the clade of three Asian bears, the Y-chromosomal topology is thus concordant with the average species tree. Due to the very long haplotypes that can theoretically be inferred, generating longer Y chromosome sequences could help us understand if Asian bears indeed evolved (almost) simultaneously, or if the unresolved grouping is an effect of insufficient phylogenetic information of the limited data available today.

Regarding the placement of the Asiatic black bear, both autosomal and Y–chromosomal phylogenies conflict with the mtDNA phylogeny (**Figure 8, left side**). The interpretation of these discordances is not straightforward. If mtDNA and the known fossil record reflected the actual species phylogeny, then hybridization between male ancestors of today´s brown bears and American black bear females could have given rise to the pattern we observe today. The ancestors of today´s American black bears would have had a genome closely related to that of the ancestor of the Asiatic black bear. This scenario implies introgression of the Y chromosome from brown into American black bears. Since then, the

Y chromosomes of American black and brown bears would have had time to accrue diversity and become the distinct lineages that we see today.

Zooming into one part of the ursid phylogeny, there is now plenty of evidence supporting the sister relationship of brown and polar bears, i.e. morphology (Kurtén & Anderson 1980), autosomal loci and genome-wide data (Hailer *et al*. 2012, Miller *et al*. 2012, Cahill *et al*. 2013, Liu *et al.* 2014), and Y-chromosomal sequences (Bidon *et al*. 2014, Kutschera *et al*. 2014). The paraphyly of brown bears is likely due to introgressive hybridization of brown bear mtDNA into polar bears some ~160.000 years ago (Lindqvist *et al*. 2010, Edwards *et al*. 2011, Hailer *et al*. 2012; **Figure 8, right side**). The clear separation of their patrilines thus likely reflects the actual organism divergence of brown and polar bears. In contrast to mtDNA, no signal of patrilineal introgression between the two species could be observed (Bidon *et al*. 2014**)**.

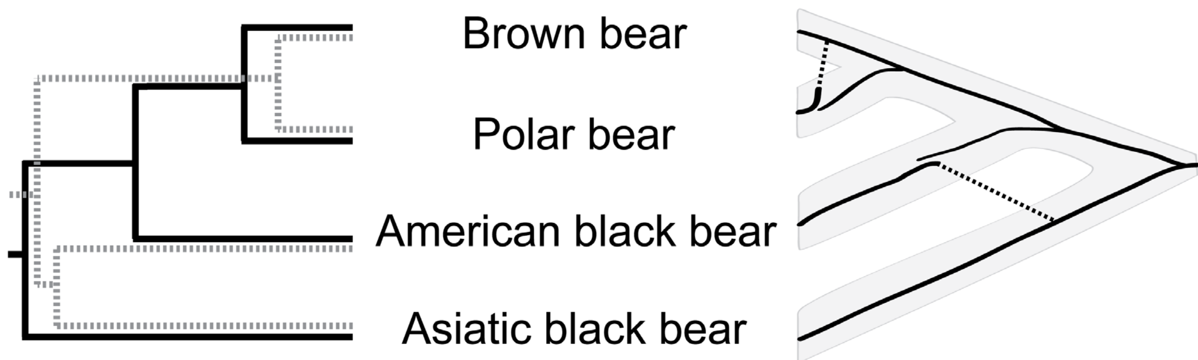


Figure 8. Phylogenetic relationship of four bear species. Left tree: the black line represents the topology of the average nuclear species tree and of the Y chromosome tree. The grey, dotted line represents the topology of the mtDNA tree. Right tree: the black lines within the cladogram (shaded grey) represent mitochondrial lineages. Dotted lines that connect branches across species boundaries indicate hypothesized hybridization events between species, accompanied by introgression of mitochondrial DNA.

Concerning the divergence times of brown and polar bears, Y chromosome data can give specific information about the age of their patrilines. In humans, the comparison of mtDNA and Y chromosome has stimulated interesting debates about the age of mitochondrial “Eve” and Y-chromosomal “Adam” (Poznik *et al.* 2013). Using two different human Y-chromosomal substitution rates for calibration (Xue *et al.* 2009; Fu *et al.* 2014), and a small (ca. 5 kb; Bidon *et al*. 2014) as well as a large dataset (ca. 500 kb; Bidon *et al*. 2015) of Y-chromosomal sequences, I estimated

that brown and polar bear patrilines have diverged ~0.3 – 0.8 million years ago. This is largely consistent with the younger of the recently published divergence time estimations based on nuclear data (e.g., Liu *et al*. 2014: 0.3-0.5 mya; Hailer *et al*. 2012: 0.3-0.9 mya), and thus might reflect the tendency of the Y chromosome to more closely track the actual population divergence than an average autosomal locus (**Figure 5**) (Moore 1995; Edwards & Beerli 2000).

Also in other species different topologies between sex-linked loci and autosomes indicated hybridization, for example among different species of Darwin´s Finches (Lamichhaney *et al.* 2015). Further, concordant signals between mtDNA and the Y chromosome uncovered two cryptic species of voles (Hellborg *et al.* 2005), and discordance between monophyletic Y-lineages and mitochondrial paraphyly was reported in macaques (Tosi *et al.* 2000). In goats, mitochondrial introgression was suggested to explain discordant patterns between mtDNA and the Y chromosome (Pidancier *et al.* 2006). Largely similar patterns between the Y-chromosomal and mt topology were found in gibbons (Chan *et al.* 2012).

Y-CHROMOSOMAL INTROGRESSION IS UNLIKELY IN BEARS

The independent Y chromosome gene tree estimate can be used to learn more about introgression in ursine bears, and to evaluate a scenario as outlined above (hybridization between American black and brown bears). Non-recombining parts of the genome are theoretically well suited to identify signals of introgression, because they remain as a whole and are not fragmented by recombination. However, reduced introgression rates compared to other loci have been described for the Y chromosome, e.g. in elephants (Roca *et al*. 2005), European rabbits (Geraldes *et al*. 2008), and common shrews (Balloux *et al.* 2000). This can be explained with genomic incompatibilities that can lead to a reduced viability of hybrids of the heterogametic sex compared to hybrids of the homogametic sex (Haldane´s rule; Haldane 1922). Males carrying an introgressed Y chromosome are thus less likely to pass on their genetic material to the next generation. Furthermore, simulation studies have shown that markers with high intraspecific gene flow show low interspecific gene flow, and thus are less likely to introgress (Petit & Excoffier 2009). Moreover, analysis of several individuals from each species would be necessary to detect haplotype sharing or species paraphyly as a support for hybridization if only some

populations show this pattern. This is the case for the brown bear matriline where only the ABC islands population is more closely related to polar bears than to other brown bears. A strong reproductive skew, i.e. few males contributing to a large number of offspring, would further decrease the effective population size of the Y chromosome, increasing the probability that the Y chromosome will track the actual species tree compared to any other locus (Moore 1995).

Taken together, this would argue that Y-chromosomal introgression as in the scenario outlined above is unlikely in bears, and that the paternal phylogeny is likely to be congruent with the actual pattern of species divergence. This would require a scenario where mtDNA introgressed from Asiatic black bears into American black bears to explain their close matrilineal relationship (**Figure 8**, **right side**; Kutschera *et al*. 2014), similar to introgression of mtDNA from brown into polar bears. The problem with this hypothesis is that sympatry of American and Asiatic black bears, and thus a chance for hybridization between the two, is not obvious from the currently available fossil record (McLellan & Reiner 1994). It will be interesting to see what whole-genome analyses accompanied by longer Y-chromosomal sequences will uncover.

THE ROLE OF THE Y CHROMOSOME IN SPECIATION MIGHT HAVE BEEN UNDERAPPRECIATED

An interesting remark is the general role of the Y chromosome in speciation. The X chromosome has been suggested to play a role in hybrid incompatibility and speciation, leading to a reduction of introgressed alleles (Masly & Presgraves 2007; Qvarnström & Bailey 2009). This effect might have played a role in the reduced introgression of genetic material from Neanderthals into modern humans (Sankararaman *et al.* 2014), and in selection against polar bear ancestry on brown bear X chromosomes (Cahill et al. 2015). Similarly, the low level of introgression of Y chromosomes could have been caused by Y-linked genes that are involved in hybrid incompatibilities, promoting isolation and speciation. Indeed, the Y chromosome is enriched for testis-specific genes with important functions in reproductive traits associated with spermatogenic failures (Skaletsky *et al*. 2003). In *Drosophila*, the Y chromosome and autosomal alleles have been shown to interact to cause male sterility (Vigneault & Zouros 1986), and the Y chromosome has also been suggested to harbor genes involved in reproductive isolation in shrews

(Balloux *et al*. 2000), and between subspecies of the house mouse (Tucker *et al.* 1992; Boissinot & Boursot 1997). The role of the Y chromosome in speciation has thus likely been underappreciated due to the lack of evolutionary studies investigating this part of the genome.

ON THE RELEVANCE OF A SINGLE LOCUS IN THE ERA OF GENOMICS

In the era of genomics, enormous amounts of whole-genome sequence data, and thus data from multiple unlinked loci, can now be obtained increasingly fast and affordable (Koepfli *et al.* 2015). This raises the question why any attention should be given to a single locus that only occurs in males, and thus only represents one side of an organisms´ evolutionary history?

One part of the answer is that this single locus is fairly large: the amount of haplotype information is only limited by the amount of Y-sequence identified (and ultimately by the size of the male-specific region; **Figure 2**). Indeed, the Y chromosome is the largest non-recombining fraction of the entire genome, and thus an almost inexhaustible source of genetic variation.

This leads to the second part of the answer: the Y chromosome can be used to trace single lineages back in time with incomparably high resolution. Uncovering phylogenetic and temporal relationships in such a finely graduated manner is not possible with multi-locus data, because variation is reshuffled by recombination in every generation, masking the record of one single phylogeny. Indeed, the resolution that can be obtained is so high that in a dataset of more than 400 human males, identical Y chromosomes have been found in only eight pairs of individuals, causing the Y-chromosomal tree to “burst into leaf” (Hallast *et al*. 2014). Similarly, increasing the amount of sequence from mtDNA enhanced the resolution in phylogeographic inferences (Keis *et al*. 2013), of course limited by the relatively smaller size of the mt genome. Notably, Hallast *et al*. (2014) analyzed 3.7 megabases from the human Y chromosome, which is in the same order as the 2 megabases that we have identified in the polar bear genome (Bidon *et al*. 2015). With the proper approaches at hand (e.g., AD-ratio approach), the increasing availability of published genome sequences now gives us the possibility to extract Y-linked sequences at a large scale. The analysis of Y-linked sequences in non-model organisms benefits from high-throughput sequencing technology not only because Y-linked sequences can be

identified from whole genome sequence data. It is also possible to use enrichment strategies to specifically and economically sequence hundreds of kilobases of Y-chromosomal sequences from many individuals.

The third part of the answer is that the male perspective can be used to complement and contrast patterns obtained from maternally and biparentally inherited markers. Although the Y chromosome may not always show the same inheritance patterns as the rest of genome, it nevertheless represents a fundamental part of a species´ evolutionary history (just like mtDNA; Rubinoff & Holland 2005). Indeed, it is the discrepancy among different markers that can reveal processes like male-biased dispersal (Bidon *et al*. 2014) and introgression (Kutschera *et al*. 2014), and can thus lead to an increased understanding of evolution. The particular male perspective remains important, no matter how much genome-wide sequence data can be generated today. Moreover, the Y chromosome is indispensable in forensic science, sex determination and paternity testing (Bidon *et al*. 2013).

Conclusion

Developing and applying Y-chromosomal markers, I have contributed to an increased understanding of different aspects of the evolutionary history of bears. My findings demonstrate the potential of Y-chromosomal markers to uncover unknown patterns (population structuring or the lack of such) and processes (male-biased gene flow, hybridization), even in genetically well-described species.

Specifically, I showed that the range-wide population structuring of brown and polar bears is shaped by male biased gene flow, which needs to be considered when drawing inferences for a species as a whole (Bidon *et al*. 2014, 2015). Besides studies on primate phylogeography, the size of the Y-chromosomal dataset analyzed is unprecedented in a non-model organism (Bidon *et al*. 2015). Moreover, I have presented evidence that the phylogeny of the patriline is largely congruent with the average autosomal species tree of bears, and that this might indicate that the topology of the mtDNA tree does not represent the actual species divergence (Kutschera *et al*. 2014). Additionally, I have shown that the Y chromosome is an indispensable element of molecular methods for sex determination, which in turn is an important tool in wildlife management of endangered bear populations (Bidon *et al*. 2013). Moreover, my thesis provides a rich resource for Y-linked sequences

(Bidon *et al*. 2015) and markers (single nucleotide polymorphisms and microsatellites; Bidon *et al*. 2014, Kutschera *et al*. 2014) that can now readily be applied in different bear species, but can also be transferred to other organisms. This dissertation thus shows and exemplifies why the Y chromosome is important and how it can provide genetic markers that can be utilized in evolutionary inferences of natural populations.

REFERENCES

Ahlering M, Hailer F, Roberts MT, Foley C (2011) A Simple and accurate method to sex savannah, forest and asian elephants using noninvasive sampling techniques. *Molecular Ecology Resources*, 11, 831-834.

Amstrup SC (2003) The Polar Bear - Ursus maritimus Biology, Management, and Conservation. In: *Wild Mammals of North America* (eds Feldhammer GA, Thompson BC, Chapman JA), pp. 589–610.

Amstrup SC, Durner GM, McDonald TL, Mulcahy DM, Garner GW (2001) Comparing movement patterns of satellite-tagged male and female polar bears. *Canadian Journal of Zoology*, 79, 2147–2158.

Amstrup SC, Garner GW, Cronin MA, Patton JC (1993) Sex identification of polar bears from blood and tissue samples. *Canadian Journal of Zoology*, 71, 2174–2177.

Anderson S, Bankier AT, Barrell BG *et al.* (1981) Sequence and organization of the human mitochondrial genome. *Nature*, 290, 457–465.

Avise JC (1986) Mitochondrial DNA and the Evolutionary Genetics of Higher Animals. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 312, 325–342.

Avise JC (1994) *Molecular Markers, Natural History and Evolution*. Chapman & Hall, New York.

Avise JC, Arnold J, Ball RM *et al.* (1987) Intraspecific Phylogeography: The Mitochondrial DNA Bridge Between Population Genetics and Systematics. *Annual Review of Ecology and Systematics*, 18, 489–522.

Avise JC, Lansman RA, Shade RO (1979) The use of restriction endonucleases to measure mitochondrial DNA sequence relatedness in natural populations. I. Population structure and evolution in the genus Peromyscus. *Genetics*, 92, 279–295.

Bachtrog D (2013) Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. *Nature Reviews. Genetics*, 14, 113–124.

Bachtrog D, Mank JE, Peichel CL *et al.* (2014) Sex determination: why so many ways of doing it? *PLoS Biology*, 12, e1001899.

Ballard JWO, Whitlock MC (2004) The incomplete natural history of mitochondria. *Molecular Ecology*, 13, 729–744.

Balloux F, Brünner H, Lugon-Moulin N, Hausser J, Goudet J (2000) Microsatellites can be misleading: an empirical and simulation study. *Evolution*, 54, 1414–1422.

Bellott DW, Hughes JF, Skaletsky H *et al.* (2014) Mammalian Y chromosomes retain widely expressed dosage-sensitive regulators. *Nature*, 508, 494–499.

Bidon T, Frosch C, Eiken HG *et al*. (2013) A sensitive and specific multiplex PCR approach for sex identification of ursine and tremarctine bears suitable for non-invasive samples. *Molecular Ecology Resources* 13, 362–368.

Bidon T, Janke A, Fain SR *et al*. (2014) Brown and Polar Bear Y Chromosomes Reveal Extensive Male-Biased Gene Flow within Brother Lineages. *Molecular Biology and Evolution* 31, 1353–1363.

Bidon T, Schreck N, Hailer F, Nilsson M, Janke A (2015) Genome-wide search identifies 1.9 megabases from the polar bear Y chromosome for evolutionary analyses. Submitted to *Genome Biology and Evolution*.

Boissinot S, Boursot P (1997) Discordant Phylogeographic Patterns Between the Y Chromosome and Mitochondrial DNA in the House Mouse: Selection on the Y Chromosome? *Genetics*, 146, 1019–1034.
Bosch E, Hurles ME, Navarro A, Jobling MA (2004) Dynamics of a human interparalog gene conversion hotspot. *Genome Research*, 14, 835–844.
Brumfield RT, Beerli P, Nickerson D, Edwards SV (2003) The utility of single nucleotide polymorphisms in inferences of population history. *Trends in Ecology & Evolution*, 18, 249–256.
Cadamuro VC, Bouakaze C, Croze M *et al.* (2015) Determined about sex: sex-testing in 45 primate species using a 2Y/1X sex-typing assay. *Forensic Science International. Genetics*, 14, 96–107.
Campagna L, Van Coeverden de Groot PJ, Saunders BL *et al.* (2013) Extensive sampling of polar bears (*Ursus maritimus*) in the Northwest Passage (Canadian Arctic Archipelago) reveals population differentiation across multiple spatial and temporal scales. *Ecology and Evolution*, 3, 3152–3165.
Cahill JA, Green RE, Fulton TL *et al.* (2013) Genomic Evidence for Island Population Conversion Resolves Conflicting Theories of Polar Bear Evolution. *PLoS Genetics*, 9, e1003345.
Cahill JA, Stirling I, Kistler L *et al.* (2015) Genomic evidence of geographically widespread effect of gene flow from polar bears into brown bears. *Molecular Ecology*, 24, 1205-1217.
Carvalho AB, Lazzaro BP, Clark AG (2000) Y chromosomal fertility factors kl-2 and kl-3 of Drosophila melanogaster encode dynein heavy chain polypeptides. *PNAS*, 97, 13239–13244.
Carvalho AB, Koerich LB, Clark AG (2009) Origin and evolution of Y chromosomes: Drosophila tales. *Trends in Genetics*, 25, 270–277.
Casanova M, Leroy P, Boucekkine C *et al.* (1985) A Human Y-Linked DNA Polymorphism and its Potential for Estimating Genetic and Evolutionary Distance. *Science*, 230, 1403–1406.
Chan Y-C, Roos C, Inoue-Murayama M *et al.* (2012) A comparative analysis of Y chromosome and mtDNA phylogenies of the Hylobates gibbons. *BMC Evolutionary Biology*, 12, 150.
Charlesworth B (1991) The Evolution of Sex Chromosomes. *Science*, 251, 1030–1033.
Charlesworth B (1996) The evolution of chromosomal sex determination and dosage compensation. *Current Biology*, 6, 149–162.
Charlesworth B, Charlesworth D (2000) The degeneration of Y chromosomes. *Philosophical transactions of the Royal Society of London. Series B, Biological Sciences*, 355, 1563–1572.
Craighead Jr. FC (1976) Grizzly Bear Ranges and Movement as Determined by Radiotracking. In: *Bears: Their Biology and Management. A selection of Papers from the Third International Conference on Bear Research and Management*, pp. 97–109. IUCN Publication New Series, New York, USA, and Moscow, Russia.
Cronin MA, Amstrup SC, Garner GW (1991) Interspecific and intraspecific mitochondrial DNA variation in North American bears (*Ursus*). *Canadian Journal of Zoology*, 69, 2985–2992.
Cronin MA, MacNeil MD (2012) Genetic relationships of extant brown bears (Ursus arctos) and polar bears (Ursus maritimus). *The Journal of Heredity*, 103, 873–881.

Cronin MA, Rincon G, Meredith RW *et al.* (2014) Molecular Phylogeny and SNP Variation of Polar Bears (*Ursus maritimus*), Brown Bears (*U. arctos*), and Black Bears (*U. americanus*) Derived from Genome Sequences. *Journal of Heredity*, 105, 312-323.

Davison J, Ho SYW, Bray SC *et al.* (2011) Late-Quaternary biogeographic scenarios for the brown bear (*Ursus arctos*), a wild mammal model species. *Quaternary Science Reviews*, 30, 418–430.

Douadi MI, Gatti S, Levrero F *et al.* (2007) Sex-biased dispersal in western lowland gorillas (*Gorilla gorilla gorilla*). *Molecular Ecology*, 16, 2247–59.

Durner GM, Amstrup SC (1995) Movements of a Polar Bear from Northern Alaska to Northern Greenland. *Arctic*, 48, 338–341.

Edwards SV, Beerli P (2000) Perspective: gene divergence, population divergence, and the variance in coalescence time in phylogeographic studies. *Evolutionary*, 54, 1839–1854.

Edwards CJ, Suchard MA, Lemey P *et al.* (2011) Ancient Hybridization and an Irish Origin for the Modern Polar Bear Matriline. *Current Biology*, 21, 1251–1258.

Ellegren H (2011) Sex-chromosome evolution: recent progress and the influence of male and female heterogamety. *Nature Reviews. Genetics*, 12, 157–166.

Ellegren H, Smeds L, Burri R *et al.* (2012) The genomic landscape of species divergence in Ficedula flycatchers. *Nature*, 491, 756-760.

Fahey AL, Ricklefs RE, Dewoody JA (2014) DNA-based approaches for evaluating historical demography in terrestrial vertebrates. *Biological Journal of the Linnean Society*, 112, 367-386.

Firestone KB, Elphinstone MS, Sherwin WB, Houlden B (1999) Phylogeographical population structure of tiger quolls *Dasyurus maculatus* (Dasyuridae: Marsupialia), an endangered carnivorous marsupial. *Molecular Ecology*, 8, 1613–1625.

Fohrmann P (2006) *Reisetagebuch eines Bären. Bruno alias JJ1*. Nicolai Verlag, Berlin.

Foster JW, Marshall Graves JA (1994) An SRY-related sequence on the marsupial X chromosome: Implications for the evolution of the mammalian testis-determining gene. *Proceedings of the National Academy of Sciences of the United States of America*, 91, 1927–1931.

Fredriksson G, Steinmetz R, Wong S, Garshelis D (2008) Helarctos malayanus. *The IUCN Red List of Threatened Species. Version 2014.3. <www.iucnredlist.org>. Downloaded on 09 February 2015.*

Freije D, Helms C, Watson MS, Donis-Keller H (1992) Identification of a Second Pseudoautosomal Region Near the Xq and Yq Telomeres. *Science*, 258, 1784–1786.

Fritts SH (1983) Record Dispersal by a Wolf From Minnesota. *Journal of Mammalogy*, 64, 166–167.

Frosch C (2014) Development and Application of Molecular Markers for Noninvasive Conservation Genetics of Brown Bears and Beavers in Europe. PhD Thesis. Goethe University, Frankfurt.

Frosch C, Dutsov A, Zlatanova D *et al.* (2014) Noninvasive genetic assessment of brown bear population structure in Bulgarian mountain regions. *Mammalian Biology*, 79, 268–276.

Fu Q, Li H, Moorjani P *et al.* (2014) Genome sequence of a 45,000-year-old modern human from western Siberia. *Nature*, 514, 445–449.

Fuchs M (2014) Entwicklung eines artspezifischen genetischen Geschlechtsmarkers für den Feldhamster (*Cricetus cricetus*). Bachelorarbeit. Goethe Universität, Frankfurt.

Garrick RC, Benavides E, Russello M *et al.* (2014) Lineage fusion in Galápagos giant tortoises. *Molecular Ecology*, 23, 5276–5290.

Garrick RC, Bonatelli IAS, Hyseni C *et al.* (2015) The evolution of phylogeographic datasets. *Molecular Ecology*, 24, 1164-1171.

Garshelis D, Crider D, van Manen F (2008a) Ursus americanus. *The IUCN Red List of Threatened Species. Version 2014.3. <www.iucnredlist.org>. Downloaded on 09 February 2015.*

Garshelis D, Ratnayeke S, Chauhan N (2008b) Melursus ursinus. *The IUCN Red List of Threatened Species. Version 2014.3. <www.iucnredlist.org>. Downloaded on 09 February 2015.*

Garshelis D, Steinmetz R (2008) Ursus thibetanus. *The IUCN Red List of Threatened Species. Version 2014.3. <www.iucnredlist.org>. Downloaded on 26 March 2015.*

Geraldes A, Carneiro M, Delibes-Mateos M *et al.* (2008) Reduced introgression of the Y chromosome between subspecies of the European rabbit (*Oryctolagus cuniculus*) in the Iberian Peninsula. *Molecular Ecology*, 17, 4489–4499.

Goldman D, Giri PR, O´Brien SJ (1989) Molecular Genetic-Distance Estimates Among the Ursidae as Indicated by One- and Two- Dimensional Protein Electrophoresis. *Evolution*, 43, 282–295.

Goldstein I, Velez-Liendo X, Paisley S, Garshelis D (2008) Tremarctos ornatus. *The IUCN Red List of Threatened Species. Version 2014.3. <www.iucnredlist.org>. Downloaded on 09 February 2015.*

Greenwood PJ (1980) Mating Systems, Philopatry and Dispersal in Birds and Mammals. *Animal Behaviour*, 28, 1140–1162.

Greminger MP, Krützen M, Schelling C, Pienkowska-Schelling A, Wandeler P (2010) The quest for Y-chromosomal markers - methodological strategies for mammalian non-model organisms. *Molecular Ecology Resources*, 10, 409–420.

Hailer F, Kutschera VE, Hallstrom BM *et al.* (2012) Nuclear Genomic Sequences Reveal that Polar Bears Are an Old and Distinct Bear Lineage. *Science*, 336, 344–347.

Hailer F (2015) Introgressive hybridization: brown bears as vectors for polar bear alleles. *Molecular Ecology*, 24, 1161–1163.

Haldane JBS (1922) Sex ratio and unisexual sterility in hybrid animals. *Journal of Genetics*, 12, 101–109.

Hallast P, Balaresque P, Bowden GR, Ballereau S, Jobling M (2013) Recombination Dynamics of a Human Y-Chromosomal Palindrome: Rapid GC-Biased Gene Conversion, Multi-kilobase Conversion Tracts, and Rare Inversions. *PLoS Genetics*, 9, e1003666.

Hallast P, Batini C, Zadik D *et al.* (2014) The Y-chromosome tree bursts into leaf: 13,000 high-confidence SNPs covering the majority of known clades. *Molecular Biology and Evolution*, 32, 661–673.

Hallström BM, Janke A (2010) Mammalian evolution may not be strictly bifurcating. *Molecular Biology and Evolution*, 27, 2804–16.

Harris H (1966) Enzyme Polymorphism in Man. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 164, 298–310.

Hawley RS (2003) The Human Y Chromosome: Rumors of Its Death Have Been Greatly Exaggerated. *Cell*, 113, 825–828.

Hellborg L, Ellegren H (2004) Low levels of nucleotide diversity in mammalian Y chromosomes. *Molecular Biology and Evolution*, 21, 158–163.

Hellborg L, Gündüz I, Jaarola M (2005) Analysis of sex-linked sequences supports a new mammal species in Europe. *Molecular Ecology*, 14, 2025–31.

Hewitt G (2000) The genetic legacy of the Quaternary ice ages. *Nature*, 405, 907–913.

Hirata D, Mano T, Abramov AV *et al.* (2013) Molecular phylogeography of the brown bear (*Ursus arctos*) in Northeastern Asia based on analyses of complete mitochondrial DNA sequences. *Molecular Biology and Evolution*, 30, 1644–1652.

Hughes JF, Rozen S (2012) Genomics and genetics of human and primate Y chromosomes. *Annual Review of Genomics and Human Genetics*, 13, 83–108.

Hughes JF, Skaletsky H, Brown LG *et al.* (2012) Strict evolutionary conservation followed rapid gene loss on human and rhesus Y chromosomes. *Nature*, 483, 82-86.

Hughes JF, Skaletsky H, Pyntikova T *et al.* (2010) Chimpanzee and human Y chromosomes are remarkably divergent in structure and gene content. *Nature*, 463, 536–539.

Hurles ME, Jobling M (2001) Haploid chromosomes in molecular ecology: lessons from the human Y. *Molecular Ecology*, 10, 1599–613.

Janke A, Feldmaier-Fuchs G, Thomas WK, von Haeseler A, Pääbo S (1994) The Marsupial Mitochondrial Genome and the Evolution of Placental Mammals. *Genetics*, 137, 243–256.

Jarne P, Lagoda PJ (1996) Microsatellites, from molecules to populations and back. *Trends in Ecology & Evolution*, 11, 424–329.

Jobling M, Hollox E, Hurles M, Kivisild T, Tyler-Smith C (2014) *Human Evolutionary Genetics*. Garland Science, Taylor & Francis Group, New York.

Jobling MA, Tyler-Smith C (1995) Fathers and sons: the Y chromosome and human evolution. *Trends in Genetics*, 11, 449–456.

Karmin M, Saag L, Vicente M *et al.* (2015) A recent bottleneck of Y chromosome diversity coincides with a global change in culture. *Genome Research*, 25, 1–8.

Kayser M (2007) Uni-parental markers in human identity testing including forensic DNA analysis. *BioTechniques*, 43, Sxv–Sxxi.

Keis M, Remm J, Ho SYW *et al.* (2013) Complete mitochondrial genomes and a novel spatial genetic method reveal cryptic phylogeographical structure and migration patterns among brown bears in north-western Eurasia. *Journal of Biogeography*, 40, 915–927.

Kocher TD, Thomas WK, Meyer A *et al.* (1989) Dynamics of mitochondrial DNA evolution in animals: Amplification and sequencing with conserved primers. *Proceedings of the National Academy of Sciences of the United States of America*, 86, 6196–6200.

Koepfli K-P, Paten B, O'Brien SJ (2015) The Genome 10K Project: A Way Forward. *Annual review of animal biosciences*, 3, 57–111.

Kondo M, Hornung U, Nanda I *et al.* (2006) Genomic organization of the sex-determining and adjacent regions of the sex chromosomes of medaka. *Genome Research*, 16, 815–826.

Kopatz A, Eiken HG, Aspi J *et al.* (2014) Admixture and gene flow from Russia in the recovering Northern European brown bear (*Ursus arctos*). *PloS one*, 9, e97558.

Kopatz A, Eiken HG, Hagen SB *et al.* (2012) Connectivity and population subdivision at the fringe of a large brown bear (*Ursus arctos*) population in North Western Europe. *Conservation Genetics*, 13, 681–692.

Krause J, Unger T, Noçon A *et al.* (2008) Mitochondrial genomes reveal an explosive radiation of extinct and extant bears near the Miocene-Pliocene boundary. *BMC Evolutionary Biology*, 8, 220.

Krzywinski J, Nusskern DR, Kern MK, Besansky NJ (2004) Isolation and Characterization of Y Chromosome Sequences From the African Malaria Mosquito *Anopheles gambiae*. *Genetics*, 166, 1291–1302.

Kurtén B, Anderson E (1980) *Pleistocene Mammals of North America*. Columbia University Press.

Kutschera VE, Bidon T, Hailer F *et al*. (2014) Bears in a forest of gene trees: Phylogenetic inference is complicated by incomplete lineage sorting and gene flow. *Molecular Biology and Evolution* 31, 2004-2017.

Lahn BT, Page DC (1997) Functional Coherence of the Human Y Chromosome. *Science*, 278, 675–680.

Lahn BT, Page DC (1999) Four Evolutionary Strata on the Human X Chromosome. *Science*, 286, 964–967.

Laidre KL, Born EW, Gurarie E *et al.* (2013) Females roam while males patrol: divergence in breeding season movements of pack-ice polar bears (*Ursus maritimus*). *Proceedings of the Royal Society B-Biological Sciences*, 280, 20122371.

Lamichhaney S, Berglund J, Almén MS *et al.* (2015) Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature*.

Landsteiner K (1900) Zur Kenntnis der antifermentativen, lytischen und agglutinierenden Wirkungen des Blutserums und der Lymphe. *Zbl Bakt*, 27, 357–362.

Lawson Handley LJ, Berset-Brändli L, Perrin N (2006) Disentangling reasons for low Y chromosome variation in the greater white-toothed shrew (*Crocidura russula*). *Genetics*, 173, 935–942.

Lawson Handley LJ, Perrin N (2007) Advances in our understanding of mammalian sex-biased dispersal. *Molecular Ecology*, 16, 1559–1578.

Leonard JA, Wayne RK, Cooper A (2000) Population genetics of ice age brown bears. *Proceedings of the National Academy of Sciences of the United States of America*, 97, 1651–1654.

Lerp H, Wronski T, Plath M, Schröter A, Pfenninger M (2013) Phylogenetic and population genetic analyses suggest a potential species boundary between Mountain (*Gazella gazella*) and Arabian Gazelles (*G. arabica*) in the Levant. *Mammalian Biology*, 78, 383–386.

Lewontin RC, Hubby JL (1966) A Molecular Approach to the Study of Genic Heterozygosity in Natural Populations. II. Amount of Variation and Degree of Heterozygostiy in Natural Populations of *Drosophila Pseudoobscura*. *Genetics*, 54, 595–609.

Li G, Davis BW, Raudsepp T *et al.* (2013) Comparative analysis of mammalian Y chromosomes illuminates ancestral structure and lineage-specific evolution. *Genome Research*, 1486–1495.

Li WH, Yi S, Makova K (2002) Male-driven evolution. *Current opinion in genetics & development*, 12, 650–6.

Li B, Zhang G, Willerslev E, Wang J (2011) Genomic data from the Polar Bear (Ursus maritimus). *GigaScience*, http://dx.doi.org/10.5524/100008.

Lindqvist C, Schuster S, Sun Y *et al*. (2010) Complete mitochondrial genome of a Pleistocene jawbone unveils the origin of polar bear. *Proceedings of the National Academy of Sciences of the United States of America*, 107, 5053-5057.

Liu S, Lorenzen ED, Fumagalli M *et al.* (2014) Population Genomics Reveal Recent Speciation and Rapid Evolutionary Adaptation in Polar Bears. *Cell*, 157, 785–794.

Lone K, Aars J, Ims RA (2012) Site fidelity of Svalbard polar bears revealed by mark-recapture positions. *Polar Biology*, 36, 27–39.

Lü Z, Wang D, Garshelis D (2008) Ailuropoda melanoleuca. *The IUCN Red List of Threatened Species. Version 2014.3. <www.iucnredlist.org>. Downloaded on 09 February 2015.*

Malaspina P, Persichetti F, Novelletto A *et al.* (1990) The human Y chromosome shows a low level of DNA polymorphism. *Annals of human genetics*, 54, 297–305.

Masly JP, Presgraves DC (2007) High-resolution genome-wide dissection of the two rules of speciation in Drosophila. *PLoS biology*, 5, e243.

Marshall Graves JA (2000) Human Y Chromosome, Sex Determination, and Spermatogenesis—A Feminist View. *Biology of Reproduction*, 63, 667–676.

Matthee CA, Eick G, Willows-Munro S *et al.* (2007) Indel evolution of mammalian introns and the utility of non-coding nuclear markers in eutherian phylogenetics. *Molecular Phylogenetics and Evolution*, 42, 827–837.

McLellan B, Hovey FW (2001) Natal dispersal of grizzly bears. *Canadian Journal of Zoology*, 79, 838–844.

McLellan B, Reiner DC (1994) A Review of Bear Evolution. *Int. Conf. Bear Res. and Manage.*, 9, 85–96.

McLellan B, Servheen C, Huber D (2008) Ursus arctos. *The IUCN Red List of Threatened Species. Version 2014.2. <www.iucnredlist.org>. Downloaded on 04 November 2014.*

McLoughlin PD, Case RL, Gau RJ, Ferguson SH, Messier F (1999) Annual and Seasonal Movement Patterns of Barren-Ground Grizzly Bears in the Central Northwest Territories. *Ursus*, 11, 79–86.

Miller W, Schuster SC, Welch AJ *et al.* (2012) Polar and brown bear genomes reveal ancient admixture and demographic footprints of past climate change. *Proceedings of the National Academy of Sciences*, 109, E2382–E2390.

Ming R, Bendahmane A, Renner SS (2011) Sex chromosomes in land plants. *Annual review of plant biology*, 62, 485–514.

Moore WS (1995) Inferring Phylogenies from mtDNA Variation: Mitochondrial-Gene Trees Versus Nuclear-Gene Trees. *Evolution*, 49, 718–726.

Muller HJ (1964) The Relation of Recombination to Mutational Advance. *Mutation Research*, 1, 2–9.

Mullis K, Faloona F, Scharf S *et al.* (1986) Specific Enzymatic Amplification of DNA in Vitro: The Polymerase Chain Reaction. *Cold Spring Harbor Symposia on Quantitative Biology*, LI, 263–273.

Murphy WJ, Pearks Wilkerson AJ, Raudsepp T *et al.* (2006) Novel gene acquisition on carnivore Y chromosomes. *PLoS genetics*, 2, e43.

Nakagome S, Pecon-Slattery J, Masuda R (2008) Unequal rates of Y chromosome gene divergence during speciation of the family Ursidae. *Molecular Biology and Evolution*, 25, 1344–56.

Nowak C, Domokos C, Dutsov A, Frosch C (2014) Molecular evidence for historic long-distance translocations of brown bears in the Balkan region. *Conservation Genetics*, 15, 743–747.

O´Brien S, Nash W, Wildt D, Bush M, Benveniste R (1985) A molecular solution ot the riddle of the giant panda´s phylogeny. *Nature*, 317, 140–144.

Ogden R, Dawnay N, McEwing R (2009) Wildlife DNA forensics—bridging the gap between conservation genetics and law enforcement. *Endangered Species Research*, 9, 179–195.

Ohno S (1969) Evolution of sex chromosomes in mammals. *Annual Review of Genetics*, 3, 495–524.

Oota H, Settheetham-Ishida W, Tiwawech D, Ishida T, Stoneking M (2001) Human mtDNA and Y-chromosome variation is correlated with matrilocal versus patrilocal residence. *Nature genetics*, 29, 20–21.

Paetkau D, Amstrup SC, Born EW *et al.* (1999) Genetic structure of the world's polar bear populations. *Molecular Ecology*, 8, 1571–1584.

Paetkau D, Waits LP, Clarkson PL, Craighead L, Strobeck C (1997) An Empirical Evaluation of Genetic Distance Statistics Using Microsatellite Data From Bear (Ursidae) Populations. *Genetics*, 147, 1943–1957.

Pagès M, Calvignac S, Klein C *et al.* (2008) Combined analysis of fourteen nuclear genes refines the Ursidae phylogeny. *Molecular Phylogenetics and Evolution*, 47, 73–83.

Pagès M, Maudet C, Bellemain E *et al.* (2009) A system for sex determination from degraded DNA: a useful tool for palaeogenetics and conservation genetics of ursids. *Conservation Genetics*, 10, 897–907.

Painter TS (1921) The Y-Chromosome in Mammals. *Science*, 53, 503–504.

Pamilo P, Nei M (1988) Relationships between gene trees and species trees. *Molecular Biology and Evolution*, 5, 568–583.

Pareek CS, Smoczynski R, Tretyn A (2011) Sequencing technologies and genome sequencing. *Journal of applied genetics*, 52, 413–435.

Pavlova A, Amos JN, Joseph L *et al.* (2013) Perched at the mito-nuclear crossroads: divergent mitochondrial lineages correlate with environment in the face of ongoing nuclear gene flow in an Australian bird. *Evolution; international journal of organic evolution*, 67, 3412–3428.

Peacock E, Sonsthagen S, Obbard ME *et al.* (2015) Implications of the circumpolar genetic structure of polar bears for their conservation in a rapidly warming arctic. *PloS one*, 10, e112021.

Pérez T, Hammer SE, Albornoz J, Domínguez A (2011) Y-chromosome phylogeny in the evolutionary net of chamois (genus Rupicapra). *BMC evolutionary biology*, 11, 272.

Petit E, Balloux F, Excoffier L (2002) Mammalian population genetics: why not Y? *Trends in Ecology & Evolution*, 17, 28–33.

Petit RJ, Excoffier L (2009) Gene flow and species delimitation. *Trends in Ecology & Evolution*, 24, 386–393.

Pidancier N, Jordan S, Luikart G, Taberlet P (2006) Evolutionary history of the genus Capra (Mammalia, Artiodactyla): discordance between mitochondrial DNA and Y-chromosome phylogenies. *Molecular phylogenetics and evolution*, 40, 739–749.

Pilot M, Jedrzejewski W, Branicki W *et al.* (2006) Ecological factors influence population genetic structure of European grey wolves. *Molecular ecology*, 15, 4533–53.

Posada D, Crandall K, Holmes EC (2002) Recombination in evolutionary genomics. *Annual Review of Genetics*, 36, 75–97.

Potter SS, Newbold JE, Hutchison CA, Edgell MH (1975) Specifc cleavage analysis of mammalian mitochondrial DNA. *Proceedings of the National Academy of Sciences of the United States of America*, 72, 4496–4500.

Poznik GD, Henn BM, Yee M-C *et al.* (2013) Sequencing Y Chromosomes Resolves Discrepancy in Time to Common Ancestor of Males Versus Females. *Science*, 341, 562–565.

Proctor MF, Mclellan BN, Strobeck C, Barclay RMR (2004) Gender-specific dispersal distances of grizzly bears estimated by genetic analysis. *Canadian Journal of Zoology*, 82, 1108–1118.

