Chromosome Res (2023) 31:13
https://doi.org/10.1007/s10577-023-09722-y

RESEARCH

Chromosome-length genome assemblies and cytogenomic analyses of pangolins reveal remarkable chromosome counts and plasticity

Marlys L. Houck · Klaus-Peter Koepfli · Taylor Hains · Ruqayya Khan · Suellen J. Charter · Julie A. Fronczek · Ann C. Misuraca · Sergei Kliver · Polina L. Perelman · Violetta Beklemisheva · Alexander Graphodatsky · Shu-Jin Luo · Stephen J. O'Brien · Norman T.-L. Lim · Jason S. C. Chin · Vanessa Guerra · Gaik Tamazian · Arina Omer · David Weisz · Kenneth Kaemmerer · Ginger Sturgeon · Joseph Gaspard · Alicia Hahn · Mark McDonough · Isabel Garcia-Treviño · Jordan Gentry · Rob L. Coke · Jan E. Janecka · Ryan J. Harrigan · Jen Tinsman · Thomas B. Smith · Erez Lieberman Aiden · Olga Dudchenko

Received: 24 October 2022 / Revised: 27 February 2023 / Accepted: 4 March 2023

Abstract We report the first chromosome-length genome assemblies for three species in the mammalian order Pholidota: the white-bellied, Chinese, and Sunda pangolins. Surprisingly, we observe extraordinary karyotypic plasticity within this order and, in female white-bellied pangolins, the largest number of chromosomes reported in a Laurasiatherian mammal: $2n = 114$. We perform the first karyotype analysis of an African pangolin and report a Y-autosome fusion in white-bellied pangolins, resulting in $2n = 113$ for males. We employ a novel strategy to confirm the fusion and identify the autosome involved by finding the pseudoautosomal region (PAR) in the female genome assembly and analyzing the 3D contact frequency between PAR sequences and the rest of the genome in male and female white-bellied pangolins. Analyses of genetic variability show that white-bellied pangolins have intermediate levels of genome-wide heterozygosity relative to Chinese and Sunda pangolins, consistent with two moderate declines of historical effective population size. Our results reveal a remarkable feature of pangolin genome biology and highlight the need for further studies of these unique and endangered mammals.

Responsible Editor: Rachel O'Neill

M.L. Houck and K.-P. Koepfli are co-first authors.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10577-023-09722-y.

M. L. Houck (✉) · S. J. Charter · J. A. Fronczek · A. C. Misuraca
Conservation Science and Wildlife Health, San Diego Zoo Wildlife Alliance, Escondido, CA 92027, USA
e-mail: mhouck@sdzwa.org

K.-P. Koepfli
Smithsonian-Mason School of Conservation, George Mason University, Front Royal, VA 22630, USA

K.-P. Koepfli
Center for Species Survival, Smithsonian's National Zoo and Conservation Biology Institute, Front Royal, VA 22630, USA

K.-P. Koepfli (✉)
Computer Technologies Laboratory, ITMO University, 197101 St. Petersburg, Russia
e-mail: kkoepfli@gmu.edu

T. Hains
Committee On Evolutionary Biology, University of Chicago, Chicago, IL 60637, USA

R. Khan · A. Omer · D. Weisz · E. L. Aiden · O. Dudchenko (✉)
The Center for Genome Architecture, Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA
e-mail: Olga.Dudchenko@bcm.edu

Published online: 12 April 2023

Keywords chromosome number · karyotype · Pholidota · *Phataginus tricuspis* · genome assembly

Introduction

The linear chromosomes found in the cells of eukaryotic taxa are characterized by a vast diversity of morphologies (defined by the position of centromeres), sizes, and numbers. Changes in chromosome number among taxa have been used to infer mechanisms of speciation (King 1993), determine phylogenetic relationships (Robinson et al. 2008; Nie et al. 2012), reconstruct ancestral karyotypes (Graphodatsky et al. 2011; Deakin and Ezaz 2014), and discern mechanisms of genome evolution and function (Bernardi 2015; Mayrose and Lysak 2021). Moreover, the highly organized way that genomes are packaged into chromosomes within the cell nucleus and the distribution of loci (synteny) among different chromosomes is likely associated with transcriptional regulation of genes (Sexton and Cavalli 2015).

Differences in chromosome number and structure have been characterized in many species of

S. Kliver
Center for Evolutionary Hologenomics, The Globe Institute, The University of Copenhagen, 5A, Oester Farimagsgade, 1353 Copenhagen, Denmark

P. L. Perelman · V. Beklemisheva · A. Graphodatsky
Department of the Diversity and Evolution of Genomes, Institute of Molecular and Cellular Biology SB RAS, 630090 Novosibirsk, Russia

S.-J. Luo
The State Key Laboratory of Protein and Plant Gene Research, Peking-Tsinghua Center for Life Sciences (CLS), School of Life Sciences, Peking University, Beijing 100871, China

S. J. O'Brien
Laboratory of Genomic Diversity, Computer Technologies Laboratory, ITMO University, 197101 St. Petersburg, Russia

S. J. O'Brien
Guy Harvey Oceanographic Center, Halmos College of Arts and Sciences, Nova Southeastern University, Fort Lauderdale, FL 33004, USA

N. T.-L. Lim
Natural Sciences and Science Education, National Institute of Education, Nanyang Technological University, Singapore 637616, Singapore

J. S. C. Chin
Taipei Zoo, No. 30 Sec. 2 Xinguang Rd., Taipei 11656, Taiwan

V. Guerra
Department of Biological Sciences, Simon Fraser University, Burnaby, BC, Canada

V. Guerra
Department of Invertebrate Zoology, National Museum of Natural History, Smithsonian Institution, Washington, DC, USA

G. Tamazian
Centre for Computational Biology, Peter the Great Saint Petersburg Polytechnic University, St. Petersburg 195251, Russia

K. Kaemmerer · G. Sturgeon · J. Gaspard · A. Hahn · M. McDonough
Pittsburgh Zoo & Aquarium, PA 15206 Pittsburgh, USA

I. Garcia-Treviño · J. Gentry · R. L. Coke
Center for Conservation and Research, San Antonio Zoo, San Antonio, TX 78212, USA

J. E. Janecka
Department of Biological Sciences, Bayer School of Natural and Environmental Sciences, Duquesne University, Pittsburgh, PA 15282, USA

R. J. Harrigan · J. Tinsman · T. B. Smith
Center for Tropical Research, Institute of the Environment and Sustainability, University of California, Los Angeles, CA 90095, USA

T. B. Smith
Department of Ecology and Evolutionary Biology, University of California, Los Angeles, CA 90095, USA

E. L. Aiden
Departments of Computer Science and Computational and Applied Mathematics, Rice University, Houston, TX 77030, USA

E. L. Aiden · O. Dudchenko
Center for Theoretical and Biological Physics, Rice University, Houston, TX 77030, USA

E. L. Aiden
Broad Institute of Harvard and Massachusetts Institute of Technology (MIT), Cambridge, MA 02139, USA

mammals using traditional comparative cytogenetic methods such as chromosome banding and painting (Graphodatsky et al. 2011) and more recently, through the application of high-throughput chromosomal conformation capture (Hi-C) using proximity ligation sequencing (Lieberman-Aiden et al. 2009) coupled with whole genome sequencing technologies (Burton et al. 2013; Dudchenko et al. 2017). These efforts have led to the development of compendiums that summarize and synthesize the evolution of mammalian chromosomes (Graphodatsky et al. 2020; DNAZoo.org). To date, the karyotypes of more than 2000 species of mammals have been characterized, representing all extant supraordinal and ordinal clades. However, this represents only about one-third of the 6399 species of extant mammals (Burgin et al. 2018). While chromosome number and evolution have been well characterized in some clades of mammals (e.g., Carnivora, Nie et al. 2012; Beklemisheva et al. 2020), the cytogenomics of other clades remains largely understudied.

One of the least studied orders is the Pholidota, which includes the pangolins, or scaly anteaters. Pangolins belong to the monotypic family Manidae, with eight extant species distributed in Asia and Africa. Four Asian species are included in the genus *Manis* (*M. crassicaudata*, *M. culionensis*, *M. javanica*, *M. pentadactyla*; Maninae) while the four African species are divided into two genera representing the ground pangolins (*Smutsia gigantea*, *S. temminckii*; Smutsiinae) and tree pangolins (*Phataginus tetradactyla*, *P. tricuspis*; Phataginae) (Gaubert et al. 2018). Pangolin species are listed as vulnerable, endangered, or critically endangered on the IUCN Red List of Threatened Species and are considered the most trafficked mammals in the world (Heinrich et al. 2016). Populations are decreasing rapidly due to habitat loss, overexploitation for bushmeat, and high demand for the large protective scales that cover their body and are used in traditional Asian medicine (Heinrich et al. 2016; Choo et al. 2022).

Cytogenetic studies of pangolins have been limited to three of the four species found in Asia. Synthesis of past studies and modern cytogenetic techniques indicate that Asian pangolin species have four diploid numbers: $2n=36$ (*M. crassicaudata* from northeast and south India), $2n=38$ (*M. javanica*, Yunnan Province, China), $2n=40$ (a single male *M. pentadactyla*, Taiwan) (Makino and Tateishi 1951), and 42 (*M. pentadactyla*, Taiwan and China) (Yang et al. 2006; Nie et al. 2009). These chromosome numbers fall within the typical range of $2n=36$–60 found in most mammalian species that have been characterized thus far (Graphodatsky et al. 2020).

None of the four pangolin species that occur in Africa has been previously karyotyped. Given the results obtained from Asian pangolin species, it seems reasonable to expect that chromosome numbers should be relatively similar, despite the ~ 38 million year divergence time separating the clades of African and Asian pangolins (Gaubert et al. 2018). To explore the cytogenomics of African pangolins, we generated the first genome assembly and karyotypes from a species within the group, the white-bellied pangolin (*P. tricuspis*). Our results revealed that this species carries one of the highest numbers of chromosomes among mammals, demonstrating a hitherto unknown karyotypic plasticity within the Pholidota.

