

1

2

3 Main Manuscript for

4 Extreme in Every Way: Exceedingly Low Genetic Diversity in Snow 5 Leopards Due to Persistently Small Population Size

6 Katherine A. Solari^{1*}, Simon Morgan¹, Andrey D. Poyarkov², Byron Weckworth³, Gustaf
7 Samelius^{4,5}, Koustubh Sharma^{4,6}, Stephane Ostrowski⁷, Uma Ramakrishnan⁸, Zairbek
8 Kubanychbekov⁹, Shannon Kachel³, Örjan Johansson^{4,10}, Purevjav Lkhagvajav¹¹, Heather
9 Hemmingmoore¹⁰, Dmitry Y. Alexandrov², Munkhtsog Bayaraa¹², Alexey Grachev¹³, Miroslav P.
10 Korablev², Jose A. Hernandez-Blanco², Bariushaa Munkhtsog¹⁴, Barry Rosenbaum¹⁵, Viatcheslav
11 V. Rozhnov², Ali Madad Rajabi¹⁶, Hafizullah Noori¹⁶, Kulbhushansingh Suryawanshi^{3,17,18}, Ellie E.
12 Armstrong^{1,19&}, Dmitri A. Petrov^{1,20,21&*}

13

14 ¹Department of Biology, Stanford University, Stanford, CA, USA

15 ²A.N. Severtsov Institute of Ecology and Evolution, Russian Academy of Science, Moscow,
16 Russia

17 ³Panthera, New York, NY, USA

18 ⁴Snow Leopard Trust, Seattle, WA, USA

19 ⁵Nordens Ark, Hunnebostrand, Sweden

20 ⁶Snow Leopard Foundation in Kyrgyzstan, Bishkek, Kyrgyz Republic

21 ⁷Wildlife Conservation Society, New York, NY, USA

22 ⁸National Centre for Biological Sciences, TIFR, Bangalore, India

23 ⁹Ilbirs Foundation, Bishkek, Kyrgyzstan

24 ¹⁰Grimsö Wildlife Research Station, Swedish University of Agricultural Sciences, Riddarhyttan,
25 Sweden

26 ¹¹Snow Leopard Conservation Foundation, Ulanbaatar, Mongolia

27 ¹²Wildlife Institute, College of Nature Conservation of the Beijing Forestry University, Beijing,
28 China

29 ¹³Institute of Zoology of Republic of Kazakhstan, Almaty, Kazakhstan

30 ¹⁴Institute of General and Experimental Biology, Academy of Sciences of Mongolia, Ulaanbaatar,
31 Mongolia

32 ¹⁵Altai Institute for Research and Conservation, Boulder, CO, USA

33 ¹⁶Wildlife Conservation Society, Afghanistan Program, Kabul, Afghanistan

34 ¹⁷Nature Conservation Foundation, Mysore, India

35 ¹⁸Canadian Institute for Advanced Research, Toronto, Ontario, Canada

36 ¹⁹Washington State University, Pullman, WA, USA

37 ²⁰Chan Zuckerberg BioHub, San Francisco, CA, USA

38 ²¹Stanford Cancer Institute, Stanford University School of Medicine, Stanford, CA, USA

39 [&]Contributed equally

40

41 * Katherine A. Solari, Dmitri A. Petrov

42 **Email:** ksolari@stanford.edu, dpetrov@stanford.edu

43

44

45 **Author Contributions:** Author contributions: K.A.S., S.M., E.E.A, and D.A.P. designed research;
46 K.A.S., E.E.A., and D.A.P. performed research; A.D.P., B.W., G.S., K.S., S.O., U.R., Z.K., S.K.,
47 Ö.J., P.L., H.H., D.Y.A., M.B., A.G., M.P.K., J.A.H., B.M., B.R., V.V.R., A.M.R., H.N., and K.S.
48 contributed new reagents/analytic tools/samples ; K.A.S. and E.E.A analyzed data; and
49 K.A.S., S.M., B.W., G.S., K.S., S.O., U.R., S.K., Ö.J., E.E.A., and D.A.P. wrote the paper.

50

51 **Competing Interest Statement:** The authors declare no competing interest.

52

Classification: Biological Sciences, Population Biology

53

Keywords: snow leopard, *Panthera uncia*, heterozygosity, structure

54

This PDF file includes:

55

Main Text

56

Figures 1-4

57

58

59 **Abstract**

60 Snow leopards (*Panthera uncia*) serve as an umbrella species whose conservation benefits their
61 high-elevation Asian habitat. Their numbers are believed to be in decline due to numerous
62 Anthropogenic threats; however, their conservation is hindered by numerous knowledge gaps. In
63 particular, the dearth of genetic data, unique among all big cat species, hinders a full
64 understanding of their population structure, historical population size, and current levels of
65 genetic diversity. Here, we use whole-genome sequencing data for 41 snow leopards (37 newly
66 sequenced) to offer new insights into these unresolved aspects of snow leopard biology. Among
67 our samples, we find evidence of a primary genetic divide between the northern and southern
68 part of the range around the Dzungarian Basin—as previously identified using landscape models
69 and fecal microsatellite markers—and a secondary divide south of Kyrgyzstan around the
70 Taklamakan Desert. Most noteworthy, we find that snow leopards have the lowest genetic
71 diversity of any big cat species, likely due to a persistently small population size throughout their
72 evolutionary history rather than recent inbreeding. We also find that snow leopards have
73 significantly less highly deleterious homozygous load compared to numerous *Panthera* species,
74 suggesting effective purging during their evolutionary history at small population sizes. Without a
75 large population size or ample standing genetic variation to help buffer them from any
76 forthcoming Anthropogenic challenges, snow leopard persistence may be more tenuous than
77 currently appreciated.

78 **Significance Statement**

79 Through an international effort, we have generated whole-genome sequencing data for 37 snow
80 leopards, increasing the number of snow leopards sequenced by an order of magnitude. We have
81 used this data to show that snow leopards have the lowest genetic diversity of any big cat
82 species, even the cheetah. We do not see evidence of high levels of inbreeding or genetic load in
83 snow leopards compared to other *Panthera* species; however, their dearth of genetic diversity
84 and small population size should be kept in mind when assessing their risk of extinction in the
85 Anthropocene.
86

87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140

Main Text

Introduction

Residing in some of the most extreme and remote areas of the world, snow leopards (*Panthera uncia*) are rarely encountered and are challenging to study, making them one of the most enigmatic of the large charismatic mammals. They are among the largest carnivores in the high-elevation habitat in which they reside and their persistence relies on healthy mountain ungulate populations (1) sometimes supplemented by livestock (2–9). Snow leopard habitat consists of mountainous areas of Asia, spanning 12 countries (Fig. 1A), habitat that offers immense ecosystem services—acting as an important source of carbon storage (10) and providing water to almost two billion people. Snow leopards serve as an umbrella species whose conservation benefits this globally crucial Asian mountain ecosystem. In spite of its apparent benefits, snow leopard conservation is impeded by the many knowledge gaps regarding this elusive species (11).

Snow leopards were listed as Endangered by the International Union for Conservation of Nature (IUCN) for 45 years but were downlisted to Vulnerable in 2017 as they did not meet specific criteria for population size (fewer than 2,500 mature individuals) and percent population decline (more than 20% over two generations) for Endangered status. This change of status has been controversial (12, 13) as snow leopard numbers are presumed to be in decline due to habitat loss, decreased availability of primary prey (high-elevation, mountain-dwelling ungulates), retaliatory killings for livestock predation (14), and poaching for their skins (15). As climate change in high mountain Asia is occurring at an even more rapid rate than elsewhere in the Northern hemisphere, excluding the Arctic (16), it is also likely to become an increasing threat to snow leopards (17).

Currently, the global snow leopard population size is estimated to be anywhere from 4,700 to 7,500 individuals and little is known about their historical population size and range (18) or their current population trends (19). While many other big cat species experienced historical declines due to range contractions during the Last Glacial Maximum (20, 21) and are facing contemporary human-driven declines (22, 23), it remains unclear if the snow leopard was previously more abundant than is currently estimated. One study using fecal microsatellite data suggests that snow leopards may have undergone a bottleneck ~8,000 years ago (24). Estimates of genetic diversity and levels of inbreeding are also limited due to a dearth of genomic data for the species. They are the least studied genetically of all the big cat species (25) with whole-genome sequencing (WGS) data available for only two wild snow leopards (26) and two captive individuals (27) prior to this study. Genetic diversity assessed using fecal microsatellite data (24, 28–31) as well as genomic data from one of the previously sequenced wild snow leopards suggests low diversity (26), but genomic data from additional samples are required to determine if this is a characteristic of the species across its range.

Additionally, there is still more to learn about snow leopard population structure and connectivity (32). Numerous studies have used microsatellite markers from fecal samples to assess genetic structure and connectivity at varying geographic scales. These studies have found connectivity at the local scale to be location dependent, with evidence of continuous habitat connectivity across at least 75 km in Pakistan (31) and >1000 km in Mongolia (29), weak genetic differentiation among snow leopards samples across 400 km of mountainous terrain in Northern China (30), and signs of genetic structure between sampling areas about 500 km apart in Russia (28). At a larger geographic scale, fecal microsatellite studies (24, 28, 29) and other lines of evidence (33, 34) suggest a geographic divide between Mongolia/Russia and the southern part of the range due to the Dzungarian Basin and Gobi Desert. However, snow leopards are known to cross long distances between mountain ranges (35, 36). In addition to this prominent north-south divide, the largest scale fecal microsatellite study also identified a second divide within the southern group separating the east and west of the Himalayas-Tibetan Plateau complex (24). Janecka et al. (24) argue that each of these three distinct groups constitute unique subspecies;

141 however, this subspecies designation and the level of connectivity across the landscape remains
142 controversial (37, 38). Population structure and connectivity can be assessed with greater
143 resolution using whole-genome data (39); however, until now, this has not been possible.