Prugnolle F, de Meeus T (2002) Inferring sex-biased dispersal from population genetic tools: a review. *Heredity*, 88, 161–165.

Pusey A (1987) Sex-biased dispersal and inbreeding avoidance in birds and mammals. *Trends in Ecology & Evolution*, 2, 295–299.

Qvarnström a, Bailey RI (2009) Speciation through evolution of sex-linked genes. *Heredity*, 102, 4–15.

Rannala B, Yang Z (2003) Bayes Estimation of Species Divergence Times and Ancestral Population Sizes Using DNA Sequences From Multiple Loci. *Genetics*, 164, 1645–1656.

Repping S, Skaletsky H, Lange J *et al.* (2002) Recombination between Palindromes P5 and P1 on the Human Y Chromosome Causes Massive Deletions and Spermatogenic Failure. *American Journal of Human Genetics*, 71, 906–922.

Rice WR (1987) The Accumulation of Sexually Antagonistic Genes as a Selective Agent Promoting the Evolution of Reduced Recombination between Primitive Sex Chromosomes. *Evolution*, 41, 911–914.

Roca AL, Georgiadis N, O'Brien SJ (2005) Cytonuclear genomic dissociation in African elephant species. *Nature genetics*, 37, 96–100.

Ross MT, Grafham DV, Coffey AJ *et al.* (2005) The DNA sequence of the human X chromosome. *Nature*, 434, 325–337.

Rubinoff D, Holland B (2005) Between Two Extremes: Mitochondrial DNA is neither the Panacea nor the Nemesis of Phylogenetic and Taxonomic Inference. *Systematic Biology*, 54, 952–961.

Sacco T, Van Valkenburgh B (2004) Ecomorphological indicators of feeding behaviour in the bears (Carnivora: Ursidae). *Journal of Zoology*, 263, 41–54.

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*, 74, 5463–5467.

Sankararaman S, Mallick S, Dannemann M *et al.* (2014) The genomic landscape of Neanderthal ancestry in present-day humans. *Nature*.

Sarich VM (1973) The Giant Panda is a Bear. *Nature*, 245, 218–220.

Schliebe S, Wiig Ø, Derocher A, Lunn N (2008) Ursus maritimus. *The IUCN Red List of Threatened Species. Version 2014.2. <www.iucnredlist.org>. Downloaded on 04 November 2014.*

Scozzari R, Massaia A, Trombetta B *et al.* (2014) An unbiased resource of novel SNP markers provides a new chronology for the human Y chromosome and reveals a deep phylogenetic structure in Africa. *Genome research*.

Seielstad MT, Minch E, Cavalli-sforza LL (1998) Genetic evidence for a higher female migration rate in humans. *Nature genetics*, 20, 278–280.

Servheen C (1983) Grizzly Bear Food Habits, Movements, and Habitat Selection in the Mission Mountains, Montana. *Journal of Wildlife Management*, 47, 1026–1035.

Shields GF, Kocher TD (1991) Phylogenetic Relationships of North American Ursids Based on Analysis of Mitochondrial DNA. *Evolution*, 45, 218–221.

Sinclair AH, Berta P, Palmer M *et al.* (1990) A Gene from the Human Sex-Determining Region Encodes a Protein with homology to a conserved DNA-binding motif. *Nature*, 346, 240–244.
Singh ND, Koerich LB, Carvalho AB, Clark AG (2014) Positive and Purifying Selection on the Drosophila Y Chromosome. *Molecular biology and evolution*, 31, 2612–2623.
Skaletsky H, Kuroda-Kawaguchi T, Minx PJ *et al.* (2003) The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature*, 423, 825–37.
Soh YQS, Alföldi J, Pyntikova T *et al.* (2014) Sequencing the Mouse Y Chromosome Reveals Convergent Gene Acquisition and Amplification on Both Sex Chromosomes. *Cell*.
Statham MJ, Murdoch J, Janecka J *et al.* (2014) Range-wide multilocus phylogeography of the red fox reveals ancient continental divergence, minimal genomic exchange and distinct demographic histories. *Molecular ecology*, 23, 4813–4830.
Steinlechner M, Berger B, Niederstätter H, Parson W (2002) Rare failures in the amelogenin sex test. *Internation Journal of Legal Medicine*, 116, 117–120.
Stern C (1957) The Problem of Complete Y-Linkage in Man. *The American Journal of Human Genetics*, 9, 147–166.
Stirling I (2012) *Polar Bears. The natural history of a threatened species*. Bloomsbury Publishing.
Stirling I, Oritsland NA (1995) Relationships between estimates of ringed seal (Phoca hispida) and polar bear (Ursus marifimus) populations in the Canadian Arctic. *Can Journal of Fisheries and Aquatic Sciences*, 52, 2594–2612.
Støen O-G, Zedrosser A, Saebø S, Swenson JE (2006) Inversely density-dependent natal dispersal in brown bears Ursus arctos. *Oecologia*, 148, 356–64.
Sundqvist A-K, Ellegren H, Olivier M, Vilà C (2001) Y chromosome haplotyping in Scandinavian wolves (Canis lupus) based on microsatellite markers. *Molecular ecology*, 10, 1959–66.
Taberlet P, Bouvet J (1994) Mitochondrial DNA polymorphism, phylogeophraphy, and conservation genetics of the brown bear Ursus acrctos in Europe. *Proceedings of the Royal Society B-Biological Sciences*, 255, 195–200.
Taberlet P, Camarra J-J, Griffin S *et al.* (1997) Noninvasive genetic tracking of the endangered Pyrenean brown bear population. *Molecular Ecology*, 869–876.
Taberlet P, Mattock H, Dubois-Paganon C, Bouvet J (1993) Sexing free-ranging brown bears Ursus arctos using hairs found in the field. *Molecular Ecology*, 2, 399–403.
Talbot S, Shields G (1996a) A Phylogeny of the Bears (Ursidae) Inferred from Complete Sequences of Three Mitochondrial Genes. *Molecular Phylogenetics and Evolution*, 5, 567–575.
Talbot S, Shields G (1996b) Phylogeography of brown bears (Ursus arctos) of Alaska and paraphyly within the Ursidae. *Molecular Phylogenetics and Evolution*, 5, 477–494.
Tammeleht E, Remm J, Korsten M *et al.* (2010) Genetic structure in large, continuous mammal populations: the example of brown bears in northwestern Eurasia. *Molecular ecology*, 19, 5359–70.
Tautz D, Schlötterer C (1994) Simple sequences. *Current Opinion in Genetics & Development*, 4, 832–837.

Taylor PG, Donnelly P, de Knijff P *et al.* (1998) Jefferson fathered slave's last child. *Nature*, 396, 27–28.

Tchaicka L, Eizirik E, De Oliveira TG, Cândido JF, Freitas TRO (2007) Phylogeography and population history of the crab-eating fox (*Cerdocyon thous*). *Molecular Ecology*, 16, 819–38.

Toews DPL, Brelsford A (2012) The biogeography of mitochondrial and nuclear discordance in animals. *Molecular Ecology*, 3907–3930.

Tosi AJ, Morales JC, Melnick DJ (2000) Comparison of Y Chromosome and mtDNA Phylogenies Leads to Unique Inferences of Macaque Evolutionary History. *Molecular Phylogenetics and Evolution*, 17, 133–144.

Trewhella WJ, Harris S, McAllister FE (1988) Dispersal Distance, Home-Range Size and Population Density in the Red Fox (Vulpes vulpes): A Quantitative Analysis. *Journal of Applied Ecology*, 25, 423–434.

Tringali MD, Davis MC, Rodriguez-Lopez M *et al.* (2008) Simultaneous use of the X- and Y-chromosome genes SMCX, SMCY, and DBY for sex determination in the Florida manatee (*Trichechus manatus latirostris*). *Marine Mammal Science*, 24, 218–224.

Tucker PK, Sage RD, Warner J, Wilson AC, Eicher EM (1992) Abrupt Cline for Sex Chromosomes in a Hybrid Zone between Two Species of Mice. *Evolution*, 46, 1146–1163.

Underhill P, Kivisild T (2007) Use of Y chromosome and mitochondrial DNA population structure in tracing human migrations. *Annual Review of Genetics*, 41, 539–64.

Vigneault G, Zouros E (1986) The Genetics of Asymmetical Male Sterility in Drosophila mojavensis and Drosophila arizonensis Hybrids: Interactions Between the Y-Chromosome and Autosomes. *Evolution*, 40, 1160–1170.

Waits L, Sullivan J, O'Brien SJ, Ward RH (1999) Rapid Radiation Events in the Family Ursidae Indicated by Likelihood Phylogenetic Estimation from Multiple Fragments of mtDNA. *Molecular Phylogenetics and Evolution*, 13, 82–92.

Waits L, Taberlet P, Swenson JE, Sandegren F, Franzén R (2000) Nuclear DNA microsatellite analysis of genetic diversity and gene flow in the Scandinavian brown bear (*Ursus arctos*). *Molecular Ecology*, 9, 421–431.

Waits L, Talbot S, Ward RH, Shields GF (1998) Mitochondrial DNA Phylogeography of the North American Brown Bear and Implications for Conservation. *Conservation Biology*, 12, 408–417.

Waters PD, Wallis MC, Marshall Graves J (2007) Mammalian sex - Origin and evolution of the Y chromosome and SRY. *Seminars in Cell & Developmental Biology*, 18, 389–400.

Wayne RK, Van Valkenburgh B, O'Brien SJ (1991) Molecular distance and divergence time in carnivores and primates. *Molecular Biology and Evolution*, 8, 297–319.

Wei W, Ayub Q, Xue Y, Tyler-Smith C (2013) A comparison of Y-chromosomal lineage dating using either resequencing or Y-SNP plus Y-STR genotyping. *Forensic science international. Genetics*, 7, 568–572.

Willard HF (2003) Tales of the Y chromosome. *Nature*, 423, 810–813.

Wilson AC, Cann RL, Carrii SM *et al.* (1985) Mitochondrial DNA and two perspectives on evolutionary genetics. *Biological Journal of the Linnean Society*, 375–400.

Wilson Sayres M, Lohmueller K, Nielsen R (2014) Natural selection reduced diversity on human y chromosomes. *PLoS genetics*, 10, e1004064.

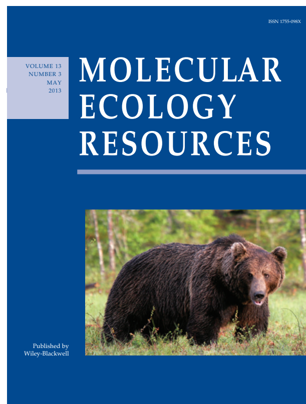
Xue Y, Wang Q, Long Q *et al.* (2009) Human Y chromosome base-substitution mutation rate measured by direct sequencing in a deep-rooting pedigree. *Current Biology ,* 19, 1453–1457.
Yamamoto K, Tsubota T, Komatsu T *et al.* (2002) Sex identification of Japanese black bear, *Ursus thibetanus japonicus*, by PCR based on amelogenin gene. *The Journal of Veterinary Medical Science*, 64, 505–8.
Yu L, Li Q, Ryder OA, Zhang Y (2004) Phylogeny of the bears (Ursidae) based on nuclear and mitochondrial genes. *Molecular Phylogenetics and Evolution*, 32, 480–94.
Yu L, Li Y-W, Ryder OA, Zhang Y (2007) Analysis of complete mitochondrial genome sequences increases phylogenetic resolution of bears (Ursidae), a mammalian family that experienced rapid speciation. *BMC Evolutionary Biology*, 7, 198.
Zedrosser A, Støen O-G, Sæbø S, Swenson JE (2007) Should I stay or should I go? Natal dispersal in the brown bear. *Animal Behaviour*, 74, 369–376.
Zhang Y-P, Ryder O (1994) Phylogenetic Relationships of Bears (the Ursidae) Inferred from Mitochondrial DNA Sequences. *Molecular Phylogenetics and Evolution*, 3, 351–359.
Zhao S, Zheng P, Dong S *et al.* (2013) Whole-genome sequencing of giant pandas provides insights into demographic history and local adaptation. *Nature Genetics*, 45, 67–71.
Zimmermann F, Breitenmoser-Würsten C, Breitenmoser U (2005) Natal dispersal of Eurasian lynx (*Lynx lynx*) in Switzerland. *Journal of Zoology*, 267, 381–395.
Zink RM, Barrowclough GF (2008) Mitochondrial DNA under siege in avian phylogeography. *Molecular Ecology*, 17, 2107–21.

PUBLICATIONS

PUBLICATION 1: A SENSITIVE AND SPECIFIC MULTIPLEX PCR APPROACH FOR SEX IDENTIFICATION OF URSINE AND TREMARCTINE BEARS SUITABLE FOR NON-INVASIVE SAMPLES

Tobias Bidon, Christiane Frosch, Hans G. Eiken, Verena E. Kutschera, Snorre B. Hagen, Siv G. Aarnes, Steven R. Fain, Axel Janke, Frank Hailer

Published in *Molecular Ecology Resources* (2013) 13:362–368



Erklärung über Anteile der Autoren/Autorinnen an den einzelnen Kapiteln der Promotionsarbeit

Titel der Publikation/des Manuskripts: A sensitive and specific multiplex PCR approach for sex identification of ursine and tremarctine bears suitable for non-invasive samples

	Was hat der/die Promovierende bzw. Was haben die Co-Autoren/Autorinnen biegetragen#	Name des/der jeweiligen Autors/Autoren/ Autorin*
(1) Entwicklung und Planung	50% 10% 40%	TB CF, HGE FH, SRF, SBH, AJ
(2) Durchführung der einzelnen Untersuchungen/ Experimente	Primer design Multiplex Entwicklung, Probenorganisation, DNA-Exktraktion, In-vitro Analysen verschieden Bären + andere Arten DNA Extraktion, Fragmentanalyse Zoo Individuen + andere Arten In-vitro Analyse andere Arten Fragmentanalyse norwegische Bären, Sensitivitätstest	TB, FH TB CF TB, VEK SGA
(3) Erstellung der Daten-sammlung und Abbildungen	Zusammenstellung / Koordination der Ergebnisse verschieden Bären und andere Arten, alle Abbildungen und Tabellen Zusammenstellung der Ergebnisse Zoo-Individuen Zusammenstellung der Ergebnisse norwegische Bären Online Archivierung der Sequenzen	TB CF HGE TB
(4) Analyse/Interpretation der Daten	Interpretation Bären und andere Arten Mikrosatellitenanalysen Sensitivitätstest Zusammenfassende Übersicht	TB CF HGE TB, FH
(5) übergeordnete Einleitung/ Ergebnisse/Diskussion	70% 30%	TB FH, HGE, AJ

#Bei 2, 3 und 4 bitte kurze inhaltliche Angaben der jeweiligen Anteile, bei 1 und 5 reichen prozentuale Angaben
*Mehrfacheintragungen möglich

Als Autoren/Autorinnen werden solche Personen bezeichnet, die an der Arbeit in Bezug auf die genannten Punkte in einer Weise mitgewirkt haben, dass sie für die ausgewiesenen Passagen (mit) verantwortlich sind. Personen, die an der Arbeit mitgewirkt haben, jedoch nicht in diese Kategorie fallen, sollten in der Danksagung Erwähnung finden.

Datum/Ort

Datum
zustimmende Bestätigung der vorgenannten Angaben

Unterschrift Promovend/Promovendin

Unterschrift Betreuer/Betreuerin

MOLECULAR ECOLOGY
RESOURCES

Molecular Ecology Resources (2013) **13**, 362–368 doi: 10.1111/1755-0998.12072

A sensitive and specific multiplex PCR approach for sex identification of ursine and tremarctine bears suitable for non-invasive samples

TOBIAS BIDON,* CHRISTIANE FROSCH,†‡ HANS G. EIKEN,§ VERENA E. KUTSCHERA,* SNORRE B. HAGEN,§ SIV G. AARNES,§ STEVEN R. FAIN,** AXEL JANKE†† and FRANK HAILER*

**LOEWE Biodiversity and Climate Research Centre (BiK-F), Senckenberg Gesellschaft für Naturforschung, Senckenberganlage 25, 60325 Frankfurt am Main, Germany, †Conservation Genetics Group, Senckenberg Research Institute and Natural History Museum, Clamecystrasse 12, 63571 Gelnhausen, Germany, ‡Frankfurt Zoological Society, Bernhard-Grzimek-Allee 1, 60316 Frankfurt am Main, Germany, §Bioforsk Soil and Environment, Svanhovd, Norwegian Institute for Agricultural and Environmental Research, NO-9925 Svanvik, Norway, **National Fish and Wildlife Forensic Laboratory, 1490 East Main Street, Ashland, 97520, OR USA, ††Evolution & Diversity, Goethe University, Institute for Ecology, Siesmayerstrasse 70, 60323 Frankfurt am Main, Germany*

Abstract

We report a new approach for molecular sex identification of extant *Ursinae* and *Tremarctinae* bears. Two Y-specific fragments (*SMCY* and *318.2*) and one X-specific fragment (*ZFX*) are amplified in a multiplex PCR, yielding a double test for male-specific amplification and an internal positive control. The primers were designed and tested to be bear-specific, thereby minimizing the risk of cross-amplification in other species including humans. The high sensitivity and small amplicon sizes (100, 124, 160 base pairs) facilitate analysis of non-invasively obtained DNA material. DNA from tissue and blood as well as from 30 non-invasively collected hair and faeces yielded clear and easily interpretable results. The fragments were detected both by standard gel electrophoresis and automated capillary electrophoresis.

Keywords: bears, molecular sexing, multiplex PCR, non-invasive, *Ursidae*, *Ursus arctos*

Received 26 October 2012; revision received 14 December 2012; accepted 22 December 2012

Introduction

Knowledge of the sex of individuals is essential in population ecology (Millar *et al.* 1996) and provides crucial data for management and conservation programs (Taberlet *et al.* 1993). However, determining sex in the field can be challenging because species might be rare, secretive or lack distinct sexual dimorphism (Rosel 2003). To avoid disturbance, catching and direct sampling of wild animals might not be feasible. Advances in molecular techniques have overcome these constraints by using non-invasively collected samples like hair or faeces as a source of DNA and determining the sex by genetic means. These methods can also allow for the sexing of forensic evidence and ambiguous or historical samples that were collected with questionable or without any gender assignation. Simple and precise molecular methods have proven expedient in large carnivores, for example, in studies of Pyrenean and Italian brown bears (*Ursus arctos*), where the ratio of males and females has been determined based on non-invasive sampling (e.g. Kohn *et al.* 1995; Taberlet *et al.* 1997).

Correspondence: Tobias Bidon and Frank Hailer, Fax: +49 (0)69 75 42 1800; E-mails: tobias.bidon@senckenberg.de; frashai@gmx.net

In eutherian mammals, discrimination between sexes based on molecular genetic methods has been done by genotyping a sex-specific length polymorphism between homologous X- and Y-chromosome-linked loci (Yamamoto *et al.* 2002; Shaw *et al.* 2003; Xu *et al.* 2008), or by the detection of a Y-chromosomal fragment that is not present in females (Waits & Paetkau 2005). As the absence or presence of the Y-fragment decides whether the specimen being tested will be interpreted as male or female, an internal PCR positive control is required (Taberlet *et al.* 1993). Amplification failure of the Y-specific fragment may otherwise lead to misidentification of males as females. Post-PCR restriction site analysis has also been used (Aasen & Medrano 1990; Amstrup *et al.* 1993), but adds another working step increasing time, cost, and augments the risk of contamination (Ahlering *et al.* 2011).

© 2013 Blackwell Publishing Ltd

There are several current methods for molecular sexing of bears, many of them suboptimal for different reasons (different molecular sexing approaches are illustrated in Fig. 1; see also Pagès *et al.* 2009). (i) Long PCR fragments impede the analysis of degraded DNA from non-invasive samples (Wasser *et al.* 1997; Shaw *et al.* 2003; Wandeler *et al.* 2003). (ii) Methods where the diagnostic Y-fragment is larger than the X-fragment risk misclassifying low-quality male samples as females, because longer fragments are more prone to amplification failure (Wandeler *et al.* 2003). (iii) Use of conserved primers that are not specific to the target species are sensitive to contamination and can lead to wrong sex assignment. For instance, primers binding to human DNA (sample collectors and lab personnel), DNA of ingested prey (when faecal samples are analysed) or DNA from other species (e.g. scent marking by several species at the collection site) may cause such errors (Wasser *et al.* 1997; Yamamoto *et al.* 2002; Murphy *et al.* 2003). Furthermore, scats of different carnivores can be confused in the field (Farrell *et al.* 2000). (iv) Homology between regions on the X- and Y-chromosome can lead to faint Y-like bands in females and can add uncertainty to the differentiation between the sexes, especially if only a low amount of DNA is available (Taberlet *et al.* 1993; Woods *et al.* 1999; Pagès *et al.* 2009). Modifications and enhancements of existing methods employing one Y- and one X-linked locus have been established (e.g. Pagès *et al.* 2009; Kopatz *et al.* 2012).

Here, we report a novel multiplex PCR approach designed to overcome the limitations of current methods, yielding reliable and efficient results for sex determination in all *Ursinae* and *Tremarctinae* bears. The method is based on primers that target two Y-specific markers (*SMCY*, intron 4, 100 bp; *318.2*, 124 bp) and one X-linked locus (*ZFX*, final intron, 160 bp). This double control of Y-specific amplification combined with an internal PCR positive control increases the reliability of result interpretation (Tringali *et al.* 2008; Ahlering *et al.* 2011).

Materials and methods

Marker development

Primer design was done as in Ahlering *et al.* (2011), focusing on the ability of the primers to bind in all extant bear species, while remaining bear-specific. A bear consensus sequence, including brown bear, polar bear (*U. maritimus*), black bear (*U. americanus*), Asian black bear (*U. thibetanus*), sun bear (*Helarctos malayanus*), sloth bear (*Melursus ursinus*) (all Ursinae), spectacled bear (*Tremarctos ornatus*, Tremarctinae) and giant panda (*Ailuropoda melanoleuca*, Ailuropodinae) was created for the *SMCY* gene (intron 4) sequences (EMBL accession numbers: AB261823.1 - AB261830.1) and the *ZFX* gene (final intron) sequences (AB261815.1 - AB261822.1). The *SMCY* gene codes for a male-specific *MHC* antigen and has been used as a sexing marker in mammals (e.g. Tringali *et al.* 2008). *ZFX* encodes a zinc finger protein that might be involved in cell proliferation and differentiation (Fang *et al.* 2012). These sequences were aligned to orthologous non-bear sequences to identify primer locations that would avoid cross-reaction in other species. The alignment of the internal positive control (*ZFX*) included, besides all eight extant bear species, sequences from humans (*Homo sapiens*), several carnivores (*Halichoerus grypus, Mustela nivalis, Martes martes, Lutra lutra, Canis lupus, Lynx lynx, Felis catus*), ungulates (*Bos taurus, Cervus elaphus, Equus caballus, Equus przewalskii*) and rodents (*Mus musculus, Rattus norvegicus*) (JN603636.1, AB491603.1, FN421124.1, AB491606, NG_021253.1, AB62-2136.1, DQ086448.1, DQ086431, AY241223.2, EF693913.1,

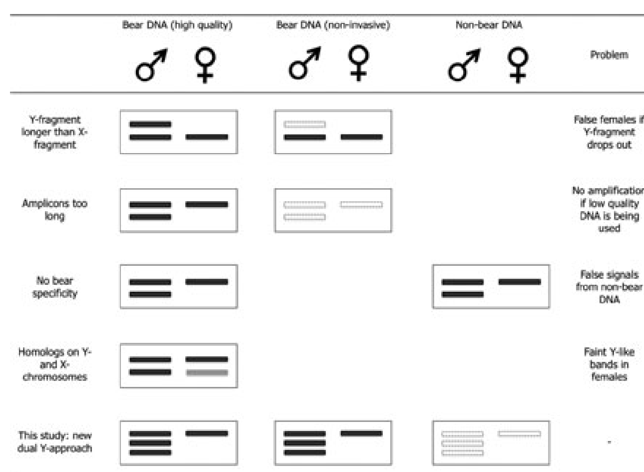


Fig. 1 Illustration of possible problems associated with molecular sexing techniques and the advantage of the novel dual-Y approach. Each panel: Banding pattern of a male (left side) and female (right side). Strong, faint and no amplification are indicated by dark grey, light grey and dashed bands respectively.

© 2013 Blackwell Publishing Ltd

FJ222502.1, DQ520647.1, X58927.1, X58933.1). All sequences of non-bear species had at least one single nucleotide or insertion-deletion (indel) mismatch near the 3′ end of at least one of the two primers. For the alignment of the Y-specific fragment (*SMCY*), cat (*Felis catus*; AY518681.1), lynx (*Lynx lynx*; AY219659.1) and horse (*Equus caballus*; AY532887.1) sequences were included as non-ursid species with mismatches in both primer sequences. On the basis of these alignments, we designed primer pairs for the *SMCY* fragment (100 bp) and the *ZFX* fragment (160 bp). An additional primer pair (*318.2*) was designed for a second locus on the Y-chromosome with an intermediate fragment size of 124 bp, serving as a double control for Y-specific amplification and PCR reliability. The latter primer pair was designed based on an anonymous Y-chromosome contig from a brown bear whole genome sequence (Björn Hallström, personal communication). All primer pairs (Table 1) were designed with the software PRIMER3 (Rozen & Skaletsky 2000).

Sampling and DNA extraction

High molecular weight DNA from all eight ursid species (brown bear, polar bear, black bear, Asian black bear, sloth bear, sun bear, spectacled bear, giant panda) was amplified with the *SMCY*, *318.2* and *ZFX* primers. The sex was known for all samples, except the giant panda. DNA was extracted from frozen tissue or blood samples using a modified Puregene (Qiagen) DNA extraction protocol. In addition, hair and scat samples from brown bears were collected in the field in the Pasvik Valley in Norway, and matched saliva, hair and faecal samples were collected from captive brown bears in a zoo. These non-invasive samples were included to test whether our method can be applied to potentially degraded DNA. The sex of the zoo brown bears was previously known and that of the remaining individuals was determined using Amelogenin primers (Kopatz *et al.* 2012). Faeces from zoo animals were sampled fresh and stored in 96% ethanol not more than 72 h after deposit. DNA was extracted from faecal samples using the PSP Spin Stool DNA Plus Kit (Invitek) or the Stool Kit (Qiagen) and from hair and saliva samples using the DNeasy Tissue Kit or Investigator Kit (both Qiagen), following the manufacturers' instructions. To test the specificity of the developed primers, DNA from 22 other species was tested, including humans, carnivores, potential prey and other co-distributed mammalian species (see Table 2). DNA concentrations for tissue, blood and hair samples were determined on a NanoDrop 1000 Spectrophotometer (Thermo Fisher Scientific) or on a NanoPhotometer (Implen), as described by the manufacturer. Primer sensitivity was tested in three replicates with a dilution series of a DNA sample extracted from tissue from one male brown bear. The different amounts of template DNA were as follows: 0.2, 0.1, 0.05, 0.04, 0.03, 0.02 and 0.01 ng.

PCR amplification and fragment analysis

Three different laboratories were involved in this study and used different reagents and PCR protocols in accordance with their particular standard laboratory procedures. These variations served to evaluate the robustness of the test to deviations in the protocol over a range of template DNA amounts (0.2–150 ng), PCR chemistries, primer annealing temperatures (i.e. 50 °C (set point) and 69–59 °C (touchdown)) and detection platforms (i.e. agarose gels, capillary electrophoresis). For sex determination of blood or tissue samples of bear species and selected non-bear species (Table 2), the three primer pairs were combined as one multiplex in a touchdown PCR approach in 15 μL reaction volumes using the following conditions: 0.35 U of GoTaq Flexi DNA Polymerase (Promega), 5x Green GoTaq Flexi Reaction Buffer (Promega), 2 mM $MgCl_2$ (Promega), 160 μM of each dNTP (TaKaRa Bio Inc.,), 0.17 μg/μL BSA (New England Biolabs), 0.33 μM each of the *ZFX* and *SMCY* primers, 0.27 μM of each of the *318.2* primers and 15–25 ng of

Table 1 Sequences of the novel primers, and corresponding EMBL accession numbers

Locus	Forward sequence (5′-3′)	Reverse sequence (5′- 3′)	Fragment size (bp)	EMBL number (brown bear sequence)
X-Chromosome (*ZFX*)	AAAGAAATCCCTCAAACACGTTAC	TCGCCACCCRCAAATAG	160	AB261815.1
Y-Chromosome (*318.2*)	AAGAAAAGTCATGCAACAGATACAG	TGATGCTTTGTGATCCTAATGTG	124	HF547901
Y-Chromosome (*SMCY*)	GTCTTCCTCCTTAGAGGGTAATTAGG	TTCGTTTGATAATGGCCTAAAACTG	100	AB261823.1

© 2013 Blackwell Publishing Ltd

Table 2 Cross-amplification test with different species, including potential prey animals, co-distributed species, closely related species and other carnivores as well as humans. Sex known (m = male, f = female) or unknown (?); no amplification (–), weak and/or inconsistent amplification (+).

Family	Species name	Common name	Sex	*ZFX* amplification	*318.2* amplification	*SMCY* amplification
Procyonidae	*Procyon lotor*	Raccoon	m	(+)	–	–
Phocidae	*Halichoerus grypus*	Grey seal	?	(+)	–	–
Mustelidae	*Lutra lutra**	European otter	?	–	–	–
	*Martes foina**	Stone marten	?	–	–	–
	Mustela nivalis	Least weasel	?	–	–	–
	*Mustela putorius**	European polecat	?	–	–	–
	Gulo gulo	Wolverine	?	–	–	–
Canidae	*Canis lupus**	Wolf	m	–	–	–
	*Canis lupus familiaris**	Dog	f	–	–	–
	Vulpes vulpes	Red fox	?	–	–	–
	Vulpes lagopus	Arctic fox	?	–	–	–
Felidae	*Lynx lynx**	Lynx	?	–	–	–
	*Felis silvestris**	Cat	m	–	–	–
			f	–	–	–
Cervidae	*Dama dama*	Fallow deer	m	–	–	–
	*Cervus elaphus**	Red deer	?	–	–	–
Bovidae	*Bos primigenius taurus**	Cow	?	–	–	–
Equidae	*Equus ferus caballus*	Horse	m	–	–	–
Castoridae	*Castor canadensis**	North American beaver	m	–	–	–
			f	–	–	–
	Castor fiber *	Eurasian beaver	m	–	–	–
			f	–	–	–
Leporidae	*Lepus europaeus* *	European hare	?	–	–	–
Muridae	*Rattus norvegicus*	Rat	m	–	–	–
Hominidae	*Homo sapiens*	Human	m	–	–	–
			f	–	–	–

*indicates non-invasively collected material.

DNA template. The amplification protocol started with 95 °C for 3 min followed by 10 cycles of 94 °C for 30 s, 69 °C (decreasing by 1 °C per cycle) for 25 s and 72 °C for 75 s. This was followed by 30 cycles of 94 °C for 30 s, 59 °C for 25 s and 72 °C for 75 s. The final elongation step was conducted for 10 min at 72 °C. Each PCR contained at least one negative control to detect possible contamination, and a positive control consisting of a known male brown or polar bear individual to ensure consistent amplification. The PCR products were electrophoresed (6 V/cm) on a 3% agarose gel stained with Gelred (Biotium Inc). Amplicon bands were evaluated and photographed under UV-light.

For sensitivity testing and sex determination of non-invasive DNA samples from Norway, the PCR was performed in 10 μL reaction volumes: 2x multiplex PCR master mix (Qiagen Multiplex kit), 0.05 μg/μL BSA (New England Biolabs), 0.3 μM each of the *ZFX* and *SMCY* primers, 0.25 μM of each of the *318.2* primers and 1 μL template DNA. PCR amplification started at 95 °C for 3 min followed by 20 cycles of 94 °C for 30 s, 69 °C (decreasing by 0.5 °C per cycle) for 25 s and 72 °C for 75 s. This was followed by 20 cycles of 94 °C for 30 s, 59 °C for 25 s and 72 °C for 75 s. Capillary electrophoresis was carried out on an ABI 3730 (Applied Biosystems), and the PCR fragments were analysed in GENEMAPPER 4.0 (Applied Biosystems). Allele sizes were determined using GENESCAN 500 LIZ standard. Negative controls were included for every seventh sample. We compared the sensitivity to a previously published method that uses the primers SE47 and R143 (described in Kopatz *et al.* 2012).

The brown bear samples from zoo animals were (in addition to sex determination) genotyped at twelve autosomal microsatellite loci (G10C, G10L, G10P, G10H, G10J, G10U, UarMU26, Mu10, Mu23, Mu50, Mu59, Mu51; see Frosch *et al.* 2011) to be able to compare the success rate (percentage of amplifying markers) in these samples to other markers. Sexing primers were labelled with NED or PET fluorescent dye (Life technologies). The PCR was performed in 10 μL volumes including 3.8 μL of DNA extract (10 ng/μL from tissue; pure extract from hair, faeces and saliva), 1.5 mM $MgCl_2$,

© 2013 Blackwell Publishing Ltd

0.2 mM dNTPs, 0.3 μM of each sexing primer and 0.5 U/μL HotStarTaq-Polymerase (Qiagen). Fragments were amplified at 95 °C for 15 min, followed by 45 cycles at 94 °C for 30 s, annealing at 50 °C 90 s, 72 °C for 60 s at and a final elongation step at 72 °C for of 30 min. PCR products were analysed on an ABI 3730 DNA Analyzer (Applied Biosystems). Fragment lengths were determined using GENEMARKER 1.6 software (SoftGenetics; size standard: LIZ).

Results

Sex determination of different bear species

For all ursine bears (brown bear, polar bear, black bear, Asian black bear, sloth bear, sun bear), at least one female and two male individuals were tested. PCR products were analyzed by agarose gel electrophoresis and the expected banding pattern (males: 100 bp, 124 bp, 160 bp; females: 160 bp) was observed in all individuals (examples for each species are shown in Figs 2a and b). Furthermore, two male individuals of the spectacled bear also yielded the expected banding pattern, although the *318.2* fragment was weaker than in ursine bears (Fig. 2b). We additionally analysed a sample from a giant panda for which the sex was unknown. A strong band from the X-linked marker was observed; however, the two Y-chromosomal markers showed only a very weak (*SMCY*; 100 bp) or no amplification (*318.2*; 124 bp) respectively (Fig. 2b).

Sensitivity tests revealed that for template DNA amounts as low as 0.02 ng, all three sexing markers amplified in at least one of three replicate PCRs and each of the three markers amplified in at least two out of three replicate PCRs. For comparison, the Amelogenin-based sexing method by Kopatz *et al.* (2012) produced overall lower signal intensity, required DNA amounts of at least 0.03 ng, and had twice the rate of amplification failure.

Specificity test

In the raccoon and grey seal, very weak bands could be observed at approximately 160 bp in most PCR replicates, indicating weak amplification of the *ZFX* marker in those species, likely due to their phylogenetic similarity to bears. In horse and rat, fragments >300 bp were visible, a size range that does not interfere with sex determination in bears. No amplification was observed in any of the other species tested (see Table 2 for a list of all species tested). An increased DNA template amount of 90 and 150 ng for some of the tested species did not alter the outcome, demonstrating the robustness of the protocol with respect to primer–template interaction.

Sex determination tests on hair and faecal DNA from brown bears

We used ten hair samples (two females, eight males, DNA amounts: 0.1–15.4 ng) and 20 faecal samples (eight females, 12 males) from brown bears collected in the field to validate our sex determination test. The correct sex was assigned to all individuals, although in one male (faecal sample collected in the wild), the ZFX-control did not amplify. Hair and faecal samples from zoo animals (three females and two males each) were sexed correctly. Additional saliva samples from the same individuals were used as controls and the results matched the known sex in all replicate tests. The relatively fresh faecal samples from the zoo individuals yielded a success rate for our sex markers that was significantly correlated to

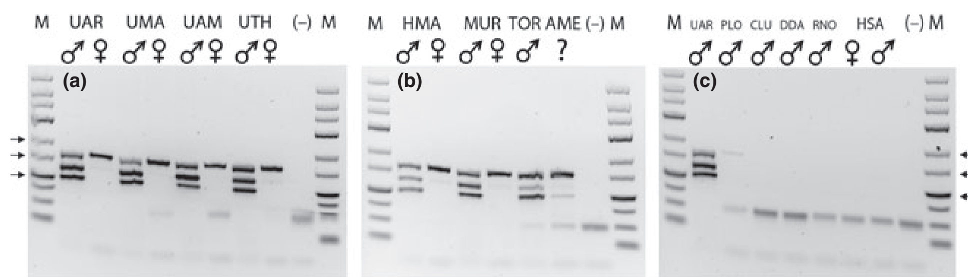


Fig. 2 Electrophoresis of PCR products from eight bear (a and b) and five non-bear species (c) on a 3% agarose gel. *Ursus arctos* (UAR), *Ursus maritimus* (UMA), *Ursus americanus* (UAM), *Ursus thibetanus* (UTH), *Helarctos malayanus* (HMA), *Melursus ursinus* (MUR), *Tremarctos ornatus* (TOR) and *Ailuropoda melanoleuca* (AME). *Procyon lotor* (PLO), *Canis lupus* (CLU), *Dama dama* (DDA), *Rattus norvegicus* (RNO) and *Homo sapiens* (HSA). Males and females are shown where both sexes were available and a negative control (no DNA template) was run on all gels (?: sex unknown). Arrows indicate 100 bp, 150 bp and 200 bp (Size marker: Gene Ruler Low Range DNA Ladder, Fermentas, St. Leon-Rot, Germany).

© 2013 Blackwell Publishing Ltd

the amplification success of the 12 microsatellites (Pearson's $R^2 = 0.930$, $P < 0.01$).

Discussion

We here establish a multiplex PCR procedure that is suitable for molecular sex identification in all extant *Ursinae* and *Tremarctinae* bear species and is applicable to both high quality DNA and degraded DNA from non-invasive samples. The method was successfully implemented in three different laboratories, using different PCR conditions (i.e. template amount, PCR chemistry, annealing temperature and detection platform), underscoring the overall robustness of our approach.

In contrast to existing sex determination methods for bears, our novel approach co-amplifies two Y-specific markers. This is advantageous, because sex determination with molecular methods has the inherent risk of non-amplification of the diagnostic, Y-specific marker. Thus, targeting two independent Y-linked loci increases the reliability of the results. To avoid preferential amplification of the internal PCR control, we chose a nuclear marker that occurs in equimolar amounts relative to the Y-marker in the heterogametic sex. Mitochondrial markers have been used before as internal PCR controls (e.g. Kohn *et al.* 1995), but the inherent risk of non-amplification of the Y-specific marker increases in such cases, as the target DNA occurs in much lower amounts relative to mitochondrial DNA in terms of copy number per cell.

Sensitivity testing is recommended for markers used in conservation genetics and wildlife forensics (see Andreassen *et al.* 2012) and our molecular sexing approach demonstrated reliable performance on DNA template amounts ≥ 0.02 ng. This corresponds to about five genome copies (assuming a diploid genome of 3×10^9 bp and an average molecular weight of 660 Da/bp), a similar threshold as observed in human forensics applications (Jobling & Gill 2004). If a higher degree of precision to measure the actual amount of bear-specific template DNA that is available in the sample is needed, quantitative PCR (qPCR) could be used instead of the photometric measurement employed here. The amount of DNA recovered from hair or faeces varies greatly and low amounts should be expected (Taberlet *et al.* 1996; Morin *et al.* 2001). The sensitivity of this new method on low-quality DNA from hair and faeces was superior to that of a similarly sized suite of microsatellite markers validated for forensic analyses in bears (Andreassen *et al.* 2012). Furthermore, it provided improved sensitivity and higher overall signal intensity compared with sex determination based on PCR amplification of the Amelogenin gene (Kopatz *et al.* 2012). Non-invasive samples are often characterized by degraded DNA, necessitating approaches that are based on small amplicon sizes (Wandeler *et al.* 2003). Experimental validation on hair collected in the field detected a decrease in peak intensity in two samples that each yielded less than 1.0 ng DNA. These results are consistent with the sensitivity threshold of ≥ 0.02 ng DNA template estimated for this test.

Several sex identification methods use trans-specific, conserved-sequence primers that can give confounding results if DNA material from mixed sources is being tested. Our primers were designed and demonstrated to be bear-specific, considerably reducing the risk of false positive results from non-bear sources in questionable samples. Weak amplification of the X-linked marker in closely related species (e.g. raccoon, grey seal; Table 2) has previously been described for other bear-specific sexing approaches (e.g. Pagès *et al.* 2009). However, our Y-linked primers do not bind efficiently in these species. For example, the raccoon sample used in our specificity tests was determined to be from a male with trans-specific Amelogenin primers (Yamamoto *et al.* 2002), but did not amplify at either diagnostic *SMCY* or *318.2* marker. The reduced amplification observed for spectacled bear and giant panda could be expected from their increasing phylogenetic distance to the polar and brown bear, upon which marker development was based. Owing to the very weak amplification of only one male-specific marker (*SMCY*) in the giant panda, the sex of this sample should be considered ambiguous and we do not recommend applying our method to samples from giant pandas without further evaluation. The usefulness of the dual-Y marker format of the test (Tringali *et al.* 2008; Ahlering *et al.* 2011) was reinforced by the results from the spectacled bear (one strong and one weak Y-specific band), because the male origin of the known-sex samples could be verified by the presence of the diagnostic *SMCY* amplicon alone.