Materials and methods

Cell cultures and chromosome preparations

We performed classical cytogenetic analysis and karyotyping on two male and one female *P. tricuspis*; all three animals were USFWS confiscations thought to have originated in Cameroon. Chromosomes from the female (Laboratory number #16406) were obtained from blood lymphocytes. Short-term lymphocyte culture, harvesting, and banding followed Kumamoto et al. (1996). Fibroblast cell lines were established from the two males using post mortem samples (skin, trachea and testes) and using a collagenase disaggregation technique previously described by Houck et al. (1994). The fibroblast cell lines were archived at the San Diego Zoo Wildlife Alliance's Frozen Zoo® (Lab #16213 and #16405). Metaphase chromosomes of all three animals were examined by non-differential staining using Giemsa and also C-banding. Male #16213 was further analyzed with G-banding. Karyotyping was done using the CytoVision Genus® system by Leica Microsystems (Wetzlar, Germany). C-banding of Sunda and Malayan pangolin chromosomes followed Sumner (1972). The nucleolar organizing region was localized by FISH with ribosome probes (Maden et al. 1987, Yang and Graphodatsky 2009).

Chromosome-length genome assembly of the white-bellied pangolin

We opportunistically collected 3 ml of whole blood into an EDTA blood tube from a female white-bellied pangolin ("Jaziri") maintained at the Pittsburgh Zoo & Aquarium, Pittsburgh, PA, USA, during a routine veterinary examination. The Pittsburgh Zoo & Aquarium is one of the seven members that comprise the Pangolin Consortium (https://pangolinconsortium.org/), which aims to establish ex situ conservation programs for pangolins in US zoos. Jaziri was one of a group of white-bellied pangolins that were rescued from the wild in Togo in 2016. The whole blood sample was stored on dry ice and then shipped on dry ice by overnight delivery to Psomagen, Inc., Rockville, MD, USA, for genomic DNA extraction, genomic library preparation, and high-throughput sequencing.

Genomic DNA was extracted from a 200-µl aliquot of the whole blood sample using a MG Blood Genomic DNA Extraction SV kit according to the manufacturer's instructions (MGmed, Seoul, Korea). The isolated genomic DNA was quantitated using the PicoGreen dsDNA assay kit (ThermoFisher Scientific, MA, USA) with a Victor 3 fluorometer (PerkinElmer, MA, USA), with a final concentration of 9.2 ng/µl. A 0.5-µl aliquot of DNA was run for 30 min at 160 V on a 1% TBE agarose gel to assess quality, with no signs of degradation observed. Finally, DNA fragment size analysis with capillary electrophoresis using the Agilent 4200 TapeStation system (Agilent Technologies, CA, USA) showed a peak of fragments around 60 Kb, indicating the genomic DNA was suitable for 10X Genomics Chromium library preparation.

Parallel libraries were prepared from 1.0 µg genomic DNA using the Chromium Genome Reagent Kit version 2 chemistry in conjunction with the 10X Genomics Chromium Controller instrument containing a micro-fluidic Genome Chip (10X Genomics, CA, USA). DNA molecules were partitioned and nick-translated using bead-specific unique molecular identifiers (UMIs) to produce gel bead-in emulsions (GEMs). The libraries were fragmented, followed by 5′ and 3′ adapter ligation, and then bridge amplified using isothermic PCR to generate clusters of each fragment size. The template size distribution of the libraries was checked with an Agilent Technologies 2100 Bioanalyzer using a DNA 1000 chip (Agilent Technologies, CA, USA). Furthermore, the prepared libraries were quantitated using qPCR on a LightCycler PCR instrument following the Illumina qPCR Quantification Protocol Guide. Libraries were then sequenced to a minimum depth of 60× using 150-bp paired-end reads with dual indexing on an Illumina NovaSeq 6000 instrument (Illumina, CA, USA). A total of 2,828,846,794 reads (GC content: 43.68%, Q30: 89.86%) were assembled into two partially phased pseudohaplotypes using Supernova version 2.0, following the manufacturer's instructions (10X Genomics, CA, USA).

Hi-C was done to generate chromosome-length scaffolding of the initial Supernova v2.0 assembly. An aliquot of whole blood from Jaziri was shipped on dry ice via overnight delivery to the Center for Genome Architecture, Baylor College of Medicine, Houston, TX, USA, where in situ Hi-C data generation and scaffolding were performed by the DNA Zoo Consortium (www.dnazoo.org) following the Rao et al. (2014) protocol by (1) crosslinking DNA and proteins in intact nuclei with formaldehyde, (2) digesting the genomic DNA using restriction enzymes, (3) incorporating biotinylated nucleotides on the 5′ overhangs followed by ligation of the blunt-end fragments, (4) shearing the genomic DNA by sonication, and (5) capturing of biotinylated DNA ligation junctions with streptavidin beads. The captured fragments were then PCR amplified and sequenced on an Illumina NovaSeq 6000 instrument. The Hi-C read data were aligned to the Supernova v2.0 assembly of Jaziri using the Juicer pipeline (Durand et al. 2016a). A candidate assembly was then generated using the 3D-DNA pipeline (Dudchenko et al. 2017). Juicebox Assembly Tools (Durand et al. 2016b; Dudchenko et al. 2018) was then used to polish and manually review the candidate assembly to obtain a final assembly (Fig. 1).

Chromosome-length scaffolding of Chinese and Sunda pangolin genomes

Primary fibroblast cell lines of Sunda (passage 4) and Chinese pangolin (passage 2) were used to produce in situ Hi-C libraries as described above for the white-bellied pangolin, Jaziri. A Sunda pangolin (*Manis javanica*) cell line (MJA-21) was established

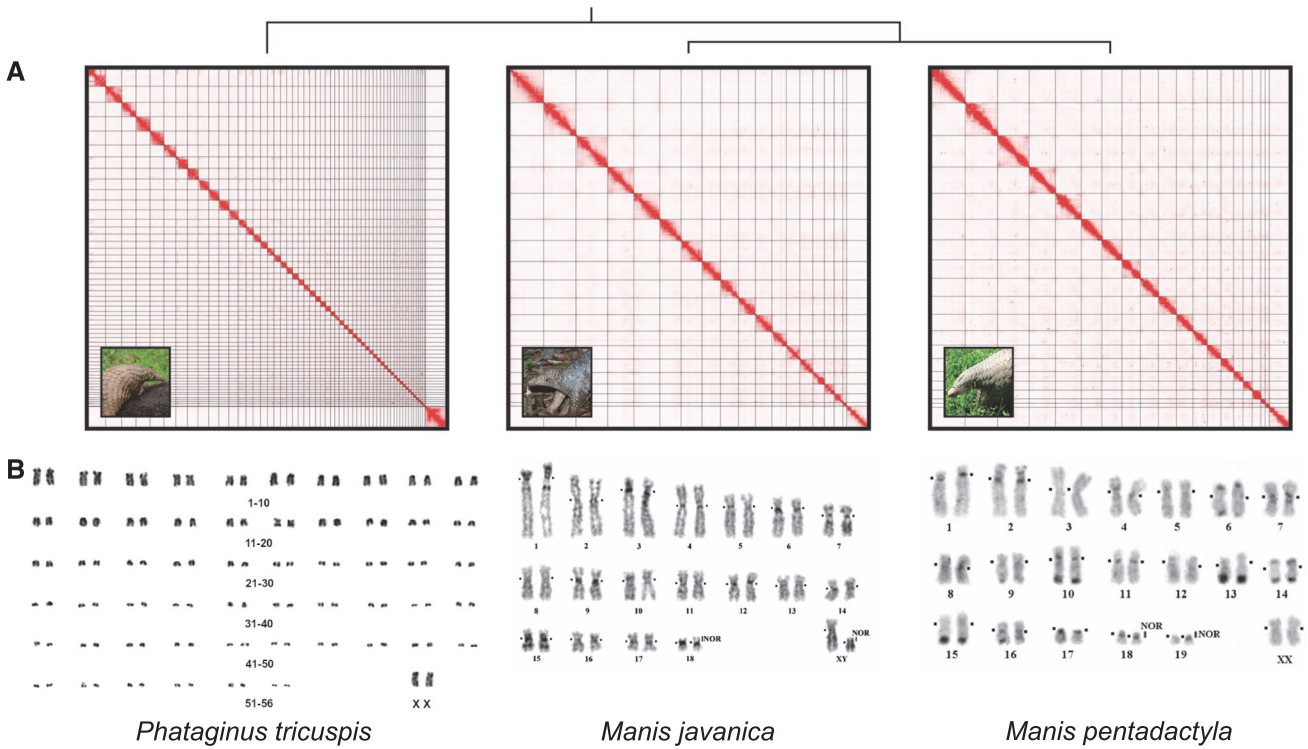


Fig. 1 Hi-C contact maps and karyotypes of three pangolin species. **A** The upper figures show the Hi-C contact map of the 57 chromosome-length scaffolds assembled from a female white-bellied pangolin *Phataginus tricuspis* (left) in comparison to the contact maps of the 19 and 20 chromosome-length scaffolds assembled for the Sunda pangolin *Manis javanica* (center) and Chinese pangolin *Manis pentadactyla* (right), respectively. Chromosome-length scaffolds for each species are ordered from the largest to the smallest autosome and the X chromosome is shown at the bottom-right corner. **B** The lower left figure displays the non-differentially Giemsa-stained karyotype of a female *P. tricuspis* showing the $2n = 114$ chromosome number and representing the complete chromosome complement of a single metaphase cell. Chromosomes are arranged based on morphology, G-banding pattern (when possible), and size, with the five largest bi-armed pairs placed first. The lower middle and right figures show the constitutive heterochromatin (C-banded karyotypes) of a male *M. javanica* and a female *Manis pentadactyla*, respectively. Small squares mark centromeres. Nomenclature follows Yang et al. (2006) and Nie et al. (2009). Nucleolar organizing regions (NOR — ribosomal genes clusters) are located in *M. pentadactyla* on chromosomes 18 and 19, and in *M. javanica* on chromosome 18, and, interestingly, on the Y chromosome. Several chromosomes in *M. pentadactyla* (2, 6, 9, 10, 13, 14, 15) carry distinct telomeric C- and DAPI-positive blocks (AT-rich). Interactive contact maps can be explored at https://tinyurl.com/2koux479

from a skin biopsy collected at the Singapore Zoo, Singapore, from a male pangolin. The primary fibroblast cell line (MPE-891) of a male Chinese pangolin (*Manis pentadactyla*; microchip #0066C1955) was established from skin biopsy collected at the Taipei Zoo, Taiwan. Skin biopsies were processed through collagenase/hyaluronidase enzymatic digestion according to the protocol by Stanyon and Galleni 1991, with modifications. Pangolin cell lines were cultured at 32 °C in AlphaMEM with nucleosides (Gibco 12,571–063; ThermoFisher Scientific, MA, USA), 15% fetal bovine serum, antibiotics (penicillin, streptomycin, gentamicin, and fungizone in concentrations recommended for cell cultures), and supplemented with bFGF at 20–50 ng/ml (Gibco PHG0264; ThermoFisher Scientific, MA, USA).