144 In addition to the wild snow leopard population, the international community has worked
145 for decades to establish a sustainable zoo population. As of 2008 there were 445 snow leopards
146 across 205 institutions globally, not including China, representing the genetic diversity of 56 wild
147 founders (40), most of which came from the wild in the 1960s-1990s, often from unknown
148 locations. As it is the goal of zoos to maintain a genetically diverse population (41) and to act as
149 reserves of genetic diversity for endangered species, it is important to know what portion of the
150 global genomic diversity of snow leopards this population represents.

151 Here, we generate WGS data for 33 wild snow leopards from multiple locations across
152 their range in addition to four captive snow leopards from the North American zoo population. We
153 combine this novel data with existing data for four individuals to achieve three main objectives: 1)
154 characterize snow leopard population structure and connectivity to see how estimates from WGS
155 data compare to fecal microsatellite data; 2) assess the current level of genomic diversity in snow
156 leopards, how this compares to other big cat species, and how this relates to historical population
157 size, inbreeding levels, and genetic load; and 3) assess the ancestry of the current zoo snow
158 leopard population. Among our samples, which do not include most of the central/southern part of
159 the range, we find evidence of the previously identified genetic divide between the north and
160 south, as well as a divide between the Kyrgyz population and populations farther south. However,
161 low levels of genetic differentiation among groups suggests some level of connectivity. We also
162 find the current North American zoo population to be dominated by Kyrgyzstan-region ancestry.
163 Most notably, we find snow leopards to be the least genetically diverse contemporary big cat
164 species, likely due to a persistently small population throughout their evolutionary history rather
165 than to recent inbreeding. Additionally, we find snow leopards to have significantly lower large-
166 effect homozygous genetic load compared to many other *Panthera* species suggesting purging of
167 highly deleterious recessive mutations throughout their evolutionary history. We believe these
168 results have significant implications for snow leopard conservation.

169 170 **Results**

171
172 After filtering samples based on sequence quality and breadth of coverage we were left
173 with a final dataset consisting of 37 snow leopard samples. This final dataset included 34
174 samples from our newly generated data with an average individual depth of coverage of 7.3X
175 (minimum of 3X and maximum of 16.8X) and three previously published samples with an
176 average depth of coverage of 23X (minimum of 12X and maximum of 28.9X). These samples
177 consisted of 32 wild born snow leopards representing seven countries (Mongolia, Russia,
178 Kyrgyzstan, Tajikistan, Afghanistan, Pakistan, and India) (Fig. 1B) and five captive samples
179 (Supplementary Table 1).

180 We mapped these data to the snow leopard reference genome (NCBI accession
181 PRJNA602938) (27) and called single nucleotide polymorphisms (SNPs) as described in the
182 methods, resulting in a final SNP set of 1,591,978. Within this final dataset, we identified one pair
183 of first-degree relatives and three pairs of second-degree relatives (all within the wild samples).
184 When conducting analyses that could be affected by the presence of related individuals, we
185 removed one representative from each of these four related pairs (the sample with lower
186 sequencing coverage), leaving a total of 28 wild and five captive snow leopards. Details of which
187 samples were included in each analysis and why are shown in Supplementary Table 5.

188 189 **Population structure and dispersal barriers:**

190 We assessed population structure among our samples using principal component
191 analysis (PCA). PCA results indicated that the one Indian sample (12X coverage), and to a lesser
192 extent, one of the Tajikistan samples (U13, 6.7X coverage) were genetically distinct from all of
193 the other samples (Fig. 2A). In order to more clearly visualize the groupings among the other

194 samples, we also visualized the PCA with the Indian and U13 sample removed. In this PCA, three
195 distinct groups are apparent—Mongolia and Russia; Kyrgyzstan and captive; and Tajikistan,
196 Afghanistan and Pakistan (Fig. 2B).

197 We also investigated population structure using Admixture (42) to test models with one to
198 ten ancestral populations ($K=1-10$) over ten independent runs. Due to issues that can arise with
199 having a small sample size for a genetically distinct group (43), we first ran Admixture without the
200 Indian sample (Fig. 2C). At $K=2$, Mongolia and Russia separated clearly from all other samples
201 with the same ancestry assignments supported in all ten iterations. At $K=3$, nine of the ten
202 iterations supported a clear separation of the samples into three groups—Mongolia and Russia;
203 Tajikistan, Afghanistan, and Pakistan; captives and Kyrgyzstan (Fig. 2C, Supplementary Fig.
204 1B)—recapitulating what was observed in the PCA. Admixture and PCA results were robust to the
205 removal of captive samples from the analyses (Supplementary Fig. 2).

206 When including the Indian sample, Admixture had difficulty assigning it at $K=2$ with five
207 runs grouping it with the southern samples and five runs showing ancestry split between the north
208 and the south (Supplementary Fig. 1A). Maximum likelihood phylogeny construction also showed
209 the Indian sample to be genetically distinct (Supplementary Fig. 3). Thus, PCA, Admixture and
210 phylogenetic results all corroborate that the one Indian sample included in this study is genetically
211 distinct from all of the other samples.

212 Among the other samples (excluding India), Admixture and PCA results suggest three
213 genetically distinct groups (Fig. 2B,C). We see a primary genetic divide between the north
214 (Mongolia and Russia) and the south (all other samples) with a secondary divide within the
215 southern group between Kyrgyzstan and populations farther south (Afghanistan, Tajikistan, and
216 Pakistan). These results also show the five captive samples, whose lineages encompass more
217 than half of all of the founders of the current captive population (33 of 56 founders are
218 represented, Supplementary Fig. 4), group most closely with the Kyrgyzstan samples.

219

220 **Assessment of gene flow:**

221 Based on Admixture and PCA results, we quantified population structure between the two
222 groups identified in Admixture at $K=2$ and among the three groups identified at $K=3$ (Fig. 2C) by
223 assessing shared versus private SNPs, F_{ST} , and rare allele sharing. We excluded captive
224 samples and the Indian sample from these analyses. We also excluded one sample from each
225 related pair (U01, U08, KGZ_F4, and AF_06) from F_{ST} and rare allele sharing analyses, but not
226 shared versus private SNP assessments.

227 At $K=2$, we compared groups that we will refer to as “North” (consisting of Russia and
228 Mongolia) and “South” (consisting of Kyrgyzstan, Tajikistan, Afghanistan and Pakistan). Each
229 group (downsampled to $n=15$) had more shared SNPs (598,449) than private SNPs (379,861
230 private to the North and 364,010 private to the South) (Supplementary Fig. 5A). At $K=3$, we
231 compared groups that we will refer to as “North” (consisting of Russia and Mongolia),
232 “Kyrgyzstan”, and “Far South” (consisting of Tajikistan, Afghanistan and Pakistan). We
233 downsampled the North and Far South groups to seven individuals such that all three groups had
234 the same sample size. Here, and in all analyses requiring downsampling, first, samples that were
235 geographically close were thinned, then samples were chosen for removal based on coverage,
236 such that the most geographically unique and highest coverage samples were retained
237 (Supplementary Table 5). We found that the Far South and North group had more than twice as
238 many private SNPs (222,569 and 219,175, respectively) as Kyrgyzstan (111,392)
239 (Supplementary Fig. 5B) and that Kyrgyzstan shared a similar number of SNPs with both the
240 North and the Far South (394,020 shared among all three groups, 65,181 shared only between
241 Kyrgyzstan and North, 64,871 shared only between Kyrgyzstan and Far South). It is worth noting
242 that the Kyrgyz group represents the smallest geographic area and this limitation in sampling
243 could contribute to our observation of fewer private SNPs in this group.

244 We calculated Weir and Cockerham’s weighted pairwise F_{ST} after removing one
245 individual for each first and second degree related pair and downsampling groups to equal size.
246 At $K=2$, F_{ST} between the North and South was 0.091. At $K=3$, the pairwise F_{ST} between the North
247 and Far South was 0.123, between North and Kyrgyzstan was 0.115, and between Kyrgyzstan

248 and the Far South was 0.093. Although F_{ST} can not be directly compared between species, for
249 rough context, these values fall below F_{ST} observed between different tiger subspecies, which
250 range from 0.164 to 0.318, and are on par with the lower bound F_{ST} observed between Bengal
251 tiger sub-populations, which range from 0.094 to 0.3 (44). Observed pairwise F_{ST} at K=3 was
252 compared to a null distribution representing panmixia created by randomly shuffling population
253 assignments 10,000 times and recalculating F_{ST} between populations. This analysis showed
254 observed F_{ST} values, although somewhat low, to be highly significant ($p < 0.0001$) (Supplementary
255 Fig. 6).