The size differences between the fragments (24 bp and 36 bp respectively) were clearly resolved on 3% agarose gels, which is a standard and economic approach. However, if evidentiary or historical samples are being analysed, exact amplicon size can be determined by using bioanalyzers or automated capillary electrophoresis. Furthermore, a qPCR approach that measures male-specific amplification in real-time (see Barbisin *et al.* 2009) can be useful, where the amount of male-specific DNA needs to be estimated from mixed-sex wildlife samples.

We have found this new method to be a valuable tool in molecular wildlife studies of bears. Our dual-Y approach, high sensitivity and specificity of the primers together with the internal, equimolar positive amplification control, ensure reliable sex determination of different bear tissues, even from low-quality non-invasively collected samples or otherwise degraded DNA.

© 2013 Blackwell Publishing Ltd

Acknowledgements

The study was supported by Hesse's 'LOEWE Landes-Offensive zur Entwicklung Wissenschaftlich-ökonomischer Exzellenz', and the Arthur und Aenne Feindt-Stiftung, Hamburg. We thank Björn Hallström (Science for Life Laboratory, School of Biotechnology, KTH, Stockholm, Sweden) for providing us with bear Y-chromosomal sequences, Alexander Kopatz (Bioforsk, Norway) for productive discussions, Carsten Nowak and the Senckenberg Conservation Reserach Group for discussions and support, and Susanne Carl, who was an encouraged supporter and collected brown bear samples from zoo individuals. We also thank two anonymous reviewers for their helpful comments on the manuscript.

References

Aasen E, Medrano JF (1990) Amplification of the ZFY and ZFX genes for sex identification in humans, cattle, sheep and goats. *Nature Biotechnology*, **8**, 1279–1281.

Ahlering M, Hailer F, Roberts M, Foley C (2011) A Simple and accurate method to sex savannah, forest and asian elephants using noninvasive sampling techniques. *Molecular Ecology Resources*, **11**, 831–834.

Amstrup SC, Garner GW, Cronin MA, Patton JC (1993) Sex identification of polar bears from blood and tissue samples. *Canadian Journal of Zoology*, **71**, 2174–2177.

Andreassen R, Schregel J, Kopatz A *et al.* (2012) A forensic DNA profiling system for Northern European brown bears (*Ursus arctos*). *Forensic Science International: Genetics*, **6**, 798–809.

Barbisin M, Fang R, O'Shea CE, Calandro LM, Furtado MR, Shewale JG (2009) Developmental validation of the quantifiler® duo DNA quantification kit for simultaneous quantification of total human and human male DNA and detection of PCR inhibitors in biological samples. *Journal of Forensic Sciences*, **54**, 305–319.

Fang J, Yu Z, Lian , M *et al.* (2012) Knockdown of zinc finger protein, X-linked (ZFX) inhibits cell proliferation and induces apoptosis in human laryngeal squamous cell carcinoma. *Molecular and Cellular Biochemistry*, **360**, 301–307.

Farrell LE, Roman J, Sunquist ME (2000) Dietary separation of sympatric carnivores identified by molecular analysis of scats. *Molecular Ecology*, **9**, 1583–1590.

Frosch C, Dutsov A, Georgiev , G , Nowak C (2011) Case report of a fatal bear attack documented by forensic wildlife genetics. *Forensic Science International: Genetics*, **5**, 342–344.

Jobling MA, Gill P (2004) Encoded evidence: DNA in forensic analysis. *Nature Reviews Genetics*, **5**, 739–751.

Kohn M, Knauer F, Stoffella A, Schröder W, Pääbo S (1995) Conservation genetics o the European brown bear - a study using excremental PCR of nuclear and mitochondrial sequences. *Molecular Ecology*, **4**, 95–103.

Kopatz A, Eiken HG, Hagen SB *et al.* (2012) Connectivity and population subdivision at the fringe of a large brown bear (*Ursus arctos*) population in North Western Europe. *Conservation Genetics*, **13**, 681–692.

Millar CD, Lambert DM, Anderson S, Halverson JL (1996) Molecular sexing of the communally breeding pukeko: an important ecological tool. *Molecular Ecology*, **5**, 289–293.

Morin PA, Chambers KE, Boesch C, Vigilant L (2001) Quantitative polymerase chain reaction analysis of DNA from noninvasive samples for accurate microsatellite genotyping of wild chimpanzees (*Pan troglodytes verus*). *Molecular Ecology*, **10**, 1835–1844.

Murphy MA, Waits LP, Kendall KC (2003) The influence of diet on faecal DNA amplification and sex identification in brown bears (*Ursus arctos*). *Molecular Ecology*, **12**, 2261–2265.

Pagès M, Maudet C, Bellemain E, Taberlet P, Hughes S, Hänni C (2009) A system for sex determination from degraded DNA: a useful tool for palaeogenetics and conservation genetics of ursids. *Conservation Genetics*, **10**, 897–907.

Rosel PE (2003) PCR-based sex determination in Odontocete cetaceans. *Conservation Genetics*, **4**, 647–649.

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

Shaw CN, Wilson PJ, White BN (2003) A reliable molecular method of gender determination for mammals. *Journal of Mammalogy*, **84**, 123–128.

Taberlet P, Mattock H, Dubois-Paganon C, Bouvet J (1993) Sexing free-ranging brown bears *Ursus arctos* using hairs found in the field. *Molecular Ecology*, **2**, 399–403.

Taberlet P, Griffin S, Goossens B (1996) Reliable genotyping of samples with very low DNA quantities using PCR. *Nucleic Acids Research*, **24**, 3189–3194.

Taberlet P, Camarra JJ, Griffin S, Uhrès E *et al.* (1997) Noninvasive genetic tracking of the endangered Pyrenean brown bear population. *Molecular Ecology*, **6**, 869–876.

Tringali MD, Davis MC, Bolen EE, Sullivan JG, Haubold EM (2008) Simultaneous use of the X- and Y-chromosome genes SMCX, SMCY, and DBY for sex determination in the Florida manatee (*Trichechus manatus latirostris*). *Marine Mammal Science*, **24**, 218–224.

Waits LP, Paetkau D (2005) Noninvasive genetic sampling tools for wildlife biologists. A review of applications and recommendations for accurate data collection.. *Journal of Wildlife Management*, **69**, 1419–1433.

Wandeler P, Smith S, Morin PA, Pettifor RA, Funk SM (2003) Patterns of nuclear DNA degeneration over time-a case study in historic teeth samples. *Molecular Ecology*, **12**, 1087–1093.

Wasser SK, Houston CS, Koehler GM, Cadd GG, Fain SR (1997) Techniques for application of faecal DNA methods to field studies of Ursids. *Molecular Ecology*, **6**, 1091–1097.

Woods JG, Paetkau D, Lewis D, McLellan BN, Proctor M, Strobeck C (1999) Genetic tagging of free-ranging black and brown bears. *Wildlife Society Bulletin*, **27**, 616–627.

Xu X, Lin L, Zhang Z, Shen F, Zhang L, Yue B (2008) A reliable, non-invasive PCR method for giant panda (*Ailuropoda melanoleuca*) sex identification. *Conservation Genetics*, **9**, 739–741.

Yamamoto K, Tsubota T, Komatsu T et al. (2002) Sex identification of Japanese black bear, *Ursus thibetanus japonicus*, by PCR based on amelogenin gene. *The Journal of Veterinary Medical Science*, **64**, 505–508.

T.B. designed and performed experiments, analysed data, and wrote the manuscript; C.F. designed and performed experiments and analysed data; V.E.K. and S.G.A. performed experiments; H.G.E. designed experiments, analysed data and wrote the manuscript; S.B.H., S.R.F. and A.J. conceptualized the study and contributed samples and reagents; F.H. conceived the study, designed the methods and experiments, analysed data and wrote the manuscript. All authors helped edit the manuscript.

Data accessibility

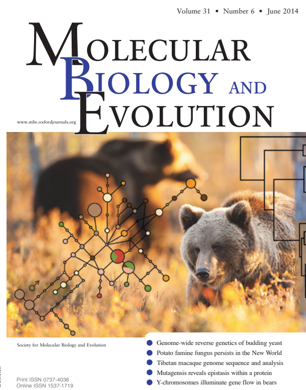
The sequence of the *318.2* locus was submitted to EMBL (accession number: HF547901).

© 2013 Blackwell Publishing Ltd

PUBLICATION 2: BROWN AND POLAR BEAR Y CHROMOSOMES REVEAL EXTENSIVE MALE-BIASED GENE FLOW WITHIN BROTHER LINEAGES

Tobias Bidon, Axel Janke, Steven R. Fain, Hans G. Eiken, Snorre B. Hagen, Urmas Saarma, Björn M. Hallström, Nicolas Lecomte, Frank Hailer

Published in *Molecular Biology and Evolution* (2014) 31(6):1353–1363



Erklärung über Anteile der Autoren/Autorinnen an den einzelnen Kapiteln der Promotionsarbeit

Titel der Publikation/des Manuskripts: Brown and Polar Bear Y Chromosomes Reveal Extensive Male-Biased Gene Flow within Brother Lineages

	Was hat der/die Promovierende bzw. Was haben die Co-Autoren/Autorinnen biegetragen#	**Name des/der jeweiligen Autors/Autoren/ Autorin***
(1) Entwicklung und Planung	60% 40%	TB FH, AJ, SRF, SBH
(2) Durchführung der einzelnen Untersuchungen/ Experimente	Probenorganisation Probenverwaltung, Primerdesign, Sequenzierungen, Fragmentanalysen In-silico Identifizierung Y-Scaffolds	TB, FH, AJ TB BH
(3) Erstellung der Daten-sammlung und Abbildungen	Erstellung der Alignments und Distanzmatrices Abbildungen und Tabellen Onlinearchivierung der Sequenzen und Allelgrößen	TB TB TB
(4) Analyse/Interpretation der Daten	Sequenzeditierung, Netzwerke, phylogen. Baum, Selektionstests Divergenzzeitberechnung NGS-Daten: Scaffoldextraktion, Scaffoldanalysen	TB TB, FH TB
(5) übergeordnete Einleitung/ Ergebnisse/Diskussion	60% 40%	TB FH, HGE, US, SRF, NL, AJ

#Bei 2, 3 und 4 bitte kurze inhaltliche Angaben der jeweiligen Anteile, bei 1 und 5 reichen prozentuale Angaben
*Mehrfacheintragungen möglich

Als Autoren/Autorinnen werden solche Personen bezeichnet, die an der Arbeit in Bezug auf die genannten Punkte in einer Weise mitgewirkt haben, dass sie für die ausgewiesenen Passagen (mit) verantwortlich sind. Personen, die an der Arbeit mitgewirkt haben, jedoch nicht in diese Kategorie fallen, sollten in der Danksagung Erwähnung finden.

Datum/Ort

Datum
zustimmende Bestätigung der vorgenannten Angaben

Unterschrift Promovend/Promovendin

Unterschrift Betreuer/Betreuerin

Brown and Polar Bear Y Chromosomes Reveal Extensive Male-Biased Gene Flow within Brother Lineages

Tobias Bidon,*[,1] Axel Janke,[1,2] Steven R. Fain,[3] Hans Geir Eiken,[4] Snorre B. Hagen,[4] Urmas Saarma,[5] Björn M. Hallström,[1,6] Nicolas Lecomte,[7] and Frank Hailer*[,1]
[1]Biodiversity and Climate Research Centre (BiK-F), Frankfurt am Main, Germany
[2]Goethe University Frankfurt, Institute for Ecology, Evolution & Diversity, Frankfurt am Main, Germany
[3]National Fish and Wildlife Forensic Laboratory, Ashland, OR
[4]Bioforsk, Norwegian Institute for Agricultural and Environmental Research, Svanvik, Norway
[5]Department of Zoology, Institute of Ecology and Earth Sciences, University of Tartu, Tartu, Estonia
[6]Science for Life Laboratory, School of Biotechnology, KTH, Stockholm, Sweden
[7]Canada Research Chair in Polar and Boreal Ecology, Department of Biology, University of Moncton, Moncton, Canada
***Corresponding author:** E-mail: tobias.bidon@senckenberg.de; frashai@gmx.net.
Associate editor: David Irwin

Abstract

Brown and polar bears have become prominent examples in phylogeography, but previous phylogeographic studies relied largely on maternally inherited mitochondrial DNA (mtDNA) or were geographically restricted. The male-specific Y chromosome, a natural counterpart to mtDNA, has remained underexplored. Although this paternally inherited chromosome is indispensable for comprehensive analyses of phylogeographic patterns, technical difficulties and low variability have hampered its application in most mammals. We developed 13 novel Y-chromosomal sequence and microsatellite markers from the polar bear genome and screened these in a broad geographic sample of 130 brown and polar bears. We also analyzed a 390-kb-long Y-chromosomal scaffold using sequencing data from published male ursine genomes. Y chromosome evidence support the emerging understanding that brown and polar bears started to diverge no later than the Middle Pleistocene. Contrary to mtDNA patterns, we found 1) brown and polar bears to be reciprocally monophyletic sister (or rather brother) lineages, without signals of introgression, 2) male-biased gene flow across continents and on phylogeographic time scales, and 3) male dispersal that links the Alaskan ABC islands population to mainland brown bears. Due to female philopatry, mtDNA provides a highly structured estimate of population differentiation, while male-biased gene flow is a homogenizing force for nuclear genetic variation. Our findings highlight the importance of analyzing both maternally and paternally inherited loci for a comprehensive view of phylogeographic history, and that mtDNA-based phylogeographic studies of many mammals should be reevaluated. Recent advances in sequencing technology render the analysis of Y-chromosomal variation feasible, even in nonmodel organisms.

Key words: **Y chromosome, phylogeography, bears, introgression, SNP, microsatellite.**

Introduction

Phylogeography describes the origin of genetic variation among closely related lineages, tracing the geographic distribution of genetic variation through time and space (Avise 2000; Hewitt 2000). Historically, phylogenetic and phylogeographic research has relied heavily on mitochondrial DNA (mtDNA), with the brown bear (*Ursus arctos*) as an extensively studied example (Taberlet et al. 1998; Purvis 2005; Davison et al. 2011). Advantages of analyzing mtDNA include its high mutation rate, availability of markers, high copy number, lack of recombination, and its haploid nature. However, the typically maternal inheritance of mtDNA implies that signatures of male-mediated dispersal cannot be detected. An approach to further investigate phylogeographic patterns is to analyze independently and biparentally inherited autosomal loci in a multilocus framework. However, recombination hampers inferences of haplotypes over long genomic regions, limiting the resolution that is available from individual autosomal loci.

The only other haploid fraction of the mammalian genome is the male-specific Y chromosome. Due to its lack of recombination, except for the small pseudoautosomal regions, haplotypes can be inferred over extended genomic regions, providing a high-resolution view of patrilineal evolutionary history. Also, both mtDNA and the Y chromosome exhibit faster lineage sorting than nuclear loci, facilitating the detection of population structuring (Avise 2000). The male-specific section of the Y chromosome therefore provides an essential complement to data from maternally inherited mtDNA and biparentally inherited loci, giving insight into the history of uniquely male-inherited lineages. Y-linked variation allows the detection of potentially contrasting patterns of male and female gene flow (Chan et al. 2012). This is particularly relevant in many mammals, where males typically disperse much

© The Author 2014. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com

farther than females (Pusey 1987). Along with other loci, Y-linked variation has therefore provided a backbone for our understanding of phylogeography in humans (Hughes and Rozen 2012; Wei et al. 2013), canids (Brown et al. 2011; Sacks et al. 2013), and domesticated animals (Meadows et al. 2006; Lippold et al. 2011).

Despite these qualities, very little data is available from mammalian nonprimate Y chromosomes, in part because it has been disregarded from many genome sequencing projects due to its repetitive nature (Willard 2003). In addition, other technical challenges, such as avoiding co-amplification of homologous X-chromosomal regions, have hampered the analysis of paternally inherited markers in natural populations (Greminger et al. 2010). The Y chromosome thus represents an understudied part of the mammalian genome, with a large potential to add valuable information to our understanding of phylogeography. In the era of genomics, it is now feasible to identify large regions on the Y chromosome and develop male-specific markers for studies of evolutionary history.

Brown and polar (*U. maritimus*) bears have been model species in phylogeography since the early 1990s (Cronin et al. 1991; Taberlet and Bouvet 1994; Kohn et al. 1995; Paetkau et al. 1998; Taberlet et al. 1998; Hewitt 2000; Waits et al. 2000), in part because these species are widely dispersing and provide the advantage of being distributed over large parts of the Northern hemisphere. Polar bears exhibit low levels of population differentiation at biparentally inherited and mitochondrial markers throughout their range (Paetkau et al. 1999; Cronin and MacNeil 2012; Miller et al. 2012; Campagna et al. 2013). Brown bears, in contrast, show considerable phylogeographic structuring at mitochondrial markers (Davison et al. 2011; Edwards et al. 2011; Hirata et al. 2013; Keis et al. 2013), and population structuring can also be discerned at biparentally inherited microsatellites (Paetkau et al. 1997; Tammeleht et al. 2010; Kopatz et al. 2012). Most mtDNA clades are confined to certain geographical regions and are not shared between continents, although one brown bear clade is widespread throughout Eurasia and extends into North America (Korsten et al. 2009; Davison et al. 2011). Surprisingly, all range-wide phylogeographic studies on brown bears have so far relied on mtDNA. Studies of autosomal markers were regionally restricted to either North America or Eurasia (Paetkau et al. 1997; Tammeleht et al. 2010; Kopatz et al. 2012; Cahill et al. 2013), and no phylogeographic study of Y chromosome markers in bears exists. However, analysis of male-specific markers is crucial to understand bear evolution in the light of their well-documented male-biased dispersal (McLellan and Hovey 2001; Zedrosser et al. 2007).

With regard to bear phylogeny, reliance on mtDNA alone has proven problematic. Polar bear mtDNA sequences are nested within the genetic diversity of brown bears, resulting in a paraphyletic matrilineal relationship (Cronin et al. 1991; Lindqvist et al. 2010). Although mtDNA is expected to attain reciprocal monophyly faster than nuclear loci (Avise 2000), recent studies utilizing autosomal markers have shown that extant brown and polar bears comprise distinct sister lineages at the species tree level, and that their divergence occurred earlier than previously estimated (Hailer et al. 2012; Miller et al. 2012; Cahill et al. 2013). Therefore, brown bear paraphyly for mtDNA is likely a consequence of past introgressive hybridization with polar bears (Edwards et al. 2011; Hailer et al. 2012; Miller et al. 2012; Hailer et al. 2013).

We mined a recently sequenced polar bear genome and developed 13 male-specific markers to sequence 5.3 kb of the Y chromosome and to analyze microsatellite variation in a broad geographic sample of 130 brown and polar bears from across Europe, Asia, and North America. We also analyzed a 390-kb-long genomic Y-chromosomal scaffold in available brown, polar, and American black bear genomes. These data allowed us to investigate 1) whether introgression between brown and polar bears can be detected at Y chromosome markers, 2) whether the male lineage shows less geographic structuring than the maternal lineage, and 3) the relative intraspecific clade depth of mtDNA and the Y chromosome.

Results

Y Chromosome Phylogeny and Lack of Introgression Signals

Male-specific sequence data revealed that brown and polar bears carry differentiated, species-specific Y chromosomes, each exhibiting a closely related group of haplotypes (fig. 2A). The clear separation and reciprocal monophyly of brown, polar, and American black bear (*U. americanus*) Y chromosomes was further supported by Bayesian phylogenetic analyses (fig. 2B), with high statistical support ($P > 0.95$) for all major nodes.

In 3,078 bp of Y chromosome sequence analyzed in 90 brown, 40 polar, and 4 black bears (fig. 1 and table 1), we found over 75% of the variable sites among species (3.1-kb data set, solid lines in fig. 2A). Only a small portion of sequence polymorphism was intraspecific. We encountered eight haplotypes, five within brown, two within polar, and one within black bears. These haplotypes were defined by a total of 21 segregating sites, 10 of which discriminate between brown and polar bears, 9 between brown and black bears, and 13 between polar and black bears. Brown and polar bears each showed one abundant haplotype that was dominant in all populations across their ranges. Haplotype BR1.1 was found in 94% of brown bears and PO1.1 in 90% of polar bears (fig. 2A). Two haplotypes found in brown bears from the ABC islands (BR5) and the Alaskan mainland (BR4) formed a joint lineage, indicative of a geographically informative clustering. Additional rare haplotypes in brown bears were found in two individuals from Kamchatka (BR2) and in one individual from the Ural Mountains (BR3). In polar bears, the rare haplotype PO2 was found in three individuals from Alaska and in one from Western Greenland (fig. 1). Results for four black bear males are described in the supplementary material, Supplementary Material online.

Increasing sequence length by ~70% (adding 2,216 bp, 5.3-kb data set, dotted lines in fig. 2A) for 63 individuals chosen to represent most populations (supplementary table S1, Supplementary Material online) increased the resolution

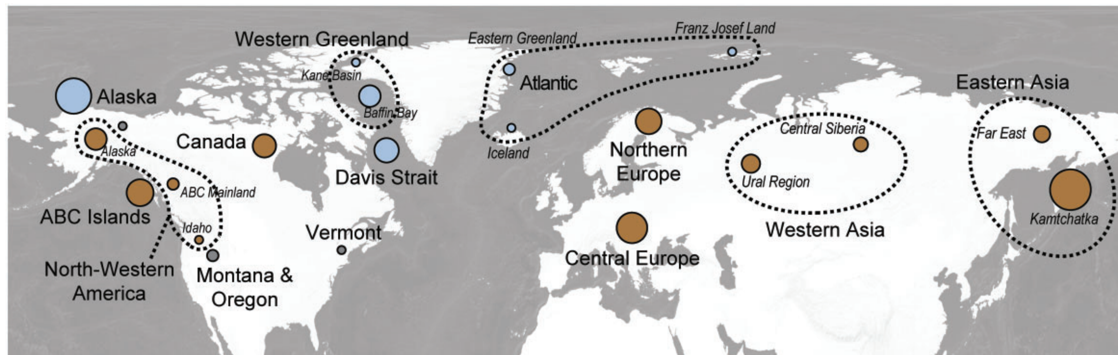


Fig. 1. Geographical distribution of analyzed bear samples. Circle area is proportional to the number of individuals. Some sampling localities (italics) were combined into groups (see table 1). Brown: brown bears; blue: polar bears; black: black bears.

Table 1. Sample Size (*n*), Number of Haplotypes (H), and Haplotype Diversity (HD) Based on the Combination of 3.1-kb Y-Chromosomal Sequence and Six Microsatellites.

Species and Population (abbreviation)	*n*	*H*	HD
Brown bear	90	41[a]	0.96 ± 0.01
Central Europe (C-EU)	14	8	0.89 ± 0.06
Northern Europe (N-EU)	10	4	0.73 ± 0.12
Western Asia (W-AS)	8	7	0.96 ± 0.08
Ural Region	5	5	
Central Siberia	3	2	
East Asia (E-AS)	29	12	0.84 ± 0.05
Far East	4	4	
Kamchatka	25	9	
North-West America (NW-A)	10	6	0.84 ± 0.10
Alaska	7	4	
ABC Mainland	2	1	
North-Western USA/Idaho	1	1	
ABC islands (ABC)	11	5	0.82 ± 0.08
Canada (CAN)	8	2	0.25 ± 0.18
Polar bear	40	17[a]	0.83 ± 0.06
Atlantic (ATL)	4	3	0.83 ± 0.22
Eastern Greenland	2	1	
Iceland	1	1	
Franz Josef Land	1	1	
Alaska (AK)	19	7	0.72 ± 0.10
Western Greenland (W-GR)	8	5	0.79 ± 0.15
Baffin Bay	7	4	
Kane Basin	1	1	
Davis Strait (DS)	9	6	0.89 ± 0.09
Black bear	4	4	1.00 ± 0.18
Alaska zoo, Oregon, Montana, Vermont	4	4	

[a]Sum of haplotypes across populations is larger than the number of haplotypes per species, due to haplotype sharing.

among species and revealed additional, rare haplotypes in brown bears (BR1.2, BR1.3), polar bears (PO1.2), and black bears (BL2). The general patterns were not substantially changed compared with the 3.1-kb data set, and still one single haplotype remained dominant across the distribution ranges in each species (BR1.1/PO1.1). Reflecting the few polymorphic sites found within species, nucleotide diversity ($\pi \pm$ SD) was low in brown (0.00007 ± 0.00002) and polar bears (0.00003 ± 0.00002) (table 2).

Using a Bayesian approach, we estimated the timing of the split between brown and polar bear male lineages ($T_{MRCA\ (B/P)}$). This was based on 5,197 bp of Y-chromosomal sequence using the spectacled bear (*Tremarctos ornatus*) as outgroup. Assuming 6 Ma for the split from the spectacled bear (a calibration based on the fossil record; Wayne et al. 1991), we estimated a $T_{MRCA\ (B/P)}$ of ~1.12 Ma (fig. 2*B*). We also constrained the analysis to a pedigree based Y-specific mutation rate (3.0×10^{-8}/site/generation [Xue et al. 2009], rendering 3.0×10^{-9}/site/year with a generation time estimate for bears of 10 years) and obtained estimates of $T_{MRCA\ (B/P)}$ of ~0.43 Ma (supplementary table S2, Supplementary Material online). The absolute timing of the split, therefore, depended strongly on the calibration prior (i.e., divergence time of the outgroup or substitution rate). Additional calibration scenarios from previous studies are examined in the supplementary material, Supplementary Material online. Our data consistently recovered the brown/polar bear split to be ~80% of the age of the older split from the black bear lineage, indicating that the divergences among different ursine species occurred relatively shortly after each other. We note, however, that the design of our Y sequence fragments targeted regions exhibiting nucleotide differences between one polar and one brown bear individual, which could lead to an upward ascertainment bias with regard to the magnitude of the brown/polar bear divergence (discussed later). Nevertheless, all variable sites on the black bear branch (fig. 2*A*) were newly discovered in our sequencing data, confirming the divergence of the black bear lineage with respect to brown and polar bears.

The findings of species-specific groups of haplotypes and the lack of haplotype sharing among species (fig. 2*A*) revealed no signal of recent Y-chromosomal introgression. In contrast, analysis of a 642-bp fragment of the mtDNA control region of the same samples showed polar bears nested within the variation of all brown bears (fig. 2*C*), as expected for this locus.

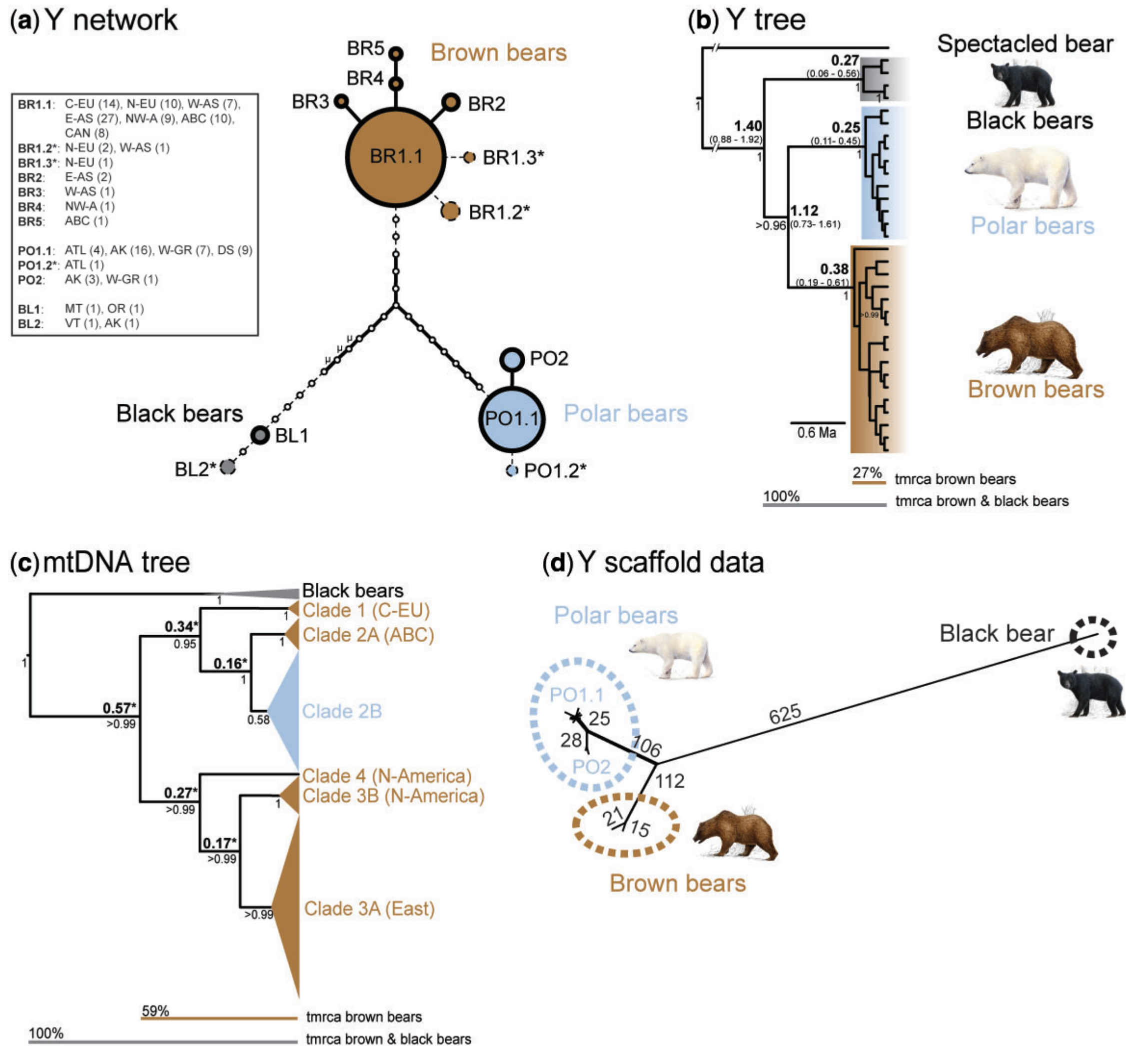


FIG. 2. Phylogenetic relationships of bears for Y-chromosomal and mitochondrial markers. (*A*) Parsimony network of Y chromosome sequences. Solid lines: variation in 3.1 kb; dashed lines: variation from additional 2.2 kb (total: 5.3 kb). Circle area is proportional to number of individuals; small, open circles: inferred, intermediate haplotypes; lines represent single mutational steps. Inset boxes: number of individuals per population. Asterisks: haplotypes found only in the 5.3-kb data set (individuals with these haplotypes have the respective common haplotypes in the 3.1-kb data set). Insertions/deletions of repeat units in microsatellite-like regions counted as number of repeat unit changes (μ). Population abbreviations as in table 1. (*B*) Maximum clade credibility tree of Y chromosomal sequence (5,197 bp), based on a divergence of the spectacled bear 6 Ma. Bold: median divergence in Ma (95% highest posterior density intervals in brackets). Numbers below nodes: posterior support >0.95. (*C*) Maximum clade credibility tree of mtDNA control region data. Sampling covers all major matrilineal brown bear clades (Davison et al. 2011) (collapsed into triangles), and polar bears (clade 2B) are nested within brown bear variation. Asterisks: divergence times obtained from complete mtDNA sequences (Hirata et al. 2013). Numbers below nodes: posterior probabilities. Below (*B*) and (*C*), brown bear clade depth (relative to the divergence from black bears) is indicated. (*D*) NeighborNet network based on a ~390 kb Y-chromosomal fragment from 12 polar bears, 2 brown bears, and 1 black bear. Numbers on branches denote numbers of variable sites. Within polar bears, two haplogroups were identified corresponding to the haplotypes PO1.1 and PO2 in figure 2*A*.

Y Chromosome Phylogeography of Bears

On the Y chromosome, we found a maximum of three variable sites separating different brown bear haplotypes (e.g., the difference between BR3 and BR5), but 14 substitutions between brown and black bears (5.3-kb dataset, not counting sites in microsatellite-like regions; see μ in fig. 2*A*). The intraspecific divergences relative to the outgroup obtained from Bayesian analyses amounted to 27% for the Y-chromosomal data and 59% for mtDNA control region data (fig. 2*B* and *C*). Similarly, estimates of mean (± SE) among-group genetic distances from mtDNA control region sequences showed that divergence between two major brown bear

Table 2. Summary Statistics Based on 5.3-kb Y-Chromosomal Sequence.

Species	n	H	f_H	S	$\pi \pm$ SD ($\times 10^{-4}$)	θ_W ($\times 10^{-4}$)	Tajima's D	D^*	F^*	F_S
Brown bear	44	6[a]	0.84	6	0.7 ± 0.2	2.6 ± 1.3	−1.94[b]	−3.01[b]	−3.13[b]	−4.659[b]
Polar bear	15	2[a]	0.93	1	0.3 ± 0.2	0.6 ± 0.6	−1.16	−1.42	−1.52	−0.649

NOTE.—Sample size (n), number of haplotypes (H), the frequency of the dominant haplotype (f_H), number of segregating sites (S), nucleotide diversity (π), Watterson's θ_W (per site), Tajima's D, Fu and Li's D^* and F^*, and Fu's F_S are given.

[a]Individuals with haplotypes BR4 and PO2 (fig. 2A) were only represented in the 3.1-kb data set (supplementary table S6, Supplementary Material online), hence these haplotypes are not counted here.

[b]$P < 0.05$.

mtDNA clades (1 and 3a) (0.036 ± 0.007) amounted to 57–60% of the mean distance between brown and black bears (0.064 ± 0.009 for clade 1, and 0.061 ± 0.009 for clade 3a). Thus, a considerable reduction in phylogeographic structuring of the patriline was detected in comparison to the established matrilineal pattern, where deeply separated mtDNA clades, most of which are region-specific, are found within brown bears.

This discrepancy in clade depth between the matri- and patriline was also obvious when analyzing a ~390-kb Y-chromosomal scaffold (scaffold number 297) from 14 published male bear genomes (Miller et al. 2012), along with the corresponding sequence from a male brown bear from northern Norway (supplementary table S3, Supplementary Material online). This alignment of 2 brown, 12 polar bears, and 1 black bear identified >1,000 high-quality variable sites, most of them distinguishing between the three bear species (fig. 2D). In this data set, the divergence between the two brown bear individuals (one from Norway and one from the ABC islands) was ~5% of the divergence of these to one black bear individual (36 substitutions between the two brown bears, 752–758 substitutions between brown and black bears), compared with ~20% between the divergence of all brown bears from the black bear based on whole mitochondrial sequences (Lindqvist et al. 2010).

The shallow clade depth on the brown bear Y chromosome could result from population expansion of one Y lineage that has replaced other clades. The pattern is also consistent with positive selection favoring a particular Y variant, and male-mediated gene flow spreading this variant across the range. To disentangle the effects of background selection, genetic hitchhiking, and recent population growth, we calculated four summary statistics to test for deviations from neutral expectations. In brown bears, all estimates were significant and negative (Tajima's $D = -1.94$, $P < 0.01$; Fu and Li's $D^* = -3.01$, $P < 0.05$; $F^* = -3.13$, $P < 0.05$; Fu's $F_S = -4.659$, $P < 0.01$; table 2), consistent with all three selective/demographic processes. The values calculated for polar bears were not significantly different from neutral expectations (Tajima's $D = -1.16$, $P > 0.1$; Fu and Li's $D^* = -1.42$, $P > 0.05$; $F^* = -1.52$, $P > 0.05$, Fu's $F_S = -0.649$, $P > 0.1$; table 2). Haplotype configuration tests (Innan et al. 2005) did not allow us to distinguish between signals of population stasis ($g = 0$), population growth ($g = 2$, $g = 10$), or selection in brown bears, because no tested scenario differed significantly from neutral expectations (cumulative $P > 0.05$ for all tests).

In addition to sequence data, we developed and analyzed six faster evolving male-specific microsatellites to obtain a high-resolution data set (fig. 3 and supplementary figs. S1–S4, Supplementary Material online). Although the overall Y-chromosomal haplotypic variability was high (table 1) and we observed a ratio of haplotypes to individuals of >40%, branches between haplotypes were short and defined by few mutational steps (fig. 3 and supplementary material, Supplementary Material online). Except for a group of three haplotypes found in Central European brown bears (fig. 1), and a group of 13 brown bears from eastern Asia (Kamchatka) exhibiting five differentiated haplotypes, all populations contained haplotypes that were distributed across the network (fig. 3A).

In polar bears, male-specific sequence data showed few rare mutations (fig. 2A), and even when combined with microsatellites, one haplotype was found to be abundant across much of the range (fig. 3B). From analysis of molecular variance (AMOVA), we obtained estimates of the proportion of variation among all populations of 0.28 for brown and 0.16 for polar bears (supplementary tables S4 and S5, Supplementary Material online). This is consistent with results from autosomal microsatellite markers which show stronger population differentiation in brown than in polar bears (Cronin and MacNeil 2012).

ABC Islands Brown Bears—Evidence for Male-Mediated Gene Flow from the Mainland

The Alaskan ABC islands are inhabited by brown bears that are unique in the close relatedness of their maternal lineage to polar bears. All polar and ABC islands brown bear samples included in our study show this expected relationship (fig. 2C). For the Y chromosome, we found five haplotypes among 11 ABC islands brown bears (fig. 3A), all clustering with brown rather than polar bears (fig. 2A). One haplotype was shared with individuals from Canada and another with individuals from northwest America and western Asia (fig. 1). Nonsignificant differentiation from brown bears on the adjacent North American mainland (ABC/NW-A: $\Phi_{ST} = 0.02$, $P > 0.05$; supplementary table S4, Supplementary Material online), but significant differentiation from all other populations further confirmed the connectivity by male-mediated gene flow. This gene flow is evidently substantial enough to maintain a high level of variability on the ABC islands: we found five haplotypes in 11 ABC islands individuals

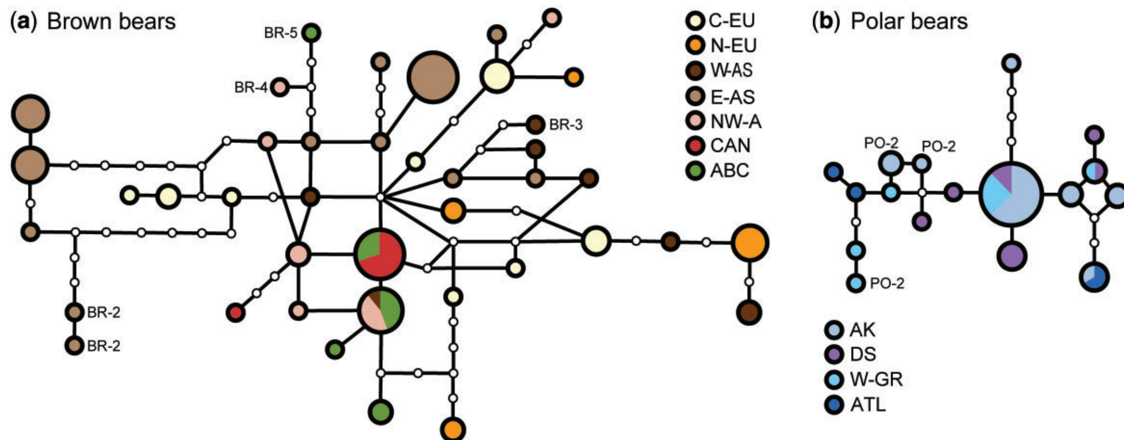


FIG. 3. Statistical parsimony networks of Y chromosome haplotypes, inferred from unweighted combination of 3.1-kb sequence data and six microsatellites, for (A) brown bears and (B) polar bears. Rare haplotype names as in figure 2A, population abbreviations as in table 1.

(haplotype diversity HD = 0.82; table 1), which is similarly high as the variability of all brown bears combined (HD = 0.96; table 1).

Discussion

Phylogeographic research has relied heavily on maternally inherited mtDNA, but male-biased dispersal in many mammals implies that mtDNA provides a highly structured (philopatric) estimate of population differentiation compared with paternally and biparentally inherited loci. Modern sequencing techniques now allow the generation of extensive genomic data, enabling large-scale identification and analysis of sequences from the male-specific Y chromosome (Bachtrog et al. 2011; Wei et al. 2013). This chromosome is especially interesting for evolutionary studies because it allows the inference of high-resolution haplotypes from long sequences, avoiding analytical challenges posed by interchromosomal recombination. Our analysis of newly developed Y-linked markers in comparison to results from maternally inherited mtDNA revealed a large impact of sex-biased gene flow on phylogeographic structuring and enabled us to examine phylogeny and introgression in brown and polar bears.