The cell cultures of the Chinese and Sunda pangolins were shipped on dry ice via overnight delivery to the Center for Genome Architecture, Baylor College of Medicine, Houston, TX, USA. Hi-C library preparation, data generation, and chromosome-length scaffolding were performed by the DNA Zoo Consortium (www.dnazoo.org) following the same protocols described above for the white-bellied pangolin. The Hi-C reads were aligned to the previously reported draft assemblies of the Chinese pangolin and Sunda pangolin (Choo et al. 2016), and candidate assemblies were prepared using 3D-DNA (Dudchenko et al. 2017). Juicebox Assembly Tools (Dudchenko et al. 2018) was used to polish and validate the assemblies of both species.

Male white-bellied pangolin Hi-C data generation and analysis

In order to identify the autosome associated with the Y-autosome fusion in white-bellied pangolins, we generated in situ Hi-C data from a 2013 necropsy sample (muscle) collected from a male individual originally housed at the San Antonio Zoo, San Antonio, TX, USA. The sample was used to generate an in situ Hi-C library following the Rao et al. (2014) protocol, and the resulting data aligned to the finalized female sample-based chromosome-length genome assembly using the Juicer pipeline (Durand et al. 2016a). The male and female maps were overlayed and compared using the "versus mode" feature in the Juicebox visualization software (Durand et al. 2016b).

Assembly completeness

We quantified the completeness of the chromosome-length assemblies of the three pangolin species using the approach of benchmarking universal single-copy orthologs (BUSCOs) with the program BUSCO v5.2.2 (Manni et al. 2021). The program was run in genome mode using the Laurasiatheria ortholog gene set (laurasiatheria_odb10; 12,234 BUSCOs), the Augustus gene predictor with the default(human) gene model (Stanke et al. 2006), and the parameter "--long," for the gene annotations.

Annotation of repetitive elements and protein-coding genes

Repeat annotations were done using RepeatMasker v4.1.1 (Smit et al. 2013-2015) using known repeats associated with the species from the RepBase Update library version 20181026 (Bao et al. 2015) and the parameters "-s -no_is -xsmall."

Homology-based gene annotation was done on the *Manis javanica* genome assembly using Gene Model Mapper (GeMoMa) v1.7 (Keilwagen et al. 2016, 2018). GeMoMa is a homology-based gene annotator that uses gene models from multiple references to annotate genomes. The reference species (with specific genome assembly reference number) used to annotate the *M. javanica* genome were *Homo sapiens* (GRCh38.p13; GCA_000001405.28), *Mus musculus* (GRCm39; GCA_000001635.9), *Canis lupus familiaris* (CanFam3.1; GCA_000002285.4), and *Felis catus* (Felis_catus_9.0; GCA_000181335.5). Complete coding DNA sequences (CDS) with no ambiguous nucleotides were extracted from the genome annotation files provided by the NCBI Eukaryotic Genome Annotation Pipeline (Fong et al. 2013) for each reference species. The extraction was performed using the "Extractor" function in GeMoMa. The soft-masked genome generated from RepeatMasker was indexed with MMseqs2 release 8-fac81 (Steinegger and Söding 2017) using the *createdb* function and the parameter "--dont-split-seq-by-len." The model reference CDS files were indexed using MMseq2 as well but without the parameter used for the reference genome. The CDS regions were mapped to the *M. javanica* genome using parameters "-e 100.0 -s 8.5 -a --comp-bias-corr 0 --max-seqs 500 -mask 0 --orf-start-mode 1 -v 2" and the resulting output was

converted to text format using the MMseq2 function "convertalis" using the parameters "--format-output 'query,target,pident,alnlen,mismatch,gapopen,qstart,qend,tstart,tend,evalue,bits,empty,raw,nident,empty,empty,empty,qframe,tframe,qaln,taln,qlen,tlen' -v 2."

The text MMseqs2 outputs were used in GeMoMa's "GeMoMa" setting the parameters "sort = True" and "Score = ReScore." Once all the MMseq2 outputs were run with GeMoMa independently, they were combined and evaluated using "GAF" with default parameters. Finally, "AnnotationFinalizer" was used to finalize and format the gene models. This was repeated for both the *Manis pentadactyla* and *Phataginus tricuspis* genome assemblies with the exception that the *M. javanica* gene models were used as a reference in GeMoMa.

Comparison of whole genome synteny

Using the repeat masked genome assemblies, pairwise genome alignments were generated using LAST v1180 (Kiełbasa et al. 2011) and filtered into one-by-one pairwise alignments and then formatted to be visualized by MCScanX (Wang et al. 2012) for the JCVI utility package (Tang et al. 2015) using custom Python scripts. MCScanX identifies intergenomic syntenic blocks from LAST hits. The *Phataginus tricuspis* X chromosome was identified from the karyotype and Hi-C data, and verified by mapping the mammalian X-linked *SOX3* gene against it using BLAST (Dutoit et al. 2017; Altschul et al. 1990).

Read mapping and variant calling

Hi-C reads were mapped against the chromosome-length scaffolds of each species using BWA-MEM v0.7.17 (Li and Durbin 2009). The resulting SAM file was then converted into a BAM alignment file with SAMtools v1.8 (Li et al. 2009) using "view" and "sort." Read groups were added to the alignment using Picard v2.24.0 (http://broadinstitute.github.io/picard/) "AddOrRemoveReadGroups." PCR and optical duplicates were marked in the alignment using Picard v2.24.0 "MarkDuplicates." PCR duplicates (exact copies of reads) were removed from the alignment. To increase base-call qualities, we realigned reads surrounding insertions and deletions (indels) using GATK v3.8.1 (Van der Auwera and O'Connor 2020) producing an alignment file which was used for all downstream analyses. Mosdepth vD4 (Pedersen and Quinlan 2018) was used in quick mode to calculate average read depth of the realigned BAM, which was used for variant filtering downstream.

Variant calling was performed on the realigned BAM, which includes the Hi-C scaffolds of each pangolin species as a reference, using bcftools v1.8 (Li 2011) mpileup and call pipeline for all sites, which annotated the output with both allele depth and read depth. We filtered SNPs with a bcftools filter using a minimum read depth of 1/3 the average depth, a maximum read depth of 2 × the average depth, and a QUAL score of 20. Variants were extracted using bcftools view, and we retained only bi-allelic sites.

Genome-wide heterozygosity

Filtered variants for each pangolin species were used as input for a custom Python script (Robinson et al. 2016, 2019) to calculate heterozygosity using a sliding window approach, with non-overlapping windows of 1 Mbp and step size of 1 Mbp per chromosome. The heterozygosity outputs were used to generate a Manhattan plot of heterozygosity across chromosomes using a custom R script (Robinson et al. 2016, 2019).

Demographic reconstruction

We used the pairwise sequentially Markovian coalescent (PSMC) model v0.6.5 (Li and Durbin 2011) to estimate past changes in effective population size in *Phataginus tricuspis*, *Manis javanica*, and *M. pentadactyla*. A psmcfa file was generated from the realigned BAM files for each species using the bcftools mpileup/call/vcfutils.pl pipeline using 1/3 the average depth as the minimum depth filter and 2 × the average depth as the maximum depth filter. This pipeline generated a fastq file which was then converted into a psmcfa using fq2psmc. The X chromosome was removed from the psmcfa file and the resulting psmcfa was split for bootstrapping using splitfa from the PSMC package. The main PSMC run was done using the parameters "-t15 -r5 -p '4 + 25*2 + 4 + 6'." The trajectories of effective population size were scaled using a generation length of 7 years and a mutation rate of 1.47×10^{-8} per generation, following Choo et al. (2016). One hundred bootstrap replicates were

performed using the same parameters as the main PSMC run but using the split psmcfa. All 100 bootstraps were run using parallel and psmc. The results for all three species were visualized together using a custom R script.

Testing for whole genome duplication

We assessed the possibility of whole genome duplication events in the white-bellied genome pangolin assembly (*Phataginus tricuspis*) relative to the genome assemblies of three other species (Chinese pangolin, *Manis pentadactyla*, GCA_014570555.1; Sunda pangolin, *Manis javanica*, GCA_001685135.1; domestic dog, *Canis familiaris*, GCA_000002285.4) using the program WGDdetector (Yang et al. 2019). The program estimates the distribution of synonymous substitutions per synonymous site (dS) of pairwise paralogs within a genome to detect dS deviations (peaks in cumulative distributions of dS) that indicate whole genome duplication.

Distribution of diploid chromosome numbers in mammals

We compiled a data set of diploid ($2n$) chromosome numbers for 2112 named species of mammals in Microsoft Excel, using a variety of sources reporting karyotypes, but relying primarily on the *Atlas of Mammalian Chromosomes* (Graphodatsky et al. 2020) and the online Chromosomes Network database (Graphodatsky et al. 2000). As many species of mammals show intraspecific differences in chromosome number according to sex or population/subspecies, we selected the sex or population/subspecies with the highest $2n$ number. We combined box plots and scatter plots to visualize the chromosome numbers grouped by mammalian orders. The final figure was produced in R (R Core Team 2018) using routines from the ggplot2 package (Wickham 2016).