256 We identified doubletons (SNPs with a minor allele count of two) where each of the two
257 minor alleles occurs in separate individuals. We then assessed doubleton sharing (45, 46) among
258 the three groups—North, Kyrgyzstan, and Far South—after downsampling each group to six
259 individuals. Rare variants, such as doubletons, are typically the most recent variants. In human
260 populations, doubletons are found most commonly within the same population, and doubleton
261 sharing between populations is interpreted to reflect recent connectivity between those
262 populations (47). Here, in addition to calculating doubleton sharing among snow leopard groups,
263 we also compared the observed doubleton sharing to a null distribution representing panmixia in
264 the same way as pairwise F_{ST} —sample population assignments were randomly shuffled and
265 observed doubleton sharing between populations was recalculated 10,000 times. Doubleton
266 sharing corroborated the pre-identified groups with individuals sharing a significantly higher
267 fraction of doubletons (0.40-0.72) with individuals within the same group (p-value North - 0.0006,
268 Kyrgyzstan - 0.008, Far South - 0.01) (Fig. 2D). However, the fraction of doubletons that each
269 individual shared with each of the other groups was 0.11-0.32. The Kyrgyzstan group showed the
270 highest fraction of doubleton sharing with outside groups— sharing enough with the Far South to
271 not be significantly different from panmixia ($p = 0.2$). Both Kyrgyzstan and Far South also share
272 enough rare alleles with the North to only be marginally significantly different from panmixia
273 ($p=0.04$ and $p=0.03$, respectively). These results support the presence of genetic divergence
274 among these three groups, but also indicate some amount of gene flow among the groups. Note
275 that our assessments of gene flow are limited by our small sample size and will benefit from
276 additional samples in future studies
277

278 **Heterozygosity and historical population size:**

279 We used publicly available data to call SNPs in all big cat species using the Genome
280 Analysis Toolkit (GATK) (48) and calculated heterozygosity using VCFtools (49). Here, we
281 defined big cat species as species with an average adult body weight of 40kg or more, which
282 includes all *Panthera* species (lion (*P. leo*), tiger (*P. tigris*), leopard (*P. pardus*), jaguar (*P. onca*),
283 and snow leopard) as well as cheetah (*Acinonyx jubatus*) and puma (*Puma concolor*) (50). We
284 found snow leopards to have the lowest heterozygosity of any big cat species, with
285 heterozygosity for every snow leopard sample included in this study falling lower than that
286 observed in any other big cat (Fig. 3A). Notably, snow leopard heterozygosity was lower than that
287 of cheetahs, which have long been considered the archetype of low heterozygosity in big cats
288 (22, 51). The relative values of observed heterozygosity among all the other species for which we
289 calculated heterozygosity were consistent with previous work (23, 52–54).

290 We used pairwise sequentially Markovian coalescent (PSMC) (55) to reconstruct
291 historical effective population size using the highest coverage individual from each genetically
292 distinct group (North-29X, Kyrgyzstan/captive-28X, Far South-9X, India-12X). Reconstructions
293 showed consistent results across all populations sampled and across all bootstrap replicates (Fig.
294 3B). All reconstructions show a consistently small effective population size over the last ~900,000
295 years with snow leopard effective population size never exceeding 28,000 individuals.
296 Reconstructions suggested that snow leopards had an effective population size of 13,000-17,000
297 individuals from ~300,000-75,000 years ago and underwent a slow decline between ~75,000-
298 30,000 years ago to an effective population size of ~600-8,000 individuals. This decline is
299 coincident with the maximum extent of glaciation in mountainous areas of Asia during the Last
300 Glacial Maximum (~40,000-100,000 years ago) which was well before the global Last Glacial
301 Maximum (~20,000 years ago) (56). Effective population size is generally smaller than census

302 size (57) and PSMC historical effective population size estimates can be impacted by numerous
303 species-specific parameters (e.g., population structure, inbreeding, mutation rate, generation
304 time) (58, 59) which have not been thoroughly characterized in snow leopards, and thus exact
305 effective population size estimates should be interpreted cautiously.

306

307 **Inbreeding and runs of homozygosity (ROH):**

308 Using the same dataset that we used to calculate genome-wide heterozygosity across
309 big cat species, we also calculated the inbreeding coefficient (F) across big cats. These results
310 showed the inbreeding coefficient of snow leopards was not significantly higher than other big
311 cats and was even significantly lower than Asian leopard and puma indicating that the lower
312 genetic diversity observed in snow leopards is not explained by higher inbreeding (Fig. 4A).

313 We also used the software GARLIC (60) to assess ROH in snow leopard samples with
314 8× sequencing coverage or more. ROH of different sizes are likely the results of different
315 processes— long ROH reflects recent inbreeding while shorter ROH can reflect shared population
316 history or background levels of relatedness due to small population size. GARLIC identified
317 130bp to be the ideal window size for our dataset. GARLIC uses a sliding window along the
318 chromosome and makes a logarithm of the odds (LOD) calculation of autozygosity for each
319 window which is an estimate of how likely homozygous regions are to be identical by descent. We
320 divided ROH into four size bins – 0.1-1Mb, 1-5Mb, 5-10Mb, and >10Mb – and the proportion of
321 the genome in each size bin for each individual was calculated by dividing the total sequence
322 length in each size bin by the total mappable length of the genome (1,818,166,894bp). These
323 proportions were compared to those observed in tigers using GARLIC output provided in
324 Armstrong et al. (54).

325 We found the general trends in the proportion of each ROH size class to be different
326 between snow leopards and tigers, which are known to have undergone recent inbreeding due to
327 small population sizes (61). Across all samples, snow leopards had a larger proportion of the
328 genome (average of 16%) in short ROH (0.1-1Mb) compared to tigers (average of 3%). On
329 average, snow leopards also had a much lower proportion of their genome in ROH longer than
330 10Mb (<1% on average) compared to tigers (9% on average). The average proportion of the
331 genome in intermediate sized ROH (1-10Mb) was similar in snow leopards (12%) and tigers
332 (11%) (Fig. 4B).

333

334 **Genetic load:**

335 We used our SNP calls from across the five *Panthera* species to assess how genetic load
336 in snow leopards compares to the other *Panthera* species. Only snow leopard samples with 8×
337 coverage or higher and one representative from each related pair were used for this analysis. We
338 used the software SnpEff (62) to annotate SNPs and the protocol outlined in (63) to identify
339 derived SNPs in each species. We used this information to filter our dataset to only derived SNPs
340 in protein coding transcripts and counted the total number of SNPs, highly deleterious SNPs, and
341 moderately deleterious SNPs in the homozygous and heterozygous state in each sample. We
342 used these counts to calculate what proportion of each individuals' total homozygous and
343 heterozygous SNPs were highly deleterious and what proportion were moderately deleterious.
344 Looking at the proportion of deleterious mutations out of the total number of protein-coding
345 mutations in this way allows us to assess the relative load across *Panthera* while accounting for
346 variability among the species, such as evolutionary rates, and is ideal for comparisons between
347 species (64).

348 We investigated homozygous and heterozygous load separately because highly
349 deleterious mutations are more likely to be recessive (65), resulting in purging of highly
350 deleterious mutations only in the homozygous state. Additionally, we expect to see a different
351 trend when looking at moderately deleterious mutations compared to highly deleterious
352 mutations—while small populations are better at removing highly deleterious mutations through

353 purifying selection, this is not the case with mildly deleterious mutations where drift will dominate
354 in small populations (66).

355 We found that snow leopards had significantly less derived highly deleterious genetic
356 load than numerous *Panthera* species/subspecies– African leopard, jaguar, lion, and Sumatran
357 tiger– in the homozygous state (Fig. 4C), and no significant differences in the heterozygous state
358 (Fig. 4D). Conversely, snow leopards had slightly more derived moderately deleterious genetic
359 load than lions and Bengal tigers in the homozygous state (Supplementary Fig. 9A) and no
360 significant differences in the heterozygous state (Supplementary Fig. 9B).

361

362 Discussion

363

364 Population structure and dispersal barriers:

365 We used WGS from 37 snow leopards to investigate population structure of the species.
366 Our results corroborate results from previous studies. Among our samples, which do not include
367 the southeast part of the range, we identify three genetically distinct groups. Admixture and PCA
368 results identify the most pronounced divide among our samples to occur between the northern
369 and southern part of the range around the Dzungarian Basin (Fig. 2B,C), consistent with previous
370 microsatellite analyses (24, 28, 29) and models (33, 34). Admixture and PCA results also identify
371 a secondary divide occurring south of Kyrgyzstan (Fig. 2B,C), around the Taklamakan Desert,
372 consistent with previous microsatellite analyses (28). Both of these genetic divides are also
373 supported by highly significant F_{ST} values between groups (Supplementary Fig. 6).

374 These results, in combination with significant F_{ST} among these regions, suggest that the
375 Dzungarian Basin and Taklamakan Desert present barriers to dispersal for snow leopards.
376 However, the level of genetic differentiation among these regions was modest as measured
377 through shared versus private alleles (Supplementary Fig. 5), and shared rare alleles (Fig. 2D)
378 suggesting that there is some level of connectivity among these groups. This is consistent with
379 previous fecal microsatellite work which has also found evidence of weak connectivity between
380 Mongolia and China (29), as well as between Kyrgyzstan and Mongolia/Russia and between
381 Kyrgyzstan and Tajikistan (28). Additional samples will be necessary to confidently estimate gene
382 flow in future analyses.

383 Population structure analyses suggest that India is genetically distinct from all other
384 samples (Fig. 2A, Supplementary Fig. 3). The uniqueness of this sample could suggest that this
385 individual is our one representative of the southeastern phylogenetic group (*P. u. uncioides*, Fig.
386 1B) suggested by Janečka et al. (24), but more samples from this area will be necessary to
387 resolve how genetically distinct Indian snow leopards may be.