Speciation and Introgression

The Y chromosome phylogeny of brown and polar bear lineages resembles the topology of species trees reconstructed from biparentally inherited autosomal markers (Hailer et al. 2012; Miller et al. 2012; Cronin et al. 2013), where the species constitute distinct sister (or rather brother) lineages, with black bears clustering outside their variation (fig. 2B). This contrasts with the pattern obtained from maternally inherited mtDNA, where polar bears cluster within the variation of brown bears, rendering the latter paraphyletic (Cronin et al. 1991; Edwards et al. 2011) (fig. 2C).

The timing of the split between brown and polar bears has been the subject of recent debates, with inferred dates ranging from ~160,000 to ~5 million years (Lindqvist et al. 2010; Edwards et al. 2011; Hailer et al. 2012; Miller et al. 2012; Cahill et al. 2013; but see Ho et al. 2008 and Davison et al. 2011 for even younger estimates depending on the calibration method used). Compared with the mtDNA divergence estimate of ~160,000 years between polar and brown bears (Lindqvist et al. 2010; Edwards et al. 2011; Hirata et al. 2013), divergence times for the Y chromosome (>0.43 Ma, supplementary table S2, Supplementary Material online) are much older, confirming earlier suggestions that mtDNA has been introgressed (Hailer et al. 2012, 2013; Miller et al. 2012; Cahill et al. 2013). Compared with divergence times estimated from autosomal data, our 1.12 Ma estimate for brown/polar bear Y chromosomes (fig. 2B; scenario B in supplementary table S2, Supplementary Material online) is older than a divergence time estimate from introns of ~0.34-0.93 Ma (Hailer et al. 2012), but younger than the 4–5 Ma estimate by Miller et al. (2012) from genomic data. When based on a rate calibration from human Y chromosomes (scenario D in supplementary table S2, Supplementary Material online), our estimate of the Y chromosome divergence (0.43 Ma) falls into the Middle Pleistocene, resembling the estimate of Hailer et al. (2012). In summary, Y chromosome evidence support the emerging understanding of brown and polar bears as distinct evolutionary lineages that started to diverge no later than the Middle Pleistocene, at least several hundreds of thousands years ago.

Although incomplete lineage sorting can hamper definite conclusions, brown and polar bears likely carry introgressed alleles at mtDNA and autosomal loci (Hailer et al. 2012; Miller et al. 2012; Cahill et al. 2013). Current hybridization levels, however, appear to be low (Cronin and MacNeil 2012; Hailer et al. 2012). Our findings of species-specific groups of Y chromosome haplotypes and a lack of haplotype sharing among species revealed no signal of patrilineal introgression. Reduced introgression of Y chromosomes has been reported previously (e.g., Geraldes et al. 2008) and can arise from several mechanisms: random effects of lineage sorting, sex-biased hybridization, reduced hybrid fitness of the heterogametic sex

due to genomic incompatibilities (Haldane's rule), or lower introgression rates at markers exhibiting high intraspecific gene flow (Petit and Excoffier 2009).

Variability on the Y Chromosome

Most variable sites on the Y chromosome in bears were found among species, while only relatively little intraspecific sequence variation was encountered. The latter is compatible with the generally low intraspecific variability observed on mammalian Y chromosomes, including field voles, elephants, chamois, and humans (Hellborg and Ellegren 2004; Roca et al. 2005; Pérez et al. 2011; Wilson Sayres et al. 2014). Nakagome et al. (2008) compared Y, X, and mtDNA phylogenies and variability in bears based on single representations per species. They found a lower than expected Y-chromosomal substitution rate within *Ursinae* as compared with the deeper nodes of the tree, possibly mirroring our findings of low variability on the Y chromosomes of brown and polar bears. After applying a standard correction factor of four to account for the smaller effective population size of the Y chromosome (but see Chesser and Baker 1996), variability on the brown bear Y chromosome was ~10% of that on the autosomes (data from Hailer et al. 2012). As shown for other mammals (Hellborg and Ellegren 2004), this discrepancy between the Y chromosome and autosomes exists despite higher male than female mutation rates. Low intraspecific variability on the Y chromosome can be explained by its haploid and uniparental inheritance, reproductive skew among males, male-biased dispersal, demographic history, but also by selection or a combination of these (Chesser and Baker 1996; Charlesworth and Charlesworth 2000; Petit et al. 2002; Wilson Sayres et al. 2014).

In polar bears, Y-linked variability patterns did not deviate significantly from neutral expectations (table 2). In brown bears, the deviation was significant, with most of the applied tests showing an excess of rare mutations (table 2), consistent with population growth and/or positive selection. However, haplotype configuration tests did not necessitate a history of ongoing or recent positive selection on the Y chromosome in brown bears. Based on SNPs from the nuclear genome, Miller et al. (2012) found a long-term decline in brown bear effective population size, particularly since the Eemian interglacial. Genome-wide data thus do not indicate recent population growth, reinforcing the particular evolutionary history of the Y chromosome in brown bears.

Despite overall low levels of intraspecific variation on the Y chromosome, our analysis of long scaffold sequences (fig. 2*D*) illustrates that application of modern genomic techniques can nevertheless recover large numbers of polymorphic sites on the Y chromosome, enabling high-resolution inferences.

Phylogeographic Structuring

mtDNA control region data show pronounced phylogeographic structuring in brown bears, with 1) deeply separated clades and 2) clades which are geographically restricted (Davison et al. 2011) (fig. 2*C*). The Y chromosome is predicted to be a geographically informative marker that shows differences among populations, because of strong genetic drift in the patriline (Petit et al. 2002). However, we observed neither of the abovementioned signals at paternally inherited markers: no deep intraspecific divergences were found, and, over evolutionary time scales, male-biased gene flow has distributed genomic variation across and among continents. Compared with mitochondrial control region data, brown bear Y chromosomes showed shallow intraspecific divergences relative to the divergence from black bears, with few substitutions differentiating among Y-chromosomal haplotypes. Despite limited sample numbers, because to date only few male bear genomes have been sequenced, ascertainment bias-free scaffold data confirm the main conclusions from our sequence data. First, patrilineal genomic divergences within brown and polar bears were considerably shallower than for mtDNA. Second, the 390-kb data set recovered the same two groups of polar bear Y haplotypes that correspond to PO1.1 and PO2. Finally, brown bear sequences were separated from each other by small genetic distances. Although increased sampling and sequencing of longer fragments might recover additional clades, our conclusions are not impacted by a strong ascertainment bias (Brumfield et al. 2003). On deeper phylogenetic scales, however, we note that the divergence of the black bear Y chromosome was likely underestimated in our 3.1- and 5.3-kb data sets.

The observed discrepancy between the matri- and patriline can be due to effects of demography and selection on the Y chromosome. In addition, mtDNA can show signals of mutational saturation (Ingman and Gyllensten 2001) and purging of slightly deleterious mutations due to purifying selection (Subramanian et al. 2009), leading to a time dependency of evolutionary rates for mtDNA (Ho et al. 2008). Whole mtDNA data from Lindqvist et al. (2010) show, relative to the divergence from black bears, a shallower clade depth in brown bears compared with data from the control region. However, our analysis of longer sequences from Y scaffold data confirmed the weaker structuring of the patriline than the matriline. Whichever the mechanism(s), a reduced phylogeographic structuring on the Y compared with well differentiated mtDNA clades has also been found in other species, for example, shrews, chamois, and gibbons (Lawson Handley et al. 2006; Pérez et al. 2011; Chan et al. 2012).

Despite known uncertainties with regard to absolute ages, our Bayesian phylogenetic analyses suggested that the most basal divergence of brown bear Y haplotypes considerably predates the last glacial maximum, with plausible dates reaching into the Middle Pleistocene (95% highest posterior density: 0.19–0.61 Ma; fig. 2*B*). This suggests that one Y chromosome lineage (BR1.1) has been maintained for a long time and at a high frequency throughout Eurasia and North America. While selection may therefore have contributed to the shallow Y-chromosomal clade depth within brown bears, our data are also consistent with a purely demographic scenario, involving extensive male gene flow across large geographical distances. Indeed, analysis of a

390-kb-long Y-chromosomal fragment showed that two brown bears from populations as far away from each other as Norway and the Alaskan ABC islands carried highly similar Y chromosomes (fig. 2*D*). This pattern in brown bears covers even larger geographic areas (throughout Eurasia and North America) than analogous findings from humans, where the Y-chromosomal lineage of Genghis Khan, founder of the Mongol Empire, was spread across much of Asia (Zerjal et al. 2003).

Our discovery of distinct Y-chromosomal haplotypes on Kamchatka mirrors previous findings of distinct mtDNA lineages (Korsten et al. 2009), highlighting the complex biogeography of this peninsula. Besides this clear signal from Kamchatka, brown bear populations in general contained a mix of different Y chromosome lineages, with the most closely related lineages of a given haplotype being located in a different geographic region. This lack of pronounced patrilineal geographic structuring is an expected consequence of male-mediated gene flow and contrasts strongly with the picture from mtDNA, where populations tend to contain region-specific lineages (Davison et al. 2011).

In polar bears, we observed weak population structuring and no clear evidence of past phylogeographic barriers on the Y chromosome. This is similar to patterns from maternally and biparentally inherited markers (Paetkau et al. 1999; Cronin et al. 2006; Miller et al. 2012; Campagna et al. 2013), reflecting the large dispersal distances described for polar bears.

Male-Biased Gene Flow and the Alaskan ABC Islands Bears

We provide the first direct evidence for male-mediated gene flow between the mainland and the Alaskan ABC islands, which host a population of bears that has long been of interest to evolutionary biologists, due to the close matrilineal relationship to extant polar bears—the extant polar bear matriline is the sister lineage of the ABC clade (Cronin et al. 1991; Davison et al. 2011). The absence of mainland brown bear mtDNA haplotypes on the ABC islands, and vice versa, shows that female-mediated gene flow is effectively zero. However, nuclear microsatellites (Paetkau et al. 1998) and comparisons of autosomal versus X chromosome variation (Cahill et al. 2013) demonstrated that ABC bears are not isolated from continental brown bear populations, postulating that connectivity between the ABC islands and the mainland stems from male-mediated gene flow. We here show that male-mediated gene flow is connecting the ABC islands to the North American mainland, and that this gene flow is substantial enough to maintain appreciable genetic variability in this island population. Cahill et al. (2013) suggested an initial polar bear ancestry of ABC islands brown bears, followed by extensive male-biased immigration of mainland brown bears. Based on this scenario, the fact that we found no polar bear Y chromosomes on the ABC islands indicates a replacement of the original polar bear Y chromosomes.

Phylogeography: Insights from Matri- and Patrilineal Markers

Since its conception, the field of phylogeography has realized the importance of sampling several statistically independent loci (reviewed in Avise 2000), but problems related to discovering intraspecific variability on the Y chromosome (Hellborg and Ellegren 2004; Luo et al. 2007) have long hampered the application of patrilineal markers in nonmodel species. Nevertheless, some studies have revealed similar paternal and maternal structuring (Hellborg et al. 2005), while others recovered discordant signals (Boissinot and Boursot 1997; Roca et al. 2005; Pidancier et al. 2006; Pérez et al. 2011). Inference of the mechanism(s) that could have led to differences in genetic structuring between the matri- and patriline is generally not straightforward, because the effects of demography and selection are difficult to disentangle (Lawson Handley et al. 2006; Pidancier et al. 2006; Nakagome et al. 2008; Pérez et al. 2011), even in humans (Wilson Sayres et al. 2014). Regardless whether demography or selection are the ultimate cause, a weaker paternal than maternal structuring is indicative of gene flow among populations, implying that mtDNA alone in such cases overestimates population structuring.

Conclusions

Bears are a prominent and widely cited example in phylogeography, with range-wide signals of pronounced population structuring reported for brown bear mtDNA (Davison et al. 2011). We reexamined this paradigm using paternally inherited markers. In strong contrast to mtDNA data, shallow divergences and lack of pronounced geographic structuring of brown bear Y chromosomes were found. mtDNA-based inferences have thus overestimated phylogeographic structuring, due to extensive male gene flow on regional and range-wide scales. Nevertheless, various adaptive traits have been linked to mtDNA (Ballard and Rand 2005), and the mtDNA of an individual may have important consequences for its phenotype and local adaptation. Phylogeographic structuring of the brown bear matriline into regional assemblages could therefore be adaptively significant. Our findings highlight that evolutionary patterns inferred from mtDNA, despite its popularity, are not representative of the entire genome and that phylogeographic histories of many species may need to be reevaluated. Y-chromosomal data are essential in any phylogeographic analyses of mammals—even in presumably well-studied species such as bears.

Materials and Methods

Identification of Y-Chromosomal Markers

A whole genome sequence assembly of a male polar bear (Li et al. 2011) was used to identify putative Y-chromosomal scaffolds by searching for matches with the sequences of known Y-linked genes (SMCY, ZFY, SRY, UBEY, RMBY). We identified five scaffolds from ~19 to ~390 kb in length (scaffold numbers: *297*, *318*, *369*, *579*, *605*). These scaffolds were extracted and compared with the corresponding sequences in a male brown bear (accession numbers: CBZK010000001–

CBZK010000005) in order to identify genomic regions containing either variable sites or microsatellite motifs, respectively, between the two individuals. To decrease the possible ascertainment bias in the subsequent application of the markers in samples from different species and populations, we did not type these variable sites, but we designed and sequenced 11 polymerase chain reaction (PCR) fragments around them with lengths of at least 500 bp (529–1,216 bp). All variable sites on the black bear branch, and most variable sites within brown and polar bears, respectively, were newly discovered by this sequencing approach (supplementary table S6, Supplementary Material online). All but three variable sites between brown and polar bears, however, were known from the ascertainment panel. Y-chromosomal sequences for each haplotype can be accessed at the EMBL data archive (accession numbers: HG423284–HG423309). The scaffold sequences were then mined for di- and tetranucleotide microsatellites that exhibited at least five uninterrupted repeat units. Primers for nine microsatellite markers are shown in supplementary table S9, Supplementary Material online. Allele size data can be accessed at the DRYAD repository (http://doi.org/10.5061/dryad.3p21q).

PCR fragments obtained from brown, polar, and black bears were then evaluated for their male specificity. This assessment resulted in seven sequence fragments and nine microsatellite markers that were ultimately used (supplementary tables S1 and S6, Supplementary Material online). Male specificity was ensured throughout all experiments by consistently including female DNA controls. See supplementary tables S7–S9, Supplementary Material online, for details on PCR conditions, sequencing, and fragment analysis.

Sampling and DNA Extraction

Tissue and DNA samples from 90 male brown and 40 male polar bears were included in this study, covering large parts of their distribution ranges (fig. 1, table 1, and supplementary table S1, Supplementary Material online). For comparison, we also analyzed four American black bear samples, covering their two previously described mitochondrial clades (supplementary fig. S4, Supplementary Material online), and a male spectacled bear as outgroup for divergence time estimations. All tissue samples originated from animals legally hunted for purposes other than this study or from zoo individuals. Individuals with unknown sex were tested as in Bidon et al. (2013). DNA was extracted using a modified Puregene (Qiagen, Hilden, Germany) DNA salt extraction protocol or DNeasy Tissue kit (Qiagen).

Analysis of Y-Chromosomal Scaffold Sequences

Genomic sequence data was used from 12 male polar bears, 1 male brown bear, and 1 male black bear (Miller et al. 2012), plus 1 male brown bear from Northern Europe (supplementary table S3, Supplementary Material online). Short reads were mapped to a >390-kb-long putative Y-linked scaffold from a male polar bear (Li et al. 2011) (scaffold 297). Consensus sequences were determined for every individual using Geneious 6.1.6 (Biomatters, Auckland, New Zealand), calling "?" for regions without coverage and "N" for bases with a Phred quality score <20. Consensus sequences of the 15 individuals were aligned and single-nucleotide variants determined in regions with coverage for all individuals. All variants were manually checked in the alignment, and we excluded all sites that contained insertions/deletions or ambiguous bases. Additionally, variants within 5 nt of ambiguous sites (? and N, respectively), variants directly adjacent to each other, and variants in microsatellite regions were excluded, in order to account for sequencing and alignment errors.

Data Analysis

PCR products were sequenced or subjected to fragment analysis (microsatellites). Sequences were aligned and edited in Geneious 5.6.2 (Biomatters, Auckland, New Zealand) and allele sizes were determined using Genemapper 4.0 (Applied Biosystems, Life Technologies GmbH, Darmstadt, Germany). To infer phylogenetic relationships among haplotypes, networks were estimated using statistical parsimony as implemented in TCS 1.21 (Clement et al. 2000), with the connection limit set to 0.95 for sequence data or fixed at 50 steps for microsatellite haplotypes. For the combined analysis of sequence and allele size polymorphisms, data from all Y-linked markers were combined into one compound haplotype per individual. A haplotype distance matrix was calculated from allele sizes with GenoDive 2.0b23 (Meirmans and Van Tienderen 2004), assuming a strictly stepwise mutation model, with single repeat unit changes counted as one mutational step. Analyses of polymorphic sites and other summary statistics, nucleotide diversity π, tests for signals of demography and selection (Tajima 1989; Fu and Li 1993; Fu 1997), and analysis of molecular variance (AMOVA) were done in DnaSP v5.10 (Librado and Rozas 2009) and Arlequin 3.5 (Excoffier and Lischer 2010). Haplotype configuration tests were performed in haploconfig and haplofreq (Innan et al. 2005), with theta values obtained from the number of segregating sites (Watterson's theta) and nucleotide diversity (π), respectively, and simulating different population expansion scenarios ($\theta = 1.38, 0.37$; growth rate $g = 0, 2, 10$; $a = 10{,}000$; $n = 44$; $s = 6$). Different weighting schemes were applied to sequence and microsatellite markers, as in Brown et al. (2011). Estimates of mean (±SE) among-group distances were obtained in MEGA5 (Tamura et al. 2011). SplitsTree4 (Huson and Bryant 2006) was used to calculate a NeighborNet network for the 390-kb-long data set. Bayesian phylogenetic analyses and divergence time estimations were performed in Beast v1.7.4 (Drummond et al. 2012).

Supplementary Material

Supplementary material, figures S1–S4, and tables S1–S9 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

The authors thank N. Schreck, D. Herbert, and C. Tobiassen for assistance, U. Arnason, M. Bálint, E.W. Born, C. Nowak,

M. Onucsán, K. Skírnisson and F. Zachos for providing samples, and the editor and three anonymous reviewers for insightful comments. This work was supported by Hesse's "LOEWE Landes-Offensive zur Entwicklung Wissenschaftlich-ökonomischer Exzellenz", by the Arthur und Aenne Feindt-Stiftung, the Estonian Research Council (IUT-2032, ESF-8525), and the European Union through the European Regional Development Fund (Centre of Excellence FIBIR). Jón Baldur Hlíðberg kindly provided the bear paintings. The findings and conclusions in this article are those of the author(s) and do not necessarily represent the views of the U.S. Fish and Wildlife Service.

References

Avise JC. 2000. Phylogeography: the history and formation of species. Cambridge (MA): Harvard University Press.

Bachtrog D, Kirkpatrick M, Mank JE, McDaniel SF, Pires JC, Rice W, Valenzuela N. 2011. Are all sex chromosomes created equal? *Trends Genet.* 27:350–357.

Ballard JWO, Rand DM. 2005. The population biology of mitochondrial DNA and its phylogenetic implications. *Annu Rev Ecol Evol Syst.* 36: 621–642.

Bidon T, Frosch C, Eiken HG, Kutschera VE, Hagen SB, Aarnes SG, Fain SR, Janke A, Hailer F. 2013. A sensitive and specific multiplex PCR approach for sex identification of ursine and tremarctine bears suitable for non-invasive samples. *Mol Ecol Resour.* 13:362–368.

Boissinot S, Boursot P. 1997. Discordant phylogeographic patterns between the Y chromosome and mitochondrial DNA in the house mouse: selection on the Y chromosome? *Genetics* 146:1019–1034.

Brown SK, Pedersen NC, Jafarishorijeh S, Bannasch DL, Ahrens KD, Wu J-T, Okon M, Sacks BN. 2011. Phylogenetic distinctiveness of Middle Eastern and Southeast Asian village dog Y chromosomes illuminates dog origins. *PLoS One* 6:e28496.

Brumfield RT, Beerli P, Nickerson DA, Edwards SV. 2003. The utility of single nucleotide polymorphisms in inferences of population history. *Trends Ecol Evol.* 18:249–256.

Cahill JA, Green RE, Fulton TL, Stiller M, Jay F, Ovsyanikov N, Salamzade R, 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.

Campagna L, Van Coeverden de Groot PJ, Saunders BL, Atkinson SN, Weber DS, Dyck MG, Boag PT, Lougheed SC. 2013. Extensive sampling of polar bears (*Ursus maritimus*) in the Northwest Passage (Canadian Arctic Archipelago) reveals population differentiation across multiple spatial and temporal scales. *Ecol Evol.* 3:3152–3165.

Chan Y-C, Roos C, Inoue-Murayama M, Inoue E, Shih C-C, Vigilant L. 2012. A comparative analysis of Y chromosome and mtDNA phylogenies of the Hylobates gibbons. *BMC Evol Biol.* 12:150.

Charlesworth B, Charlesworth D. 2000. The degeneration of Y chromosomes. *Philos Trans R Soc Lond B Biol Sci.* 355:1563–1572.

Chesser RK, Baker RJ. 1996. Effective sizes and dynamics of uniparentally and diparentally inherited genes. *Genetics* 144:1225–1235.

Clement M, Posada D, Crandall KA. 2000. TCS: a computer program to estimate gene genealogies. *Mol Ecol.* 9:1657–1660.

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

Cronin MA, Amstrup SC, Scribner KT. 2006. Microsatellite DNA and mitochondrial DNA variation in polar bears (*Ursus maritimus*) from the Beaufort and Chukchi seas, Alaska. *Can J Zool.* 660:655–660.

Cronin MA, MacNeil MD. 2012. Genetic relationships of extant brown bears (*Ursus arctos*) and polar bears (*Ursus maritimus*). *J Hered.* 103: 873–881.

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.

Davison J, Ho SYW, Bray SC, Korsten M, Tammeleht E, Hindrikson M, Østbye K, Østbye E, Lauritzen S-E, Austin J, et al. 2011. Late-Quaternary biogeographic scenarios for the brown bear (*Ursus arctos*), a wild mammal model species. *Quat Sci Rev.* 30:418–430.

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'Conell 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.

Fu Y-X. 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* 147: 915–925.

Fu Y-X, Li W-H. 1993. Statistical tests of neutrality of mutations. *Genetics* 133:693–709.

Geraldes A, Carneiro M, Delibes-Mateos M, Villafuerte R, Nachman MW, Ferrand N. 2008. Reduced introgression of the Y chromosome between subspecies of the European rabbit (*Oryctolagus cuniculus*) in the Iberian Peninsula. *Mol Ecol.* 17:4489–4499.

Greminger MP, Krützen M, Schelling C, Pienkowska-Schelling A, Wandeler P. 2010. The quest for Y-chromosomal markers - methodological strategies for mammalian non-model organisms. *Mol Ecol Resour.* 10:409–420.

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* 339:1522–1522.

Hailer F, Kutschera VE, Hallstrom 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.

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

Hellborg L, Gündüz I, Jaarola M. 2005. Analysis of sex-linked sequences supports a new mammal species in Europe. *Mol Ecol.* 14: 2025–2031.

Hewitt G. 2000. The genetic legacy of the Quaternary ice ages. *Nature* 405:907–913.

Hirata D, Mano T, Abramov AV, Baryshnikov GF, Kosintsev PS, Vorobiev AA, Raichev EG, Tsunoda H, Kaneko Y, Murata K, et al. 2013. 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.

Ho SYW, Saarma U, Barnett R, Haile J, Shapiro B. 2008. The effect of inappropriate calibration: three case studies in molecular ecology. *PLoS One* 3:e1615.

Hughes JF, Rozen S. 2012. Genomics and genetics of human and primate Y chromosomes. *Annu Rev Genomics Hum Genet.* 13:83–108.

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

Ingman M, Gyllensten U. 2001. Analysis of the complete human mtDNA genome: methodology and inferences for human evolution. *J Hered.* 92:454–461.

Innan H, Zhang K, Marjoram P, Tavaré S, Rosenberg NA. 2005. Statistical tests of the coalescent model based on the haplotype frequency distribution and the number of segregating sites. *Genetics* 169: 1763–1777.

Keis M, Remm J, Ho SYW, Davison J, Tammeleht E, Tumanov IL, Saveljev AP, Männil P, Kojola I, Abramov AV, et al. 2013. Complete mitochondrial genomes and a novel spatial genetic method reveal cryptic phylogeographical structure and migration patterns among brown bears in north-western Eurasia. *J Biogeogr.* 40: 915–927.

Kohn M, Knauer F, Stoffella A, Schröder W, Pääbo S. 1995. Conservation genetics o the European brown bear—a study using excremental PCR of nuclear and mitochondrial sequences. *Mol Ecol.* 4:95–103.
Kopatz A, Eiken HG, Hagen SB, Ruokonen M, Esparza-Salas R, Schregel J, Kojola I, Smith ME, Wartiainen I, Aspholm PE, et al. 2012. Connectivity and population subdivision at the fringe of a large brown bear (*Ursus arctos*) population in North Western Europe. *Conserv Genet.* 13:681–692.
Korsten M, Ho SYW, Davison J, Pähn B, Vulla E, Roht M, Tumanov IL, Kojola I, Andersone-Lilley Z, Ozolins J, et al. 2009. Sudden expansion of a single brown bear maternal lineage across northern continental Eurasia after the last ice age: a general demographic model for mammals? *Mol Ecol.* 18:1963–1979.
Lawson Handley LJ, Berset-Brändli L, Perrin N. 2006. Disentangling reasons for low Y chromosome variation in the greater white-toothed shrew (*Crocidura russula*). *Genetics* 173:935–942.
Li B, Zhang G, Willerslev E, Wang J. 2011. Genomic data from the Polar Bear (*Ursus maritimus*) Gigascience. [cited 2014 Mar 7]. Available from: http://dx.doi.org/10.5524/100008.
Librado P, Rozas J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25:1451–1452.
Lindqvist C, Schuster SC, Sun Y, Talbot SL, Qi J, Ratan A, Tomsho LP, Kasson L, Zeyl E, Aars J, et al. 2010. Complete mitochondrial genome of a Pleistocene jawbone unveils the origin of polar bear. *Proc Natl Acad Sci U S A.* 107:5053–5057.
Lippold S, Knapp M, Kuznetsova T, Leonard JA, Benecke N, Ludwig A, Rasmussen M, Cooper A, Weinstock J, Willerslev E, et al. 2011. Discovery of lost diversity of paternal horse lineages using ancient DNA. *Nat Commun.* 2:450.
Luo S-J, Johnson WE, David VA, Menotti-Raymon M, Stanyon R, Cai QX, Beck T, Yuhki N, Pecon-Slattery J, Smith JLD, et al. 2007. Development of Y chromosome intraspecific polymorphic markers in the Felidae. *J Hered.* 98:400–413.
McLellan BN, Hovey FW. 2001. Natal dispersal of grizzly bears. *Can J Zool.* 79:838–844.
Meadows JRS, Hanotte O, Drögemüller C, Calvo J, Godfrey R, Coltman D, Maddox JF, Marzanov N, Kantanen J, Kijas JW. 2006. Globally dispersed Y chromosomal haplotypes in wild and domestic sheep. *Anim Genet.* 37:444–453.
Meirmans PG, Van Tienderen PH. 2004. GENOTYPE and GENODIVE: two programs for the analysis of genetic diversity of asexual organisms. *Mol Ecol Notes.* 4:792–794.
Miller W, Schuster SC, Welch AJ, Ratan A, Bedoya-Reina OC, Zhao F, Kim HL, 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.
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.
Paetkau D, Amstrup SC, Born EW, Calvert W, Derocher AE, Garner GW, Messier F, Stirling I, Taylor MK, Wiig Ø, et al. 1999. Genetic structure of the world's polar bear populations. *Mol Ecol.* 8:1571–1584.
Paetkau D, Shields GF, Strobeck C. 1998. Gene flow between insular, coastal and interior populations of brown bears in Alaska. *Mol Ecol.* 7:1283–1292.
Paetkau D, Waits LP, Clarkson PL, Craighead L, Strobeck C. 1997. An Empirical Evaluation of Genetic Distance Statistics Using Microsatellite Data From Bear (Ursidae) Populations. *Genetics* 147: 1943–1957.
Pérez T, Hammer SE, Albornoz J, Domínguez A. 2011. Y-chromosome phylogeny in the evolutionary net of chamois (genus Rupicapra). *BMC Evol Biol.* 11:272.
Petit E, Balloux F, Excoffier L. 2002. Mammalian population genetics: why not Y? *Trends Ecol Evol.* 17:28–33.
Petit RJ, Excoffier L. 2009. Gene flow and species delimitation. *Trends Ecol Evol.* 24:386–393.
Pidancier N, Jordan S, Luikart G, Taberlet P. 2006. Evolutionary history of the genus Capra (Mammalia, Artiodactyla): discordance between mitochondrial DNA and Y-chromosome phylogenies. *Mol Phylogenet Evol.* 40:739–749.
Purvis A. 2005. Phylogeny and conservation. Cambridge: Cambridge University Press.
Pusey A. 1987. Sex-biased dispersal and inbreeding avoidance in birds and mammals. *Trends Ecol Evol.* 2:295–299.
Roca AL, Georgiadis N, O'Brien SJ. 2005. Cytonuclear genomic dissociation in African elephant species. *Nat Genet.* 37:96–100.
Sacks BN, Brown SK, Stephens D, Pedersen NC, Wu J-T, Berry O. 2013. Y chromosome analysis of dingoes and southeast asian village dogs suggests a neolithic continental expansion from Southeast Asia followed by multiple austronesian dispersals. *Mol Biol Evol.* 30: 1103–1118.
Subramanian S, Denver DR, Millar CD, Heupink T, Aschrafi A, Emslie SD, Baroni C, Lambert DM. 2009. High mitogenomic evolutionary rates and time dependency. *Trends Genet.* 25:482–486.
Taberlet P, Bouvet J. 1994. Mitochondrial DNA polymorphism, phylogeography, and conservation genetics of the brown bear *Ursus arctos* in Europe. *Proc R Soc Lond B Biol Sci.* 255:195–200.
Taberlet P, Fumagalli L, Wust-Saucy A, Cosson J. 1998. Comparative phylogeography and postglacial colonization routes in Europe. *Mol Ecol.* 7:453–464.
Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123:585–595.
Tammeleht E, Remm J, Korsten M, Davison J, Tumanov I, Saveljev A, Männil P, Kojola I, Saarma U. 2010. Genetic structure in large, continuous mammal populations: the example of brown bears in northwestern Eurasia. *Mol Ecol.* 19:5359–5370.
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.
Waits L, Taberlet P, Swenson JE, Sandegren F, Franzén R. 2000. Nuclear DNA microsatellite analysis of genetic diversity and gene flow in the Scandinavian brown bear (*Ursus arctos*). *Mol Ecol.* 9: 421–431.
Wayne RK, Van Valkenburgh B, O'Brien SJ. 1991. Molecular distance and divergence time in carnivores and primates. *Mol Biol Evol.* 8: 297–319.
Wei W, Ayub Q, Chen Y, McCarthy S, Hou Y, Carbone I, Xue Y, Tyler-Smith C. 2013. A calibrated human Y-chromosomal phylogeny based on resequencing. *Genome Res.* 23:388–395.
Willard HF. 2003. Tales of the Y chromosome. *Nature* 423:810–813.
Wilson Sayres MA, Lohmueller KE, Nielsen R. 2014. Natural selection reduced diversity on human y chromosomes. *PLoS Genet.* 10: e1004064.
Xue Y, Wang Q, Long Q, Ng BL, Swerdlow H, Burton J, Skuce C, Taylor R, Abdellah Z, Zhao Y, et al. 2009. Human Y chromosome base-substitution mutation rate measured by direct sequencing in a deep-rooting pedigree. *Curr Biol.* 19:1453–1457.
Zedrosser A, Støen O-G, Sæbø S, Swenson JE. 2007. Should I stay or should I go? Natal dispersal in the brown bear. *Anim Behav.* 74: 369–376.
Zerjal T, Xue Y, Bertorelle G, Wells RS, Bao W, Zhu S, Qamar R, Ayub Q, Mohyuddin A, Fu S, et al. 2003. The genetic legacy of the Mongols. *Am J Hum Genet.* 72:717–721.

Supplementary Information

Brown and polar bear Y chromosomes reveal extensive male-biased gene flow within brother lineages

Tobias Bidon*, Axel Janke, Steven R. Fain, Hans Geir Eiken, Snorre B. Hagen, Urmas Saarma, Björn M. Hallström, Nicolas Lecomte, Frank Hailer*

* Corresponding authors: Tobias Bidon & Frank Hailer. Biodiversity and Climate Research Centre (BiK-F), Senckenberg Gesellschaft für Naturforschung, Senckenberganlage 25, 60325 Frankfurt am Main, Germany. Fax: +49 6975421800; E-mails: tobias.bidon@senckenberg.de; frashai@gmx.net

Supplementary information includes

Analysis of fast evolving male-specific microsatellites 2
Combination of sequence data and microsatellites 2
Population differentiation in brown and polar bears 3
American black bear phylogeography 3
Details of laboratory methods: PCR conditions, sequencing, and fragment analysis 4
Details of sampling and DNA extraction 5
Details for divergence time estimation 5
Figure S1. Statistical parsimony networks of Y chromosome haplotypes in bears. 7
Figure S2. Allele size distribution histograms of six Y-linked microsatellite markers. 7
Figure S3. Median joining networks (Bandelt et al. 1999) of weighted Y haplotypes within species, inferred from the combination of 3.1 kb sequence data and six microsatellite markers. 8
Figure S4. Statistical parsimony network of American black bear haplotypes. 8
Table S1 List of individuals, their particular haplotypes, and source populations. 9
Table S2. Divergence time estimates of Y chromosome lineages in bears, obtained from different Beast runs. 14
Table S3. Localities and accession numbers of bears used in the scaffold-wide analyses. 14
Table S4. Pairwise ΦST-values among brown bear populations. 14
Table S5. Pairwise ΦST-values among polar bear populations. 15
Table S6. Segregating sites of Y-chromosome haplotypes in the 3.1 kb (A) and 5.3 kb (B) datasets. 16
Table S7 Primer sequences, annealing temperatures and fragment sizes of Y-chromosomal sequence fragments. 17
Table S8. Primer sequences, annealing temperature, and fragment size of small, overlapping Y-chromosomal sequence markers used to amplify DNA from lower-quality samples. 18
Table S9. Primer sequences, fluorescent dyes, and repeat unit of Y-chromosomal microsatellite markers. 19
References 20

Analysis of fast evolving male-specific microsatellites

Microsatellite haplotype networks (fig. S1*A*) showed the three species (brown, polar and black bears) separated from each other, but with only few mutational differences between brown and polar bears compared to their intraspecific variability. However, as expected from the higher mutation rate of microsatellite markers compared to point mutations (Ellegren 2004), compound haplotypes of the Y-linked microsatellites provided greatly increased intraspecific resolution in comparison to the sequence data (fig. 2*A*). Combining the microsatellite results with the sequence data from the 3.1 kb dataset (solid lines in fig. 2*A*; the 5.3 kb dataset includes only a subset of individuals) yielded a phylogenetic network not only with high power to differentiate among species, but also with large intraspecific resolution due to the microsatellite allele size polymorphism data (fig. S1*B*). The sequence data therefore served as a backbone to infer phylogenetic relationships, and the addition of more rapidly mutating microsatellite markers yielded power to resolve intraspecific structuring (Brown et al. 2011; Sacks et al. 2013).

All samples were genotyped at six Y-linked microsatellite markers. Of these, *318.9* was monomorphic in polar bears, and *318.4* and *318.2* were monomorphic in both polar and black bears (fig. S2). No marker was monomorphic in brown bears. Generally, the allele size ranges of brown and polar bears overlapped (*318.9*, *369.1*, *318.1*, and *318.6*) and brown bears showed more alleles per marker (range: 2-9) than polar bears (range: 1-6). Differences in allele sizes between brown and polar bears were fixed only at *318.4*, rendering this marker useful for screening purposes aiming at the discrimination between Y chromosomes of the two species. Brown bears had unique alleles that were absent from polar and black bears at marker *318.4*, and black bears had unique alleles at *318.2* and *318.6*. Despite the small sample size, black bears exhibited a large allelic variation, e.g., four black bear males had a higher number of alleles than 40 polar bears at markers *318.9* and *318.1*, respectively. Further, black bears carried alleles that were separated from brown/polar bears by substantial allele size differences at several markers. Interestingly, 13 brown bears from Eastern Asia (Kamchatka) carried divergent alleles at *369.1* (fig. 3). Sequencing of one individual from each species for the markers (*318.9*, *318.2*, *369.1*, *318.1*, *318.6*), plus two additional brown bears for marker *369.1*, confirmed that the substantial allele size variation indeed arose from repeat number polymorphism within the microsatellite region and not from insertions/deletions in the flanking regions.

Combination of sequence data and microsatellites

Combining Y chromosome sequences of the 3.1 kb dataset with microsatellite data provided discrimination of 41 haplotypes in 90 brown bear individuals, and 17 haplotypes in 40 polar bears (figs. 3*A* and *B*).

Two haplotypes were shared by brown bears from different populations (ABC-islands and Canada; northwest America, ABC-islands and western Asia). All other haplotypes were exclusively shared with individuals from the same population. The overall high haplotypic variability allows detection of several geographical groupings in the network, but branches between these groups are short. Major conclusions drawn from these unweighted networks

were not affected by weighting the six microsatellite markers inversely to their variance and, additionally, weighting sequence markers differentially (10 times the highest microsatellite weight) than microsatellites (Brown et al. 2011; Sacks et al. 2013) (fig. S3). The clustering of North and Central European haplotypes as well as the clustering of East Asian (Kamchatkan) brown bear haplotypes near the Northwest American haplotypes was more pronounced when markers were weighted (figs. 3*A* and S3*A*), as expected given likely dispersal across the Beringian landbridge around the LGM (last glacial maximum).
As shown in fig. 3*B*, of 17 polar bear haplotypes, one was found in 15 individuals (38%) from regions as far apart as Alaska, western Greenland and Davis Strait. All surveyed populations contained at least three different haplotypes. No different conclusions were reached when a weighting scheme was applied to the markers (fig. S3*B*), although the compound haplotypes based on PO2 now clustered next to each other.

Population differentiation in brown and polar bears

Estimation of population differentiation within brown and polar bears (Weir and Cockerham 1984) (Φ_{ST}) was calculated in Arlequin 3.5 (Excoffier and Lischer 2010) based on a distance matrix from combination of six microsatellites and seven sequence fragments (same as for the networks in fig. 3, figs. S1*B*, S4). In brown bears Φ_{ST} –values ranged from 0.02 to 0.38 (table S4), indicating low to moderate population differentiation. We found a stronger differentiation between populations with larger geographic separation (fig. 1), e.g. Northern Europe compared to Eastern Asia or Canada and less differentiation between populations that are spatially closer, e.g. NW-America and Canada. Non-significant and low values were only obtained for population pairs that are geographic neighbors (fig. 3), e.g. NW-America and the ABC-Islands. The proportion of variation among all brown bear populations was 0.28, as estimated by analysis of molecular variance (AMOVA) in Arlequin. In polar bears Φ_{ST} -values ranged from <-0.001 to 0.51 (table S5), with three out of six comparisons being non-significant. The proportion of variation among all polar bear populations was 0.16. We note the low number of samples for some polar bear populations, rendering the obtained differentiation indices less reliable in this species.

American black bear phylogeography

In American black bears, two distinct mtDNA lineages have been described: a coastal lineage that is restricted to a zone along the North American Pacific coast, and a second, more widespread continental lineage that occurs from north and east of the Rocky Mountains to Florida (Wooding and Ward 1997; Stone and Cook 2000). In the Y-chromosomal 5.3 kb dataset (fig. 2*A*), the four male black bears showed two haplotypes, separated by two mutations. Nucleotide diversity (π ±S.D.) was higher in black bears (0.00025±0.00008; n=4) compared to brown (0.00007±0.00002; n=44) and polar bears (0.00003±0.00002; n=15). Combined Y-chromosomal data from sequences and microsatellites also showed two distinct lineages (fig. S4), a pattern that resembles findings from mtDNA. However, discordance between maternal and paternal lineages occurred in a black bear individual from Montana: while its mitochondrial haplotype clustered with two other black bear individuals (Vermont

and C122), potentially representing a continental lineage, its Y-chromosomal haplotype clustered with the individual from Oregon (which had a distinct mitochondrial haplotype representing a coastal lineage (Wooding and Ward 1997)). Sex-biased introgression among lineages may explain this pattern (Roca et al. 2005; Hailer and Leonard 2008). We note, however, that many more than four individuals are needed to investigate the phylogeography of black bears - we report these findings to demonstrate that our newly developed markers can also be used for Y chromosome studies in other bear species.