Results and discussion

The linked-read/Hi-C-guided chromosome-length genome assembly of a female white-bellied pangolin had a contig and scaffold N50 = 112.9 Kb and 46.35 Mbp, respectively, and a completeness of 90.2%, based on 11,043 complete single-copy and duplicated BUSCOs out of a total of 12,234 BUSCOs evaluated (Table 1). Remarkably, the Hi-C interaction matrix visualized against the chromosome-length assembly showed a total of 57 contact blocks, which represents the haploid number of chromosomes and translates into $2n = 114$ (Fig. 1). This high diploid number was validated by the karyotype generated from blood lymphocytes collected from a another female white-bellied pangolin (Fig. 1). The X chromosome was identified as the largest scaffold in females according to the karyotype, and was further verified by mapping the X-linked *SOX3* gene against the genome assembly.

To conduct comparative genomic analyses of the *P. tricuspis* assembly, we generated Hi-C data from primary fibroblasts for *Manis pentadactyla* ($2n = 40$) and *M. javanica* ($2n = 38$) to scaffold the previously published draft genome assemblies of these species (Choo et al. 2016) to chromosome-length (Fig. 1). The haploid number of chromosome-length scaffolds was concordant with the diploid number of chromosomes based on previously established karyotypes for both species. Interestingly, we identified a 30-Mbp heterozygous inversion in *M. pentadactyla* (Fig. S2), suggesting that pangolin chromosomes may be plastic at the level of individual species as well as at the level of the whole order.

Genome assembly size varied among the three species, with 2.47 Gbp in *P. tricuspis*, 2.56 Gbp in *M. javanica*, and 2.22 Gbp in *M. pentadactyla* (Table 1). The scaffold N50 of *P. tricuspis* is the lowest among the three pangolin species, but in the case of chromosome-length assemblies, this reflects the size distribution of the numerous small C-scaffolds found in this species rather than assembly quality. Repeat content also varied among the species: 36.63% in *P. tricuspis*, 32.53% in *M. javanica*, and 30.10% in *M. pentadactyla*. Consistent with the increased repeat content and differences in contigging strategy, the contig N50s were lower for *M. javanica* and *M. pentadactyla* (16,350 bp and 20,721 bp compared to 112,908 bp in *P. tricuspis*).

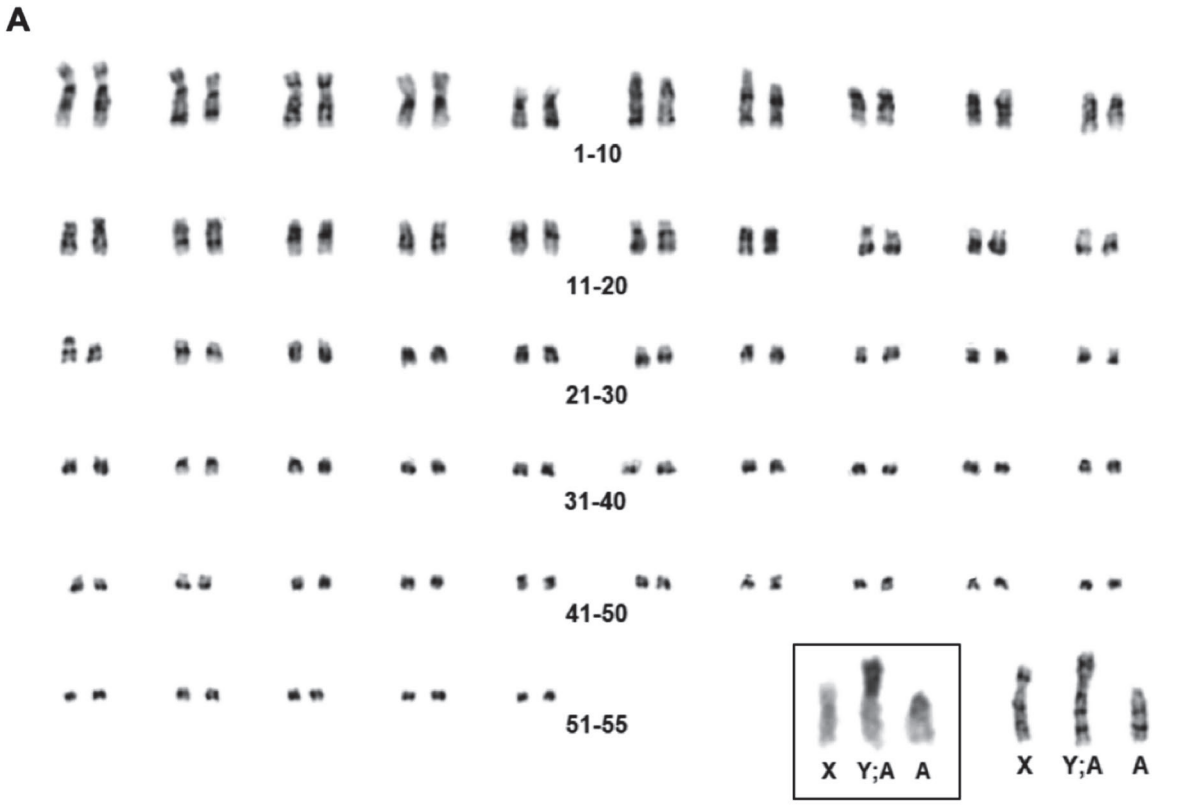
Given the expanded number of chromosomes found in *P. tricuspis*, we tested for whole genome duplication (WGD) events unique to *P. tricuspis* based on deviations in the distribution of

Table 1 Comparison of genome assembly statistics, assembly completeness, and repeat content for white-bellied, Sunda, and Chinese pangolins. *Number of annotated gene models were determined using GeMoMa v.17 (Keilwagen et al. 2016, 2018)

	Phataginus tricuspis	*Manis javanica*	*Manis pentadactyla*
Length	2,473,187,469	2,555,637,331	2,215,491,672
Contig N50 (bp)	112,908	16,350	20,721
Number of contigs	67,643	1,066,770	353,949
Longest contig	892,280	225,448	292,755
Scaffold N50 (bp)	46,351,422	131,848,799	111,940,472
Number of scaffolds	27,299	64,789	71,390
Longest scaffold (bp)	129,801,464	216,609,843	195,883,037
GC content (%)	40.85	40.74	40.97
Complete and single-copy BUSCO genes (%)	10,931 (89.3)	11,122 (90.9)	9688 (79.2)
Complete and duplicated BUSCO genes (%)	112 (0.9)	77 (0.6)	91 (0.7)
Fragmented BUSCO genes (%)	258 (2.1)	288 (2.4)	663 (5.4)
Missing BUSCO genes (%)	933 (7.7)	747 (6.1)	1792 (14.7)
Number of annotated genes*	23,245	20,222	21,113
Repeat content (% of genome assembly)	36.63	32.53	30.1
Repeat content breakdown (occupied bp and %):			
SINEs	66,138,782 (2.84)	65,879,076 (2.86)	61,157,621 (2.96)
LINEs	599,505,025 (25.71)	495,716,016 (21.54)	392,529,724 (19.02)
LINE 1	508,569,328 (21.81)	406,443,083 (17.66)	310,227,307 (15.03)
LINE 2	79,244,281 (3.4)	77,768,702 (3.38)	71,763,251 (3.48)
L3/CR1	8,458,211 (0.36)	8,346,367 (0.36)	7,649,784 (0.37)
RTE	3,045,820 (0.13)	2,965,180 (0.13)	2,716,243 (0.13)
LTR elements	104,418,821 (4.48)	103,542,891 (4.5)	92,527,957 (14.48)
DNA elements	60,955,336 (2.61)	60,392,695 (2.62)	55,310,783 (2.68)
Small RNA	451,460 (0.02)	488,787 (0.02)	379,094 (0.02)
Simple repeats	17,254,632 (0.74)	17,884,914 (0.78)	15,337,834 (0.74)
Low complexity	4,445,377 (0.19)	3,986,510 (0.17)	3,205,883 (0.16)

synonymous substitutions per synonymous site (dS) among pairwise paralogs within a genome. Evidence for recent or ancient WGD events is inferred from the presence of distinct peaks in the cumulative density of pairwise dS among gene families, which violate the assumption that the rate of gene duplication is constant (Maere et al. 2005). For this analysis, we calculated the pairwise dS among the subset of total annotated gene models that constitute duplicated genes (paralogs) inferred from the GeMoMa annotations of the three pangolin species and the annotated assembly of the domestic dog. The number of duplicated/total number of annotated gene models for the four species was as follows: 7370/23,245 for *P. tricuspis*, 6125/21,113 for *M. pentadactyla*, 6929/20,222 for *M. javanica*, and 5143/20,257 for *Canis lupus familiaris*. We found no evidence of independent whole genome duplication in *P. tricuspis* relative to *M. pentadactyla and M. javanica* or the more distantly related domestic dog (Fig. S1). The distributions of the pairwise dS values formed one shallow and broad peak in all four species, suggesting the absence of WGD. This is consistent with analysis of another mammal with an unusually high chromosome count, the red vizcacha rat (*Tympanoctomys barrerae*), with $2n = 102$ and a *C*-value of 8.4 picograms, for which no signal of WGD was similarly found (Evans et al. 2017).