388

389 Heterozygosity, demographic history, ROH, and genetic load:

390 Our results show that snow leopards have the lowest genetic diversity of any big cat
391 species (Fig. 3A), lower than previously appreciated (23, 67). Demographic history and ROH
392 assessments indicate the exceptionally low heterozygosity in snow leopards is likely due to a
393 persistently small population size over the last 900,000 years (Fig. 3B) rather than recent
394 inbreeding events. Our demographic history assessment (Fig. 3B) did not pick up the more recent
395 population bottleneck ~8,000 years ago suggested by Janečka et al. (24) from microsatellite data
396 likely because PSMC analyses lose power at more recent time scales.

397 With a persistently small population size and low genetic diversity throughout their
398 evolutionary history, snow leopards serve as yet another example that high genetic diversity is
399 not a requirement for the long-term persistence of a species (68–73). Such long term persistence
400 has been suggested to be facilitated by purging of deleterious mutations over long time scales in
401 relatively small populations (68–70, 73). Our assessment of genetic load across *Panthera*
402 supports this hypothesis, showing that snow leopards have a significantly smaller proportion of
403 highly deleterious homozygous SNPs than many other *Panthera* species (Fig. 4C). We also find
404 that snow leopards have slightly more moderately deleterious homozygous SNPs than some
405 other *Panthera* species (Supplementary Fig. 9), which is also consistent with a long-term small
406 population size resulting in less effective purging of mildly deleterious mutations.

407 Consistent with demographic history assessments, ROH analyses show snow leopards
408 have a greater proportion of their genome in shorter ROH (Fig. 4B), which likely reflects shared
409 population history and small historic population size (74). Conversely, many tiger subspecies,
410 which are known to have high levels of recent inbreeding due to small isolated populations (61),
411 show the opposite trend, with the highest proportion of their genome in longer ROH (54) (Fig. 4B).
412 Taken together, all of our results—heterozygosity, demographic history, ROH, and load—are all
413 consistent with snow leopards having a persistently small population size.

414 Note that during the review process of this manuscript, a different study (75), using
415 partially overlapping samples with ours, published results consistent with what we have presented
416 here – very low heterozygosity in snow leopards, purging of highly deleterious variants, and
417 lengths of ROH indicating historic inbreeding due to long-term small population size rather than
418 recent inbreeding.

420 **Origin of captive samples:**

421 Studbook-based pedigrees for the five captive individuals sequenced in this study show
422 that the origin of their wild ancestors is mostly unknown. The few ancestors of known origin are
423 documented as originating from the USSR, Kyrgyzstan, Kazakhstan, and the western Himalayas
424 (Supplementary Fig. 4). Admixture and PCA results both show a strong signal of genetic similarity
425 between the five captive individuals and the current Kyrgyzstan population (Fig. 2A-C). The
426 current Kyrgyzstan population is likely genetically similar to that in Kazakhstan and northeastern
427 China, for which we have no genomic data. Thus, these results suggest that the ancestors of the
428 captive individuals largely came from what is currently Kyrgyzstan, or the surrounding area, and
429 fewer ancestors have come from other portions of the range. Given that it is a goal of zoos to
430 maintain a genetically diverse population and potentially be a reserve of genetic diversity, the fact
431 that the current North American zoo population only represents a small subset of the current wild
432 diversity indicates that this captive population should not be perceived as a reserve of rangewide
433 wild genetic diversity.

435 **Conservation implications:**

436 Snow leopards live in arid, cold, low-productivity, high-elevation habitats where few
437 species can persist – an environment that has evidently only ever been able to support a limited
438 number of snow leopards (Fig. 3B). As a result, our data shows that they have likely always had
439 an effective population size much lower than other big cats (as suggested in Pečnerová et al. (23)
440 and Cho et al. (26)), and harbor less genetic diversity than even the cheetah (Fig. 3A).

441 Thanks to their extreme environment, snow leopards have not yet been exposed to the
442 same level of acute Anthropogenic pressures as have big cats living in habitats more easily
443 accessible to humans. Yet, even having been spared the most intense human impacts, our data
444 indicates that they already have extremely low genomic diversity and population sizes compared
445 to other big cats. Although these characteristics have not hindered their long-term persistence
446 thus far, and have likely resulted in the purging of highly deleterious mutations from the
447 population (Fig. 4C), this means that snow leopards cannot rely on a large population size or
448 standing genetic variation to help them survive any forthcoming Anthropogenic challenges, as
449 other big cats have done (76, 77). Additionally, snow leopards carry similar amounts of highly
450 deleterious heterozygous load as other *Panthera* (Fig. 4D), load that would become unmasked
451 with future inbreeding, suggesting that snow leopards are just as vulnerable to future inbreeding
452 depression as other *Panthera* species. Unfortunately, the most intense Anthropogenic pressures
453 may lie ahead for snow leopards. Anthropogenic climate change threatens to shrink snow leopard
454 range through habitat change (17) and more intense interspecific competition (e.g. (78), (79));
455 shifting grazing practices risk facilitating spillover of novel pathogens from domestic animals into
456 snow leopards and their prey (80); and accelerating mining, energy, and infrastructure
457 development threaten to fragment and degrade previously remote snow leopard habitats (81, 82).
458 Protection of snow leopards and their habitat, to the greatest degree possible, will be pivotal as
459 this low-density carnivore appears genetically and demographically ill-equipped to bounce back
460 from Anthropogenic perturbations (83).

461

462

Materials and Methods

463

464

Sample collection and sequencing:

465

466

467

468

469

470

471

472

Through an international collaborative effort, we collected a total of 37 snow leopard blood or tissue samples, composed of 32 wild caught samples from five countries (Afghanistan, Kyrgyzstan, Tajikistan, Mongolia, and Russia), one currently captive but wild-born individual from Pakistan (included in the wild group throughout the manuscript), and four captive samples from mixed/unknown ancestry. Sample details, including DNA extraction and library preparation methods used, can be found in Supplementary Table 1. We sequenced all samples on an Illumina sequencing platform using paired end 150bp reads with an aim of sequencing each sample to ~5-8X coverage.

473

474

475

476

477

All samples from captive individuals were collected as part of routine animal care in the respective Association of Zoos and Aquariums zoos and all wild samples were collected as part of ongoing monitoring of snow leopards by local conservation groups with all necessary collection permits. Samples were processed in labs in the United States, France, and Russia and were transported to respective locations with appropriate transport permits.

478

479

480

481

482

Additionally, we collected all currently published WGS data for snow leopards, which included data for two captive individuals (NCBI accessions SRR16227515 (27) and SRR12437590), one wild individual from Mongolia (NCBI accession SRR836372 (26)), and one wild individual from India (NCBI BioProject PRJNA1051290). In total, we gathered WGS data for 41 snow leopards.

483

484

485

486

Note that there is WGS data for an additional sample identified as a snow leopard on Genbank (biosample SAMN17432540) that we did not use in this study because we concluded that this data was an Asian leopard (*P. pardus*) (Supplementary Methods and Supplementary Fig. 7).

487

488

489

We were unable to include any samples from a large portion of the snow leopard range in the southeast, but are hopeful that data from this area will be available in future analyses.

490

Variant calling:

491

492

493

494

495

496

497

Reads were mapped to the snow leopard reference genome (NCBI accession PRJNA602938 (27)). Mapping, using BWA-MEM (84), and SNP calling, using GATK (85), were performed by Gencove Inc., a service provider. We calculated the depth and breadth of coverage for each sample from BAM files using SAMtools (86) (Supplementary Table 1). Variants were filtered for quality as described in detail in the Supplementary Methods resulting in a final set of 1,591,978 SNPs.

498

Relatedness assessment:

499

500

501

502

503

504

505

506

507

We estimated relatedness among samples using SNPrelate (87) in R (88) as described in detail in the Supplementary Methods. We identified one sample pair with a kinship coefficient greater than that expected from first-order relatives—U01 and U09 (kinship coefficient of 0.335), and we identified three sample pairs with kinship coefficients consistent with second-order related pairs—SL_KGZ_F1 and SL_KGZ_F4 (0.148), AF_SL_07 and AF_SL_06 (0.135), and U14 and U08 (0.124). We removed the sample with lower sequencing coverage from each of these pairs (U01, SL_KGZ_F4, AF_SL_06, and U08) from indicated analyses. All sample pairs with non-zero kinship coefficients are listed in Supplementary Table 2.

508

Pedigrees for captive-bred samples:

509

510

511

512

513

514

Using snow leopard studbooks (40, 89), we compiled information on all of the ancestors of the five captive-born individuals included in this study and drew pedigrees with the kinship2 package (90) in R (Supplementary Fig. 4). The wild origin of their ancestors is mostly unknown; however, each sample has a few ancestors for which there is a wild origin location listed (Supplementary Fig. 4), these origin locations include the USSR (between 1964-1974), Kazakhstan (in 1972), Kyrgyzstan (between 1974-1980), Przewalsk (between 1974-1975, which

515 we believe to refer to Karakol, Kyrgyzstan which was previously named Przewalsk), and Aksai (a
516 contested region between China and India in the Western Himalayas, in 1979).

517

518 **Population structure assessments:**

519 All population structure assessments were conducted on the dataset after filtering for first
520 and second degree relatives (N=33, 1,448,657 SNPs). We conducted PCA using PLINK2 (91),
521 admixture analyses using Admixture (42), and constructed a phylogenetic tree using IQtree (92).
522 PCA and admixture analyses were also conducted without the India samples as well as without
523 the captive samples. Details can be found in the Supplementary Methods.