Details of laboratory methods: PCR conditions, sequencing, and fragment analysis

Touchdown polymerase chain reactions (PCRs) were performed to amplify all sequence markers in 15 µl reaction volumes containing 2x Taq DNA Polymerase mix (VWR International GmbH, BDH Prolabo, Darmstadt, Germany), 0.17 µg/µl BSA (New England Biolabs, Ipswich, MA, USA), 0.27 µM each of forward and reverse primers, and 10-15 ng template DNA. Re-runs of individual failed PCRs were conducted with GoTaq Flexi DNA Polymerase (Promega, Madison, WI, USA), using 3.2 mM $MgCl_2$ (Promega) and 10-100 ng of DNA template in a 20 µl volume. The amplification protocol started with 95°C for 3 min followed by 10 cycles at 94°C for 30 s, a touchdown step starting at 69°C, 68°C, or 66°C (marker-specific, see table S7) for 25 s, and 72°C for 75 s. This was followed by 25 cycles of 94°C for 30 s, 64°C, 58°C, or 61°C (table S7) for 25 s, and 72°C for 75 s. A final elongation step of 10 min at 72°C followed. Each PCR setup contained no-template and female controls. For samples with low DNA quality, amplicons were subdivided into smaller, overlapping fragments (each 334-465 bp long; table S8). PCR products were checked on agarose gels and sequenced in both directions with the BigDye Terminator v3.1 chemistry on an ABI 3730 (Applied Biosystems, Life Technologies GmbH, Darmstadt, Germany). Sequences were aligned and edited in Geneious 5.6.2 (Biomatters, Auckland, New Zealand). Insertions/deletions longer than one bp were shortened to one bp in the alignments and regarded as a single mutational step in downstream network analyses. Two of the PCR fragments (*318.2C* and *318.7C*) overlap by 142 bp and were trimmed accordingly in order to remove doubly represented sequence in the alignments.

Nine Y-linked microsatellite markers were combined in two multiplex PCRs (table S9). Forward primers were labeled with fluorescent dyes PET, 6-FAM, VIC, or NED (Applied Biosystems, Life Technologies GmbH) and mixed with unlabeled forward primers (1:8). PCRs were performed in 10 µl volumes using 5 µl of 2x Multiplex PCR Mastermix (Qiagen, Hilden, Germany), 0.05 µg/µl BSA (New England Biolabs), and 10-15 ng of template DNA. Primer concentrations are listed in table S9. Amplification started at 95°C for 3 min followed by 20 touchdown cycles of 94°C for 30 s, 68°C (decreasing by 0.5°C per cycle) for 25 s, and 72°C for 75 s. Another 20 standard PCR cycles followed at 94°C for 30 s, 58°C for 25 s, and 72°C for 75 s, completed by a final elongation step at 72°C for 10 min. PCR products were subjected to fragment analysis on an ABI 3730 (Applied Biosystems, Life Technologies GmbH) instrument using the 500(-250) LIZ size standard. Negative controls (no-template controls and DNA from females) were included. Allele sizes were determined using Genemapper 4.0 (Applied Biosystems, Life Technologies GmbH).

Details of sampling and DNA extraction

Brown bear samples originated from Central and Northern Europe, several regions throughout Asia as well as North America including the Alaskan Admiralty, Baranof and Chichagof (ABC) Islands and two individuals from the adjacent mainland. Polar bears from Iceland, Franz Josef Land (Arctic Ocean, north of Siberia), Alaska, Baffin Bay, Davis Strait, and Kane Basin were included. Black bears were sampled from the Pacific northwest (Oregon), central Rocky Mountains (Montana), and northeastern USA (Vermont). A fourth black bear individual came from a zoo in Alaska with uncertain geographic origin. After extraction, nucleic acid concentration was determined on a NanoPhotometer instrument (Implen, München, Germany).

Sequences from a 642 bp fragment of the mitochondrial control region, amplified and sequenced as in Hailer et al. (2012), confirmed that our samples represent major matrilineal (mtDNA) lineages of brown and polar bears (Davison et al. 2011) (clades 1, 2A, 2B, 3A, 3B, and 4; fig. 2*C*; EMBL accession numbers for newly obtained mitochondrial sequences (n=116): HG426316-HG426431; 18 sequences from Hailer et al. (2012). Our sampling thus covers the major Eurasian and North American lineages of brown bears, and no mismatches occurred between clade assignment and geographic/taxonomic origin.

Six brown bears from central Europe formed clade 1, 11 individuals from the ABC-islands represented clade 2A. 59 bears from central and northern Europe, western and eastern Asia and northwestern America belonged to clade 3A. Clade 3B was represented by five individuals from northwestern America (including two individuals from the mainland adjacent to the ABC-islands) and all eight bears from Canada. One individual from Idaho represented clade 4. All 40 polar bears belonged to clade 2B, nested within the variation of brown bears. The four black bears clustered outside the brown/polar bear variation. Individuals from Montana and Vermont (as well as the individual with uncertain geographic origin) clustered together, while the individual from Oregon was more distantly related.

Details for divergence time estimation

Sequence data from a male spectacled bear was included as an outgroup in the divergence time estimation in Beast v1.7.4 (Drummond et al. 2012). The alignment contained 5,197 bp of concatenated Y-chromosomal sequence, which is 97 bp less than in the alignment used for the other phylogenetic analyses because of missing data for the spectacled bear. For computational reasons, we included a maximum of ten individuals per haplotype, yielding 33 sequences from four species: six haplotypes in brown bears, two haplotypes each in polar and black bears, and one in the spectacled bear. We used the HKY substitution model, as indicated by the Bayesian Information Criterion procedure implemented in jModelTest 2.1.1 (Darriba et al. 2012). For time calibration, we used a divergence time prior on the spectacled bear of 6 Ma (Wayne et al. 1991) and explored other calibration scenarios as well (scenarios A and C in table S2), a Yule process as tree prior, empirical base frequencies, and a strict clock with an uninformative, flat prior (substitution rate 0-0.1). Additionally, we restricted the analysis to a mutation rate of $3x10^{-9}$ mutations/site/year, assuming a generation time of 10 years (scenario D, table S2). The program was run for $2x10^{9}$ generations that were sampled

every 100,000 generations. Tracer 1.5 (Rambaut and Drummond 2007) was used to confirm convergence. Maximum clade credibility trees were obtained from TreeAnnotator 1.7.4, discarding a burn-in of 10%.

Median estimates of $T_{MRCA\ (B/P)}$ for the different scenarios ranged from ca. 0.43 million years ago (Ma) assuming a fixed mutation rate of 3.0 x 10^{-9} mutations/site/year to ca. 2.06 Ma, assuming a split of the spectacled bear more than 10 Ma (table S2).

For divergences within brown bears, we additionally explored the analysis of Y-chromosomal microsatellite data alone to infer the age of the most recent common ancestor using the Rho-statistics as implemented in Network 4.6 (Bandelt et al. 1999). By choosing one node as ancestral and the other nodes as descendent the T_{MRCA} can be estimated. The outcome, however, depends directly on the assumed microsatellite mutation rate, which is known to vary within two orders of magnitude in humans (Ellegren 2004). Unfortunately, there is no information available on the mutation rate of the herein analyzed Y-specific microsatellites in bears. This uncertainty regarding microsatellite mutation rates translated directly into a wide range of conceivable T_{MRCA} values for our data. Based on a generation time of 10 years for brown and polar bears, and the divergence between a central, unsampled haplotype and the most divergent haplotype (found on Kamchatka), T_{MRCA} estimates for all brown bear haplotypes ranged from 615,392 years ago (± 38,462), assuming a mutation rate of 2.6 x 10^{-5} /site/year (Forster et al. 2000), to 12,304 years ago (± 769), assuming a 50-fold faster rate of 1.3 x 10^{-3} /site/year.

The mtDNA control region sequences (642 bp) from all individuals were aligned and a Bayesian maximum clade credibility tree was reconstructed with Beast using a HKY substitution model, constant size tree prior, empirical base frequencies and sampling 500 million steps every 100,000 generations.

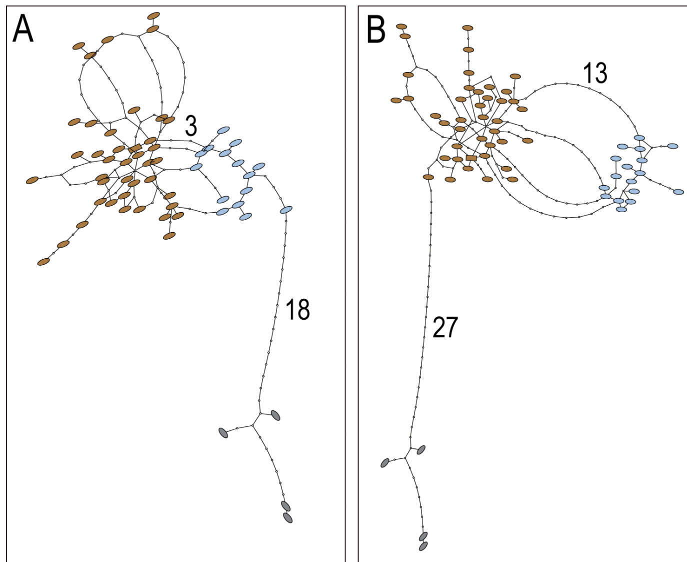


Figure S1. Statistical parsimony networks of Y chromosome haplotypes in bears.
(A) Haplotypes inferred from six Y-linked microsatellite markers. (B) Compound haplotypes inferred from a combination of 3.1 kb Y-chromosomal sequence data and six Y-linked microsatellite markers. All markers were weighted equally. Numbers on branches refer to the number of substitutions between the species. Colors: brown (brown bears), blue (polar bears), dark gray (black bears). The connection limit was set to 50 steps. At the 95% connectivity limit, haplotypes of the three species were not connected to each other.

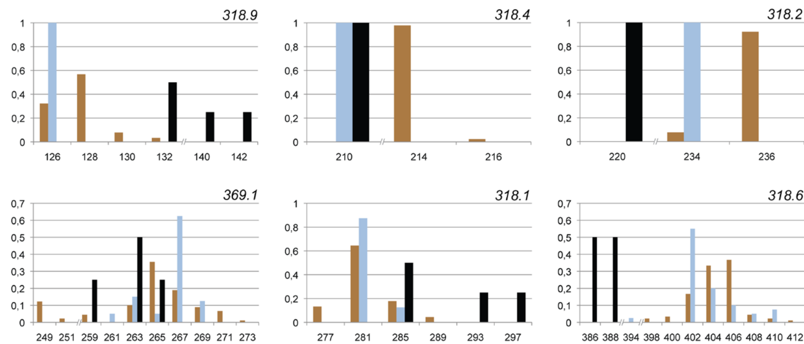


Figure S2. Allele size distribution histograms of six Y-linked microsatellite markers.
Genotyped individuals comprise 90 brown bears, 40 polar bears, and 4 black bears, represented by brown, blue, and black bars, respectively. Note discontinuities in the allele size distributions for *318.9*, *318.4*, *318.2*, *369.1*, and *318.6* (indicated by slashes).

7

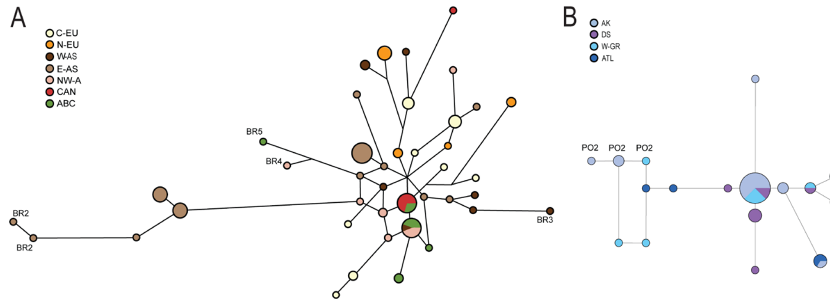


Figure S3. Median joining networks (Bandelt et al. 1999) of weighted Y haplotypes within species, inferred from the combination of 3.1 kb sequence data and six microsatellite markers.

Microsatellite markers were weighted inversely to their variance and sequence markers weighted ten times the highest microsatellite weight (Brown et al. 2011). (A) Brown bears: *318.9*=6, *318.4*=9, *318.2*=9, *369.1*=1, *318.1*=6 and *318.6*=2. (B) Polar bears: *318.9*=9, *318.4*=9, *318.2*=9, *369.1*=5, *318.1*=8 and *318.6*=4. Population abbreviations are as in table 1.

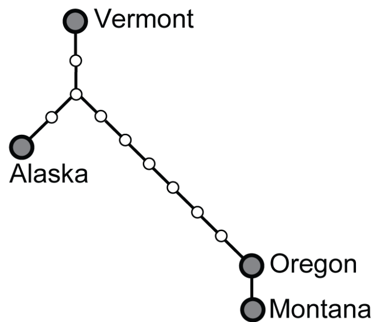


Figure S4. Statistical parsimony network of American black bear haplotypes.

Haplotypes were inferred from the unweighted combination of 3.1 kb Y-chromosomal sequence data and six microsatellites.

Table S1 List of individuals, their particular haplotypes, and source populations.

For a subset of individuals (indicated by “x”), additional ~2.2 kb of Y-chromosomal sequence data was obtained. Fragment sizes at 9 microsatellite markers are shown (http://doi.org/10.5061/dryad.3p21q). Asterisks indicate microsatellite markers not included in downstream analyses due to pseudoheterozygous genotypes for some individuals. Background shading (white or gray) identifies individuals that share identical haplotypes for the 3.1 kb sequence data set combined with genotypes from six microsatellite loci.

	Species	Haplotype (sequence fragments and microsatellites)	Haplotype (sequence fragments)	Population	Mitochondrial clade	Additional Y-chromosomal sequence obtained? (total: 5.3 kb)	Genebank accession numbers for Y-haplotypes	318.9	318.4	318.2	369.1	318.1	318.6	15020.1*	369.4*	69217.1*
1	*U. arctos*	**BR1-1**	BR1.1	NW-A (AK)	3B	x	HG423290, HG423291	128	214	236	265	281	402	185	197	240, 242
2	*U. arctos*	**BR1-1**	BR1.1	NW-A (AK)	3B			128	214	236	265	281	402	185	201	240, 242
3	*U. arctos*	**BR1-1**	BR1.1	NW-A (AK)	3A	x		128	214	236	265	281	402	185	197	240, 242
4	*U. arctos*	**BR1-1**	BR1.1	NW-A (AK)	3A			128	214	236	265	281	402	185	201	240, 242
5	*U. arctos*	**BR1-1**	BR1.1	ABC	2A	x		128	214	236	265	281	402	185	197	240, 242
6	*U. arctos*	**BR1-1**	BR1.1	ABC	2A	x		128	214	236	265	281	402	185	197	240, 242
7	*U. arctos*	**BR1-1**	BR1.1	ABC	2A	x		128	214	236	265	281	402	185	197	240, 242
8	*U. arctos*	**BR1-1**	BR1.1	ABC	2A	x		128	214	236	265	281	402	185	197	240, 242
9	*U. arctos*	**BR1-1**	BR1.1	W-AS (Ural)	3A			128	214	236	265	281	402	189	201	242
10	*U. arctos*	**BR1-10**	BR1.1	C-Eur	3A	x		128	214	236	259	281	406	187, 189	193	242
11	*U. arctos*	**BR1-11**	BR1.1	C-Eur	1	x		128	214	236	259	281	402	187, 189	193	242
12	*U. arctos*	**BR1-11**	BR1.1	C-Eur	1			128	214	236	259	281	402	187, 189	193	242
13	*U. arctos*	**BR1-12**	BR1.1	C-Eur	1			128	214	236	259	281	400	187, 189	193	242
14	*U. arctos*	**BR1-13**	BR1.1	E-AS (Kam)	3A			126	214	236	249	281	408	185	189	242
15	*U. arctos*	**BR1-14**	BR1.1	E-AS (Kam)	3A	x		126	214	236	249	281	404	185	195	242
16	*U. arctos*	**BR1-14**	BR1.1	E-AS (Kam)	3A			126	214	236	249	281	404	185	195	242
17	*U. arctos*	**BR1-14**	BR1.1	E-AS (Kam)	3A	x		126	214	236	249	281	404	185	195	242
18	*U. arctos*	**BR1-14**	BR1.1	E-AS (Kam)	3A	x		126	214	236	249	281	404	185	195	242
19	*U. arctos*	**BR1-14**	BR1.1	E-AS (Kam)	3A	x		126	214	236	249	281	404	185	195	242
20	*U. arctos*	**BR1-15**	BR1.1	E-AS (Kam)	3A			126	214	236	249	277	404	185	195	242
21	*U. arctos*	**BR1-15**	BR1.1	E-AS (Kam)	3A	x		126	214	236	249	277	404	185	195	242
22	*U. arctos*	**BR1-15**	BR1.1	E-AS (Kam)	3A	x		126	214	236	249	277	404	185	195	242
23	*U. arctos*	**BR1-15**	BR1.1	E-AS (Kam)	3A	x		126	214	236	249	277	404	185	195	242
24	*U. arctos*	**BR1-15**	BR1.1	E-AS (Kam)	3A	x		126	214	236	249	277	404	185	195	242
25	*U. arctos*	**BR1-16**	BR1.1	NW-A (AK)	3A	x		126	214	236	263	281	404	185	195	242
26	*U. arctos*	**BR1-17**	BR1.1	W-AS (Ural)	3A			128	214	236	263	281	406	189	193	242
27	*U. arctos*	**BR1-18**	BR1.1	E-AS (Kam)	3A			126	214	236	263	281	406	185	195	242
28	*U. arctos*	**BR1-19**	BR1.1	E-AS (Kam)	3A			126	214	236	265	281	412	185	195	242
29	*U. arctos*	**BR1-2**	BR1.1	NW-A (AK)	3B	x		128	214	236	263	281	402	185	201	240, 242
30	*U. arctos*	**BR1-20**	BR1.1	E-AS (Kam)	3A			126	214	236	265	281	406	185	195	242
31	*U. arctos*	**BR1-21**	BR1.1	E-AS (Far-E)	3A			126	214	236	267	281	406	185	195	242
32	*U. arctos*	**BR1-21**	BR1.1	E-AS (Kam)	3A			126	214	236	267	281	406	185	195	242
33	*U. arctos*	**BR1-21**	BR1.1	E-AS (Kam)	3A	x		126	214	236	267	281	406	185	199	242
34	*U. arctos*	**BR1-21**	BR1.1	E-AS (Kam)	3A			126	214	236	267	281	406	185	195	242
35	*U. arctos*	**BR1-21**	BR1.1	E-AS (Kam)	3A			126	214	236	267	281	406	185	195	242
36	*U. arctos*	**BR1-21**	BR1.1	E-AS (Kam)	3A			126	214	236	267	281	406	185	195	242
37	*U. arctos*	**BR1-21**	BR1.1	E-AS (Kam)	3A			126	214	236	267	281	406	185	195	242
38	*U. arctos*	**BR1-21**	BR1.1	E-AS (Kam)	3A			126	214	236	267	281	406	185	195	242
39	*U. arctos*	**BR1-21**	BR1.1	E-AS (Kam)	3A			126	214	236	267	281	406	185	195	242
40	*U. arctos*	**BR1-21**	BR1.1	E-AS (Kam)	3A	x		126	214	236	267	281	406	185	199	242
41	*U. arctos*	**BR1-22**	BR1.1	C-Eur	3A	x		128	214	234	265	281	406	189	199	242
42	*U. arctos*	**BR1-23**	BR1.1	C-Eur	3A	x		128	214	234	269	281	406	187, 189	199	242
43	*U. arctos*	**BR1-23**	BR1.1	C-Eur	3A			128	214	234	269	281	406	187, 189	199	242
44	*U. arctos*	**BR1-23**	BR1.1	C-Eur	3A			128	214	234	269	281	406	187, 189	199	242
45	*U. arctos*	**BR1-23**	BR1.1	C-Eur	3A			128	214	234	269	281	406	187, 189	199	242
46	*U. arctos*	**BR1-24**	BR1.1	E-AS (Far-E)	3A			128	214	234	269	277	406	181, 187	195	242
47	*U. arctos*	**BR1-25**	BR1.1	NW-A (ID)	4	x		128	214	234	273	281	406	185	193	242
48	*U. arctos*	**BR1-26**	BR1.1	N-Eur	3A			128	214	236	269	281	406	189	193	242
49	*U. arctos*	**BR1-27**	BR1.1	E-AS (Far-E)	3A	x		128	214	236	265	277	406	181, 187	195	242
50	*U. arctos*	**BR1-28**	BR1.1	W-AS (C-Sib)	3A	x		128	214	236	267	277	408	187	193	242
51	*U. arctos*	**BR1-29**	BR1.1	E-AS (Far-E)	3A	x		128	214	236	267	277	406	181, 187	193	242
52	*U. arctos*	**BR1-3**	BR1.1	ABC	2A	x		128	214	236	265	277	402	185	201	240, 242
53	*U. arctos*	**BR1-30**	BR1.1	W-AS (Ural)	3A	x		130	214	236	267	277	406	181, 187	193	242
54	*U. arctos*	**BR1-31**	BR1.1	N-Eur	3A			128	214	236	265	285	406	187	187	242
55	*U. arctos*	**BR1-31**	BR1.1	N-Eur	3A			128	214	236	265	285	406	187	187	242
56	*U. arctos*	**BR1-32**	BR1.1	C-Eur	3A			130	214	236	267	281	404	187, 189	201	242
57	*U. arctos*	**BR1-33**	BR1.1	C-Eur	1	x		130	214	236	267	285	406	189	197	242
58	*U. arctos*	**BR1-33**	BR1.1	C-Eur	3A	x		130	214	236	267	285	406	189	197	242
59	*U. arctos*	**BR1-33**	BR1.1	C-Eur	1			130	214	236	267	285	406	189	197	242

60	*U. arctos*	**BR1-34**	BR1.1	W-AS (Ural)	3A			130	214	236	271	285	406	185, 187	195	242
61	*U. arctos*	**BR1-35**	BR1.1	N-Eur	3A			128	214	236	271	285	404	187	187	242
62	*U. arctos*	**BR1-35**	BR1.1	N-Eur	3A			128	214	236	271	285	404	187	187	242
63	*U. arctos*	**BR1-35**	BR1.1	N-Eur	3A			128	214	236	271	285	404	187	187	242
82	*U. arctos*	**BR1-35**	**BR1.2**	N-Eur	3A	x	HG423296, HG423297	128	214	236	271	285	404	187	187	242
83	*U. arctos*	**BR1-35**	**BR1.2**	N-Eur	3A	x		128	214	236	271	285	404	187	187	242
64	*U. arctos*	**BR1-36**	BR1.1	W-AS (C-Sib)	3A			128	214	236	269	285	402	185, 187	187	242
84	*U. arctos*	**BR1-36**	**BR1.2**	W-AS (C-Sib)	3A	x		128	214	236	269	285	402	185, 187	187	242
65	*U. arctos*	**BR1-4**	BR1.1	ABC	2A	x		128	214	236	265	281	398	185	197	240, 242
66	*U. arctos*	**BR1-4**	BR1.1	ABC	2A	x		128	214	236	265	281	398	185	197	240, 242
67	*U. arctos*	**BR1-5**	BR1.1	N-Eur	3A			132	216	236	265	285	400	187	193	242
85	*U. arctos*	**BR1-5**	**BR1.3**	N-Eur	3A	x	HG423294, HG423295	132	216	236	265	285	400	187	193	242
68	*U. arctos*	**BR1-6**	BR1.1	C-Eur	1			132	214	236	265	281	406	189	191	242
69	*U. arctos*	**BR1-7**	BR1.1	ABC	2A	x		128	214	236	265	281	404	185	199	240, 242
70	*U. arctos*	**BR1-7**	BR1.1	ABC	2A	x		128	214	236	265	281	404	185	199	240, 242
71	*U. arctos*	**BR1-7**	BR1.1	ABC	2A	x		128	214	236	265	281	404	185	199	240, 242
72	*U. arctos*	**BR1-7**	BR1.1	CAN	3B			128	214	236	265	281	404	185	197	240, 242
73	*U. arctos*	**BR1-7**	BR1.1	CAN	3B			128	214	236	265	281	404	185	197	240, 242
74	*U. arctos*	**BR1-7**	BR1.1	CAN	3B			128	214	236	265	281	404	185	197	240, 242
75	*U. arctos*	**BR1-7**	BR1.1	CAN	3B			128	214	236	265	281	404	185	197	240, 242
76	*U. arctos*	**BR1-7**	BR1.1	CAN	3B			128	214	236	265	281	404	185	199	240, 242
77	*U. arctos*	**BR1-7**	BR1.1	CAN	3B			128	214	236	265	281	404	185	197	240, 242
78	*U. arctos*	**BR1-7**	BR1.1	CAN	3B			128	214	236	265	281	404	185	197	240, 242
79	*U. arctos*	**BR1-8**	BR1.1	NW-A (ABC-Main)	3B	x		128	214	236	263	281	404	185	199	240, 242
80	*U. arctos*	**BR1-8**	BR1.1	NW-A (ABC-Main)	3B	x		128	214	236	263	281	404	185	199	240, 242
81	*U. arctos*	**BR1-9**	BR1.1	CAN	3B			130	214	236	263	289	404	185, 187	193	242
86	*U. arctos*	**BR2-1**	BR2	E-AS (Kam)	3A		HG423298, HG423299	126	214	236	251	289	408	185	195	242
87	*U. arctos*	**BR2-2**	**BR2**	E-AS (Kam)	3A	x		126	214	236	251	289	410	185	195	242
88	*U. arctos*	**BR3**	BR3	W-AS (Ural)	3A	x	HG423300, HG423301	128	214	236	265	277	410	181, 187	193	242
89	*U. arctos*	**BR4**	**BR4**	NW-A (AK)	3A		HG423306, HG423307	126	214	236	263	285	408	185	197	242
90	*U. arctos*	**BR5**	BR5	ABC	2A	x	HG423292, HG423293	126	214	236	263	289	406	185, 187	195	242
91	*U. maritimus*	**PO1-1**	PO1.1	AK	2B	x	HG423302, HG423303	126	210	234	267	281	402	185	191	242
92	*U. maritimus*	**PO1-1**	PO1.1	AK	2B	x		126	210	234	267	281	402	185	191	242

11

93	*U. maritimus*	**PO1-1**	PO1.1	AK	2B	x		126	210	234	267	281	402	185	191	242
94	*U. maritimus*	**PO1-1**	PO1.1	AK	2B			126	210	234	267	281	402	185	191	242
95	*U. maritimus*	**PO1-1**	PO1.1	AK	2B			126	210	234	267	281	402	185	191	242
96	*U. maritimus*	**PO1-1**	PO1.1	AK	2B			126	210	234	267	281	402	185	191	242
97	*U. maritimus*	**PO1-1**	PO1.1	AK	2B			126	210	234	267	281	402	185	191	242
98	*U. maritimus*	**PO1-1**	PO1.1	AK	2B			126	210	234	267	281	402	185	197	242
99	*U. maritimus*	**PO1-1**	PO1.1	AK	2B			126	210	234	267	281	402	185	191	242
100	*U. maritimus*	**PO1-1**	PO1.1	AK	2B			126	210	234	267	281	402	185	191	242
101	*U. maritimus*	**PO1-1**	PO1.1	W-GR (BB)	2B			126	210	234	267	281	402	185	199	242
102	*U. maritimus*	**PO1-1**	PO1.1	W-GR (BB)	2B	x		126	210	234	267	281	402	185	199	242
103	*U. maritimus*	**PO1-1**	PO1.1	W-GR (BB)	2B	x		126	210	234	267	281	402	185	199	242
104	*U. maritimus*	**PO1-1**	PO1.1	W-GR (BB)	2B			126	210	234	267	281	402	185	191	242
105	*U. maritimus*	**PO1-1**	PO1.1	DS	2B	x		126	210	234	267	281	402	185	199	242
106	*U. maritimus*	**PO1-1**	PO1.1	DS	2B			126	210	234	267	281	402	185	189	242
107	*U. maritimus*	**PO1-10**	PO1.1	DS	2B			126	210	234	263	285	402	185	199	242
108	*U. maritimus*	**PO1-11**	PO1.1	W-GR (Kane)	2B	x		126	210	234	263	281	404	183, 185	191	242
109	*U. maritimus*	**PO1-12**	PO1.1	ATL (Ice)	2B	x		126	210	234	263	281	406	185	191	242
110	*U. maritimus*	**PO1-13**	PO1.1	ATL (FJL)	2B	x		126	210	234	265	281	406	185	191	242
111	*U. maritimus*	**PO1-14**	PO1.1	W-GR (BB)	2B			126	210	234	261	281	408	185	197	242
112	*U. maritimus*	**PO1-2**	PO1.1	AK	2B			126	210	234	267	281	394	185	199	242
113	*U. maritimus*	**PO1-3**	PO1.1	DS	2B	x		126	210	234	267	285	402	185	199	242
114	*U. maritimus*	**PO1-3**	PO1.1	DS	2B			126	210	234	267	285	402	185	199	242
115	*U. maritimus*	**PO1-3**	PO1.1	DS	2B			126	210	234	267	285	402	185	199	242
116	*U. maritimus*	**PO1-4**	PO1.1	DS	2B			126	210	234	269	285	404	185	195	242
117	*U. maritimus*	**PO1-5**	PO1.1	W-GR (BB)	2B			126	210	234	269	281	404	185	197	242
118	*U. maritimus*	**PO1-5**	PO1.1	DS	2B			126	210	234	269	281	404	185	197	242
119	*U. maritimus*	**PO1-6**	PO1.1	AK	2B			126	210	234	267	281	404	185	197	242
120	*U. maritimus*	**PO1-6**	PO1.1	AK	2B			126	210	234	267	281	404	185	197	242
121	*U. maritimus*	**PO1-7**	PO1.1	AK	2B	x		126	210	234	269	281	406	185	197	242
122	*U. maritimus*	**PO1-7**	PO1.1	AK	2B			126	210	234	269	281	406	185	197	242
123	*U. maritimus*	**PO1-8**	PO1.1	AK	2B	x		126	210	234	267	281	410	185	199	242
124	*U. maritimus*	**PO1-8**	PO1.1	ATL (E-Gr)	2B	x		126	210	234	267	281	410	185	199	242
126	*U. maritimus*	**PO1-8**	**PO1.2**	ATL (E-Gr)	2B	x	HG423304, HG423305	126	210	234	267	281	410	185	199	242
125	*U. maritimus*	**PO1-9**	PO1.1	DS	2B	x		126	210	234	265	281	402	185	197	242

12

127	*U. maritimus*	**PO2-1**	**PO2**	AK	2B		HG423308, HG423309	126	210	234	263	281	402	185	201	242
128	*U. maritimus*	**PO2-2**	**PO2**	AK	2B			126	210	234	263	281	404	185	201	242
129	*U. maritimus*	**PO2-2**	**PO2**	AK	2B			126	210	234	263	281	404	185	201	242
130	*U. maritimus*	**PO2-3**	**PO2**	W-GR (BB)	2B			126	210	234	261	281	408	185	197	242
131	*U. americanus*	**BL1-1**	BL1	Montana		x	HG423286, HG423287	142	210	221	263	285	386	185	195	-
132	*U. americanus*	**BL1-2**	BL1	Oregon		x		140	210	221	263	285	386	185	197	-
133	*U. americanus*	**BL2-1**	BL2	Alaska zoo		x	HG423288, HG423289	132	210	221	259	293	388	185	189, 195	-
134	*U. americanus*	**BL2-2**	BL2	Vermont		x		132	210	221	265	297	388	183	193, 201	-
135**	*T. ornatus*		SP1	Zoo Basel		x	HG423284, HG423285									

**The spectacled bear sample was used as outgroup only for the divergence time estimation in Beast.

13

Table S2. Divergence time estimates of Y chromosome lineages in bears, obtained from different Beast runs.

Scenario	Prior [Ma]	$T_{MRCA\ (Bears)}$ [Ma]	$T_{MRCA\ (BLACK/B/P)}$ [Ma]	$T_{MRCA\ (B/P)}$ [Ma]	Reference for calibration
A	0.95 (±10%)^{BLACK}	3.75 (2.54 – 5.28)	0.94 (0.86 – 1.03)	0.77 (0.56 – 0.94)	(Hailer et al. 2012)
B	6.0 (±10%)^{SPECTACLED}	5.91 (5.4 – 6.52)	1.40 (0.88 – 1.92)	1.12 (0.73 – 1.61)	(Wayne et al. 1991)
C	10.91 (±10%)^{SPECTACLED}	10.74 (9.82 – 11.85)	2.54 (1.7 – 3.6)	2.06 (1.31 – 2.95)	(Yu et al. 2007)
D	3.0 x 10^{-9}/site/year	2.22 (1.76 – 2.73)	0.53 (0.36 – 0.72)	0.43 (0.29 – 0.61)	(Xue et al. 2009)

We explored various scenarios, using a prior on the divergence time of the American black bear lineage from brown/polar bears (^{BLACK}; scenario A), priors on the divergence time of the spectacled bear lineage (^{SPECTACLED}; scenarios B and C), or a pedigree-based mutation rate estimate (per generation) from human Y chromosomes assuming a generation time of 10 years (scenario D). Ma = million years. Fig. 2 shows scenario B.

Table S3. Sample localities and accession numbers of bear genome data, used in the scaffold-wide analyses.

Species	Locality	Number (individuals)	Accession numbers
Ursus maritimus	Spitsbergen, Svalbard	9	SRX155945, SRX155949, SRX155951, SRX155953, SRX155954, SRX155955, SRX155957, SRX155960, SRX155961
	Alaska	3	SRX156102, SRX156103, SRX156105
Ursus arctos	ABC-Islands	1	SRX156108
	Northern Europe	1	CBZK010000001 - CBZK010000005
Ursus americanus	Alaska	1	SRX156137

Table S4. Pairwise ΦST-values among brown bear populations.

ABC: ABC-islands, NW-A: North-western America, CAN: Nunavut, Canada, C-EU: Central Europe, N-EU: Northern Europe, W-AS: Western-Asia, E-AS: E-Asia (fig. 1, table 1).

	ABC	NW-A	CAN	C-EU	N-EU	W-AS	E-AS
ABC							
NW-A	*0.02**						
CAN	0.14	0.11					
C-EU	0.24	0.15	0.23				
N-EU	0.35	0.29	0.36	0.20			
W-AS	0.20	0.13	0.25	*0.04**	*0.09**		
E-AS	0.34	0.27	0.32	0.25	0.38	0.27	

* not significant (p>0.05; >1,000 permutations)

Table S5. Pairwise ΦST-values among polar bear populations.

AK: Alaska, DS: Davis Strait, W-GR: Western Greenland, ATL: Atlantic (fig. 1, table 1).

	AK	DS	W-GR	ATL
AK				
DS	0.09			
W-GR	*<-0.001**	*0.13**		
ATL	0.35	0.51	*0.17**	

* not significant (p>0.05; >1,000 permutations)

Table S6. Segregating sites of Y-chromosome haplotypes in the 3.1 kb (A) and 5.3 kb (B) datasets.

Polymorphic sites (within species variation) shown with grey background. Asterisks in (B) indicate variation that was already known from the ascertainment panel.

A	Position in alignment																				
PCR Fragment	318.7C											318.10B		579.3C		318.11C					
Haplotype	41	228	234	242	372	445	611	769	774	788	891	916	1284	1308	1963	2136	2207	2568	2686	2753	3075
BR1.1	C	A	C	C	-	T	C	G	G	G	-	G	C	T	A	A	C	A	T	C	T
BR2	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
BR3	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
BR4	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
BR5	.	.	T	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.
PO1.1	.	.	.	.	.	C	.	A	.	A	.	.	G	A	.	G	T	.	A	T	C
PO2	.	.	.	.	.	C	.	A	.	A	.	.	G	A	.	G	T	G	A	T	C
BL1	.	.	.	T	T	C	T	.	.	.	A	-	.	A	G	.	.	.	A	.	.

B	Position in alignment																																	
PCR fragment	318.2C					318.3C					318.7C												579.1B				318.10B		579.3C		318.11C			
Haplotype	17	325	333	560	630	823	998	1037	1363	1542	1661	1848	1854	1862	1992	2065	2231	2389	2394	2408	2511	2536	2653	2669	2790	3070	3500	3524	4179	4352	4423	4902	4069	5291
Identified in ascertainment panel			*					*										*		*						*	*	*		*	*	*	*	*
BR1.1	G	C	G	G	A	G	G	C	C	A	C	A	C	C	-	T	C	G	G	G	-	G	G	T	T	T	C	T	A	A	C	T	C	T
BR1.2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.
BR1.3	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
BR2	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
BR3	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
BR5	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
PO1.1	.	.	A	.	.	.	.	T	.	T	.	.	.	.	.	C	.	A	.	A	.	.	.	C	.	.	G	A	.	G	T	A	T	C
PO1.2	.	.	A	.	.	A	.	T	.	T	.	.	.	.	.	C	.	A	.	A	.	.	.	C	.	.	G	A	.	G	T	A	T	C
BL1	A	G	A	C	.	.	A	.	T	T	.	.	.	T	T	C	T	.	.	.	A	-	.	C	.	.	.	A	G	.	.	A	.	.
BL2	A	G	A	C	.	.	A	.	T	T	.	.	.	T	T	C	T	.	.	.	A	-	A	C	G	.	.	A	G	.	.	A	.	.

16

Table S7 Primer sequences, annealing temperatures and fragment sizes of Y-chromosomal sequence fragments.

Marker	Forward primer (5´-3´)	Reverse primer (5´-3´)	Fragment size (bp) including primers	T_A (°C) 10 cycles*	T_A (°C) 25 cycles
318.2C	AATGCAGATGCCACCATACC	GTTCATGCAGTTCTGTGTACTCG	1157**	66	61
318.3C	CGACCTTGACCAACAAGAGG	GAGATGGTCTCTGCAAGATGG	1216	66	61
318.7C	TCTTCGTCTTCATGCTGTGG	CCAGCTCCTTATATGCTGAACC	1095**	68	58
318.10B	TGCACAGTTCAATGGCTACAG	TCAGCAGACATTTTCTTGGAAC	529	66	61
318.11C	GATGATGCATAAGCAATCCTTG	TGCAACCATAACTTGTTTACTTCC	1012	69	64
579.1B	CTGCAGGCCTGTCAATGTTA	TGTGTATCGACCCCATACTTTG	660	66	61
579.3C	TTAACTGCTCTGACCTTCATCG	GTGCACAGGCAAGTGTTAGG	1157	68	58

*decreasing by 0.5°C/1°C in order to reach the second annealing temperature after ten cycles

**these two fragments overlap by 142 bp

17

Table S8. Primer sequences, annealing temperature, and fragment size of small, overlapping Y-chromosomal sequence markers used to amplify DNA from lower-quality samples.

Marker	Forward primer (5´-3´)	Reverse primer (5´-3´)	Fragment size (bp) including primers	T_A (°C) 10 cycles*	T_A (°C) 25 cycles*
318.3C_1	CGACCTTGACCAACAAGAGG	ACCACCTACAGTCTGGGTTTG	372	68	61
318.3C_3	TGTCTTATTTCAAAAGCAAAATGG	CTGTGTTTGAGTGCCATGAAT	365	68	61
318.3C_4	CAGGTTTCCTACAAGTCTCAGTG	GAGATGGTCTCTGCAAGATGG	387	68	61
318.7C_1	TCTTCGTCTTCATGCTGTGG	TTCAGTTCATGCAGTTCTGTG	368	68	61
318.7C_2	AATGAAACTGGGAACACACTTC	CACAGTGGTTCACTGACATGG	391	68	61
318.7C_3	GACTTGAAGACACTTAAGTAGCATTTG	AAGCATCTATACTTCATGTAGCTTGTG	341	68	58
318.7C_4	TCACTCACGTAAGCCACTCTC	CCAGCTCCTTATATGCTGAACC	385	68	58
318.10B_1	TGCACAGTTCAATGGCTACAG	CATGTTTACAAAGGGATCAGCA	334	72	65
318.10B_2	TTCAACTGACCCTTGAAACACT	TCAGCAGACATTTTCTTGGAAC	345	72	65
318.11C_1	TGCATAAGCAATCCTTGTAATATACC	TTGTTAAAGTCTCCTTTCCTTGC	378	68	61
318.11C_2	GGTTACCAGGGAACTGATGG	CACAAATTCTGTGGTTTGTATAAGG	367	69	62
318.11C_3	AGACTGCACCCAAGATCTTACA	AAGATTCTTCAGTCCTTAAACACTAGC	408	69	62
579.1B_1	CTGCAGGCCTGTCAATGTTA	TGTTTACATGAAATATTCAGAGGAGA	335	68	61
579.1B_2	GAACCACTAATGTAGTTCCTGTCTCT	TGTGTATCGACCCCATACTTTG	386	68	61
579.3C_1	TTAACTGCTCTGACCTTCATCG	GCAGATATCCATTACAAAAAGCAA	395	68	61
579.3C_2	AGGGCTTTTTGGCTTTTTGT	TCCTTTACTAATGCCATCCTCTG	397	69	62
579.3C_3	CCTCACTGGGATGTTGTGAG	AATTAAAATATTAGAACCCTCTGTTGA	351	65	58
579.3C_4	CAGAGGATGGCATTAGTAAAGGA	GTGCACAGGCAAGTGTTAGG	465	70	63

* Amplification followed a touchdown PCR procedure, with 10 initial cycles at a higher (more specific) annealing temperature, followed by 25 cycles at a lower annealing temperature.