Karyotypes generated by using fibroblast cell lines established for two male white-bellied pangolins revealed



a chromosome number of $2n = 113$, with evidence of a fusion event between the Y chromosome and one of the autosomes (Fig. 2A). To confirm the Y chromosome fusion in males and identify the autosome involved in the fusion, we generated Hi-C data for a male white-bellied pangolin and mapped it against the female sample-based chromosome-length genome assembly. The resulting contact map shows enrichment between a portion of the X chromosome and the 5th largest autosome not observed in females. Notably the same portion of the X chromosome has ~2× coverage as compared to the rest of the sex chromosome in a male sample, consistent with

◂**Fig. 2** Chromosome arrangement in the male white-bellied pangolin *P. tricuspis*. **A** G-banded karyotype of a male showing the $2n = 113$ chromosome number with inset of C-banded sex chromosomes. The karyotype represents the complete chromosome complement of a single metaphase cell. As in the female (Fig. 1), chromosomes are arranged based on morphology, G-banding pattern (when possible), and size, with the five largest bi-armed pairs placed first. The Y chromosome contains a large constitutive heterochromatin block (see inset: C-bands of the sex chromosome group from another metaphase spread of the same individual). **B** Chr Y-autosome fusion identification from Hi-C data. For this analysis, we generated Hi-C data from a male white-bellied pangolin and mapped it to the female genome assembly (center), comparing the resulting contact pattern with Hi-C data from a female sample (left and right, the latter showing the male and female data in the "versus" Juicebox view mode (Durand et al. 2016b), with the male data in the upper right corner, and the female data in the lower left). For simplicity, only two chromosomes are shown: the fifth largest autosome (HiC_scaffold_5) and chr X. Note the increased contact frequency between a ~ 5-Mbp region at the proximal end of the assembled chr X and HiC_scaffold_5 in the male sample but not the female sample. The same 5-Mbp region also shows an elevated coverage (indicated by the blue coverage plots above the contact maps) as compared to the rest of chromosome X in the male sample, consistent with this being a PAR, and the overall pattern reflecting the Y-autosome fusion. Interactive contact maps can be explored here: https://tinyurl.com/2oknbnn2

it matching the pseudoautosomal region (PAR) and the enriched Hi-C signal resulting from a Y-specific fusion event (Fig. 2B).

The karyotypes of both sexes of the white-bellied pangolin are composed of five pairs of submetacentric autosomes and 51 acrocentric/telocentric pairs in the female, while the males have 50 acrocentric/telocentric pairs and one unpaired element that would pair with the autosome fused to the Y. The X is a submetacentric element roughly equal in size to the largest autosome. The complex Y chromosome is the largest element in the karyotype and is nearly metacentric (one arm represents the Y and the other arm is the fused autosome). Patterns of C-banding show the smaller arm of the complex Y stains almost entirely dark indicating constitutive heterochromatin often seen in Y chromosomes of other mammalian species. The larger arm of the complex Y is C-band negative and represents the fused autosome. C-bands are also present at most centromere locations. We note that further standardization of the karyotype of both sexes by molecular techniques such as chromosome painting is required as many of the smaller chromosomes cannot be identified reliably based on G-bands.

It is often the case that the karyotypic differences between closely related species boil down to fusion events where acrocentric chromosomes combine to form bi-armed chromosomes or by fissioning of metacentric chromosomes into acrocentrics, while the fundamental number (FN) of autosomal arms remains the same. The karyotype analysis suggests $FN = 122$ for *P. tricuspis* whereas the highest documented FNa (FN excluding sex chromosomes) in *Manis* is 70, suggesting that chromosomal rearrangements in the order involved not only Robertsonian translocations, but also fission and fusion within the chromosomal arms. No interstitial telomeric sites that could mark recent fusion events were found in Sunda and Malayan pangolins via FISH, supporting that fissioning of the ancestral karyotype happened in *P. tricuspis*. Remarkably, several chromosomes in *M. pentadactyla* (2, 6, 9, 10, 13, 14, 15) carry distinct telomeric C- and DAPI-positive blocks (AT-rich).

Consistent with the inferences based on the karyotype data, synteny analysis of the three pangolin assemblies suggests a predominance of fissioning of chromosome blocks in *P. tricuspis* relative to the two Asian species (Fig. 3). For example, while the large chromosome 3 in *M. pentadactyla* and chromosome 2 in *Manis javanica* show a one-to-one correspondence in synteny, this chromosome has been divided across at least five smaller and distinct chromosome scaffolds in *P. tricuspis*. This chromosome scaffold in the two Asian species is > 150 Mbp in length, yet has been split into C-scaffolds < 50 Mbp in *P. tricuspis*. Such a pattern of fissioning in *P. tricuspis* is consistently observed across all but the smallest C-scaffold, which shows complete synteny with chromosomes 18 and 19 in *M. javanica* and *M. pentadactyla*, respectively. However, the analysis also reveals instances of chromosomal fusion in *P. tricuspis* relative to the Asian species, such as part of chromosome 3 and 4 in *M. javanica* (corresponding to chromosomes 2 and 4 in *M. pentadactyla*) being joined in *P. tricuspis* (Fig. 3). The synteny of the X chromosome is largely intact, in agreement with the conservation in size and gene content of this chromosome across placental mammals (Brashear et al. 2021).

Based on analyses of chromosomal rearrangements and syntenic relationships in mammals using chromosome-scale genome assemblies, the

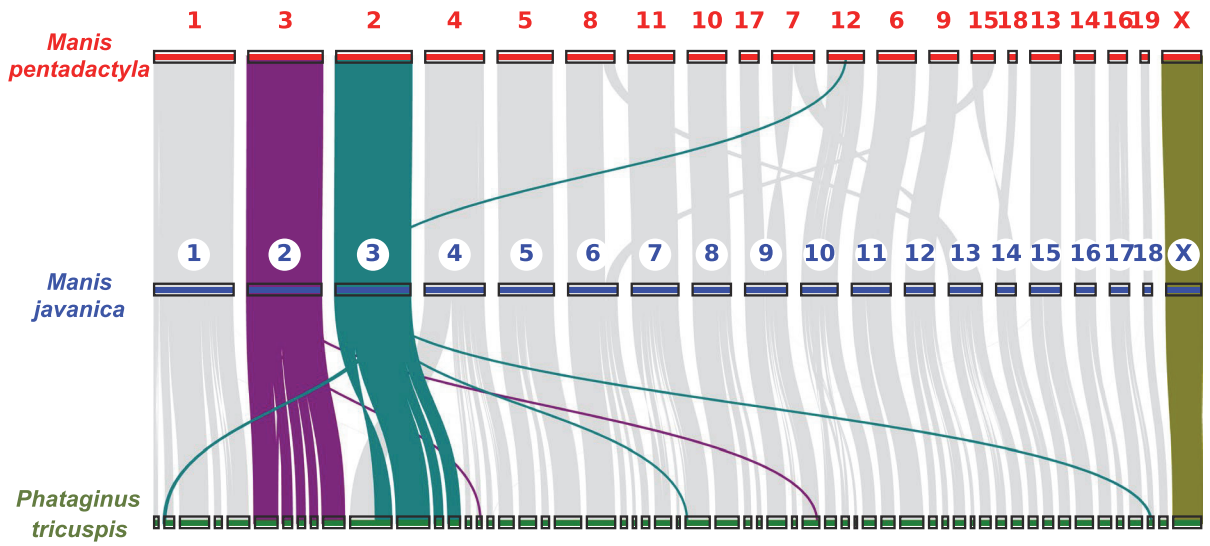


Fig. 3 Synteny among the autosomes and X chromosomes of Chinese (top), Sunda (bottom), and white-bellied (bottom) pangolins. Chromosomes 2 and 3 in the former two species are highlighted to show the extensive fissioning of syntenic blocks in the white-bellied pangolin. Rearrangements of smaller syntenic blocks, likely representing translocations, are apparent as well. The X chromosome of the three species is also highlighted to show the relative conservation of this sex chromosome

ancestral karyotype of superorder Laurasiatheria, of which Pholidota is a member, was estimated to be $2n = 48$ ($n = 23 + X$) (Damas et al. 2022). Moreover, the ancestral karyotype for the Carnivora, sister lineage to the Pholidota, was estimated as $2n = 38$ based on chromosome painting (Nash et al. 2008). Given these estimates, we hypothesize that the high diploid chromosome number in *P. tricuspis* is primarily the result of fissioning of a smaller ancestral pangolin karyotype. The karyotypes of *P. tetradactyla*, *Smutsia gigantea*, and *S. temminckii* need to be obtained to determine whether this occurred in the Phataginiae lineage (i.e., prior to the split between *P. tetradactyla* and *P. tricuspis*) or in the ancestor of the Phataginiae + Smutsiinae (i.e., after the split from the Asian pangolins, Maninae).

We found that *P. tricuspis* was intermediate in genome-wide heterozygosity (mean = 0.0015) relative to *M. pentadactyla* (mean = 0.0009) and *M. javanica* (mean = 0.0022) (Fig. 4A). These levels of genetic diversity are notable given these species are categorized as either endangered (*P. tricuspis*) or critically endangered (*M. pentadactyla* and *M. javanica*). Reconstruction of demographic history showed a trajectory of effective population size (*Ne*) punctuated by two moderate declines for *P. tricuspis*, one ~ 100 Kya and the other more recently (Fig. 4B). In contrast, *Ne* has declined sharply during the last hundred thousand years in *M. pentadactyla*, resulting in a current effective population size that is only approximately one-fifth or less than that of *P. tricuspis* or *M. javanica*. As pangolins are largely found in tropical and subtropical regions, fluctuations and declines in population size may have been affected by climate and habitat changes that occurred during the Last Glacial Period (115–11.7 Kya) (Hu et al. 2020).