524

525 **Quantifying population structure:**

526 We further characterized population divides identified in Admixture and PCA by
527 calculating the number of shared versus private SNPs among groups using BCFtools (93),
528 pairwise F_{ST} using VCFtools, and the rate of rare variant sharing among groups using VCFtools
529 and PLINK. We excluded captive samples from these analyses as well as the sample from India
530 as the PCA indicates that this sample is genetically distinct from all groups. Details of these
531 analyses can be found in the Supplementary Methods.

532

533 **PSMC:**

534 We used PSMC (55) to estimate snow leopard effective population size back in time
535 using the highest coverage sample for each population cluster. Starting with BAM files filtered to
536 include only putative autosomes, we used SAMtools *mpileup* and BCFtools *call* to generate a
537 VCF. We then used *vcfutils.pl vcf2fq* to generate diploid FASTQ files using the '-D' flag to set the
538 maximum read depth to twice the average depth for each sample and the '-d' flag to set the
539 minimum depth to a third of the average depth for each sample, as recommended by PSMC for
540 generating PSMC input. We then ran PSMC with the default settings. We also performed 30
541 rounds of bootstrapping using random sampling with replacement for each sample. For plotting,
542 we used a mutation rate of 3.62×10^{-9} per site per generation as calculated by Armstrong et al.
543 2025 (94) and a generation time of five years as suggested for snow leopards by Cho et al. (26).

544

545 **Heterozygosity in other big cats:**

546 We calculated heterozygosity in all big cat species using publicly available data in order
547 to see how snow leopard heterozygosity levels compared to other big cat species. We included
548 all other species in the genus *Panthera* (leopard, lion, tiger, jaguar) as well as cheetah and puma.
549 Accession numbers of publicly available WGS data and reference genomes used in this analysis
550 are listed in Supplementary Table 3 and 4, respectively. We mapped all FASTQ data to the
551 corresponding reference genome using BWA-MEM (84), called SNPs using GATK (85), and
552 filtered the resultant VCF as described in detail in the supplementary methods.

553 We calculated the mappable length of each genome using mappability BED files to
554 calculate the number of nucleotides with a mappability score of one. Using the filtered VCF for
555 each species or group, we calculated observed homozygosity for each sample using VCFtools
556 (49) with the flag '--het'. In R, we calculated the number of heterozygous sites by subtracting the
557 number of observed homozygous sites column ((O)HOM) from the total number of sites column
558 (NSITES). We then calculated heterozygosity by dividing the number of heterozygous sites by the
559 length of the genome consisting of putative autosomes with a mappability score of one. We used
560 ggplot2 in R to create boxplots of heterozygosity results.

561

562 **Heterozygosity in snow leopards:**

563 Although SNPs had already been called by Gencove Inc. (as described above), we re-
564 called SNPs from the snow leopard dataset using the same pipeline used to call SNPs in all of
565 the other big cat species to ensure comparability of heterozygosity calculations among species.
566 We calculated observed heterozygosity in snow leopards using both SNP datasets (Gencove's
567 and ours) in the same way as described above. We calculated heterozygosity for all individuals,

568 regardless of relatedness; however, because heterozygous SNPs are most accurately called with
569 higher coverage data (95), we only used samples with 8× coverage or more.

570 We compared snow leopard heterozygosity calculated from SNPs called using our
571 pipeline to heterozygosity calculated from SNPs called by Gencove using Pearson correlation
572 coefficient calculated using the ggpubr (96) package in R (Supplementary Fig. 8). We found
573 heterozygosity calculated from the two different SNP calling pipelines to be extremely correlated
574 ($R = 0.97$, $p=5.1E-9$); however, all heterozygosity estimates calculated from our SNP calls were
575 slightly higher (average of ~ 0.000056) than that estimated from Gencove SNP calls.

576

577 **Runs of homozygosity:**

578 We converted VCF files to transposed PLINK files using PLINK1.9 with the flags ‘--allow-
579 extra-chr --const-fid 0 --recode transpose’. We then inferred ROH in each individual using the
580 software GARLIC (60) with an error rate of 0.001 and the centromere location of each
581 chromosome set to 0, 0; since centromere location is unknown. In addition, we used ‘--auto-win-
582 size, --auto-overlap-frac, --winsize 100’ to allow GARLIC to determine the ideal window size to
583 use.

584 We divided ROH into four size bins—0.1-1Mb, 1-5Mb, 5-10Mb, and >10Mb, and then
585 calculated the proportion of the genome in each size bin by dividing the total sequence length in
586 each size bin by the total sequence length of putative autosomes with a mappability score greater
587 than one. The proportion of the genome in each size bin was visualized in boxplots using ggplot2
588 in R.

589 We pulled ROH values for tigers, calculated using GARLIC, from Armstrong et al. (54)
590 and calculated the proportion of the genome in each size bin of ROH for each individual in the
591 same way as described above, but using the total mappable length across all autosomes in the
592 tiger reference genome.

593 As with heterozygosity assessments in snow leopards, in order to limit any biases caused
594 by allelic drop out in lower coverage samples, we only used samples with 8× coverage or higher
595 in ROH assessments in both snow leopards and tigers.

596

597 **Inbreeding across big cats:**

598 We calculated the inbreeding coefficient F in all big cat species using the same SNP
599 datasets used to calculate heterozygosity. As with heterozygosity assessments, we only
600 calculated inbreeding for snow leopards samples with 8× coverage or higher. We calculated the
601 inbreeding coefficient F using the method of moments ($(\text{observed homozygous count} - \text{expected}$
602 $\text{count})/(\text{total observations} - \text{expected count})$) using VCFtools with the flag ‘-het’.

603

604 **Genetic load across *Panthera*:**

605 We assessed genetic load across *Panthera* using the same SNP datasets used to
606 calculate heterozygosity. We first built a new database for each *Panthera* species in SnpEff (62)
607 using an annotated genome for each species. We then used SnpEff to characterize the effect of
608 each SNP in each species. In the case of tiger and lion, the reference genome used to call SNPs
609 was not annotated, so we first used LiftOverVCF (97) to project our SNP calls onto the annotated
610 tiger and lion genome, respectively. We used the pipeline from (63) to identify derived SNPs by
611 comparing across the *Panthera* clade. We used this information to limit our SNP dataset to only
612 derived SNPs in protein coding transcripts and counted the number of total SNPs, high impact
613 SNPs, and moderate impact SNPs in the heterozygous and homozygous state in each individual.
614 Highly deleterious homozygous load in each individual was calculated as the number of high
615 impact homozygous SNPs divided by the total number of homozygous SNPs in that individual.
616 Highly deleterious heterozygous load was calculated for each individual by dividing the number of
617 high impact heterozygous SNPs by the total number of heterozygous SNPs. Moderately
618 deleterious homozygous and heterozygous load was calculated in the same way. Additional
619 details can be found in the Supplementary Methods.

620

621 **Samples included in each analysis:**

622 We provide a detailed list of which samples were included in each analysis in
623 Supplementary Table 5.

624

625

626 **Data Availability:**

627 Data associated with this study has been deposited into bioproject PRJNA1048427 and will be
628 released upon publication.

629

630 **Code Availability:**

631 The code used for analyses in this project is available on the project's github:
632 https://github.com/ksolari/SL_WGS

633

634 **Acknowledgments**

635

636 This work was supported by funding from the Snow Leopard Trust. We would like to thank those
637 who helped us in acquiring the captive samples used in this study, including Karen Ingerman and
638 Susie Bartlett from the Bronx Zoo. We would like to thank Haiqing Xu for his help in conducting
639 simulations to produce the null distribution for the shared doubleton analysis.