18

Table S9. Primer sequences, fluorescent dyes, and repeat unit of Y-chromosomal microsatellite markers.
The allele size range includes alleles from brown, polar, and American black bears.

Marker	Fluorescent Dye	Forward primer (5´-3´)	Reverse primer (5´-3´)	Concentration in multiplex PCR A or B (µM)	Repeat Unit	Allele size range (bp)
Y318.1	PET	GGGATCAAGCCCCACATCAA	ACTTGTAGATGCACATCTGTGGT	0.6 (B)	AAAT	277-297
Y318.2	PET	CAGGCTGACACTGGGGATTT	AAGAGGGAGTCATCTGGGGT	0.7 (A)	TA	221-236
Y318.4	6FAM	TACCTGGCTGGCTTTCTTGG	CACTGTTGGTTTTGGCTCCG	0.05 (A)	GA	210-216
Y318.6	PET	GCTGGCTGTCTCTCTCTCTGA	AAATTCCTTTGGAAACGTCCT	0.6 (A)	TG	386-412
Y318.9	VIC	CACTCAGGCACCCCTCTATC	TGGCCAGGATACAGAAACAAC	0.05 (B)	AC	126-142
Y369.1	NED	TCCCTGAATGAGCAGTAGCC	GGGGTATTGCGTTGCATTGG	0.1 (A)	GT	249-273
Y369.4	VIC	AGGCATCCATTCTATCACCAC	TGTGGATGTATCTGCCCAAC	0.1 (A)	AC	187-201
Y69217.1	VIC	CTCCACCTTGTCTGCCACTC	TTCCCCTCCCTTTCTGTCCT	0.08 (B)	TG	242
Y15020.1	PET	TGCAATTTCTCTCAAACAACTTCCT	GCGATGAAGGTCAGAGCAGT	0.25 (B)	TG	183-185

19

References

Bandelt HJ, Forster P, Röhl A. 1999. Median-joining networks for inferring intraspecific phylogenies. Mol. Biol. Evol. 16:37–48.
Brown SK, Pedersen NC, Jafarishorijeh S, Bannasch DL, Ahrens KD, Wu J-T, Okon M, Sacks BN. 2011. Phylogenetic distinctiveness of Middle Eastern and Southeast Asian village dog Y chromosomes illuminates dog origins. PLoS One 6:e28496.
Darriba D, Taboada GL, Doallo R, Posada D. 2012. jModelTest 2: more models, new heuristics and parallel computing. Nat. Methods 9:772.
Davison J, Ho SYW, Bray SC, et al. 2011. Late-Quaternary biogeographic scenarios for the brown bear (*Ursus arctos*), a wild mammal model species. Quat. Sci. Rev. 30:418–430.
Drummond AJ, Suchard MA, Xie D, Rambaut A. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol. Biol. Evol. 29:1969–1973.
Ellegren H. 2004. Microsatellites: simple sequences with complex evolution. Nat. Rev. Genet. 5:435–445.
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.
Forster P, Röhl A, Lünnemann P, Brinkmann C, Zerjal T, Tyler-Smith C, Brinkmann B. 2000. A short tandem repeat-based phylogeny for the human Y chromosome. Am. J. Hum. Genet. 67:182–196.
Hailer F, Kutschera VE, Hallstrom 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.
Hailer F, Leonard J. 2008. Hybridization among three native North American Canis species in a region of natural sympatry. PLoS One 3:e3333.
Rambaut A, Drummond AJ. 2007. Tracer v1.4, Available from http://beast.bio.ed.ac.uk/Tracer.
Roca AL, Georgiadis N, O'Brien SJ. 2005. Cytonuclear genomic dissociation in African elephant species. Nat. Genet. 37:96–100.
Sacks BN, Brown SK, Stephens D, Pedersen NC, Wu J-T, Berry O. 2013. Y Chromosome Analysis of Dingoes and Southeast Asian Village Dogs Suggests a Neolithic Continental Expansion from Southeast Asia Followed by Multiple Austronesian Dispersals. Mol. Biol. Evol. 30:1103–1118.
Stone KD, Cook JA. 2000. Phylogeography of black bears (*Ursus americanus*) of the Pacific Northwest. Can. J. Zool. 78:1218–1223.
Wayne RK, Van Valkenburgh B, O'Brien SJ. 1991. Molecular distance and divergence time in carnivores and primates. Mol. Biol. Evol. 8:297–319.
Weir BS, Cockerham CC. 1984. Estimating F-Statistics for the Analysis of Population Structure. Evolution. 38:1358–1370.
Wooding S, Ward R. 1997. Phylogeography and Pleistocene Evolution in the North American Black Bear. Mol. Biol. Evol. 14:1096–1105.
Xue Y, Wang Q, Long Q, Ng BL, Swerdlow H, Burton J, Skuce C, Taylor R, Abdellah Z, Zhao Y, et al. 2009. Human Y chromosome base-substitution mutation rate measured by direct sequencing in a deep-rooting pedigree. Curr. Biol. 19:1453–1457.
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.

Publication 3: 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

Published in *Molecular Biology and Evolution* (2014) 31(8):2004-2017

Erklärung über Anteile der Autoren/Autorinnen an den einzelnen Kapiteln der Promotionsarbeit

Titel der Publikation/des Manuskripts: Bears in a Forest of Gene Trees: Phylogenetic Inference Is Complicated by Incomplete Lineage Sorting and Gene Flow

	Was hat der/die Promovierende bzw. Was haben die Co-Autoren/Autorinnen biegetragen#	Name des/der jeweiligen Autors/Autoren/ Autorin*
(1) Entwicklung und Planung	20% 40% 40%	TB VEK FH, AJ
(2) Durchführung der einzelnen Untersuchungen/ Experimente	Probenorganisation und -verwaltung DNA Extraktion, Primerdesign, PCRs, Sequenzierung (Y) DNA Extraktion, Primerdesign, PCRs, Sequenzierung (Introns)	TB, VEK, AJ, SRF, FH TB VEK, JLR
(3) Erstellung der Daten-sammlung und Abbildungen	Sequenzeditierung, Alignments (Y) Sequenzeditierung, Alignments (Introns) Alignments mtDNA Onlinearchivierung der Sequenzen	TB VEK, JLR VEK TB, VEK
(4) Analyse/Interpretation der Daten	Statistische Analyse Y-Chromosom Statistische Analysen Intron und mtDNA Statistische Analysen Intron+Y Zusammenfassende Interpretation	TB VEK, FH TB, VEK TB, VEK, FH, AJ
(5) übergeordnete Einleitung/ Ergebnisse/Diskussion	20% 60% 20%	TB VEK FH, AJ

#Bei 2, 3 und 4 bitte kurze inhaltliche Angaben der jeweiligen Anteile, bei 1 und 5 reichen prozentuale Angaben
*Mehrfacheintragungen möglich

Als Autoren/Autorinnen werden solche Personen bezeichnet, die an der Arbeit in Bezug auf die genannten Punkte in einer Weise mitgewirkt haben, dass sie für die ausgewiesenen Passagen (mit) verantwortlich sind. Personen, die an der Arbeit mitgewirkt haben, jedoch nicht in diese Kategorie fallen, sollten in der Danksagung Erwähnung finden.

Datum/Ort

Datum
zustimmende Bestätigung der vorgenannten Angaben

Unterschrift Promovend/Promovendin

Unterschrift Betreuer/Betreuerin

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.**

Article

Fast Track

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

© The Author 2014. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/3.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Open Access

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

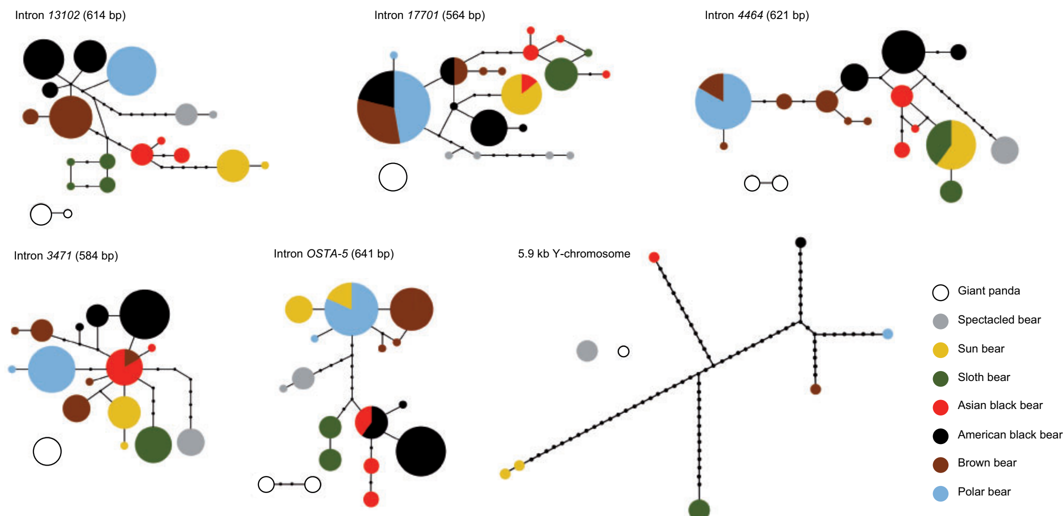


Fig. 1. Statistical parsimony networks for five autosomal intron markers and 5.9 kb of Y-chromosomal sequence in bears. Circle areas are proportional to haplotype frequencies and inferred intermediate states are shown as black dots. For some loci, spectacled bear and giant panda haplotypes were too divergent 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

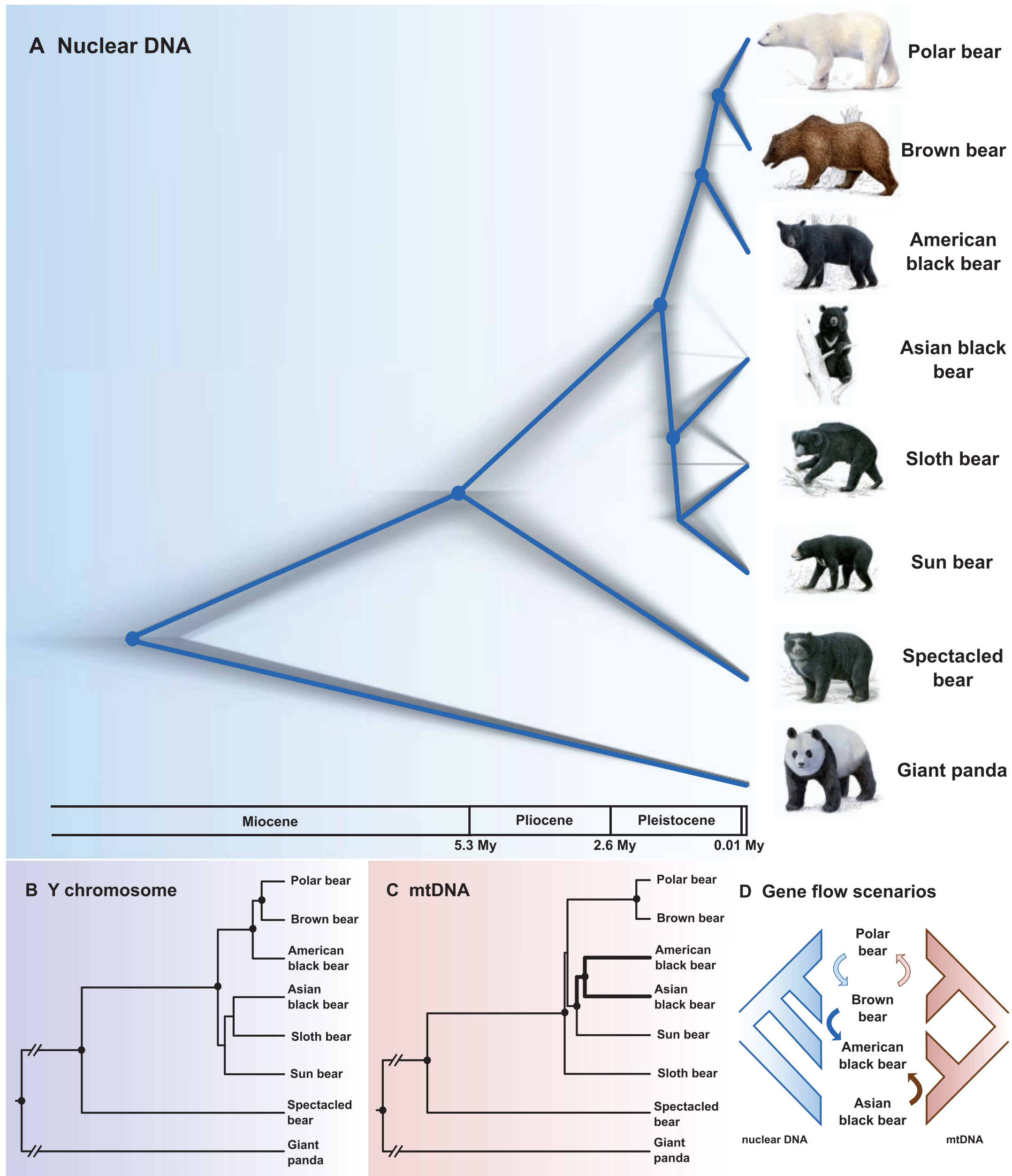


Fig. 2. (*A*) Cloudogram of species trees from *BEAST analysis, based on 14 autosomal introns and 5.9 kb of Y-chromosomal sequence (90,000 species trees). The consensus tree of the most frequently occurring topology in the posterior distribution is superimposed onto the cloudogram in blue. Blue dots 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. S3B, 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 $\log_{10}(\mathrm{BF}) > 2$ (or $\mathrm{BF} > 100$) as “decisive” (Kass and Raftery 1995), the nuclear species tree topology was favored over the mitochondrial gene tree topology with high statistical support ($\log_{10}[\mathrm{BF}] = 4.2$, or $\mathrm{BF} = 15{,}811$).

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

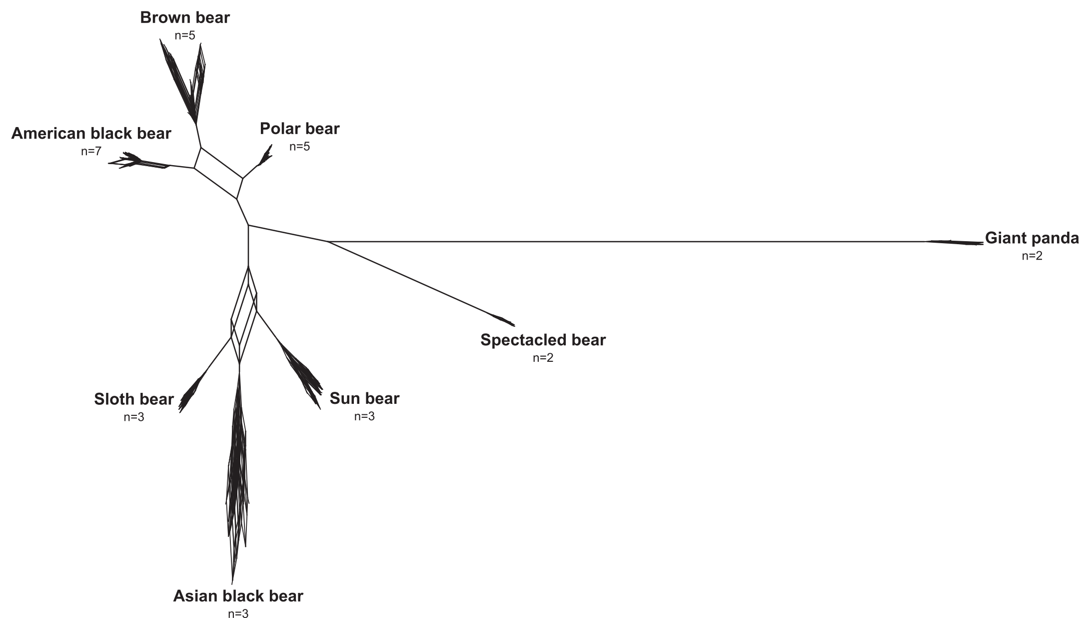


Fig. 3. Consensus network of 14 autosomal gene trees obtained from a *BEAST analysis of 14 nuclear introns. All splits found in at least two gene trees (2/14, threshold = 0.14) are shown. *n*, number of individuals analyzed per species.

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	MO, 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

8

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	---

9

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.

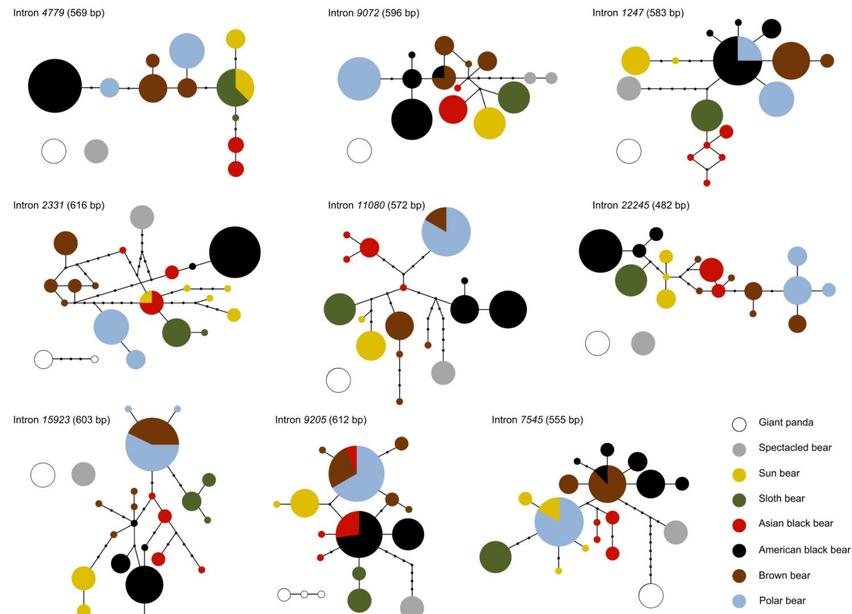
[d] Touchdown PCR, during which the annealing temperature was lowered by 0.5°C in each of 14 cycles, followed by 26 normal cycles.

[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.

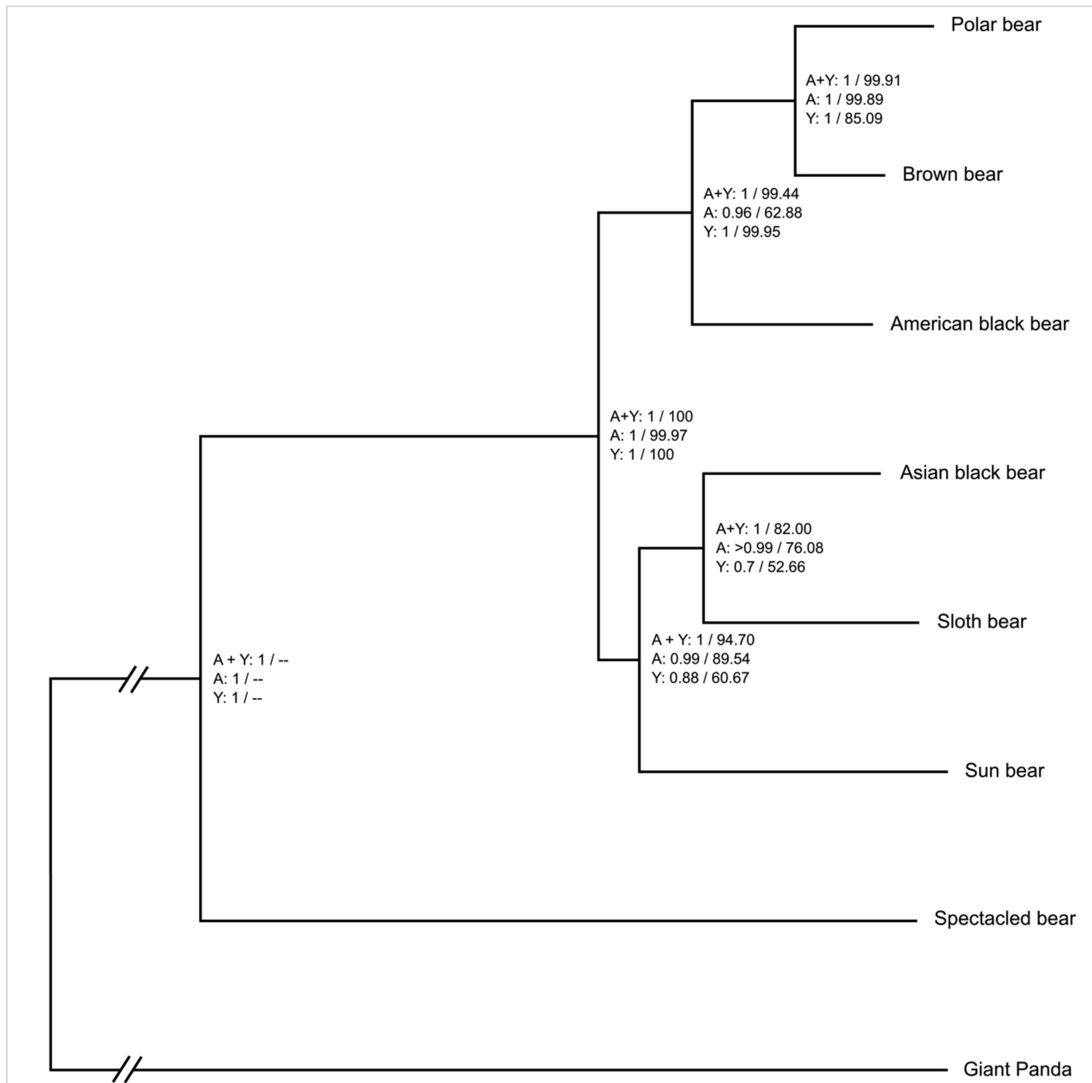
[g] This marker includes exon 4 of *Usp9Y*.

Supplementary figures

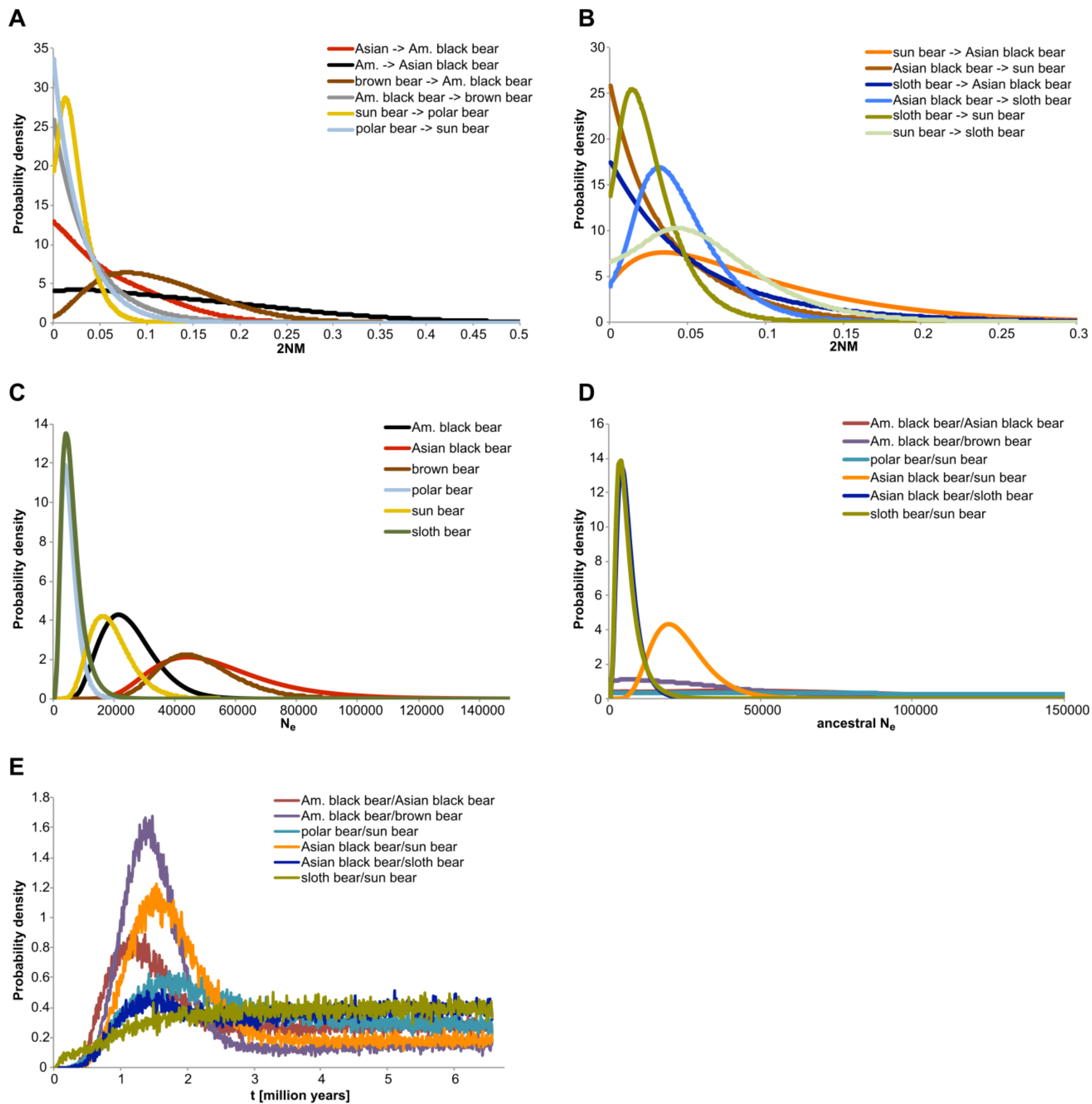


Supplementary figure S1: Statistical parsimony networks for nine autosomal intron markers in bears. Circle areas are proportional to haplotype frequencies and inferred intermediate states are shown as black dots. For some loci, spectacled bear and giant panda haplotypes were too divergent to be connected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

14



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.

APPENDIX

GENOME-WIDE SEARCH IDENTIFIES 1.9 MEGABASES FROM THE POLAR BEAR Y CHROMOSOME FOR EVOLUTIONARY ANALYSES

Tobias Bidon*, Nancy Schreck*, Frank Hailer, Maria Nilsson, Axel Janke

*Shared first-authorship

Submitted to *Genome Biology and Evolution* on 10/02/2015. A revised version was submitted on 30/03/2015.

Genome-wide search identifies 1.9 megabases from the polar bear Y chromosome for evolutionary analyses

Tobias Bidon[1,#], Nancy Schreck[1,#], Frank Hailer[1,2] , Maria Nilsson[1] and Axel Janke[1, 3,*]

[1] Senckenberg Biodiversity and Climate Research Centre Frankfurt, 60325 Frankfurt am Main, Germany
[2] School of Biosciences, Cardiff University, Cardiff CF10 3AX, Wales, UK
[3] Goethe University Frankfurt, Institute for Ecology, Evolution & Diversity, 60439 Frankfurt am Main, Germany

# These authors have contributed equally

* **Author for Correspondence:** Axel Janke, Institute for Ecology, Evolution & Diversity, Goethe University Frankfurt, Frankfurt am Main, Germany. +49 69 7542 1842, ajanke@senckenberg.de

ABSTRACT

The male-inherited Y chromosome is the major haploid fraction of the mammalian genome, rendering Y-linked sequences an indispensable resource for evolutionary research. However, despite recent large-scale genome sequencing approaches, only a handful of Y chromosome sequences have been characterized to date, mainly in model organisms. Using polar bear (*Ursus maritimus*) genomes, we compare two different *in-silico* approaches to identify Y-linked sequences: a) similarity to known Y-linked genes and b) difference in the average read depth of autosomal versus sex chromosomal scaffolds. Specifically, we mapped available genomic sequencing short reads from a male and a female polar bear against the reference genome and identify 112 Y-chromosomal scaffolds with a combined length of 1.9 megabases (Mb). We verified the *in-silico* findings for the longer polar bear scaffolds by male specific *in-vitro* amplification, demonstrating the reliability of the average read depth approach. The obtained Y chromosome sequences contain protein coding sequences, single nucleotide polymorphisms, microsatellites, and transposable elements that are useful for evolutionary studies. A high-resolution phylogeny of the polar bear patriline shows two highly divergent Y chromosome lineages, obtained from analysis of the identified Y scaffolds in twelve previously published male polar bear genomes. Moreover, we find evidence of gene conversion among *ZFX* and *ZFY* sequences in the giant panda lineage and in the ancestor of ursine and tremarctine bears. Thus, the identification of Y-linked scaffold sequences from unordered genome sequences yield valuable data to infer phylogenomic and population-genomic patterns in bears.

Keywords: sex chromosome, patriline, male inheritance, Ursidae, Y chromosome, divergence

INTRODUCTION

Genomic sequence data have become an important resource for evolutionary biology, and new sequenced genomes are becoming available at increasing speed. The mammalian genome consists of autosomes, sex chromosomes and mitochondrial (mt) DNA, which are differentially inherited. These parts of the genome can thus provide information about distinctive aspects of a species' evolutionary history (Chesser & Baker 1996; Veeramah & Hammer 2014).

For technical reasons, the maternally inherited mtDNA has been a standard tool to study evolutionary processes in model and non-model organisms (Wilson et al. 1985). Consequently, the first available genomic resources for evolutionary studies were fully sequenced mt genomes (Anderson et al. 1981; Janke et al. 1994). The paternally inherited counterpart of mtDNA is the male-specific Y chromosome, one of two sex chromosomes in the mammalian genome. Similar to mtDNA, the Y chromosome is haploid, lacks interchromosomal recombination for most of its length, and is uniparentally inherited. These properties allow the inference of long and high-resolution haplotypes, enabling researchers to trace the evolutionary history of male lineages over time (Jobling & Tyler-Smith 2003; Wei et al. 2013). Biparentally inherited autosomes provide the largest amount of sequence data, but their phylogenetic analysis can be complicated by reticulate evolution (Posada et al. 2002).

Polar bears have recently been in the focus of genome scale evolutionary analyses and genomic sequences have been used to address the evolution, population history, and unique adaptations of this high arctic mammal (Hailer et al. 2012; Miller et al. 2012; Cahill et al. 2013; 2014; Cronin et al. 2014; Liu et al. 2014). Furthermore, short Y-chromosomal sequences and six Y-linked microsatellites in polar and brown bears were used to investigate the distribution of male-specific genetic variation across their ranges (Bidon et al. 2014). The application of Y-chromosomal markers is particularly interesting in many mammals, because sex-specific differences in dispersal behavior are predicted to affect phylogeographic and population genetic conclusions that have so far been drawn almost exclusively from female-inherited mtDNA.

Y-chromosomal sequences are important in studies of evolutionary history, chromosome structure, and forensic applications (Jobling & Tyler-Smith 2003; Kayser 2007; Hallast et al. 2013). The Y chromosome´s unique evolutionary viewpoint has been used to investigate patterns of domestication and migration, e.g. in horses and dogs (Sacks et al. 2013; Wallner et al. 2013), and to study human phylogeography and migration (Wei et al. 2013; Scozzari et al. 2014; van Oven et al. 2014).

Despite the wealth of genomic data, identification of large amounts of Y-chromosomal sequences from high throughput sequencing data is rarely done. Genome sequences are usually ordered into scaffolds, without information about their relative orientation or chromosomal origin, because thoroughly annotated reference genomes and physical maps are still lacking for most taxonomic groups. In addition, many mammalian genomes have been sequenced from female individuals, to obtain equal coverage of autosomes and the X chromosome (Hughes & Rozen 2012), but also for technical difficulties relating to the assembly and the high amount of repetitive and ampliconic sequences on Y chromosomes (Bachtrog 2013; Willard 2003). This has hampered sequencing, assembly, identification and application of Y-chromosomal markers (Greminger et al. 2010). As a consequence, complete Y chromosome sequences are only published for four mammalian species: human,

chimpanzee, rhesus macaque, and mouse (Skaletsky et al. 2003; Hughes et al. 2010, 2012; Soh et al. 2014). This list is complemented by large-scale analyses of Y-chromosomal sequences for dog, cat, marmoset, rat, bull, opossum, *Drosophila* and medaka fish (Kondo et al. 2006; Carvalho et al. 2009; Li et al. 2013; Bellott et al. 2014).

In this study, we utilize a previously published polar bear reference assembly that is based on a male sequenced at high coverage (Li et al. 2011), and available short sequence reads from additional male and female polar bears (Miller et al. 2012). We identify Y-linked scaffolds by a) searching for sequence similarity using known mammalian Y-linked gene sequences, and b) identifying scaffolds with sex-specific sequencing coverage characteristics indicative of Y linkage. The latter approach makes use of differences in the expected sequence coverage of male versus female sequence reads on autosomal, X-chromosomal, and Y-chromosomal reference scaffolds. We apply stringent quality filters to minimize false positives, i.e., scaffolds wrongly identified as Y-linked. In addition, *in-vitro* amplification of the longest candidate scaffolds confirmed the *in-silico* findings. We demonstrate that genome scale Y-chromosomal sequences can be reliably identified from high-throughput sequencing data, also in organisms lacking a chromosome-based physical map of the genome.

METHODS

We used two different approaches to identify Y-chromosomal sequences in the recently published polar bear genome assembly from a male individual that was sequenced at 101-fold coverage (Li et al. 2011). This assembly has a size of 2.3 gigabases (Gb) and is arranged into 72,214 scaffolds with an N50 value of 15.9 megabases (Mb). Information regarding chromosomal locations and the relative orientation of the scaffolds is not available. Thus, it is unknown which of the scaffolds are of Y-chromosomal origin. In the following, we refer to this genome assembly as the “polar bear assembly”, and to the scaffolds of this assembly by their respective scaffold ID numbers.

Similarity search of Y-linked genes lists candidate scaffolds

The first approach was to use 32 genes known to be Y-linked in other mammals as queries for a similarity search in the polar bear assembly. Exon sequences from human (*Homo sapiens*), mouse (*Mus musuculus*), chimpanzee (*Pan troglodytes*), and dog (*Canis lupus familiaris*) were downloaded from Genbank for these genes (Table S1). Similarity between exon sequences and the scaffolds of the polar bear assembly was identified using BLAST, analyzing one exon at a time.

Scaffolds from the polar bear assembly were extracted from the list of BLAST hits according to the following criteria: (1) the scaffold with the lowest E-value (expect value) for a particular exon relative to all other scaffolds in the list, (2) scaffolds with ≥95% sequence similarity compared to the scaffold with the lowest E-value, with the additional constraint that the difference in alignment length of exon and scaffold (compared to the scaffold with the lowest E-value) must not exceed 5%. We then obtained the exact position of each exon on its respecitve scaffold by realigning exon and scaffold using CLUSTAL W. Only scaffolds with a sequence identity of ≥ 80% between scaffold and exon were kept (Table 1).

In addition, *in-vitro* validated male-specific polar bear sequences from five known Y-linked genes (exons and introns, *AMELY*, *KDM5D* (*SMCY*), *SRY*, *UBA1Y*, *ZFY*, Table S1) (Pagès et al. 2008, 2009; Nakagome et al. 2008) were downloaded and used as a query against the polar bear assembly using BLAT, with default parameters. Polar bear sequences from the two X-linked genes *ZFX* and *AMELX* (Pagès et al. 2009) were used to differentiate between Y and X gametologs, i.e. homologous gene copies on the X and the Y chromosome (Table S1).

<u>Average depth (AD) ratio for identification of Y-linked, X- linked and autosomal scaffolds</u>

In a second approach to identify Y-linked scaffolds, we utilized previously published short sequence reads from whole-genome sequence data of one female (SRX155950/PB06) and one male (SRX155954/PB10) polar bear. The two polar bear individuals had been sequenced at similar sequence depth (~12X) on an Illumina HiSeq 2000 platform, generating paired-end reads (101 bp) with an insert size of about 400 bp (Miller et al. 2012). The AD ratio approach is based on differences in the relative numbers of X and Y chromosomes between females (2-0) and males (1-1), while both sexes carry two copies of each autosome. As unique Y-chromosomal sequences are not present in a female genome, reads obtained from genome sequencing of female and male individuals should map with characteristic sex-specific patterns to scaffolds from the Y chromosome, the X chromosome and the autosomes. The expected differences in sequencing coverage were utilized primarily to identify Y-chromosomal scaffolds in the polar bear assembly, but our approach also allowed the assignation of anonymous scaffolds from the polar bear assembly as autosomal or X-linked.

Short read sequences were evaluated for residual adapter sequences and low-quality bases were clipped off the read-ends using FastQC v 0.10.0 (Andrews 2010) and sickle (Joshi & Fass 2011). BWA (Li & Durbin 2009) was used for the reference guided mapping of the cleaned reads against the polar bear assembly. Using Samtools (Li et al. 2009), we merged read data from separate sequencing runs of the same sample into one single BAM file per individual. Picard (http://picard.sourceforge.net/) was used to mark duplicated reads, and realignment of reads was performed in GATK v2.3 (McKenna et al. 2010).

After mapping, the mpileup modul of samtools was used to calculate the read depth at each position on a given scaffold for the male and the female genome. Scaffolds without mapped reads or with low mapping quality (n=614), and scaffolds that were <1 kb (n=68,017; ~15 Mb) were disregarded and not considered in the downstream analyses. For the remaining scaffolds (≥1 kb, n=3,583), the average read depth was calculated: we determined the sum of the depth values at ambiguity-free scaffold positions (no “N”) with ≤50 reads per position, and divided this by the number of ambiguity free scaffold positions.

Finally, the AD-ratio of each scaffold was calculated by dividing the average read depth in the female individual by the average read depth in the male individual (1). A normalization factor adjusted the number of female and male reads to each other (2): to this end we divided the total number of reads (quality ≥ 20) in the female BAM file by that of the male BAM file.

For each given scaffold, average sequencing depth for the female and male genome was calculated using the following formulas:

$$\textbf{AD-ratio} = \text{average-depth}_{female} / (\text{average-depth}_{male} * \text{norm}) \quad (1)$$

$$\textbf{norm} = \text{total number of reads}_{female} / \text{total number of reads}_{male} \quad (2)$$

The normalization factor is used to enable comparison of read depth of individual scaffolds among individuals, despite possible differences in genome-wide sequencing coverage between them. Using this normalization factor, the male and female genomes are standardized to the same genome-wide average coverage. The AD-ratio is zero for perfectly mapped Y chromosome scaffolds, one for autosomal and two for X-linked scaffolds. For graphical representation we combined scaffolds with different AD-ratios into bins of size 0.02.

In-vitro validation of putative Y-linked scaffolds in different bear species

To verify the male-specificity of scaffolds identified by the in-silico analysis, we PCR amplified fragments from 20 Y-scaffolds (Table 2) in at least one male and one female individual of each of three closely related ursine bears: polar bear, brown bear, and American black bear (U. americanus). In addition, amplification of fragments from two X-linked and two autosomal scaffolds as identified by the AD-ratio approach were PCR amplified in both male and female bears to verify their non-Y-chromosomal origin. Before amplification, newly designed primers (Table S7) were tested *in-silico* for unique binding by aligning the forward and reverse sequences against the scaffolds of the polar bear assembly using blastn. Scaffolds were defined as being Y-chromosome specific when one clear amplification product was detected in males, but no amplicons or only low-intensity bands of different sizes were observed in females. *In-vitro* experiments included touchdown PCRs (see supplementary material) and agarose gel-electrophoresis to verify the expected size of the amplicons. Each PCR setup contained a no-template control.

Repetitive element estimation in the polar bear genome and on Y-linked scaffolds

The amount of transposable elements (TE) on 14 of the larger validated scaffolds (scaffold IDs: 297, 309, 318, 322, 369, 393, 389, 403, 420, 519, 579, 605, 646, 657; 1.6 Mb) was identified using RepeatMasker (http://www.repeatmasker.org) using the carnivore library (Smit et al. 1996). RepeatMasker with the carnivore library was also used to identify microsatellites with a minimum of 15 repeat units (Table S6).

Analysis of X-Y gene conversion in bears

The partial ZFY and ZFX exon sequences of all ursid species from Pagès et al. (2009) were downloaded from Genbank and aligned with homologous sequences from other mammals (397 bp) in Geneious 8.0.3 (Biomatters, Auckland, New Zealand). Based on the model suggested by jModeltest2 (Darriba et al. 2012), HKY+4G, phylogenetic trees were constructed in Geneious, and a statistical parsimony network was generated in TCS (Clement et al. 2000).