The current highest recorded diploid numbers of chromosomes in mammals are found in two species of South American rodents, the Bolivian bamboo rat (*Dactylomys boliviensis*) with $2n = 118$ (Dunnum et al. 2001), and the plains viscacha rat (*Tympanoctomys barrerae*) with $2n = 102$ (Contreras et al. 1990), belonging to the superorder

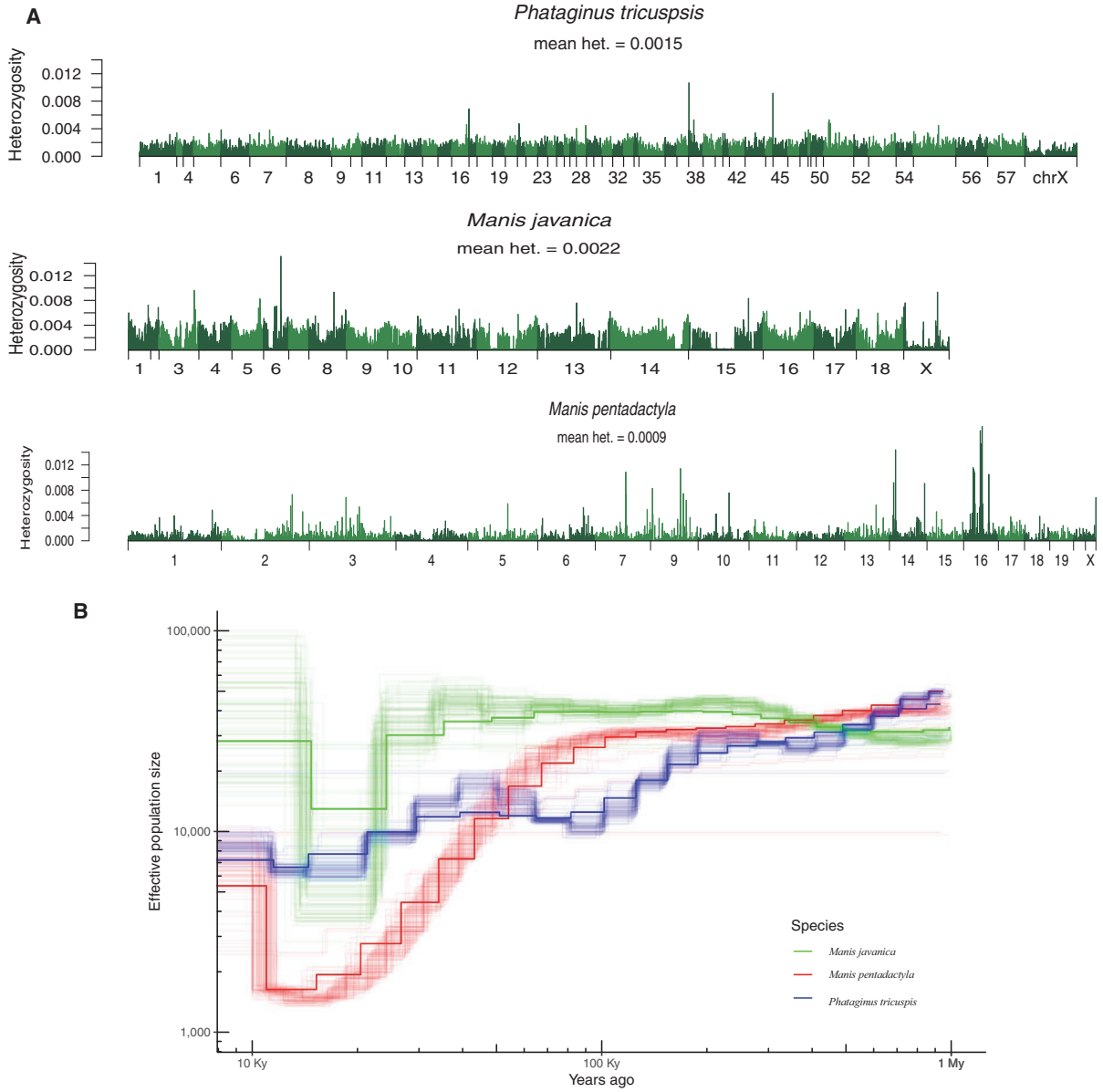


Fig. 4 **A** Manhattan plot of genome-wide heterozygosity of a female white-belled pangolin (top), female Chinese pangolin (middle), and female Sunda pangolin (bottom). Heterozygosity was estimated using non-overlapping windows of 1 Mbp and step size of 1 Mbp per chromosome-length scaffold. C-scaffolds are numbered for *Phataginus tricuspis*, with c-scaffold 2 corresponding to the X chromosome. For *Manis javanica* and *M. pentadactyla*, chromosomes are numbered according to assignment of C-scaffolds to each species' karyotype. Mean heterozygosity of each species is shown. **B** Effective population size trajectories derived from the instantaneous inverse coalescent rate through time of white-bellied pangolin (blue line), Chinese pangolin (red line), and Sunda pangolin (green line), estimated using the pairwise sequentially Markovian coalescent (PSMC) model. Finer lines around bold lines represent trajectories estimated from 100 bootstrap replicates. The trajectories were scaled using a generation length of 7 years and a mutation rate of 1.47×10^{-08} per generation

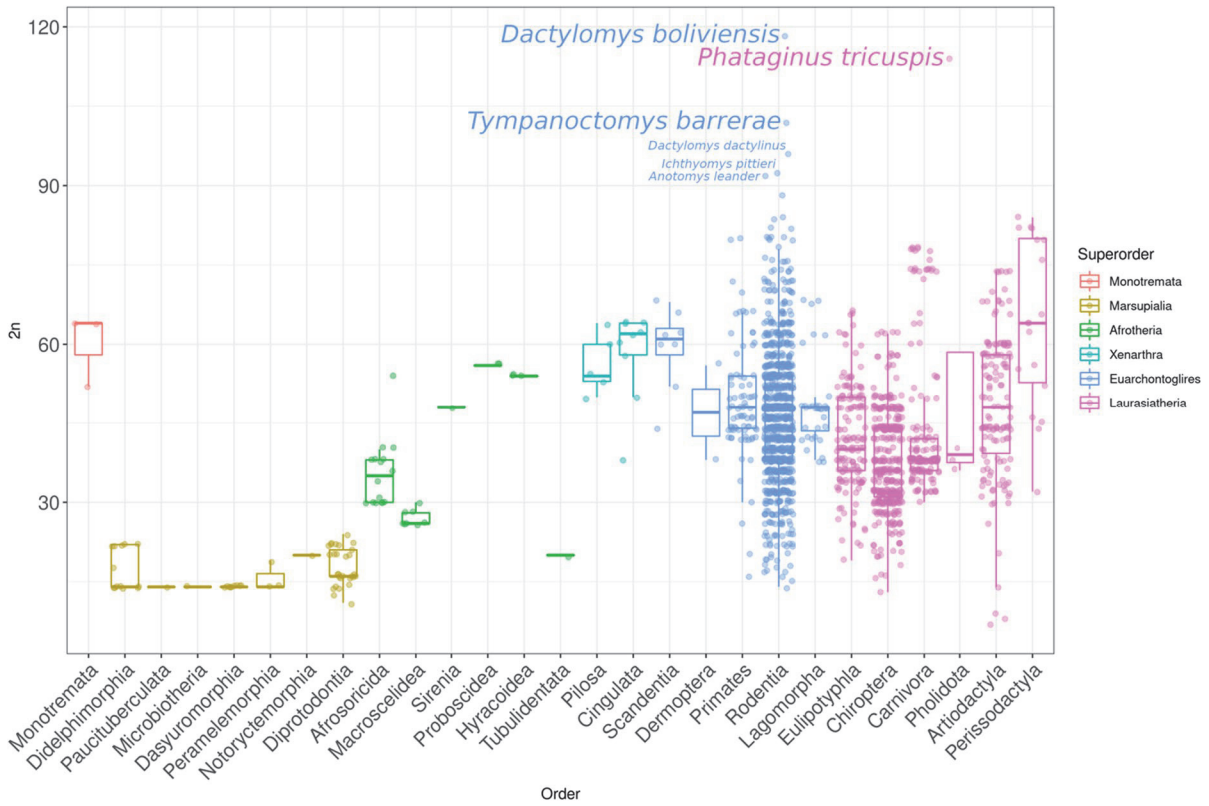


Fig. 5 Box plot showing diploid chromosome number among 2112 named species of mammals, representing 27 orders of mammals. The three species with the highest known $2n$ are indicated and include *Dactylomys boliviensis* ($2n = 118$, Rodentia), *P. tricuspis* ($2n = 114/113$, Pholidota), and *Tympanoctomys barrera* ($2n = 102$, Rodentia). Three other species from the order Rodentia with $2n > 90$ are shown in smaller letter size. The boxes contain the 25th to 75th percentiles, with the inside lines indicating the median value of $2n$ numbers. The vertical lines represent the 5th (lower) and 95th (upper) percentiles and dots outside of these bounds are outliers. Box and dot colors correspond to mammalian superorders: Monotremata (red), Marsupialia (gold), Afrotheria (green), Xenarthra (teal), Euarchontoglires (blue), and Laurasiatheria (magenta)

Euarchontoglires. Our finding of $2n = 113/114$ in *P. tricuspis* is the highest documented among Laurasiatheria and the second highest for all mammals (Fig. 5). Our results show that Pholidota, despite being a relatively small order, displays one of the largest disparities in diploid chromosome numbers ($2n = 36$–114), on par to what is found in much more speciose mammalian orders such as Rodentia ($2n = 14$–118), Carnivora ($2n = 30$–78), Artiodactyla (whales and terrestrial artiodactyls) ($2n = 7$–74), Chiroptera ($2n = 13$–62), Eulipotyphla ($2n = 19$–66), and Primates ($2n = 16$–80). Furthermore, to our knowledge, this is only the second example, after the Indian and the Chinese muntjacs with $2n = 6/7$ and $2n = 46$ respectively (Graphodatsky et al. 2020; Hoencamp et al. 2021), when closely related species with large karyotypic differences have been assembled to chromosome-length. The extreme karyotype variation along with the availability of chromosome-length genome assemblies positions pangolins as a promising model group for studying the historical and mechanistic processes behind extraordinary karyotype plasticity.