640

641 **References**

- 642 1. [K. Jumabay-Uulu, P. Wegge, C. Mishra, K. Sharma, Large carnivores and low diversity of](#)
643 [optimal prey: a comparison of the diets of snow leopards *Panthera uncia* and wolves *Canis lupus*](#)
644 [in Sarychat-Ertash Reserve in Kyrgyzstan. *Oryx* 48, 529–535 \(2014\).](#)
- 645 2. [P. Wegge, R. Shrestha, Ø. Flagstad, Snow leopard *Panthera uncia* predation on livestock](#)
646 [and wild prey in a mountain valley in northern Nepal: implications for conservation management.](#)
647 [*Wildlife Biol.* 18, 131–141 \(2012\).](#)
- 648 3. [W. Shehzad, *et al.*, Prey preference of snow leopard \(*Panthera uncia*\) in South Gobi,](#)
649 [Mongolia. *PLoS One* 7, e32104 \(2012\).](#)
- 650 4. [A. Aryal, *et al.*, Multipronged strategy including genetic analysis for assessing conservation](#)
651 [options for the snow leopard in the central Himalaya. *J. Mammal.* 95, 871–881 \(2014\).](#)
- 652 5. [M. Chetri, M. Odden, P. Wegge, Snow Leopard and Himalayan Wolf: Food Habits and Prey](#)
653 [Selection in the Central Himalayas, Nepal. *PLoS One* 12, e0170549 \(2017\).](#)
- 654 6. [K. R. Suryawanshi, *et al.*, Impact of wild prey availability on livestock predation by snow](#)
655 [leopards. *R Soc Open Sci* 4, 170026 \(2017\).](#)
- 656 7. [K. Thapa, N. Schmitt, N. M. B. Pradhan, H. R. Acharya, S. Rayamajhi, No silver bullet?](#)
657 [Snow leopard prey selection in Mt. Kangchenjunga, Nepal. *Ecol. Evol.* 11, 16413–16425 \(2021\).](#)
- 658 8. [Q. Lu, *et al.*, Snow Leopard Dietary Preferences and Livestock Predation Revealed by](#)
659 [Fecal DNA Metabarcoding: No Evidence for Apparent Competition Between Wild and Domestic](#)
660 [Prey. *Frontiers in Ecology and Evolution* 9 \(2021\).](#)
- 661 9. [Ö. Johansson, *et al.*, Snow leopard predation in a livestock dominated landscape in](#)
662 [Mongolia. *Biol. Conserv.* 184, 251–258 \(2015\).](#)
- 663 10. [C. Mu, *et al.*, The status and stability of permafrost carbon on the Tibetan Plateau. *Earth-*](#)
664 [*Sci. Rev.* 211, 103433 \(2020\).](#)
- 665 11. [T. McCarthy, D. Mallon, Eds., *Snow Leopards: Biodiversity of the World: Conservation from*](#)
666 [*Genes to Landscapes* \(Academic Press, 2016\).](#)
- 667 12. [S. B. Ale, C. Mishra, The snow leopard’s questionable comeback. *Science* 359, 1110](#)
668 [\(2018\).](#)
- 669 13. [A. Aryal, Poaching: Is snow leopard tally underestimated? *Nature* 550, 457 \(2017\).](#)
- 670 14. [R. M. Jackson, C. Mishra, T. M. McCarthy, S. B. Ale, Snow leopards: conflict and](#)
671 [conservation. *The biology and conservation of wild felids* 2, 417–430 \(2010\).](#)
- 672 15. [S. L. Network, Snow leopard survival strategy. *Seattle, Washington, USA* 1–145 \(2014\).](#)
- 673 16. [M. Lalande, M. Ménégou, G. Krinner, K. Naegeli, S. Wunderle, Climate change in the High](#)
674 [Mountain Asia in CMIP6. *Earth Syst. Dyn.* 12, 1061–1098 \(2021\).](#)
- 675 17. [J. Li, *et al.*, Climate refugia of snow leopards in High Asia. *Biol. Conserv.* 203, 188–196](#)
676 [\(2016\).](#)
- 677 18. [T. McCarthy, D. Mallon, P. Zahler, “Chapter 3 - What is a snow leopard? Biogeography](#)
678 [and status overview” in *Snow Leopards \(Second Edition\)*, D. Mallon, T. McCarthy, Eds. \(Academic](#)
679 [Press, 2024\), pp. 31–41.](#)
- 680 19. [T. McCarthy, D. Mallon, R. Jackson, P. Zahler, K. McCarthy, *Panthera uncia*. *The IUCN*](#)
681 [*Red List of Threatened Species* e.T22732A50664030. \(2017\).](#)
- 682 20. [D. M. Cooper, A. J. Dugmore, A. C. Kitchener, M. J. Metzger, A. Trabucco, A kingdom in](#)
683 [decline: Holocene range contraction of the lion \(*Panthera leo*\) modelled with global environmental stratification.](#)
684 [*PeerJ* 9, e10504 \(2021\).](#)
- 685 21. [D. M. Cooper, *et al.*, Predicted Pleistocene–Holocene range shifts of the tiger \(*Panthera*](#)
686 [*tigris*\). *Divers. Distrib.* 22, 1199–1211 \(2016\).](#)
- 687 22. [P. Dobrynin, *et al.*, Genomic legacy of the African cheetah, *Acinonyx jubatus*. *Genome*](#)
688 [*Biol.* 16, 277 \(2015\).](#)
- 689 23. [P. Pečnerová, *et al.*, High genetic diversity and low differentiation reflect the ecological](#)
690 [versatility of the African leopard. *Curr. Biol.* 31, 1862–1871.e5 \(2021\).](#)
- 691 24. [J. E. Janečka, *et al.*, Range-Wide Snow Leopard Phylogeography Supports Three](#)
692 [Subspecies. *J. Hered.* 108, 597–607 \(2017\).](#)
- 693 25. [B. Weckworth, Snow Leopard \(*Panthera uncia*\) Genetics: The Knowledge Gaps, Needs,](#)
694 [and Implications for Conservation. *J. Indian Inst. Sci.* 101, 279–290 \(2021\).](#)

- 695 26. [Y. S. Cho, et al., The tiger genome and comparative analysis with lion and snow leopard](#)
696 [genomes. *Nat. Commun.* 4, 2433 \(2013\).](#)
- 697 27. [E. E. Armstrong, et al., Genome report: chromosome-level draft assemblies of the snow](#)
698 [leopard, African leopard, and tiger \(*Panthera uncia*, *Panthera pardus pardus*, and *Panthera tigris*\).](#)
699 [*G3* 12 \(2022\).](#)
- 700 28. [M. P. Korablev, et al., Large-scale and fine-grain population structure and genetic diversity](#)
701 [of snow leopards \(*Panthera uncia* Schreber, 1776\) from the northern and western parts of the range](#)
702 [with an emphasis on the Russian population. *Conserv. Genet.* 22, 397–410 \(2021\).](#)
- 703 29. [C. Hacker, et al., Genetic diversity and spatial structures of snow leopards \(*Panthera uncia*\)](#)
704 [reveal proxies of connectivity across Mongolia and northwestern China. *Landsc. Ecol.* 38, 1013–](#)
705 [1031 \(2023\).](#)
- 706 30. [L. Atzeni, et al., Evidence of spatial genetic structure in a snow leopard population from](#)
707 [Gansu, China. *Heredity* 127, 522–534 \(2021\).](#)
- 708 31. [S. Aruge, H. Batool, F. M. Khan, Fakhari-Abbas, S. Janjua, A pilot study—genetic diversity](#)
709 [and population structure of snow leopards of Gilgit-Baltistan, Pakistan, using molecular techniques.](#)
710 [*PeerJ* 7, e7672 \(2019\).](#)
- 711 32. [T. McCarthy, D. Mallon, E. W. Sanderson, P. Zahler, K. Fisher, “What is a Snow Leopard?](#)
712 [Biogeography and Status Overview” in *Snow Leopards*, T. McCarthy, D. Mallon, Eds. \(Elsevier,](#)
713 [2016\), pp. 23–42.](#)
- 714 33. [J. Li, et al., Defining priorities for global snow leopard conservation landscapes. *Biol.*](#)
715 [*Conserv.* 241, 108387 \(2020\).](#)
- 716 34. [P. Riordan, S. A. Cushman, D. Mallon, K. Shi, J. Hughes, Predicting global population](#)
717 [connectivity and targeting conservation action for snow leopard across its range. *Ecography* 39,](#)
718 [419–426 \(2016\).](#)
- 719 35. [K. Sharma, et al., Vigorous dynamics underlie a stable population of the endangered snow](#)
720 [leopard *Panthera uncia* in Tost Mountains, South Gobi, Mongolia. *PLoS One* 9, e101319 \(2014\).](#)
- 721 36. [T. M. McCarthy, T. K. Fuller, B. Munkhtsog, Movements and activities of snow leopards in](#)
722 [Southwestern Mongolia. *Biol. Conserv.* 124, 527–537 \(2005\).](#)
- 723 37. [H. Senn, G. Murray-Dickson, A. C. Kitchener, P. Riordan, D. Mallon, Response to Janečka](#)
724 [et al. 2017. *Heredity* 120, 581–585 \(2018\).](#)
- 725 38. [J. E. Janečka, M. J. Janečka, K. M. Helgen, W. J. Murphy, The validity of three snow](#)
726 [leopard subspecies: response to Senn et al. *Heredity* 120, 586–590 \(2018\).](#)
- 727 39. [C. Gärke, et al., Comparison of SNPs and microsatellites for assessing the genetic](#)
728 [structure of chicken populations. *Anim. Genet.* 43, 419–428 \(2012\).](#)
- 729 40. [L. Blomqvist, International Pedigree Book for Snow leopards, *Uncia uncia* Volume 9.](#)
730 [\[Preprint\] \(2008\).](#)
- 731 41. [J. Tetzloff, K. R. Schwartz, “Chapter 26 - Role of zoos in snow leopard conservation: The](#)
732 [Species Survival Plan® in North America” in *Snow Leopards \(Second Edition\)*, D. Mallon, T.](#)
733 [McCarthy, Eds. \(Academic Press, 2024\), pp. 333–346.](#)
- 734 42. [D. H. Alexander, J. Novembre, K. Lange, Fast model-based estimation of ancestry in](#)
735 [unrelated individuals. *Genome Res.* 19, 1655–1664 \(2009\).](#)
- 736 43. [A. G. Nazareno, J. B. Bemmels, C. W. Dick, L. G. Lohmann, Minimum sample sizes for](#)
737 [population genomics: an empirical study from an Amazonian plant species. *Mol. Ecol. Resour.* 17,](#)
738 [1136–1147 \(2017\).](#)
- 739 44. [E. E. Armstrong, et al., Recent Evolutionary History of Tigers Highlights Contrasting Roles](#)
740 [of Genetic Drift and Selection. *Mol. Biol. Evol.* 38, 2366–2379 \(2021\).](#)
- 741 45. [Anopheles gambiae 1000 Genomes Consortium, et al., Genetic diversity of the African](#)
742 [malaria vector *Anopheles gambiae*. *Nature* 552, 96–100 \(2017\).](#)
- 743 46. [1000 Genomes Project Consortium, et al., An integrated map of genetic variation from](#)
744 [1,092 human genomes. *Nature* 491, 56–65 \(2012\).](#)
- 745 47. [S. Gravel, et al., Demographic history and rare allele sharing among human populations.](#)
746 [*Proc. Natl. Acad. Sci. U. S. A.* 108, 11983–11988 \(2011\).](#)
- 747 48. [G. A. Van der Auwera, B. D. O'Connor, *Genomics in the Cloud: Using Docker, GATK, and*](#)
748 [*WDL in Terra* \(O'Reilly Media, Inc., 2020\).](#)