Calculation of polar bear patrilineal phylogeny

Five Y-linked scaffolds (IDs: 309, 322, 389, 393, 403) with a combined size of 743 kb were used to reconstruct the phylogenetic relationship of 12 polar bear individuals sampled in Svalbard (Norway) and Alaska. The Y-linked sequences were used to estimate the divergence time of the lineages within polar bears, using one American black and one brown bear as outgroup. Short reads of all 14 bear individuals (Miller et al. 2012) were retrieved

from databases (Table S8) and mapped to the polar bear assembly as described above. The individuals have been labeled according to their respective description in the short read archive (Table S8). The five respective scaffolds together with previously mapped short reads were extracted using Samtools and loaded into Geneious 8.0.3 (Biomatters, Auckland, New Zealand). Geneious was then used to create a consensus sequence for each individual, to align those consensus sequences, and to remove alignment columns containing ambiguous sites and gaps, respectively. Additionally, the alignments were manually inspected to find and remove columns where only one individual contained multiple differentiating sites adjacent to each other. This strict filtering reduced the size of the alignments by ~30% (see below).

A NeighborNet network was calculated in SplitsTree 4.12.6 (Huson & Bryant 2006) based on a 511 kb-long alignment of the concatenated Y-sequences of 12 polar bears. BEAST 2.1.3 (Bouckaert et al. 2014) was used to estimate divergence times among polar bears, using a strict clock model, a Yule tree model, and a uniform prior of 343-479 kya, based on the relatively young population divergence between brown and polar bears (Liu et al. 2014). An additional calibration scenario employed a fixed mutation rate obtained from human Y chromosomes: $0.76x10^{-9}$/site/year (Fu et al. 2014). We used the GTR+I substitution model as indicated by the Bayesian Information Criterion in jModeltest 2.1.1. Convergence was checked in Tracer (ESS>200). The concatenated alignment comprising 506 kb included 12 polar bears, one brown bear, and one black bear. This alignment was thus slightly shorter than the polar bear alignment, due to ambiguous sites and gaps introduced by the inclusion of additional individuals/taxa.

RESULTS

We identified a total of 1.9 Mb of Y-chromosomal sequence data in the polar bear assembly, located on 112 different scaffolds. The scaffolds were identified by applying two different approaches: a) the search for similarity of known Y-linked genes, and b) comparison of the AD-ratio of reads from male and female genomes.

The similarity search identified 23 putative Y-chromosomal scaffolds

The first approach identified scaffolds in the polar bear assembly that showed similarity to known Y-linked gene sequences from four different mammals (human, mouse, chimpanzee, dog). Exons from 18 of 32 Y-linked candidate genes that were blasted against the polar bear assembly identified polar bear scaffold sequences above a threshold of 80% identity (Table 1, Table S1). The hits were distributed across 23 scaffolds, ranging from 0.7-26,707 kb in size (Table 1, Figure 1). The full sequence length of scaffold *3836* (1,069 bp) had an identical sequence stretch on scaffold *318* (237 kb), with no nucleotide mismatches. Thus, we do not report scaffold *3836* as a distinct scaffold, although it is a separate entry in the current polar bear assembly.

Six sequences of five Y-linked genes from polar bear Y chromosomes (Nakagome et al. 2008; Pagès et al. 2008, 2009) aligned to the polar bear assembly with 98.9%-100% identity. A 227 bp fragment from *ZFY*, and a 49 bp fragment from *KDM5D* (*SMCY*) were uncharacterized (“N”) in the polar bear assembly.

We found that 10 query genes had similarity to two or more different scaffolds in the polar bear assembly, thereby creating combinations of scaffolds with stretches of homologous sequence (Table S4). These scaffold combinations consisted of one *in-vitro* validated Y-scaffold and one (or two) scaffold(s) with an AD-ratio expected for X-chromosomal linkage, indicating sequential homology between the Y-chromosomal and other scaffolds. For instance, the *ZFY* exon sequences mapped to both scaffold *318* and scaffold *20* with similar identity (99.2% vs. 99.5%). The gametologous polar bear *ZFX* sequence also mapped to both these scaffolds, at the same location as *ZFY*. However, when using less conserved intronic sequences from polar bears (Table S1) in a BLAT search against the polar bear assembly, scaffolds *318* (containing *ZFY*, Y-linked) and scaffold *20* (containing *ZFX*, X-linked) were clearly diagnosable.

Phylogenetic analyses of *ZFX/ZFY* sequences in mammals showed that the X- and Y-linked copies of giant panda (*Ailuropoda melanoleuca*) form a cluster, and that all ursine and tremarctine *ZFX/ZFY* sequences form a second cluster of closely related sequences (Figure S1). Ursid sequences thus clustered together, regardless of their X- or Y-chromosomal origin. Other mammals clustered outside the ursid variation.

Based on the similarity of known Y-linked candidate gene sequences from different mammals, the similarity search provided us with a list of 23 scaffolds that might potentially be located on the polar bear Y chromosome. However, 12 of these scaffolds were identified to be autosomal or X-linked, due to their respective AD-ratios (see below, Figure 1, Table S4).

The average-depth ratio identified 112 Y-chromosomal scaffolds

Most scaffolds had an AD-ratio of either ~1 or ~2, indicative of autosomal and X-chromosomal scaffolds, respectively (Figure 2, Figure S2). The combined sequence length of all putative autosomal scaffolds ≥1kb (0.7 < AD-ratio < 1.3; n=2,618) was ~2.18 Gb, and putative X-linked scaffolds (1.7 < AD-ratio < 2.3; n=214) amounted to ~109 Mb. At an AD-ratio of zero, which is the expected AD-ratio for Y-linked scaffolds, we detected 90 scaffolds with a combined sequence length of 686 kb (Figures 2 and S2, Tables 2 and S2). An additional 22 scaffolds with a combined sequence length of 1.21 Mb showed AD-ratios ≤0.3, of which eleven were amplified *in-vitro*, all showing male-specific amplification (Table 2). Thus, applying a relaxed AD-ratio cut-off of ≤0.3, thereby allowing for a certain proportion of wrongly mapped reads, identified 112 Y-linked scaffolds, comprising 1.9 Mb of Y-chromosomal sequence.

Nine scaffolds totaling 1.24 Mb were identified by both approaches (Figure 1, Table S3). Among the scaffolds obtained exclusively from the similarity search, one had an AD-ratio of exactly zero, but it was <1 kb (*scaffold ID 6612*; 794 bp). Four putative Y-chromosomal scaffolds from the similarity search had an AD-ratio of ~1, indicating autosomal origin. Eight scaffolds from the similarity search had an AD-ratio of ~2, indicating X-linked origin. For one scaffold identified by the similarity search, neither male nor female reads mapped with sufficient quality (scaffold ID *4889*), precluding any linkage classification.

In-vitro amplification validates all tested Y-linked scaffolds as being male-specific

The male-specificity of the longest putative Y-linked scaffolds (n=20) was additionally evaluated *in-vitro* by PCR amplification (Table 2). At least one fragment of 635-800 bp sequence length of each of the scaffolds was PCR amplified using male DNA samples along with female DNA controls of each brown, polar and black bears. All 20 scaffolds showed

male-specific PCR amplification, defined as the occurrence of a clear amplicon of a distinct size in males but not in females. In female DNA samples, the Y-chromosomal fragments could either not be amplified (scaffold IDs *309, 318, 322, 369, 389, 393, 420, 579, 596, 605, 613, 632, 813*) or the observed amplicons were smaller, with multiple low-intensity (unspecific) bands/smears on agarose gels (scaffold IDs *297, 403, 519, 646, 657, 771, 795*). For comparison, we validated two fragments with putative autosomal (scaffold IDs *236, 267*) or X-linkage (scaffold IDs *301, 253*), based on results from the AD-ratio approach. Markers on these putatively non Y-linked scaffolds could be PCR amplified in both male and female DNA samples, and showed clear amplicons of the same sizes in both sexes.

High abundance of repetitive elements on the Y-linked scaffolds

Overall, TEs covered 54.38% of the total length of the 14 Y-scaffolds used in this analysis (Supplementary Table S5 and S6, Figures 3 and S3). The majority of the transposable element sequences represent placental mammalian LINE-1 (38%) or the carnivore CAN-SINEs (7.8%). The average LINE-1 coverage of the polar bear genome is 16.93%, thus LINE-1 covered nearly twice as much sequence on the Y chromosome scaffolds compared to the entire genome. In addition, one full-length LINE-1 copy, the L1-1_AMe, with a length=6,021 bp was found on scaffold 297 (Figure 3). The full-length L1-1_AMe is likely to have been recently active, due to the presence of only two stop codons in the endonuclease/reverse transcriptase encoding ORF2. The abundance of repetitive regions along the Y chromosome, and the positions of homologous regions to candidate gene sequences are shown in Figure 3, exemplary for two long Y-chromosomal scaffolds (scaffold IDs 297, 318). Corresponding maps for 12 additional Y-scaffolds are provided in supplementary Figure S3. We found a higher abundance of LINE-1 elements and a lower abundance of older LINE-2 and LINE-3 elements on the Y-chromosome compared to the whole genome. Moreover, a higher abundance of carnivore-specific SINEs as compared to ancestral Mammalian Interspersed Repeats (MIR) was detected. We identified a similar amount of LTRs/ERVs on the Y chromosome and the whole genome while less DNA transposons were identified on the Y chromosome compared to the whole genome. We identified 115 microsatellites with at least 15 repeat units, that are likely to show intraspecific polymorphism and are thus useful for population genetic studies, covering 0.3% of the combined length of all Y-scaffolds (Tables S6).

Phylogenetic analyses identify two distinct male polar bear lineages

Phylogenetic analysis of 511 kb Y chromosome sequence in 12 polar bears identified two highly divergent paternal lineages (Figure 4*A*), with two individuals (AK4 and PB16) being clearly separated from the remaining 10 polar bears. This separation does not correspond to geography, since both major lineages occur in Alaska and Svalbard (Norway), respectively. Some individuals have a considerable number of unique substitutions (e.g. PB16: 55 substitutions) relative to 101 substitutions separating the two lineages. Our Bayesian analysis yielded a phylogenetic tree with high posterior support for all major nodes (Figure 4*B*), showing two distinct patrilineal clades within polar bears. Based on the demographic split of brown and polar bears at 343-479 kya (Liu et al. 2014), we obtained a median divergence time estimate for the split of these two clades at 0.12 mya (million years ago) (95% HPD: 0.10–0.15). The split between brown and polar bears was estimated at 0.40 mya

(0.34–0.47), and the divergence of the black bear at 1.190 mya (0.99-1.44). Using a fixed mutation rate as an alternative calibration scenario, older divergence time estimates were obtained: the split within polar bears was estimated at 0.22 mya (0.19-0.25), the split between brown and polar bears at 0.70 mya (0.65-0.76), and the divergence of the black bear 2.13 mya (2.03-2.23).

DISCUSSION

The Y chromosome is poorly characterized in most mammals, including the carnivoran bear family. We used the polar bear reference assembly to identify a large amount of Y-linked sequence, totaling 1.9 Mb distributed across 112 Y-linked scaffolds. We did so by applying a similarity search with mammalian Y-linked genes and by analyzing differences in sequencing coverage (AD-ratio) on Y-chromosomal, autosomal, and X-linked scaffolds. We backed up these *in-silico* results by validating male-specificity by PCR amplification for 20 of the largest scaffolds, corresponding to 1.7 Mb of Y-chromosomal sequence (Figure 1, Tables 1 and 2).

The AD-ratio approach can reliably identify scaffolds from autosomes, the X and the Y chromosome

Using the AD-ratio approach, the majority of scaffolds in the polar bear assembly could be assigned to one of three chromosomal classes. We identified 2.18 Gb of autosomal sequence, which is close to the total size of the polar bear assembly of 2.3 Gb (Figure 2). The X-linked scaffolds amounted to 109 Mb, which approximates two thirds the size of the human X chromosome (Ross et al. 2005). This scaffold class included the twelve scaffolds that were previously identified as being X-linked in bears (Cahill et al. 2013). The amount of identified autosomal and X-linked sequences thus fit the expectations for a typical mammalian genome of ~2-3 Gb (Rogers & Gibbs 2014). Only ~0.2% (~5.3 Mb) of the polar bear assembly remained unassigned, because AD-ratios for these scaffolds were beyond our thresholds for autosomal, X-chromosomal, and Y-chromosomal sequences. This illustrates the reliability of the AD-ratio approach, and its suitability to screen a genome assembly for the three chromosome classes. The 1.9 Mb identified to be Y-linked are a considerable amount of Y-chromosomal sequence, given the lack of Y-linked genomic sequences for many mammals, the generally small size of the mammalian Y chromosome, and its highly repetitive nature that impedes assembly. We likely underestimate the total amount of Y-linked sequences (see below), and 1.9 Mb represent only a small fraction of the entire polar bear Y chromosome. The size of Y chromosomes differs considerably among mammals and even among carnivores, but Y chromosomes are typically longer than 20 Mb (e.g., dog: 20 Mb, cat: 45 Mb (Li et al. 2013)). While the size of the polar bear Y chromosome has not yet been determined, it appears to be about half the physical size of the X chromosome in metaphase spreads (O´Brien et al. 2006).

The similarity search identified 23 scaffolds as being Y-linked, however, later inspection indicated that only nine of these had an AD ratio indicative of Y-linkage (Figure 1, Tables 1, S3). The AD-ratio approach yielded 112 Y-linked scaffolds and thus proved to be more efficient than the similarity search in terms of scaffold numbers. However, the nine scaffolds identified by both approaches total 1.24 Mb, which is more than 60% of the entire Y-linked sequence data. Although a similarity search technically simple, successful, and easily applied, several drawbacks are associated with this approach.

We based our selection of query genes on their previous description as being Y-linked in other mammals, implicitly assuming the presence of these genes also on the polar bear Y chromosome. However, Y chromosomes can differ in their gene content across taxa, and lineage-specific sets of Y-linked genes exist (Murphy et al. 2006; Cortez et al. 2014). Indeed, we found that Y-linked genes that are absent in carnivores, e.g. *NLGN4Y* (Cortez et al. 2014), were also absent in polar bears. In contrast, genes that are widespread throughout placental mammals, and occur in other carnivores (dog and cat; e.g. *ZFY*, *UTY*, *EIF1AY*) (Cortez et al. 2014), are those genes that are actually found on Y-linked scaffolds in the polar bear assembly (Figures 3, S3). Currently limited knowledge of gene contents on the Y chromosomes of different mammalian lineages is therefore still restricting the efficiency of similarity-based approaches for the identification of Y-chromosomal scaffolds.

In several cases, the similarity search produced hits to more than one scaffold (Table S4). For example, a search with *ZFY* sequences yielded similarity to scaffold *318* (containing *ZFY*) and scaffold *20* (containing *ZFX*). The scaffolds in such groups all had AD-ratios characteristic of either Y- or X-linkage (Table S4). Most Y-linked genes on these scaffolds are classified as X-degenerate in humans (Table S1). These genes are relics of the ancient autosomes from which the mammalian X and Y chromosomes evolved, and are thus expected to show homology between the X and Y chromosome (Skaletsky et al. 2003). In contrast, *RBMY* (scaffold IDs *369*, *105*; Table S4), is classified as ampliconic in humans, and such genes normally lack X-linked counterparts. *RBMY*, however, is one of the two ampliconic genes with an X-linked homolog in humans (*RBMX*), explaining its detection on an X-linked scaffold. These findings illustrate the high degree of sequence similarity between some sex chromosome gametologs (homologous genes on the two sex chromosomes), and the common evolutionary history of Y and X chromosomes, deriving from an ancestral pair of autosomes. Moreover, four scaffolds with sequence similarity to Y-linked genes, but an AD-ratio indicative of autosomal (or pseudoautosomal) origin were identified (Table S3). Interestingly, one of these genes (*RPS4Y*, on scaffold: *13*) is in close proximity to the pseudoautosomal region (PAR) on the small arm of the human Y chromosome (Skaletsky et al. 2003). The location of the PAR is not known in polar bears, but genes in regions recombining with the X chromosome would hinder correct identification of Y-linked scaffolds by the AD-ratio approach.

The similarity search is further complicated by the high degree of similarity between some gametologous genes. This is exemplified by *ZFY*/*ZFX* genes, for which we were initially not able to differentiate between the respective Y- and X-scaffolds based on exon sequences. The more rapidly evolving intron sequences, however, allowed us to differentiate between Y- and X-linked scaffolds (Figure S1). With a more stringent set of candidate genes, i.e. carnivore-specific Y-linked genes, the reliability of the similarity search can be improved, and the search for intronic sequences would allow for a better differentiation between gametologs on the two sex chromosomes. A drawback of a similarity search based on Y-linked gene sequences from other taxa is that scaffolds consisting of exclusively intergenic sequence cannot be identified. This is an important limitation of similarity search approaches, because mammalian Y chromosomes are generally gene poor, with only 78 protein-coding genes in humans (Bachtrog 2013). Indeed, 103 out of 112 Y-linked scaffolds were solely identified by their AD-ratio. Nevertheless, four out of five Y-linked scaffolds with a size of >100 kb were also identified by the similarity search (Figure 1, Table 2). Assemblies with

fewer but larger scaffolds will thus be more amenable to accurate detection of Y-linkage by a similarity search approach.

The structure of the Y chromosome complicates identification of Y-linked sequences

The heterochromatic, highly repetitive regions of a genome usually remain unassembled in whole-genome sequencing projects. Some Y chromosomes contain extended regions of largely uncharacterized heterochromatin, e.g. human and *Drosophila* (Bachtrog 2013). Other Y chromosomes are largely euchromatic, e.g. mouse and chimpanzee, but even the euchromatic regions are enriched for ampliconic sequences containing duplicated genes (Skaletsky et al. 2003; Hughes et al. 2010; Soh et al. 2014). Accurate sequence assembly is therefore inherently difficult for the Y chromosome, and sequence similarity to the X and possibly other chromosomes further complicate the identification of a distinct Y-linked sequence. Therefore, high-quality Y chromosome reference sequence assemblies are so far lacking from most mammalian genome sequencing projects.

The identification of Y-linked scaffolds has previously been achieved by *in-silico* search for known Y-linked genes and massive *in-vitro* PCR-based verification in *Drosophila* and *Anopheles* (Carvalho et al. 2000; Krzywinski et al. 2004). Moreover, Y-linked sequences can be retrieved by subtracting the scaffolds of the homogametic from the heterogametic assembly (Chen et al. 2014). Approaches based on some measure of the coverage depth of sequence reads on Y-linked scaffolds (Carvalho et al. 2003, Chen et al. 2012), e.g. the “Y chromosome genome scan” (Carvalho & Clark 2013), or on the number of alignments in males and females, the “chromosome quotient” (Hall et al. 2013), have also been applied.

The occurrence of gene conversion, where a gene copy on one chromosome is overwritten by the information from the other chromosome, further complicates identification of chromosome-specific sequences. This process appears also to occur in the bear lineage (Figure S1). Compared to the human and dog outgroups, the tremarctine and ursine *ZFY* and *ZFX* sequences cluster together, and not with human and dog *ZFY* and *ZFX*, respectively. Additionally, the *ZFY* and *ZFX* sequences from giant panda (*Ailuropoda melanoleuca*), are more closely related to each other than any gene copy is to those from ursine and tremarctine bears. A likely explanation for these observations is that gene conversion has occurred in the ancestral giant panda lineage as well as in the lineage leading to tremarctine and ursine bears. Considering the divergence times of ursid lineages, these two conversion events occurred in the Miocene, more than 12 and six mya, respectively (Kutschera et al. 2014). The occurrence of gene conversion between sex chromosomes has been described in various mammalian lineages such as primates and felids (Slattery et al. 2000; Rosser et al. 2009; Trombetta et al. 2014), including *ZFX*/*ZFY*.

Transposable elements (TEs) on the sex chromosomes pose yet another challenge for accurate assembly and identification of chromosome-specific sequences. Mammalian genomes contain large amounts of TEs that propagate via different mechanisms. The human genome has over 44% of TEs (Lander et al. 2001), while the polar bear genome consists of 39.2% TEs (Table S5). Previous studies have shown that there is a preferential insertion of some TEs (primate-specific LINE1 and Alu elements) on the human and chimpanzee X and Y chromosomes (Kvikstad & Makova 2010). The same distribution is observed on polar bear Y chromosome scaffolds, as there is a high abundance of LINE-1 and the carnivore-specific Can-SINEs (Walters-Conte et al. 2011) compared to the autosomes (Table S5). The ancestral TEs, such as LINE-2, LINE-3 and MIR elements which were active before the split

between marsupial and placental mammals (Smit & Riggs 1995), are found in very low numbers on the polar bear Y scaffolds. The endogenous retrovirus (ERV) and DNA transposons seem to accumulate more evenly across the genome than LINE-1 and Can-SINEs, as there are only small differences between Y-chromosomal and autosomal scaffolds. The reason for the preferential accumulation on the sex chromosomes has been attributed to male and female germline TE integrations occurring before meiotic sex chromosome inactivation (Kvikstad & Makova 2010).

Due to the repetitive nature of the Y chromosome, assembly methods will likely produce numerous smaller scaffolds and collapse repetitive sequences into chimeric scaffolds that actually comprise multi-copy sequences. Indeed, stretches of very high sequence coverage were found on many of the Y-linked scaffolds. Moreover, long and highly repetitive regions of the Y chromosome might be entirely missing from the assembly. TEs and X-transposed sequences on the Y chromosome likely cause a proportion of female reads from polar bear X chromosome and autosomes to be falsely mapped to Y-linked scaffolds, due to the high similarity among such regions. This produces AD-ratios greater than zero for these true Y-scaffolds. Moreover, due to the paucity of information on bear sex chromosomes, we cannot exclude the possibility of recent stratum formation, with the existence of segments that have not yet attained a high level of divergence between the Y and X chromosome. However, recent stratum formation or added genes from autosomes has not been reported for the well-studied Y chromosomes of two other carnivores, cat and dog (Cortez et al. 2014). A strict AD-ratio threshold of exactly zero is therefore likely to produce many false negatives. Our employed relaxed AD-ratio threshold of ≤0.3 yielded an additional 22 scaffolds, eleven of them tested and verified *in-vitro* (Table 2) to be of Y-chromosomal origin.

Assembly artifacts resulting from the repetitive nature of the Y chromosome imply that we likely underestimate the actual number and length of the identified Y-linked sequences. The AD-ratio approach should thus not be seen as an attempt to identify all Y-linked sequences in bears, nor to determine the size of the polar bears´ Y chromosome. Rather, the approach is an effective means to identify sequences that demonstratively have a high probability of being Y-linked and that can be used for evolutionary studies.

<u>Y-chromosomal sequences provide a high resolution patrilineal perspective on polar bear evolutionary history</u>

Our phylogenetic analyses of Y-linked scaffold sequences provide a patrilineal view on polar bear evolution that support a previously identified pattern of two distinct Y-chromosomal lineages in polar bears, PO1.1 and PO2 (Bidon et al 2014) (Figure 4). The large amount of analyzed sequence data provides high resolution of individual lineages, with many haplotype-specific substitutions. Our divergence time estimation places the split of these two polar bear clades around the Eemian interglacial period (0.12 - 0.13 mya), implying that the two lineages separated long before the last glacial maximum (ca. 18–25 kya). The clear separation into two paternal lineages indicates an ancient population structuring in polar bears, possibly due to the separation into multiple refugia during glaciation cycles, similar to other arctic species (Flagstad & Røed 2003).

The divergence time was estimated using a recently published date on the population split between brown and polar bears (343-479 kya (Liu et al. 2014)). This demographic split is expected to be younger than estimates based on the coalescence of allelic lineages, e.g.

the 338-934 kya estimated by Hailer et al. (2012). The lower effective population size of the Y chromosome implies that coalescence of Y-lineages occurs faster than that of autosomal lineages. Therefore, the Y-chromosomal gene tree might track the demographic splits of the species more closely.

It is noteworthy that our divergence estimate of the black bear patriline (0.99 – 1.44 mya 95% HPD) is relatively young in this calibration scenario. The fossil record suggests a first occurrence of the black bear lineage at least 1.8 mya (Kurtén & Anderson 1980). In principle, a Y-specific mutation rate would be a reasonable alternative calibration method. However, an independent mutation rate for the ursid Y chromosome has not yet been determined, and lineage-specific rates in mammals make the adoption of a Y-specific rate from another taxon unreliable. Applying a recent estimate for the mutation rate of the human Y chromosome, we obtained even older divergence time estimates for the patrilines of polar bears (0.19 - 0.25 mya), of brown and polar bears (0.65-0.76 mya), and of the black bear lineage (2.03-2.23 mya). These dates are broadly consistent with other estimates of genomic divergence times for these splits (Hailer et al. 2012; Cahill et al. 2013; Cronin et al. 2014), and more in line with the fossil record of American black bears.

Short Y-linked sequences were recently used as markers for sex determination in bears (Bidon et al. 2013), phylogeographic analyses of brown and polar bear brother lineages (Bidon et al. 2014), and phylogenetic analyses of all eight bear species (Kutschera et al. 2014). Sequences on Y-chromosomal scaffolds have thus already proven to be a reliable resource for studying the evolutionary history of polar bears and other members of the ursid family.

Conclusions

The analyses of Y-chromosomal scaffolds provided a high-resolution view on the patrilineal relationship within polar bears, identifying two highly distinct clades that separated during the middle Pleistocene. A preferential accumulation of younger TEs on the polar bear Y-chromosome could be shown. As more and more genomes become available in the form of reference assemblies and short read archives, straightforward *in-silico* strategies to identify sex-linked sequences from these data can now be applied in many species. Overall, the AD-ratio approach seems to be highly specific and preferable for a reliable identification of Y chromosome scaffolds. It can be used as long as a reference assembly of the heterogametic sex, and short reads of one male and one female are available.

ACKNOWLEDGEMENTS

This work was supported by LOEWE Landes-Offensive zur Entwicklung Wissenschaftlich-ökonomischer Exzellenz; the Arthur und Aenne Feindt-Stiftung, Hamburg; and the Leibniz-Association. We thank Steven Fain, Nicolas Lecomte, and Frank Zachos for providing samples, and the anonymous reviewers for their constructive comments. Jón Baldur Hlidberg kindly provided the polar bear painting.

REFERENCES

Anderson S et al. 1981. Sequence and organization of the human mitochondrial genome. Nature. 290:457–465.
Andrews S. 2010. FastQC: A Quality Control tool for High Throughput Sequence Data.
Bachtrog D. 2013. Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. Nat. Rev. Genet. 14:113–24.
Bellott DW et al. 2014. Mammalian Y chromosomes retain widely expressed dosage-sensitive regulators. Nature. 508:494–499.
Bidon T et al. 2013. A sensitive and specific multiplex PCR approach for sex identification of ursine and tremarctine bears suitable for non-invasive samples. Mol. Ecol. Resour. 13:362–368.
Bidon T et al. 2014. Brown and Polar Bear Y Chromosomes Reveal Extensive Male-Biased Gene Flow within Brother Lineages. Mol. Biol. Evol. 31:1353–1363.
Bouckaert R et al. 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. PLoS Comput. Biol. 10:e1003537.
Cahill JA et al. 2013. Genomic Evidence for Island Population Conversion Resolves Conflicting Theories of Polar Bear Evolution. PLoS Genet. 9:e1003345.
Cahill JA et al. 2014. Genomic evidence of geographically widespread effect of gene flow from polar bears into brown bears. Mol. Ecol. doi: 10.1111/mec.13038.
Carvalho AB et al. 2003. Y chromosome and other heterochromatic sequences of the Drosophila melanogaster genome: how far can we go? Genetica. 117:227–237.
Carvalho AB, Clark AG. 2013. Efficient identification of Y chromosome sequences in the human and Drosophila genomes. Genome Res. 23:1894–907.
Carvalho AB, Koerich LB, Clark AG. 2009. Origin and evolution of Y chromosomes: Drosophila tales. Trends Genet. 25:270–277.
Carvalho AB, Lazzaro BP, Clark AG. 2000. Y chromosomal fertility factors kl-2 and kl-3 of Drosophila melanogaster encode dynein heavy chain polypeptides. PNAS. 97:13239–13244.
Chen N, Bellott DW, Page DC, Clark AG. 2012. Identification of avian W-linked contigs by short-read sequencing. BMC Genomics. 13:183.
Chen S et al. 2014. Whole-genome sequence of a flatfish provides insights into ZW sex chromosome evolution and adaptation to a benthic lifestyle. Nat. Genet. 46.
Chesser RK, Baker RJ. 1996. Effective sizes and dynamics of uniparentally and diparentally inherited genes. Genetics. 144:1225–1235.
Clement M, Posada D, Crandall K. 2000. TCS: a computer program to estimate gene genealogies. Mol. Ecol. 9:1657–9.
Cortez D et al. 2014. Origins and functional evolution of Y chromosomes across mammals. Nature. 508:488–493.
Cronin MA et al. 2014. Molecular Phylogeny and SNP Variation of Polar Bears (*Ursus maritimus*), Brown Bears (*U . arctos*), and Black Bears (*U . americanus*) Derived from Genome Sequences. J. Hered.
Darriba D, Taboada GL, Doallo R, Posada D. 2012. jModelTest 2: more models, new heuristics and parallel computing. Nat. Methods. 9:772.
Flagstad Ø, Røed KH. 2003. Refugial Origins of Reindeer (*Rangifer tarandus L .*) Inferred from Mitochondrial DNA Sequences. Evolution. 57:658–670.
Fu Q et al. 2014. Genome sequence of a 45,000-year-old modern human from western Siberia. Nature. 514:445–449.
Greminger MP, Krützen M, Schelling C, Pienkowska-Schelling A, Wandeler P. 2010. The quest for Y-chromosomal markers - methodological strategies for mammalian non-model organisms. Mol. Ecol. Resour. 10:409–420.
Hailer F et al. 2012. Nuclear Genomic Sequences Reveal that Polar Bears Are an Old and Distinct Bear Lineage. Science. 336:344–347.
Hall AB et al. 2013. Six novel Y chromosome genes in Anopheles mosquitoes discovered by independently sequencing males and females. BMC Genomics. 14:273.

Hallast P, Balaresque P, Bowden GR, Ballereau S, Jobling M. 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.

Hughes JF et al. 2010. Chimpanzee and human Y chromosomes are remarkably divergent in structure and gene content. Nature. 463:536–539.

Hughes JF et al. 2012. Strict evolutionary conservation followed rapid gene loss on human and rhesus Y chromosomes. Nature. 1–6.

Hughes JF, Rozen S. 2012. Genomics and genetics of human and primate Y chromosomes. Annu. Rev. Genomics Hum. Genet. 13:83–108.

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

Janke A, Feldmaier-Fuchs G, Thomas WK, von Haeseler A, Pääbo S. 1994. The Marsupial Mitochondrial Genome and the Evolution of Placental Mammals. Genetics. 137:243–256.

Jobling M, Tyler-Smith C. 2003. The human Y chromosome: an evolutionary marker comes of age. Nat. Rev. Genet. 4:598–612.

Joshi NA, Fass JN. 2011. Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files. Available at https://github.com/najoshi/sickle.

Kayser M. 2007. Uni-parental markers in human identity testing including forensic DNA analysis. Biotechniques. 43:Sxv–Sxxi.

Kondo M et al. 2006. Genomic organization of the sex-determining and adjacent regions of the sex chromosomes of medaka. Genome Res. 16:815–826.

Krzywinski J, Nusskern DR, Kern MK, Besansky NJ. 2004. Isolation and Characterization of Y Chromosome Sequences From the African Malaria Mosquito Anopheles gambiae. Genetics. 166:1291–1302.

Kurtén B, Anderson E. 1980. *Pleistocene Mammals of North America*. Columbia University Press.

Kutschera VE et al. 2014. Bears in a Forest of Gene Trees: Phylogenetic Inference Is Complicated by Incomplete Lineage Sorting and Gene Flow. Mol. Biol. Evol.

Kvikstad EM, Makova KD. 2010. The (r)evolution of SINE versus LINE distributions in primate genomes : Sex chromosomes are important. Genome Res. 20:600–613.

Lander ES et al. 2001. Initial sequencing and analysis of the human genome. Nature. 409:860–921.

Li B, Zhang G, Willerslev E, Wang J. 2011. Genomic data from the Polar Bear (*Ursus maritimus*). Gigascience. http://dx.doi.org/10.5524/100008.

Li G et al. 2013. Comparative analysis of mammalian Y chromosomes illuminates ancestral structure and lineage-specific evolution. Genome Res.

Li H et al. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 25:2078–2079.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 25:1754–1760.

Liu S et al. 2014. Population Genomics Reveal Recent Speciation and Rapid Evolutionary Adaptation in Polar Bears. Cell. 157:785–794.

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

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

Murphy WJ et al. 2006. Novel gene acquisition on carnivore Y chromosomes. PLoS Genet. 2:e43.

Nakagome S, Pecon-Slattery J, Masuda R. 2008. Unequal rates of Y chromosome gene divergence during speciation of the family Ursidae. Mol. Biol. Evol. 25:1344–56.

O´Brien S, Menninger J, Nash W. 2006. *Atlas of Mammalian Chromosomes*. John Wiley & Sons: Hoboken, NJ.

Van Oven M, Van Geystelen A, Kayser M, Decorte R, Larmuseau MHD. 2014. Seeing the wood for the trees: a minimal reference phylogeny for the human Y chromosome. Hum. Mutat. 35:187–191.

Pagès M et al. 2008. Combined analysis of fourteen nuclear genes refines the Ursidae phylogeny. Mol. Phylogenet. Evol. 47:73–83.

Pagès M et al. 2009. A system for sex determination from degraded DNA: a useful tool for palaeogenetics and conservation genetics of ursids. Conserv. Genet. 10:897-907.

Posada D, Crandall K, Holmes EC. 2002. Recombination in evolutionary genomics. Annu. Rev. Genet. 36:75–97.

Rogers J, Gibbs R. 2014. Comparative primate genomics: emerging patterns of genome content and dynamics. Nat. Rev. Genet. 15:347–59.

Ross MT et al. 2005. The DNA sequence of the human X chromosome. Nature. 434:325–337.

Rosser ZH, Balaresque P, Jobling M. 2009. Gene conversion between the X chromosome and the male-specific region of the Y chromosome at a translocation hotspot. Am. J. Hum. Genet. 85:130–4.

Sacks BN et al. 2013. Y Chromosome Analysis of Dingoes and Southeast Asian Village Dogs Suggests a Neolithic Continental Expansion from Southeast Asia Followed by Multiple Austronesian Dispersals. Mol. Biol. Evol. 30:1103–1118.

Scozzari R et al. 2014. An unbiased resource of novel SNP markers provides a new chronology for the human Y chromosome and reveals a deep phylogenetic structure in Africa. Genome Res.

Skaletsky H et al. 2003. The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. Nature. 423:825–37.

Slattery JP, Sanner-wachter L, O´Brien SJ. 2000. Novel gene conversion between X-Y homologues located in the nonrecombining region of the Y chromosome in Felidae (Mammalia). PNAS. 97:5307–5312.

Smit AFA, Riggs AD. 1995. MIRs are classic, tRNA-derived SINEs that amplified before the mammalian radiation. Nucleic Acids Res. 23:98–102.

Soh YQS et al. 2014. Sequencing the Mouse Y Chromosome Reveals Convergent Gene Acquisition and Amplification on Both Sex Chromosomes. Cell.

Trombetta B, Sellitto D, Scozzari R, Cruciani F. 2014. Inter- and intraspecies phylogenetic analyses reveal extensive X-Y gene conversion in the evolution of gametologous sequences of human sex chromosomes. Mol. Biol. Evol. 31:2108-2123.

Veeramah KR, Hammer MF. 2014. The impact of whole-genome sequencing on the reconstruction of human population history. Nat. Rev. Genet.

Wallner B et al. 2013. Identification of Genetic Variation on the Horse Y Chromosome and the Tracing of Male Founder Lineages in Modern Breeds. PLoS One. 8:e60015.

Walters-Conte KB, Johnson DLE, Allard MW, Pecon-Slattery J. 2011. Carnivore-specific SINEs (Can-SINEs): distribution, evolution, and genomic impact. J. Hered. 102 Suppl :S2–10.

Wei W et al. 2013. A calibrated human Y-chromosomal phylogeny based on resequencing. Genome Res. 23:388–95.

Willard HF. 2003. Tales of the Y chromosome. Nature. 423:810–813.

Wilson AC et al. 1985. Mitochondrial DNA and two perspectives on evolutionary genetics. Biol. J. Linn. Soc. 375–400.

FIGURE LEGENDS

Fig 1: Identified scaffolds in the polar bear assembly

A Scaffolds identified by the similarity search, the AD-ratio, and by both approaches (overlap). Scaffold ≥10 kb are shown by their ID numbers. Details for 92 additionaly Y-linked scaffolds (<10 kb, combined length: ~170 kb) are listed in Table S2. Some scaffolds identified by the similarity search showed AD-ratios characteristic of autosomal linkage (red) or X-linkage (blue). Scaffolds with an asterisk (*) have been verified *in-vitro* to be male-specific. Two asterisks indicate scaffolds that show PCR amplification in both sexes. No reads mapped with sufficient mapping quality to scaffold *4889*, so its AD-ratio could not be calculated, and scaffold *6612* was <1 kb. ***B*** AD-ratios of X-linked (blue), autosomal (red), and Y-linked (green) scaffolds.

Fig. 2: AD-ratio histogram of polar bear scaffolds

A Distribution of AD-ratios of scaffolds ≥1 kb, and their combined size shown in bins of width 0.02. Autosomal scaffolds cluster around an AD-ratio of 1 (red), X-linked scaffolds around 2 (blue). The stippled box highlights the region shown enlarged in ***B***. Enlargement of the box in A. Scaffolds below the threshold of 0.3 are identified as Y-linked (green). Scaffolds unassigned to chromosomal classes are shown in black.

Fig. 3: Annotations of Y-linked scaffolds 297 and 318.

Exons homologous to mouse and human are shown in red. Previously published Y-linked polar bear sequences are shown in pink. The repeat unit of each microsatellite is indicated and regions with >200 bp of consecutive "N" are highlighted in gray. Due to the high abundance, only placental mammalian non-LTR retrotransposons ≥500 bp (LINEs) and ≥100 bp (SINEs) were plotted. The maps of additional scaffolds are shown in Figure S3.

Fig. 4: Phylogenomic analysis of ~0.5 megabases Y-chromosomal sequence from 12 polar bears.

Geographic origins of the polar bear individuals is denoted by AK (Alaska) and PB (Svalbard). ***A*** NeighborNet analysis. ***B*** Time-calibrated Bayesian coalescent-based phylogeny from BEAST. Numbers at nodes indicate the median of the divergence time in million years ago, with 95% highest posterior density in brackets. Dots at nodes indicate posterior probability >0.99. The scale axis is in units of million years ago. Note that ~80% older dates were retrieved from an alternative calibration scenario (see text).