Abbreviation *PAR*: Pseudo-autosomal region

Acknowledgements We thank Kang Sa and the laboratory and bioinformatic teams at Psomagen, Inc. for their excellent services and support in generating the linked-read assembly for the white-bellied pangolin, Jaziri. We are grateful to Russ Hansen (National Cancer Institute-Frederick) for assistance with CITES paperwork. We thank Melody Roelke (Leidos Inc), Carlos Driscoll, Christina Barr, Stephen Lindell, Cheryl Marietta, and David Goldman (National Institute on Alcohol Abuse and Alcoholism) for preservation of pangolin cell lines as part of the former NCI-Laboratory of Genomic Diversity collection. We thank Saul Godinez Puido and Zane Colaric for their help with Hi-C sample acquisitions and processing. Hi-C data were created by the DNA Zoo Consortium (www.dnazoo.org). DNA Zoo is supported by Illumina, Inc.; IBM; and the Pawsey Supercomputing Center. Figure 1 photo credits: *P. tricuspis*: copyright: ehbidault, some rights reserved (CC-BY-NC), via inaturalist.org; *M. javanica*: Sunda pangolin, photo by budak (CC BY-NC-ND 2.0), via flickr.com; *M. pentadactyla*: *Manis pentadactyla pentadactyla* (2010.07.15), photo by Skink Chen (CC BY-NC-ND 2.0), via flickr.com.

Author contribution M.L.H., K-P.K., T.B.S., E.L.A., and O.D.: conceived the study. A.G., S-L.J., S.J.O., N.T-L.L., J.S.C.C., K.K., G.S., J.G., A.H., M.M., I.G.-T., J.G., R.L.C., J.E.J., R.J.H., J.T.: facilitated and provided experimental materials. M.L.H., S.J.C., J.A.F., A.C.M., P.L.P, V.B.: performed the karyotyping. K-P.K., J.E.J., R.J.H., J.T., T.B.S: led the effort to generate the linked-read genome assembly. R.K., A.O., D.W., E.L.A., O.D.: performed the Hi-C experiments and generated the chromosome-length assemblies. M.L.H., K-P.K., T.H., S.K., A.G., V.G., G.T., E.L.A., O.D.: analyzed the data. M.L.H., K-P.K., P.L.P., E.L.A., O.D.: drafted the manuscript, with subsequent input from all the authors.

Funding Work was supported by grants from the National Geographic Society (NGS-418C-18) and US Gov. INLEC (S-INLEC-17-GR-1006). E.L.A. was supported by the Welch Foundation (Q-1866), a McNair Medical Institute Scholar Award, an NIH Encyclopedia of DNA Elements Mapping Center Award (UM1HG009375), a US-Israel Binational Science Foundation Award (2019276), the Behavioral Plasticity Research Institute (NSF DBI-2021795), an NSF Physics Frontiers Center Award (NSF PHY-2019745), and an NIH CEGS (RM1HG011016-01A1). A.S. Graphodatsky, P.L. Perelman, and V.R. Beklemisheva were supported by the Russian Science Foundation (Grant No. 19–14-00034p to A.S.G.). G. Tamazian was supported by Peter the Great St. Petersburg Polytechnic University in the framework of the Russian Federation's Priority 2030 Strategic Academic Leadership Programme (Agreement 075–15-2021–1333).

Data availability The Whole Genome Shotgun project for *P. tricuspis* has been deposited at DDBJ/ENA/GenBank under the accession JAQQAC000000000 within the BioProject PRJNA905096. The version described in this paper is version JAQQAC010000000. The short-read sequencing data generated for the Hi-C scaffolding of the three pangolin species have been deposited into the NCBI Sequence Read Archive under the accessions SRX8933295 (*Phataginus tricuspis*), SRX9606524 (*Manis javanica*), and SRX9606525 (*Manis pentadactyla*), part of BioProject PRJNA512907. The *P. tricuspis* assembly and the upgraded *M. javanica* and *M. pendactyla* assemblies are available at https://www.dnazoo.org/assemblies?search=pangolins. Custom scripts used for this study have been deposited on GitHub at https://github.com/thainsCEB/PangolinCytogenomics.

Declarations

Ethical approval The blood sample from the white-bellied pangolin, Jaziri, was collected opportunistically in April 2019 during a routine veterinary examination by the veterinary staff at the Pittsburgh Zoo & Aquarium, following all required and standard animal handling protocols and with approval by the Pittsburgh Zoo & Aquarium institutional animal care and use committee.

Consent to participate Not applicable.

Consent for publication Not applicable.

Competing interests The authors declare no competing interests.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215(3):403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Bao W, Kojima KK, Kohany O (2015) Repbase Update, a database of repetitive elements in eukaryotic genomes. Mob DNA 6:11. https://doi.org/10.1186/s13100-015-0041-9

Beklemisheva VR, Perelman PL, Lemskaya NA, Proskuryakova AA, Serdyukova NA, Burkanov VN, Gorshunov MB, Ryder O, Thompson M, Lento G, O'Brien SJ, Graphodatsky AS (2020) Karyotype evolution in 10 pinniped species: variability of heterochromatin versus high conservatism of euchromatin as revealed by comparative molecular cytogenetics. Genes 11(12):1485. https://doi.org/10.3390/genes11121485

Bernardi G (2015) Chromosome architecture and genome organization. PLoS One 10(11):e0143739. https://doi.org/10.1371/journal.pone.0143739

Brashear WA, Bredemeyer KR, Murphy WJ (2021) Genomic architecture constrained placental mammal X chromosome evolution. Genome Res 31(8):1353–1365. https://doi.org/10.1101/gr.275274.121

Burgin CJ, Collela JP, Kahn PL, Upham NS (2018) How many species of mammals are there? J Mammal 99(1):1–14. https://doi.org/10.1093/jmammal/gyz052

Burton JN, Adey A, Patwardhan RP, Qiu R, Kitzman JO, Shendure J (2013) Chromosome-scale scaffolding of *de novo* genome assemblies based on chromatin interactions. Nat Biotechnol 31(12):1119–1125. https://doi.org/10.1038/nbt.2727

Choo SW, Rayko M, Tan TK, Hari R, Komissarov A, Wee WY, Yurchenko AA, Kliver S, Tamazian G, Antunes A, Wilson RK, Warren WC, Koepfli KP, Minx P, Krasheninnikova K, Kotze A, Dalton DL, Vermaak E, Paterson IC, Dobrynin P, … Wong GJ (2016) Pangolin genomes and the evolution of mammalian scales and immunity. Genome Res 26(10), 1312–1322 https://doi.org/10.1101/gr.203521.115

Choo SW, Chong JL, Gaubert P, Hughes AC, O'Brien S, Chaber AL, Antunes A, Platto S, Sun NC, Yu L, Koepfli KP, Suwal TL, Thakur M, Ntie S, Panjang E, Kumaran JV, Mahmood T, Heighton SP, Dorji D, Gonedelé BS, … Aziz MA (2022) A collective statement in support of saving pangolins. Sci Total Environ 824:153666. https://doi.org/10.1016/j.scitotenv.2022.153666

Contreras LC, Torres-Mura JC, Spotorno AE (1990) The largest known chromosome number for a mammal, in a South American desert rodent. Experientia 46(5):506–508. https://doi.org/10.1007/BF01954248

Damas J, Corbo M, Kim J, Turner-Maier J, Farré M, Larkin DM, Ryder OA, Steiner C, Houck ML, Hall S, Shiue L, Thomas S, Swale T, Daly M, Korlach, J, Uliano-Silva M, Mazzoni CJ, Birren BW, Genereux DP, Johnson J, … Lewin HA (2022) Evolution of the ancestral mammalian karyotype and syntenic regions. Proc Natl Acad Sci US America 119(40):e2209139119. https://doi.org/10.1073/pnas.2209139119

Deakin JE, Ezaz T (2014) Tracing the evolution of amniote chromosomes. Chromosoma 123(3):201–216. https://doi.org/10.1007/s00412-014-0456-y

Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand NC, Shamim MS, Machol I, Lander ES, Aiden AP, Aiden EL (2017) De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. Science 356(6333):92–95. https://doi.org/10.1126/science.aal3327

Dudchenko O, Shamim MS, Batra S, Durand NC, Musial NT, Mostofa R, Pham M, Hilaire BGS, Yao W, Stamenova E, Hoeger M, Nyquist SK, Korchina V, Pletch K, Flanagan JP, Tomaszewicz A, McAloose D, Estrada CP, Novak BJ, Omer AD, Aiden EL (2018) The Juicebox Assembly Tools module facilitates de novo assembly of mammalian genomes with chromosome-length scaffolds for under $1000. bioRxiv.https://doi.org/10.1101/254797

Dunnum JL, Salazar-Bravo J, Yates TL (2001) The Bolivian bamboo rat *Dactylomys boliviensis* (Rodentia: Echimyidae), a new record for chromosome number in a mammal. Mamm Biol 66(2):121–126

Durand NC, Shamim MS, Machol I, Rao SS, Huntley MH, Lander ES, Aiden EL (2016a) Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. Cell Syst 3(1):95–98. https://doi.org/10.1016/j.cels.2016.07.002

Durand NC, Robinson JT, Shamim MS, Machol I, Mesirov JP, Lander ES, Aiden EL (2016b) Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. Cell Syst 3(1):99–101. https://doi.org/10.1016/j.cels.2015.07.012

Dutoit L, Vijay N, Mugal CF, Bossu CM, Burri R, Wolf J, Ellegren H (2017) Covariation in levels of nucleotide diversity in homologous regions of the avian genome long after completion of lineage sorting. Proc R Soc B: Biol Sci 284(1849):20162756. https://doi.org/10.1098/rspb.2016.2756

Evans BJ, Upham NS, Golding GB, Ojeda RA, Ojeda AA (2017) Evolution of the largest mammalian genome. Genome Biol Evol 9(6):1711–1724. https://doi.org/10.1093/gbe/evx113

Fong JH, Murphy TD, Pruitt KD (2013) Comparison of RefSeq protein-coding regions in human and vertebrate genomes. BMC Genomics 14:654. https://doi.org/10.1186/1471-2164-14-654

Gaubert P, Antunes A, Meng H, Miao L, Peigné S, Justy F, Njiokou F, Dufour S, Danquah E, Alahakoon J, Verheyen E, Stanley WT, O'Brien SJ, Johnson WE, Luo SJ (2018) The complete phylogeny of pangolins: scaling up resources for the molecular tracing of the most trafficked mammals on Earth. J Hered 109(4):347–359. https://doi.org/10.1093/jhered/esx097