749 49. [P. Danecek, et al., The variant call format and VCFtools. *Bioinformatics* 27, 2156–2158](#)
750 [\(2011\).](#)

751 50. [Ö. Johansson, et al., The timing of breeding and independence for snow leopard females](#)
752 [and their cubs. *Mamm. Biol.* 101, 173–180 \(2021\).](#)

753 51. [S. Prost, et al., Genomic analyses show extremely perilous conservation status of African](#)
754 [and Asiatic cheetahs \(*Acinonyx jubatus*\). *Mol. Ecol.* 31, 4208–4223 \(2022\).](#)

755 52. [J. L. A. Paijmans, et al., African and Asian leopards are highly differentiated at the genomic](#)
756 [level. *Curr. Biol.* 31, 1872–1882.e5 \(2021\).](#)

757 53. [C. Wang, et al., Population genomic analysis provides evidence of the past success and](#)
758 [future potential of South China tiger captive conservation. *BMC Biol.* 21, 64 \(2023\).](#)

759 54. [E. E. Armstrong, et al., Unraveling the genomic diversity and admixture history of captive](#)
760 [tigers in the United States. *Proc. Natl. Acad. Sci. U. S. A.* 121, e2402924121 \(2024\).](#)

761 55. [H. Li, R. Durbin, Inference of human population history from individual whole-genome](#)
762 [sequences. *Nature* 475, 493–496 \(2011\).](#)

763 56. [P. D. Hughes, P. L. Gibbard, J. Ehlers, Timing of glaciation during the last glacial cycle:](#)
764 [evaluating the concept of a global “Last Glacial Maximum” \(LGM\). *Earth Sci. Rev.* 125, 171–198](#)
765 [\(2013\).](#)

766 57. [R. Frankham, Effective population size/adult population size ratios in wildlife: a review.](#)
767 [*Genet. Res.* 89, 491–503 \(2007\).](#)

768 58. [O. Mazet, W. Rodríguez, S. Grusea, S. Boitard, L. Chikhi, On the importance of being](#)
769 [structured: instantaneous coalescence rates and human evolution--lessons for ancestral](#)
770 [population size inference? *Heredity* 116, 362–371 \(2016\).](#)

771 59. [N. Mather, S. M. Traves, S. Y. W. Ho, A practical introduction to sequentially Markovian](#)
772 [coalescent methods for estimating demographic history from genomic data. *Ecol. Evol.* 10, 579–](#)
773 [589 \(2020\).](#)

774 60. [Z. A. Szpiech, A. Blant, T. J. Pemberton, GARLIC: Genomic Autozygosity Regions](#)
775 [Likelihood-based Inference and Classification. *Bioinformatics* 33, 2059–2062 \(2017\).](#)

776 61. [A. Khan, et al., Genomic evidence for inbreeding depression and purging of deleterious](#)
777 [genetic variation in Indian tigers. *Proc. Natl. Acad. Sci. U. S. A.* 118 \(2021\).](#)

778 62. [P. Cingolani, et al., A program for annotating and predicting the effects of single nucleotide](#)
779 [polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-](#)
780 [3. *Fly \(Austin\)* 6, 80–92 \(2012\).](#)

781 63. [R. S. Taylor, M. Manseau, P. Liu, P. J. Wilson, Protocol for genetic load analysis in caribou](#)
782 [using a modified genomic evolutionary rate profiling. *STAR Protoc* 6, 103789 \(2025\).](#)

783 64. [J. A. Robinson, et al., The critically endangered vaquita is not doomed to extinction by](#)
784 [inbreeding depression. *Science* 376, 635–639 \(2022\).](#)

785 65. [C. D. Huber, A. Durvasula, A. M. Hancock, K. E. Lohmueller, Gene expression drives the](#)
786 [evolution of dominance. *Nat. Commun.* 9 \(2018\).](#)

787 66. [N. Dussex, H. E. Morales, C. Grossen, L. Dalén, C. van Oosterhout, Purging and](#)
788 [accumulation of genetic load in conservation. *Trends Ecol. Evol.* 38, 961–969 \(2023\).](#)

789 67. [S. Kim, et al., Comparison of carnivore, omnivore, and herbivore mammalian genomes](#)
790 [with a new leopard assembly. *Genome Biol.* 17, 211 \(2016\).](#)

791 68. [P. A. Morin, et al., Reference genome and demographic history of the most endangered](#)
792 [marine mammal, the vaquita. *Mol. Ecol. Resour.* 21, 1008–1020 \(2021\).](#)

793 69. [M. V. Westbury, et al., Extended and Continuous Decline in Effective Population Size](#)
794 [Results in Low Genomic Diversity in the World’s Rarest Hyena Species, the Brown Hyena. *Mol.*](#)
795 [*Biol. Evol.* 35, 1225–1237 \(2018\).](#)

796 70. [M. V. Westbury, B. Petersen, E. Garde, M. P. Heide-Jørgensen, E. D. Lorenzen, Narwhal](#)
797 [Genome Reveals Long-Term Low Genetic Diversity despite Current Large Abundance Size.](#)
798 [*iScience* 15, 592–599 \(2019\).](#)

799 71. [Y. Xue, et al., Mountain gorilla genomes reveal the impact of long-term population decline](#)
800 [and inbreeding. *Science* 348, 242–245 \(2015\).](#)

801 72. [A. Benazzo, et al., Survival and divergence in a small group: The extraordinary genomic](#)
802 [history of the endangered Apennine brown bear stragglers. *Proc. Natl. Acad. Sci. U. S. A.* 114,](#)
803 [E9589–E9597 \(2017\).](#)

804 73. [C. C. Kyriazis, et al., Genomic underpinnings of population persistence in Isle Royale](#)
805 [moose. *Mol. Biol. Evol.* 40 \(2023\).](#)

806 74. [T. J. Pemberton, et al., Genomic patterns of homozygosity in worldwide human](#)
807 [populations. *Am. J. Hum. Genet.* 91, 275–292 \(2012\).](#)

808 75. [L. Yang, et al., Genomic evidence for low genetic diversity but purging of strong deleterious](#)
809 [variants in snow leopards. *Genome Biol.* 26, 94 \(2025\).](#)

810 76. [E. Dinerstein, et al., The Fate of Wild Tigers. *Bioscience* 57, 508–514 \(2007\).](#)

811 77. [H. Bauer, et al., Lion \(*Panthera leo*\) populations are declining rapidly across Africa, except](#)
812 [in intensively managed areas. *Proc. Natl. Acad. Sci. U. S. A.* 112, 14894–14899 \(2015\).](#)

813 78. [R. Pal, A. Panwar, S. P. Goyal, S. Sathyakumar, Changes in ecological conditions may](#)
814 [influence intraguild competition: inferring interaction patterns of snow leopard with co-predators.](#)
815 [PeerJ 10, e14277 \(2022\).](#)

816 79. [S. Lovari, S. Kachel, L. Xueyang, F. Ferretti, “Chapter 13 - Snow leopard, common leopard,](#)
817 [and wolf: Are they good neighbors?” in *Snow Leopards \(Second Edition\)*, D. Mallon, T. McCarthy,](#)
818 [Eds. \(Academic Press, 2024\), pp. 137–147.](#)

819 80. [S. Ostrowski, M. Gilbert, “Chapter 9 - Diseases of wild snow leopards and their wild](#)
820 [ungulate prey” in *Snow Leopards \(Second Edition\)*, D. Mallon, T. McCarthy, Eds. \(Academic Press,](#)
821 [2024\), pp. 95–111.](#)

822 81. [M. Heiner, J. Oakleaf, G. Davaa, J. Kiesecker, “Chapter 10 - Emerging threats to snow](#)
823 [leopards from energy and mineral development” in *Snow Leopards \(Second Edition\)*, D. Mallon, T.](#)
824 [McCarthy, Eds. \(Academic Press, 2024\), pp. 113–121.](#)

825 82. [P. Zahler, R. Victurine, “Chapter 11 - Linear infrastructure and snow leopard conservation”](#)
826 [in *Snow Leopards \(Second Edition\)*, D. Mallon, T. McCarthy, Eds. \(Academic Press, 2024\), pp.](#)
827 [123–128.](#)

828 83. [J. A. Markert, et al., Population genetic diversity and fitness in multiple environments. *BMC*](#)
829 [Evol. Biol. 10, 205 \(2010\).](#)

830 84. [H. Li, Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.](#)
831 [arXiv \[q-bio.GN\] \(2013\).](#)

832 85. [A. McKenna, et al., The Genome Analysis Toolkit: a MapReduce framework for analyzing](#)
833 [next-generation DNA sequencing data. *Genome Res.* 20, 1297–1303 \(2010\).](#)

834 86. [H. Li, et al., The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–](#)
835 [2079 \(2009\).](#)