FIGURES

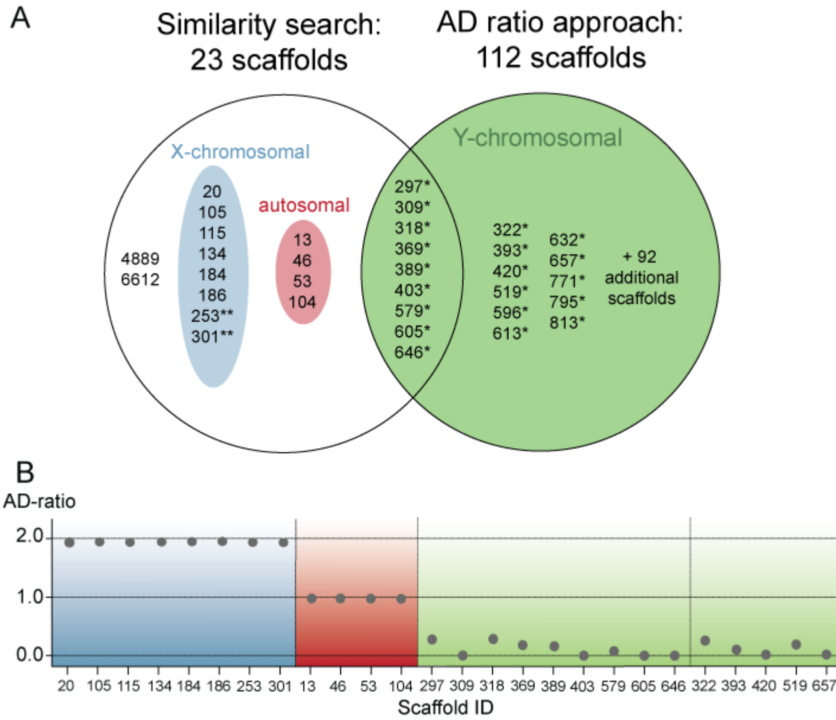


Fig. 1

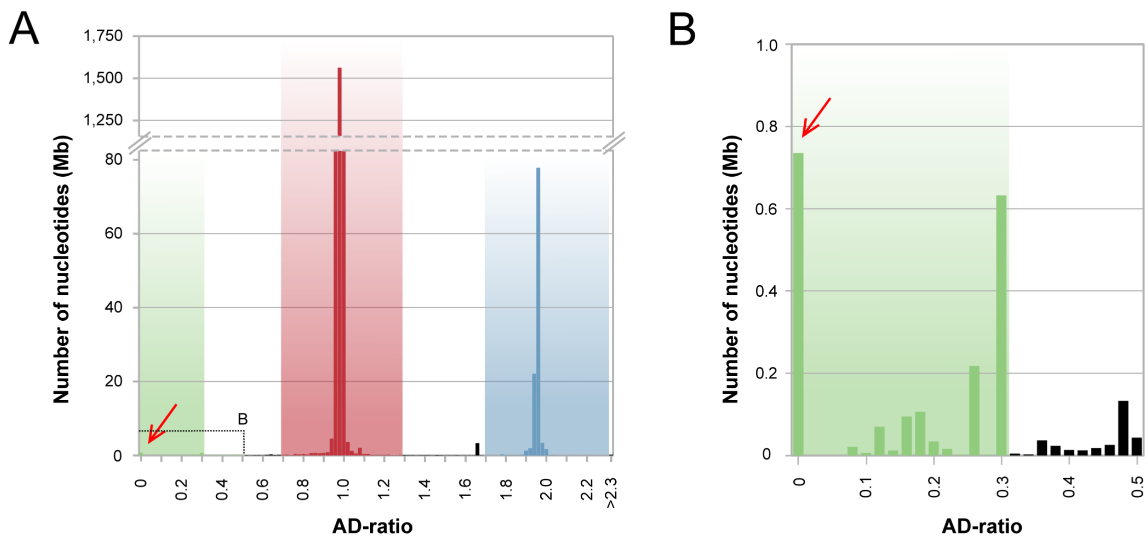


Fig. 2

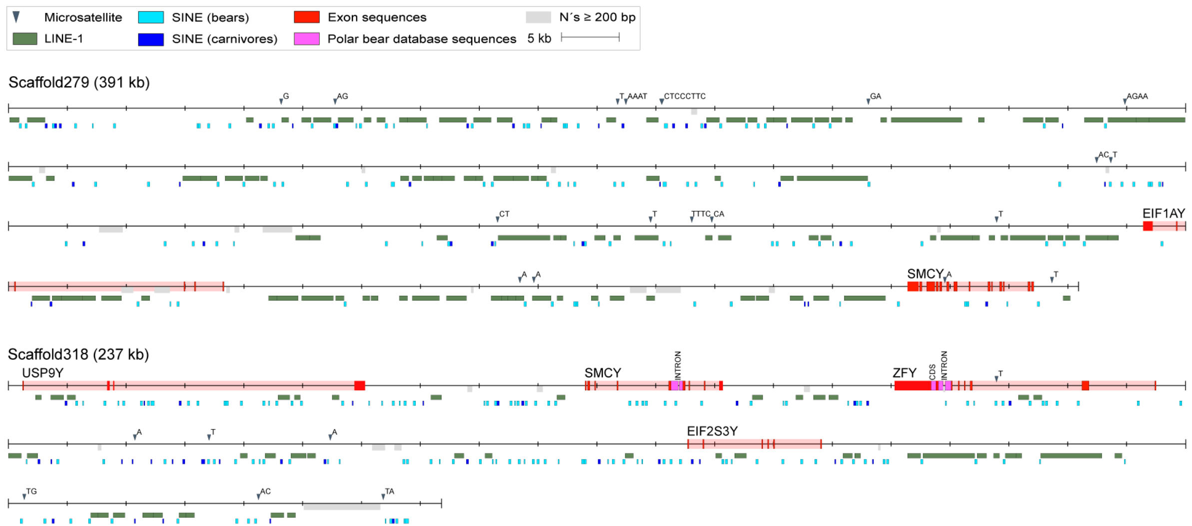


Fig. 3

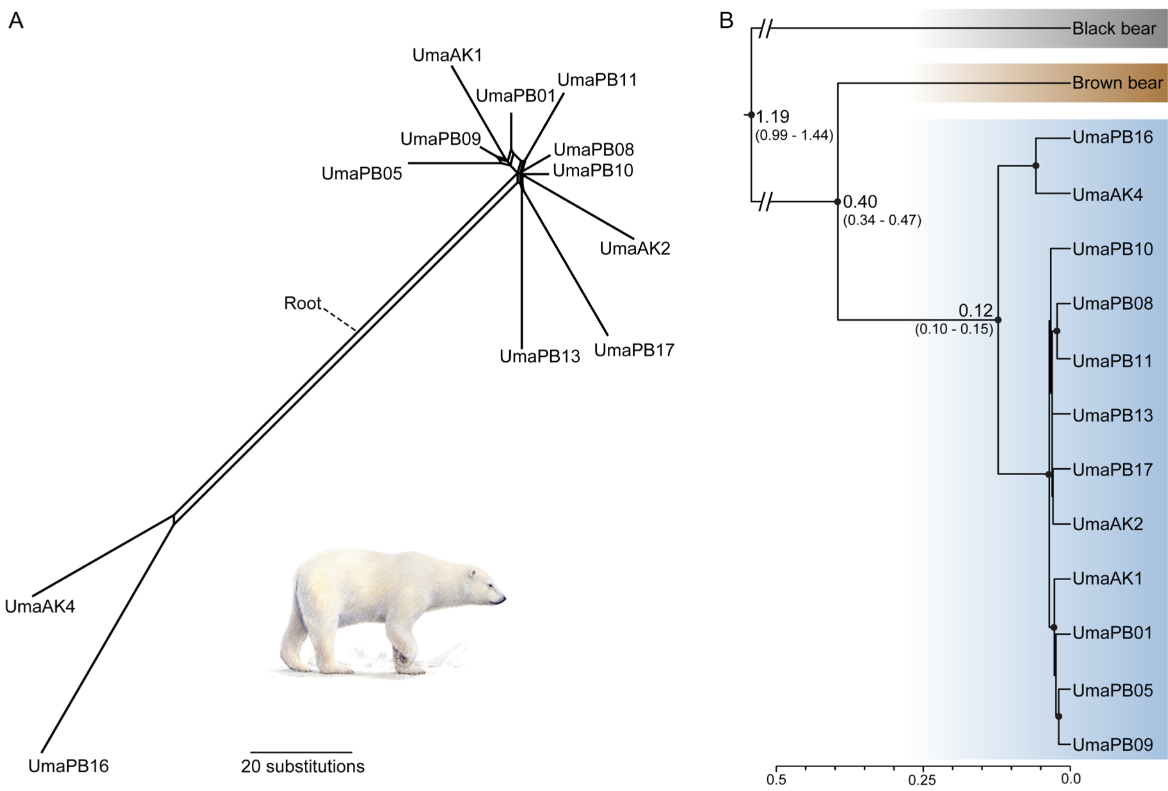


Fig. 4

TABLES

Table 1: Polar bear scaffolds showing similarity to 18 mammalian Y-linked genes.

Nr.	Scaffold (ID)	Scaffold size (kb)	Gene
1	*13*	26,707	*RPS4Y*
2	*20*	22,125	*EIF1AY, EIF2S3Y, USP9Y, ZFY*
3	*46*	15,941	*TBL1Y*
4	*53*	14,458	*SLY*
5	*104*	6,801	*NLGN4Y, PRKY, TBL1Y*
6	*105*	6,717	*RBMY1A1*
7	*115*	5,608	*AMELY*
8	*134*	4,673	*UBA1Y, UTY*
9	*184*	2,589	*DDX3Y, USP9Y*
10	*186*	2,578	*PCDH11Y*
11	*253*	821	*RPS4Y, RPS4Y2*
12	*297**	391	*EIF1AY, KDM5D*
13	*301*	351	*KDM5D*
14	*309**	317	*DDX3Y, USP9Y, UTY,*
15	*318* (3836)***	237	*EIF2S3Y, KDM5D, USP9Y, ZFY***
16	*369**	104	*RBMY1A1*
17	*389**	77	*AMELY*
18	*403**	63	*UBA1Y*
19	*579**	21	*SRY*
20	*605**	19	*UBA1Y*
21	*646**	15	*EIF2S3Y*
22	*4889*	0.9	*UBA1Y*
23	*6612*	0.7	*AMELY*

* Scaffolds have an AD-ratio indicative of Y-linkage and were validated *in-vitro* to be male-specific.
** The entire length of scaffold *3836* (1 kb; with similarity to *ZFY*) is included within scaffold *318* with 100% identity.

Table 2: Y-chromosomal scaffolds ≥10 kb identified by the AD-ratio
The male-specificity of all scaffolds listed here has been validated *in-vitro*. Additional scaffolds (<10 kb) are shown in Table S2.

Nr.	Scaffold ID	Size (kb)	AD-ratio	Similarity to Y-linked gene
1	*297*	391	0.284	*EIF1AY, KDM5D*
2	*309*	317	0	*DDX3Y, USP9Y, UTY*
3	*318*	237	0.285	*EIF2S3Y, KDM5D, USP9Y, ZFY*
4	*322*	217	0.252	-
5	*369*	104	0.18	*RBMY1A1*
6	*389*	77	0.16	*AMELY*
7	*393*	70	0.12	-
8	*403*	63	0	*UBA1Y*
9	*420*	54	0	-
10	*519*	31	0.198	-
11	*579*	21	0.075	*SRY*
12	*596*	20	0	-
13	*605*	19	0	*UBA1Y*
14	*613*	18	0.158	-
15	*632*	16	0.205	-
16	*646*	15	0	*EIF2S3Y*
17	*657*	14	0	-
18	*771*	10	0.135	-
19	*795*	10	0	-
20	*813*	10	0	-

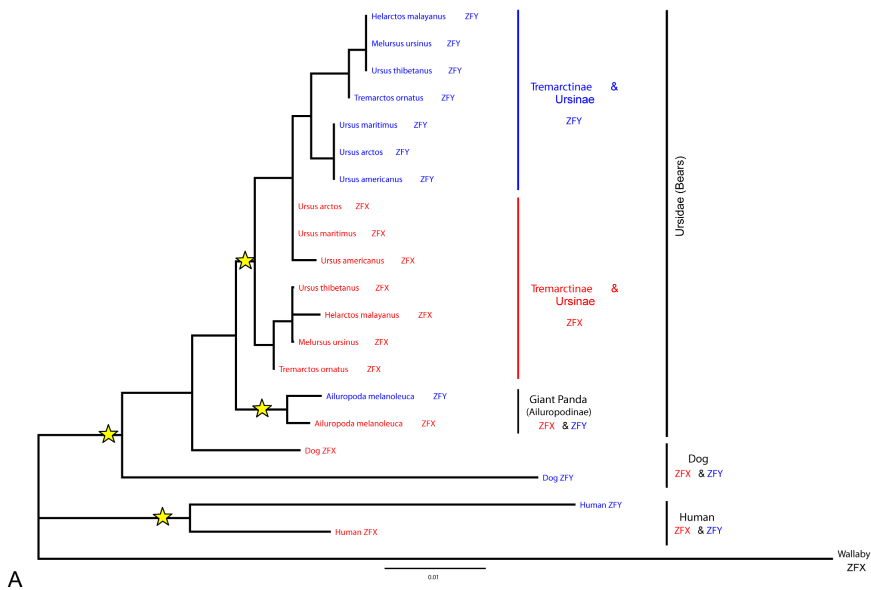
Supplementary information

Genome-wide search identifies 1.9 megabases from the polar bear Y chromosome for evolutionary analyses

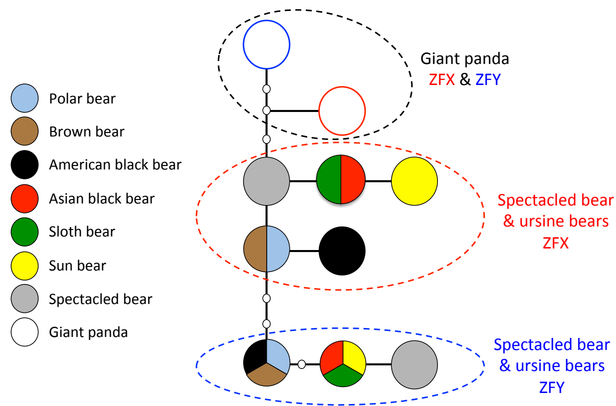
Tobias Bidon, Nancy Schreck, Frank Hailer , Maria Nilsson and Axel Janke

Details on *in-vitro* experiments

Polymerase chain reaction (PCRs) were performed in 15 µl reaction volumes containing 2x Taq DNA Polymerase mix (VWR International GmbH, BDH Prolabo, Darmstadt, Germany), 0.17 µg/µl BSA (New England Biolabs, Ipswich, MA, USA), 0.27 µM each of forward and reverse primer, and 10-15 ng template DNA. The amplification protocol started with 95°C for 3 min followed by 14 cycles at 94°C for 30 s, a touchdown step for 25 s (see Table S7 for specific starting temperatures) and 72°C for 75 s. This was followed by 25 cycles of 94°C for 30 s, a specific annealing temperature for 25 s (Table S7) and 72°C for 75 s. Final elongation was conducted for 10 min at 72°C.



A



B

Figure S1: Evidence for X/Y gene conversion events (yellow stars) in bears.
A Rooted phylogeny of ~397 bp from *ZFY* and *ZFX* sequences in bears and other mammals.
B Statistical haplotype network of bear *ZFY* and *ZFX* exon sequences (bear sequences from ***A***).

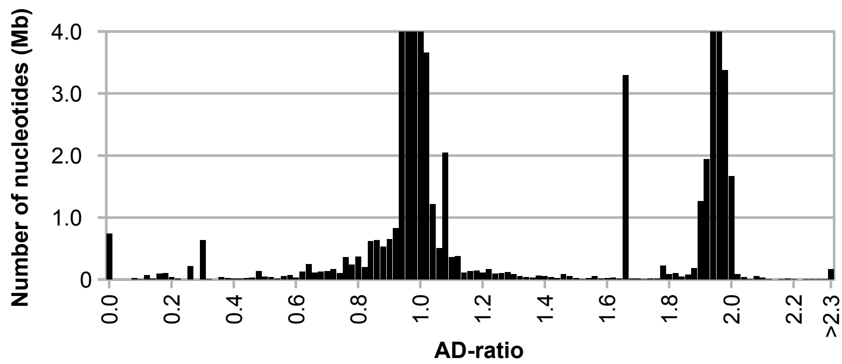


Figure S2: Average-depth (AD) ratio for scaffolds in the polar bear assembly (enlarged version of Fig. 2A).

Figure S3: Annotations of Y-linked scaffolds.

Exons homologous to mouse and human are shown in red. Previously published Y-linked polar bear sequences are shown in pink. The repeat unit of each microsatellite is indicated and regions with >200 bp of consecutive "N" are highlighted in gray. Due to the high abundance, only placental mammalian non-LTR retrotransposons ≥500 bp (LINEs) and ≥100 bp (SINEs) were plotted. The maps of additional scaffolds are shown in Figure 3.

Table S1: Gene sequences used in the similarity search. Previously published polar bear sequences are in bold. Asterisk indicates the 18 mammalian genes that were found to have identity ≥80% to scaffolds of the polar bear assembly (see Table 1).

No.	Gene	Sequence class in humans (according to Skaletsky et al. 2003)	Available sequences*	Fragment extracted	Species	GenBank accession number	Scaffold identified	Number of gene fragments recovered on respective scaffold
Y-linked								
1*	*AMELY*		Exon 5	exon	***Ursus maritimus***	AM941064.1	115	1 exon
	AMELY	X-degenerate	RefSeqGene	exon	*Homo sapiens*	NG_008011.1	389	2 exons
2	*BPY2*	Ampliconic	RefSeqGene	exon	*Homo sapiens*	NG_009862.1	-	-
3	*CDY1*	Ampliconic	RefSeqGene	exon	*Homo sapiens*	NG_011754.1	-	-
4*	*DDX3Y (DBY)*	X-degenerate	Intron 5 + partial CDS	exon	*Pan troglodytes*	JF293113.1	309	1 exon
	DDX3Y (DBY)		RefSeqGene	exon	*Homo sapiens*	NG_012831.1	184, 309	12 exons, 17 exons
	DDX3Y (DBY)		RefSeqGene	exon	*Mus musculus*	NM_012008.2	184, 309	9 exons, 15 exons
5*	*EIF1AY*	X-degenerate	mRNA	exon	*Homo sapiens*	NM_004681.2	20, 297	3 exons, 5 exons
	EIF1AY		Partial exon	exon	*Pan troglodytes*	AB176583.1	-	-
6*	*EIF2S3Y*		mRNA	exon	*Mus musculus*	NM_012011.1	20, 318, 646	8 exons, 6 exons, 2 exons
7	*HSFY1*	Ampliconic	RefSeqGene	exon	*Homo sapiens*	NG_012030.1	-	-
8*	*KDM5D (SMCY)*	X-degenerate	RefSeqGene	exon	*Homo sapiens*	NG_032920.1	297, 301, 318	12 exons, 9 exons, 8 exons
	KDM5D (SMCY)		mRNA	exon	*Mus musculus*	NM_011419.3	297, 301, 318	8 exons, 4 exons, 7 exons
	KDM5D (SMCY)		intron 4	complete (intron)	***Ursus maritimus***	AB261824.1	318	1 intron
9*	*NLGN4Y*	X-degenerate	RefSeqGene	exon	*Homo sapiens*	NG_028212.1	104	4 exons
10*	*PCDH11Y*	X-transposed	RefSeqGene	exon	*Homo sapiens*	NG_011652.1	186	3 exons
11*	*PRKY*	X-degenerate	mRNA	exon	*Homo sapiens*	Y15801.1	104	4 exons
12	*PRY*	Ampliconic	RefSeqGene	exon	*Homo sapiens*	NG_032916.1	-	-
13	*RBM31Y*		mRNA	exon	*Mus musculus*	NM_028970.1	-	-
14	*RBMY*		mRNA	exon	*Mus musculus*	NM_011253.2	-	-
15*	*RBMY1A1*	Ampliconic	RefSeqGene	exon	*Homo sapiens*	NG_012805.1	105, 369	1 exon, 6 exons
16*	*RPS4Y*		Partial CDS	exon	*Pan troglodytes*	AH012491.2	13, 253	1 exon, 6 exons
17*	*RPS4Y2*	X-degenerate	RefSeqGene	exon	*Homo sapiens*	NG_032924.1	253	5 exons
18*	*SLY*		mRNA	exon	*Mus musculus*	NM_201530.2	53	1 exon
19*	*SRY*		Exon 1	complete (exon +UTR)	***Ursus maritimus***	AM748305.1	579	1 exon + UTR
	SRY	X-degenerate	RefSeqGene	exon	*Homo sapiens*	NG_011751.1	-	-
	SRY		mRNA	exon	*Mus musculus*	NM_011564.1	-	-
20	*SSTY1*		mRNA	exon	*Mus musculus*	NM_009220.2	-	-
21	*SSTY2*		mRNA	exon	*Mus musculus*	NM_023546.3	-	-
22*	*TBL1Y*	X-degenerate	mRNA	exon	*Homo sapiens*	NM_033284.1	46, 104	1 exon, 12 exons
23	*TGIF2LY*	X-transposed	mRNA	exon	*Homo sapiens*	NM_139214.2	-	-
24	*TMSB4Y*	X-degenerate	mRNA	exon	*Homo sapiens*	NM_004202.2	-	-
25	*TSPY1*	Ampliconic	RefSeqGene	exon	*Homo sapiens*	NG_027958.1	-	-
26*	*UBA1Y*		mRNA	exon	*Mus musculus*	NM_011667.2	134, 403, 605, 4889	12 exons, 1 exon, 13 exons, 1 exon
	UBA1Y		Exon 18	complete (exon)	***Ursus maritimus***	AM748329.1	605	1 exon
27*	*USP9Y*	X-degenerate	RefSeqGene	exon	*Homo sapiens*	NG_008311.1	20, 184, 309, 318	1 exon, 34 exons, 38 exons, 1 exon
	USP9Y		mRNA	exon	*Mus musculus*	NM_148943.2	20, 184, 309, 318	1 exon, 29 exons, 35 exons, 1 exon

28*	*UTY*		mRNA	exon	*Mus musculus*	NM_009484.2	134, 309	19 exons, 21 exons
29	*VCY*	Ampliconic	RefSeqGene	exon	*Homo sapiens*	NG_032915.1	-	-
30	*XKRY*	Ampliconic	RefSeqGene	exon	*Homo sapiens*	NG_032917.1	-	-
31*	*ZFY*		Final intron	complete (intron)	***Ursus maritimus***	AB261808.1	318	1 intron
	ZFY		partial gene	complete (partial gene)	***Ursus maritimus***	AM748297.1	20, 318	1 exon, 1 exon
	ZFY		Partial CDS	exon	*Canis lupus*	JX475923.1	20, 318	3 exons, 4 exons
	ZFY	X-degenerate	RefSeqGene	exon	*Homo sapiens*	NG_008113.1	20, 318	5 exons, 6 exons
32	*ZFY1*		mRNA	exon	*Mus musculus*	NM_009570.4	-	-
X-linked								
1	*AMELX*		Exon 5	exon	***Ursus maritimus***	AM941056.1	115, 6612	1 exon, 1 exon
2	*ZFX*		Final intron	complete (intron)	***Ursus maritimus***	AB261816.1	20	1 intron
3	*ZFX*		partial gene	complete (partial gene)	***Ursus maritimus***	AM941048.1	20, 318, 3838	1 exon, 1 exon, 1 exon

Table S2: Y-linked scaffolds <10 kb (n=92) identified by their AD-ratio.

Scaffold ID	Size (kb)	AD-ratio
882	9	0
949	6	0
951	6	0
955	6	0.085
983	5	0
1057	5	0
1166	4	0
1214	3	0
1229	3	0.3
1232	3	0
1235	3	0
1237	3	0
1251	3	0
1261	3	0
1273	3	0
1329	3	0
1346	3	0
1359	3	0
1484	3	0
1621	2	0
1629	2	0
1665	2	0
1687	2	0
1727	2	0
1741	2	0
1756	2	0
1760	2	0
1772	2	0
1850	2	0
1880	2	0
1885	2	0
1911	2	0
1934	2	0
1956	2	0
2053	2	0

Scaffold ID	Size (kb)	AD-Ratio
2560	1	0
2698	1	0
2697	1	0
2720	1	0
2812	1	0
2853	1	0.136
2856	1	0
2882	1	0
2879	1	0
3005	1	0
3008	1	0.225
3026	1	0
3027	1	0
3043	1	0
3094	1	0
3105	1	0
3158	1	0
3258	1	0
3283	1	0.299
3349	1	0
3381	1	0.271
3481	1	0
3487	1	0
3578	1	0
3585	1	0
3598	1	0.196
3608	1	0
3649	1	0
3662	1	0.163
3691	1	0
3693	1	0
3723	1	0
3838	1	0
3849	1	0
3886	1	0

2099	2	0	3889	1	0
2138	2	0	3894	1	0
2227	2	0	3958	1	0.184
2256	2	0	3965	1	0
2285	2	0	4014	1	0
2428	2	0	4021	1	0.182
2434	2	0	4107	1	0.172
2433	2	0	4138	1	0
2467	1	0	4147	1	0
2483	1	0	4146	1	0
2502	1	0	4157	1	0

Table S3: Scaffolds identified by the similarity search to be Y-linked, but with AD-ratios indicative of being autosomal or X-linked.

Scaffold ID	Size [Kbp]	AD-ratio	Inferred chromosomal location	Homology search
13	26,707	0.962	A	*RPS4Y*
20	22,125	1.942	X	*EIF1AY, EIF2S3Y, USP9Y, ZFY*
46	15,941	0.978	A	*TBL1Y*
53	14,458	0.973	A	*SLY*
104	6,801	0.969	A	*NLGN4Y, PRKY, TBL1Y*
105	6,717	1.944	X	*RBMY1A1*
115	5,608	1.934	X	*AMELY*
134	4,672	1.943	X	*UBA1Y, UTY*
184	2,589	1.949	X	*DDX3Y, USP9Y*
186	2,578	1.956	X	*PCDH11Y*
253	821	1.933	X	*RPS4Y, RPS4Y2*
301	351	1.928	X	*KDM5D*
4889 *	0.9	-	-	*UBA1Y*
6612 **	0.7	0	-	*AMELY*

* Reads mapped to scaffold *4889* with quality <20, so this AD-ratio could not be determined.

** Scaffold 6612 (749 bp) has an AD-ratio of zero but was filtered out due to its size <1kb.

Table S4: Scaffold combinations containing Y- and X-linked scaffolds
Details from the 10 query genes that were observed on ≥2 scaffolds are shown (see also Table S1).

Scaffold with AD-ratio indicative of Y-linkage and male-specific amplification in-vitro	Scaffold with putative X-linkage (AD-ratio in brackets)	Query Gene
297	*20* (1.94)	*EIF1AY*
297	*301* (1.94)	*KDM5D*
309	*134* (1.94)	*UTY*
309	*184* (1.95)	*DDX3Y*
309	*20* (1.94), *184* (1.95)	*USP9Y*
318	*20* (1.94)	*EIF2S3Y, ZFY*
318	*301* (1.94)	*KDM5D*
318	*20* (1.94),	*USP9Y*

	184 (1.95)	
369	*105* (1.94)	*RBMY1A1*
389	*115* (1.93)	*AMELY*
403	*134* (1.94)	*UBA1Y*
605	*134* (1.94)	*UBA1Y*
646	*20* (1.94)	*EIF2S3Y*
6612	*115* (1.93)	*AMELY*

Table S5: Percentage of transposable elements in 1.6 Mb Y-chromosomal scaffolds in comparison to the genome-wide average in polar and brown bears.

	Polar bear		**Brown bear**
	Y-scaffolds	**Genome**	**Genome**
SINE	8.23	8.18	8.15
MIRs	0.42	3.03	3.05
CanSINE	7.81	5.05	5.10
LINE	38.88	21.03	21.09
LINE-1	38.42	16.93	16.96
LINE-2	0.44	3.58	3.60
LINE-3	0.03	0.39	0.39
RTE	0.00	0.12	0.13
ERV	5.29	5.40	5.45
DNA transposons	1.98	3.00	3.03
Total	54.38	39.20	39.27

Table S6: Sequence characteristics of microsatellites with ≥ 15 repeat units on Y-linked scaffolds. Bases in lowercase denote regions with imperfect tandem repeats.

Scaffold	**Repeat unit**	**Number of repeat units**	**Start position on scaffold**	**Repeat sequence (obtained from Repeatmasker)**
scaffold297	G	21	23088	GGGGGGGGGGGGGGGGGGaGG
scaffold297	AG	23	27644	AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGacAG AGAG
scaffold297	T	29	51610	TTTTTTTTTTTTTTTTTTTTTTTTTTTTT
scaffold297	AAAT	15	52281	AAATaaagaaAAATatAAATAAATAAATAAATAAATAAATAAATAAATA AATAAATAAATAAATAAAT
scaffold297	CTCCCTTC	49	55269	CTCCCTTCCTCCCTTCCTCCCTTCCTCCCTTCCTCCCTTCCTCCCT TCCTCCCTTCCTCCCTTCCTCCCTTCCTCCCTTCCTCCCTTCCTCC CTTCCTCCCTTCCTCCCTTCCTCCCTTCCTCCCTTCCTCCCTTCCT CCCTTCCTCCCTTCCTCCCTTCCTCCCTTCCTCCCTTCCTCCCTTC CTCCCTTCCTCCCTTCCTCCCTTCCTCCCTTCCTCCCTTCCTCCCT TCCTCCCTTCCTCCCTTCCTCCCTTCCTCCCTTCCTCCCTTCCTCC CTTCCTCCCTTCCTCCCTTCCTCCCTTCCTCCCTTCCTCCCTTCCT CCCTTCCTCCCTTCCTCCCTTCCTCCCTTCCTCCCTTCCTCCCTTC CTCCCTTCCTCCCTTCCTCCCTTCcttccttccttccttcttctttctctccttcccttcctc cctccctccctccctccct
scaffold297	GA	21	72854	GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAAaGAGAGAGA
scaffold297	AGAA	16	94667	AGAAAGAAaaAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGA AAGAAAGAAAGAAAGAAAGAAa
scaffold297	AC	23	192322	ACACaaACACACACACACACACACACACACACACACACACA CAC
scaffold297	T	23	193512	TTTTTTTTTTTTTTTTTTTTTTT
scaffold297	CT	18	241447	CTCTCTCTCTCTcaCTCTCTCTCTCTCTCTCTCTCTc

scaffold297	T	17	254442	TTTTTTTTTTTTTTTTT
scaffold297	TTTC	24	257890	TTTCTTTCTTTCTTTCTTTCTTTCccTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCtt
scaffold297	CA	19	259587	CACACACACACACACACACACACACACACACACAcCACA
scaffold297	T	46	283813	TTTTcTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT
scaffold297	A	27	343300	AAAAgAAAAAAAAAAAAAAAAAAAAAA
scaffold297	A	23	344483	AAAAAAAAAAAAAAAAAAAAAAA
scaffold297	A	21	379383	AAAAAAAAAAAAAAAAAAAAA
scaffold297	T	17	388460	TTTTTTTTTTTTTTTTT
scaffold309	AC	19	23799	ACACACACACACACACACACACACACACACACACACACa
scaffold309	A	24	62764	AAAAAAAAAAAAAAAAAAAAAAAA
scaffold309	ATAG	22	64857	ATAGATAGATAGATAGATAGgtagATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAG
scaffold309	T	20	79617	TTTTTTTTTTTTTTTTTTTT
scaffold309	A	21	95327	AAAAAAAAAAAAAAAAAAgAA
scaffold309	T	19	109061	TTTTTTTTTTTTTTTTTTT
scaffold309	A	17	111760	AAAAAAAAAAAAAAAAA
scaffold309	T	18	116040	TTTTTTTTTTTTTTTTTT
scaffold309	A	19	127694	AAAAAAAAAAAAAAAAAAA
scaffold309	C	18	145706	CCCCCCCCCCCCCCCCCC
scaffold309	T	17	155868	TTTTTTTTTTTTTTTTT
scaffold309	A	18	159378	AAAAAAAAAAAAAAAAAA
scaffold309	T	21	185488	TTTTTTTTTTTTTTTTTTTTT
scaffold309	A	20	195689	AAAAAAAAAAAAAAAAAgAA
scaffold309	A	19	197350	AAAAAAAAAAAAAAAAAAA
scaffold309	T	22	206595	TTTTTTTTTTTTTTTTTTTTTT
scaffold309	A	29	209969	AAAAAAAAAAAAAAAAAAAAAAAAAAAAA
scaffold309	T	24	212205	TTTaTTTTTTTTTTTTTTTTTTTTT
scaffold309	TG	22	285274	TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG
scaffold309	TG	21	304646	TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGt
scaffold309	T	19	307203	TTTTTTTTTTTTTTTTTTT
scaffold318	T	17	83831	TTTTTTTTTTTTTTTTT
scaffold318	A	19	110623	AAAAAAAAAAAAAAAAAAA
scaffold318	T	17	116906	TTTTTTTTTTTTTTTTT
scaffold318	A	17	127184	AAAAAAAAAAAAAAAAA
scaffold318	TG	24	201228	TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGt
scaffold318	AC	20	221123	ACACatACACACACACACACACACACACACACACACACACa
scaffold318	TA	15	231655	TATATATATATATATATAcaTATATATATA
scaffold322	T	23	6778	TTTTTTTTTTTTTTTTTTTTTTT
scaffold322	AGAA	18	13277	AGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAa
scaffold322	T	17	39419	TTTTTTTTTTTTTTTTT
scaffold322	C	19	44156	CCCCCCCCCCCCCCCCCCC
scaffold322	A	40	53605	AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAgAAA
scaffold322	GAA	204	104231	GAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGA AGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAG AAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAA GAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGA AGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAG AAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAA GAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGA AGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAG AAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAA GAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGA AGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAG AAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAA GAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGA AGAAGAAGAAGAAGAAGAAggaGAAGAAGAAGAAGAAaaaGAAaaa GAAg
scaffold322	TG	17	180871	TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGttTGTGt
scaffold322	AC	17	197380	ACACACACACACACACACACACACACACACACACa

scaffold322	CA	20	212493	CACACACACACACACACACACAaaCACACACACACACACA
scaffold322	AC	19	213052	ACACACACACACACACACACACACACACACACACACACa
scaffold322	CA	16	213587	CACAaaCACACACACACACACACACACACACA
scaffold322	CA	20	214622	CACACACACACACACACACACACACACACACACACACACAc
scaffold369	GT	23	1696	GTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT
scaffold369	CA	25	63947	CACACACACACACACACACACACACACACACACACACACACActCAtaCACACACAc
scaffold369	AC	19	64158	ACACACACACACACACACACACACACACtcACaACACACa
scaffold369	AC	23	64880	ACACACACACACACACACACACACACACACACACACACACACACACa
scaffold369	CA	17	65986	CACACACACACACACACACACACACACACACA
scaffold369	CA	26	66252	CACACACACACACACACACACACACAgaCACACACACACACACACACACA
scaffold369	AC	20	67968	ACACaaACACaACACACACACACACACACACACACACACAC
scaffold369	CA	20	71098	CACACACACACACACACACACACACACACACACACActCACAc
scaffold369	CA	18	72002	CACACACACACACACACACACACACACACACACACA
scaffold369	AC	18	72937	ACACACACACACACACACACACACACACACACACAC
scaffold369	CA	21	73479	CACACAcgCACACACACACACACACACACACACACACACACA
scaffold369	CA	17	75832	CACACACACACACACACACACACACACACACACAc
scaffold369	AC	16	76906	ACACgcACACACACACACACACACACACACAC
scaffold369	AC	19	78941	ACACACACACACACACACACACACACACACACACACACa
scaffold369	AC	18	80402	ACACACACACatACACACACACACACACACaACACACa
scaffold369	AC	16	81716	ACACACACACACACACACACACACACACACAC
scaffold369	AC	22	83146	ACACACACACACACACACatACACACACACACACACACACACa
scaffold369	CA	24	87519	CACACACACACACACACACACACACACACACACACACACACACACACAc
scaffold369	CA	17	88953	CACACACAgaCACACACACACACACACACACACAc
scaffold369	CA	21	90054	CACACACACACACACACACACACACACACACACAaCACACACAc
scaffold369	CA	17	100413	CACACACACAtaCACACACACACACACACACACA
scaffold369	AC	17	100984	ACACACACACACACACACACACACACACACACACa
scaffold369	AC	14	101708	ACACACACACACACACACagACACACAC
scaffold369	AC	18	103579	ACACACACACACACACACACACACACACgACACACACa
scaffold389	AC	16	7087	ACACACACACACACACACACACACACACACACa
scaffold389	TG	20	12691	TGTGTGTGtcTGTGTGTGTGtaTGTGTGTGTGTGTGTGTGTG
scaffold389	TG	20	22172	TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGt
scaffold389	TG	21	26235	TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGt
scaffold389	TG	20	52851	TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGt
scaffold389	AC	17	73174	ACACACACACACACACACACACACACACACACAC
scaffold389	TG	18	74158	TGTGTGTGgcTGTGTGTGTGTGTGTGTGTGTGTGTGt
scaffold389	TG	19	76543	TGTGTGtagTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG
scaffold393	AC	20	1822	ACACACACACACACACACACACACACACACACACACACAC
scaffold393	GT	15	13861	GTGTGTGTGTGTGTGTGTGTGTGTGTGTGT
scaffold393	CT	16	22325	CTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTc
scaffold393	GT	18	48001	GTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTg
scaffold393	CT	14	54007	CTCTttCTCTCTCTCTCTCTCTCTCTCT
scaffold393	CA	24	54039	CACACACACACACACACACACACACACACACACACAtaCACACACACACACA
scaffold393	GT	16	61378	GTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTg
scaffold403	GT	17	16279	GTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTg
scaffold403	GT	19	26610	GTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT
scaffold403	AC	15	44337	ACACACACACACACACACACACACACACACa
scaffold403	AC	15	51140	ACACACACACACACACACACACACACACACa
scaffold420	TG	16	7834	TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGt
scaffold420	CAA	15	29461	CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAA
scaffold519	AAGG	23	1420	AAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGaaaaaaagAAGGAAGGAAGGAAGG

scaffold579	T	26	10424	TTgTTTTTTTTTTTTTTTTTTTTTTTT
scaffold579	T	33	11001	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT
scaffold579	AAGA	35	13351	AAGAaggaAAGAaggaAAGAAAGAAAGAaAAGAaaaAAGAAAGAAAG AagaAAGAAAGAAAGAAAGAAAGAAAGAAAGAagaAAGAAAG AAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGA AAGAAAGAAAGAtaaAAGAAAGAaa
scaffold579	A	17	14764	AAAAAAAAAAAAAAAAA
scaffold579	TC	17	16925	TCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC
scaffold657	AC	24	6191	ACACaaACACACACACACACACACACACACACACACACACA CACACa

Table S7. Primer sequences and PCR conditions of fragments used for in-vitro validation of Y-scaffolds.

Scaffold	Forward primer (5´->3´)	Reverse primer (5´->3´)	Fragment size (bp) incl. primers	T_{A1} (°C) 14 cycles	T_{A2} (°C) 25 cycles
297	ACAGAGAGATGCGGTCCTG	GGAGACTCTGCTGACCAAGG	783	69	62
309	TCAGCCTTGCTTCATTGTTC	ATTGTGTGCCTGTTCCATTC	696	68	61
318.20	CAATGCTAGAGATGCAAGTGG	AAGCATATGAAGACACACAGTGG	757	67	60
318.21	ATGCTATGATCCAGCAATCC	TCCTTGTTAATTCTTCTCTGATGC	800	67	60
318.22	CTGACATACGTGGCTCACAG	ATGTGGAGGTTCAGGAGGAC	800	67	60
318.23	GTGGTCAGATCCAGCTCCTC	GACCTTGCTTCTGCCTTCTC	786	67	60
*322***	GAGTAGAGCTGGTGCTTGTGAG	GAAGCAGAGCTCAAGTCTGAAG	687	68	61
369	CATTGAAACAAGGGCACATC	TCCTCTAGGAACCCAACTGC	780	68	61
393	AATTGTGAAGGATGAGGAAGG	CCAGAGAAGCAGAGGATGG	788	70	63
*389***	ACCCACTGCTGTTCTGTATCC	CCAACAGTGTAGTGGTTGTGC	679	68	61
*403***	CACTCAGGAGAGCACAGGTC	TGTGTGTCGTAAGCAGAGGTC	796	68	61
420	TTCAGAGGGAAGGGTGTAGG	AATGGTAGTGAGGGAGGTATGG	635	69	62
519	CCCAAGTTAGGGAAGTTTGC	TGTTTCAGCTTCAGCTGTCC	734	68	61
*579**	CTGCAGGCCTGTCAATGTTA	TGTGTATCGACCCCATACTTTG	660	66	61
605	TTTGACGGCTGAGCAATATC	GATGCAGCCACAATGAAATC	706	68	61
646	TGTCTAGCCATCTGGTCCTG	CATGGACATTGTTGCATTGA	778	69	62
657	TCCAGGTTGTCAAGCACATG	CTTGCTCCATCCACATGCTG	769	69	62
596	ATCACCCTCTCCACTCACAA	CCACCTCTTTGACTTCCTGG	735	69	62
613	TTGTGGACATTGCTGCTTCT	TCTGAACATAGGCTGCAACC	674	69	62
632	CTATCGTGGACATTGCTGCT	CTATCGTGGACATTGCTGCT	621	69	62
771	AGAACTTCCTGTACCTGGCT	GCTGATTCCGTGGATGTTCA	753	69	62
795	GGAACACGACAAGGATCTCC	TTGTGCCAATACCATGCTGT	705	69	62
813	ATCTTGCTGCTCCTCTGTTG	CCACGTCACATCATACTGCA	767	69	62
236a	GACTCAAGCTCAGCGTTCAC	GTCCAGTGGAAGAGGTAGGC	760	68	61
267a	ATACATACGCGCACACACAC	GCACATACCTCGTCAAGGAC	655	68	61
253x	CTGGAGGTCTGTGTGTGGAG	AACCTAATGCGTGTGTCTGC	796	68	61
301x	TGGATAAGGCATTCTGCAAG	CTGAAGGCAGTGGTGAAGAG	717	68	61

* Primers from (Bidon et al. 2014) (fragment *579.1B*)
** Primers from (Kutschera et al. 2014)
a Autosomal scaffold based on AD-ratio
x X-linked scaffold based on AD-ratio

Table S8. Accession numbers and sample origin of polar, brown and black bear genomes

Species	Sample origin	Number of individuals	Accession numbers
Ursus maritimus	Spitsbergen, Svalbard	9	SRX155945, SRX155949, SRX155951, SRX155953, SRX155954, SRX155955, SRX155957, SRX155960, SRX155961
	Alaska	3	SRX156102, SRX156103, SRX156105
Ursus arctos	ABC-Islands	1	SRX156108
Ursus americanus	Alaska	1	SRX156137