Graphodatsky AS, Sharshov A, Lavryushov S, Sablina OV, Biltueva LS, Perelman PL, Orlov VN, Kozlovsky AI, Nadjafova RS, Bulatova NSH (2000) Chromosomes network. http://www.bionet.nsc.ru/labs/chromosomes/

Graphodatsky AS, Trifonov VA, Stanyon R (2011) The genome diversity and karyotype evolution of mammals. Mol Cytogenet 4:22. https://doi.org/10.1186/1755-8166-4-22

Graphodatsky A, Perelman PL, O'Brien SJ (eds). (2020) *Atlas of* mammalian *chromosomes, Second Edition*. Hoboken, NJ: John Wiley & Sons, Inc. 1008

Heinrich S, Wittman TA, Prowse TAA, Ross JV, Delean S, Shepherd CR, Cassey P (2016) Where did all the pangolins go? International CITES trade in pangolin species. Glob Ecol Conserv 8:241–253. https://doi.org/10.1016/j.gecco.2016.09.007

Hoencamp C, Dudchenko O, Elbatsh A, Brahmachari S, Raaijmakers JA, van Schaik T, Sedeño Cacciatore Á, Contessoto VG, van Heesbeen R, van den Broek B, Mhaskar AN, Teunissen H, St Hilaire BG, Weisz D, Omer AD, Pham M, Colaric Z, Yang Z, Rao S, Mitra N, … Rowland BD (2021) 3D genomics across the tree of life reveals condensin II as a determinant of architecture type. Science 372(6545):984–989. https://doi.org/10.1126/science.abe2218

Houck ML, Ryder OA, Váhala J, Kock RA, Oosterhuis JE (1994) Diploid chromosome number and chromosomal variation in the white rhinoceros (*Ceratotherium simum*). J Hered 85(1):30–34

Hu J-Y, Hao Z-Q, Frantz L, Wu S-F, Chen W, Jiang Y-F, Wu H, Kuang W-M, Li H, Zhang Y-P, Yu L (2020) Genomic consequences of population decline in critically endangered pangolins and their demographic histories. Natl Sci Rev 7(4):798–814. https://doi.org/10.1093/nsr/nwaa031

Keilwagen J, Wenk M, Erickson JL, Schattat MH, Grau J, Hartung F (2016) Using intron position conservation for homology-based gene prediction. Nucleic Acids Res 44(9):e89. https://doi.org/10.1093/nar/gkw092

Keilwagen J, Hartung F, Paulini M, Twardziok SO, Grau J (2018) Combining RNA-seq data and homology-based gene prediction for plants, animals and fungi. BMC Bioinformatics 19(1):189. https://doi.org/10.1186/s12859-018-2203-5

Kiełbasa SM, Wan R, Sato K, Horton P, Frith MC (2011) Adaptive seeds tame genomic sequence comparison. Genome Res 21(3):487–493. https://doi.org/10.1101/gr.113985.110

King M (1993) Species evolution – the role of chromosome change. Cambridge University Press, Cambridge, p 336

Kumamoto AT, Charter SJ, Houck ML, Frahm M (1996) Chromosomes of *Damaliscus* (Artiodactyla, Bovidae): simple and complex centric fusion rearrangements. Chromosome Res 4(8):614–621. https://doi.org/10.1007/BF02261724

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25(14):1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li H, Durbin R (2011) Inference of human population history from individual whole-genome sequences. Nature 475(7357):493–496. https://doi.org/10.1038/nature10231

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup (2009) The Sequence Alignment/Map format and SAMtools. Bioinformatics 25(16):2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Li H (2011) A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics 27(21):2987–2993. https://doi.org/10.1093/bioinformatics/btr509

Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, Sandstrom R, Bernstein B, Bender MA, Groudine M, Gnirke A, Stamatoyannopoulos J, Mirny LA, Lander ES, Dekker J (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science 326(5950):289–293. https://doi.org/10.1126/science.1181369

Maden BEH, Dent CL, Farrell TE, Garde J, McCallum FS, Wakeman JA (1987) Clones of human ribosomal DNA containing the complete 18 S-rRNA and 28 S-rRNA genes. Characterization, a detailed map of the human ribosomal transcription unit and diversity among clones. Biochem J 246(2):519–527. https://doi.org/10.1042/bj2460519

Maere S, De Bodt S, Raes J, Casneuf T, Van Montagu M, Kuiper M, & Van de Peer Y (2005) Modeling gene and genome duplications in eukaryotes. Proc Natl Acad Sci U S A 102(15):5454–5459. https://doi.org/10.1073/pnas.0501102102

Makino S, Tateishi S (1951) Notes on the chromosomes of the pangolin, *Manis pentadactyla* (Edentata). J Fac Sci Hokkaido Univ Series VI Zool 10(3–4):319–323

Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM (2021) BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. Mol Biol Evol 38(10):4647–4654. https://doi.org/10.1093/molbev/msab199

Mayrose I, Lysak MA (2021) The evolution of chromosome numbers: mechanistic models and experimental approaches. Genome Biol Evolution 13(2):evaa220. https://doi.org/10.1093/gbe/evaa220

Nash WG, Menninger JC, Padilla-Nash HM, Stone G, Perelman PL, O'Brien SJ (2008) The ancestral carnivore karyotype (2n = 38) lives today in ringtails. J Hered 99(3):241–253. https://doi.org/10.1093/jhered/esm130

Nie W, Wang J, Su W, Wang Y, Yang F (2009) Chromosomal rearrangements underlying karyotype differences between Chinese pangolin (*Manis pentadactyla*) and Malayan pangolin (*Manis javanica*) revealed by chromosome painting. Chromosome Res 17(3):321–329. https://doi.org/10.1007/s10577-009-9027-0

Nie W, Wang J, Su W, Wang D, Tanomtong A, Perelman PL, Graphodatsky AS, Yang F (2012) Chromosomal rearrangements and karyotype evolution in carnivores revealed by chromosome painting. Heredity 108(1):17–27. https://doi.org/10.1038/hdy.2011.107

Pedersen BS, Quinlan AR (2018) Mosdepth: quick coverage calculation for genomes and exomes. Bioinformatics (Oxford, England) 34(5):867–868. https://doi.org/10.1093/bioinformatics/btx699

R Core Team (2018) R: a language and environment for statistical computing. R Foundation for Statistical Computing. https://www.R-project.org/

Rao SS, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES, Aiden EL (2014) A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell 159(7):1665–1680. https://doi.org/10.1016/j.cell.2014.11.021

Robinson JA, Ortega-Del Vecchyo D, Fan Z, Kim BY, vonHoldt BM, Marsden CD, Lohmueller KE, Wayne RK (2016) Genomic flatlining in the endangered island fox. Curr Biol 26(9):1183–1189. https://doi.org/10.1016/j.cub.2016.02.062

Robinson JA, Räikkönen J, Vucetich LM, Vucetich JA, Peterson RO, Lohmueller KE, Wayne RK (2019) Genomic signatures of extensive inbreeding in Isle Royale wolves, a population on the threshold of extinction. Sci Adv 5(5):eaau0757. https://doi.org/10.1126/sciadv.aau0757

Robinson TJ, Ruiz-Herrera A, Avise JC (2008) Hemiplasy and homoplasy in the karyotypic phylogenies of mammals. Proc Natl Acad Sci U S A 105(38):14477–14481https://doi.org/10.1073/pnas.0807433105

Sexton T, Cavalli G (2015) The role of chromosome domains in shaping the functional genome. Cell 160(6):1049–1059. https://doi.org/10.1016/j.cell.2015.02.040

Smit AFA, Hubley R, Green P (2013–2015) RepeatMasker Open-4.0.2013–2015. http://www.repeatmasker.org

Stanke M, Schöffmann O, Morgenstern B, Waack S (2006) Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. BMC Bioinformatics 7:62. https://doi.org/10.1186/1471-2105-7-62

Stanyon R, Galleni L (1991) A rapid fibroblast culture technique for high resolution karyotypes. Bollettino Di Zoologia 58(1):81–83. https://doi.org/10.1080/11250009109355732

Steinegger M, Söding J (2017) MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. Nat Biotechnol 35(11):1026–1028. https://doi.org/10.1038/nbt.3988

Sumner AT (1972) A simple technique for demonstrating centromeric heterochromatin. Exp Cell Res 75(1):304–306. https://doi.org/10.1016/0014-4827(72)90558-7

Tang H, Krishnakumar V, Li J (2015) jcvi: JCVI utility libraries. Zenodo. https://doi.org/10.5281/zenodo.31631

Van der Auwera GA, O'Connor BD (2020) Genomics in the cloud: using Docker, GATK, and WDL in Terra. O'Reilly Media Inc., Sebastopol, California

Wang Y, Tang H, Debarry JD, Tan X, Li J, Wang X, Lee TH, Jin H, Marler B, Guo H, Kissinger JC, Paterson AH (2012) MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res 40(7):e49. https://doi.org/10.1093/nar/gkr1293

Wickham H (2016) ggplot2: Elegant graphics for data analysis 2nd ed. Springer-Verlag, New York

Yang F, Graphodatsky AS (2009) Animal probes and ZOO-FISH. In: Liehr T (ed) Fluorescence in situ hybridization (FISH) - application guide. Springer-Verlag, Berlin, pp 323–346

Yang F, Graphodatsky AS, Li T, Fu B, Dobigny G, Wang J, Perelman PL, Serdukova NA, Su W, O'Brien PC, Wang Y, Ferguson-Smith MA, Volobouev V, Nie W (2006) Comparative genome maps of the pangolin, hedgehog, sloth, anteater and human revealed by cross-species chromosome painting: further insight into the ancestral karyotype and genome evolution of eutherian mammals. Chromosome Res 14(3):283–296. https://doi.org/10.1007/s10577-006-1045-6

Yang Y, Li Y, Chen Q, Sun Y, Lu Z (2019) WGDdetector: a pipeline for detecting whole genome duplication events using the genome or transcriptome annotations. BMC Bioinformatics 20(1):75. https://doi.org/10.1186/s12859-019-2670-3