836 87. [X. Zheng, et al., A high-performance computing toolset for relatedness and principal](#)
837 [component analysis of SNP data. *Bioinformatics* 28, 3326–3328 \(2012\).](#)

838 88. [R Core Team, *R: A language and environment for statistical computing* \(2020\).](#)

839 89. [L. Tupa, North American Regional Snow Leopard Studbook. \[Preprint\] \(2011\).](#)

840 90. [J. P. Sinnwell, T. M. Therneau, D. J. Schaid, The kinship2 R package for pedigree data.](#)
841 [Hum. Hered. 78, 91–93 \(2014\).](#)

842 91. [S. Purcell, et al., PLINK: a tool set for whole-genome association and population-based](#)
843 [linkage analyses. *Am. J. Hum. Genet.* 81, 559–575 \(2007\).](#)

844 92. [L.-T. Nguyen, H. A. Schmidt, A. von Haeseler, B. Q. Minh, IQ-TREE: a fast and effective](#)
845 [stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–274](#)
846 [\(2015\).](#)

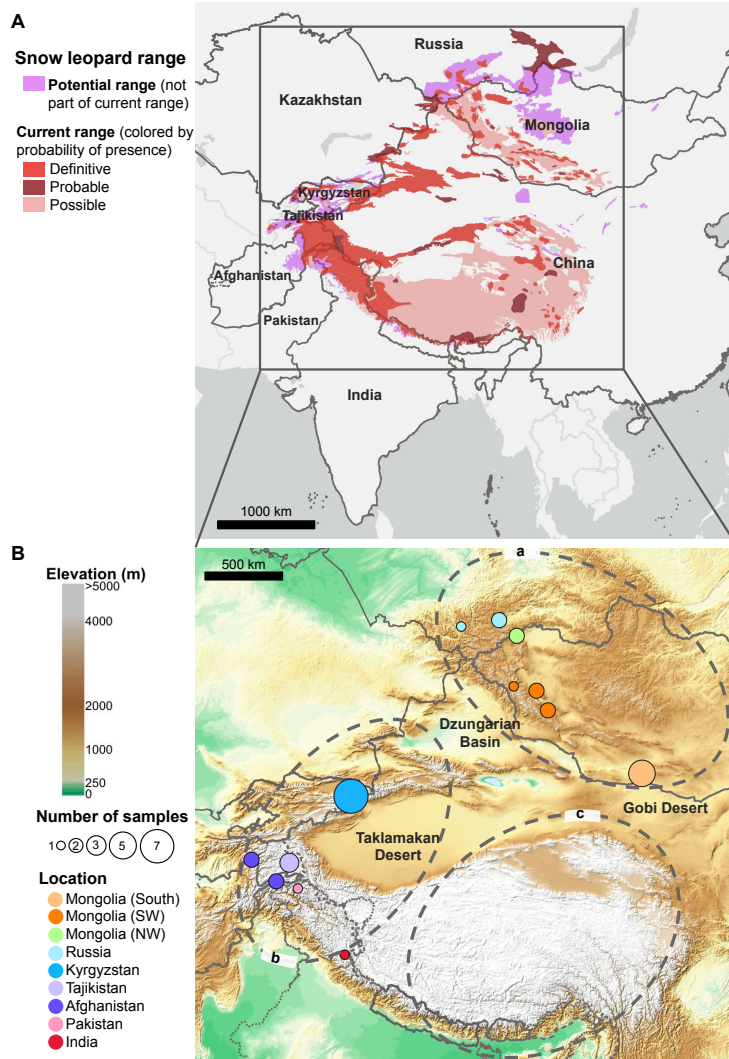
847 93. [H. Li, A statistical framework for SNP calling, mutation discovery, association mapping and](#)
848 [population genetical parameter estimation from sequencing data. *Bioinformatics* 27, 2987–2993](#)
849 [\(2011\).](#)

850 94. [E. E. Armstrong, et al., Parameterizing Pantherinae: De Novo mutation rate estimates from](#)
851 [Panthera and Neofelis pedigrees. *Genome Biol. Evol.* 17 \(2025\).](#)

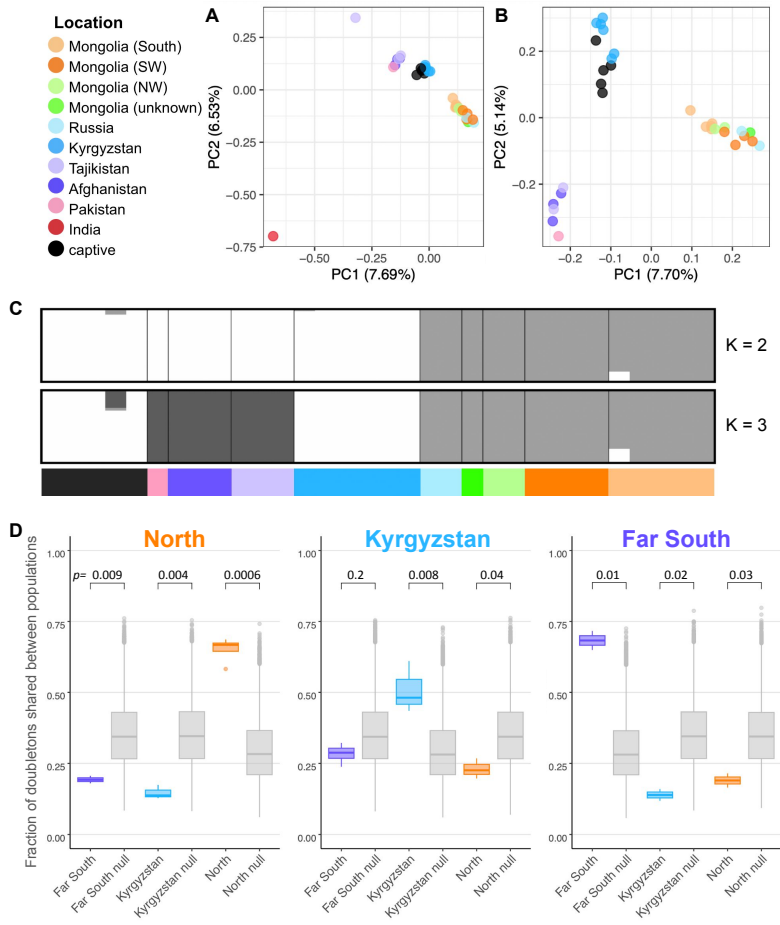
852 95. [T. Maruki, M. Lynch, Genotype Calling from Population-Genomic Sequencing Data. *G3* 7,](#)
853 [1393–1404 \(2017\).](#)

854 96. [A. Kassambara, ggpubr: “ggplot2” based publication ready plots. \(2020\).](#)

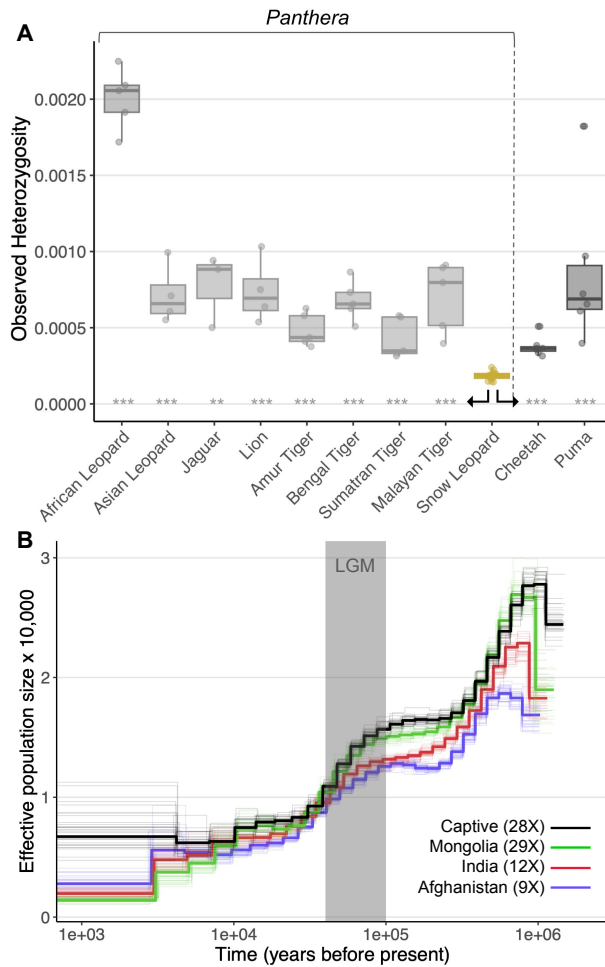
855 97. [A. S. Hinrichs, et al., The UCSC Genome Browser Database: update 2006. *Nucleic Acids*](#)
856 [Res 34, D590–8 \(2006\).](#)
857 98. [ESRI, Terrain: Elevation Tinted Hillshade. \(2013\).](#)
858 99. [The Snow Leopard Trust, Gallery for Snow Leopard Trust. *Arcgis Gallery for Snow Leopard*](#)
859 [Trust. Available at: snowleopard.maps.arcgis.com \[Accessed 1 June 2024\].](#)
860
861
862
863
864
865



867
 868 **Figure 1. Snow leopard distribution and sample maps.** A) The IUCN snow leopard distribution
 869 (19), taken from (32), is shown and the largest snow leopard range countries are labeled. The
 870 total potential snow leopard range (all shades of red and pink) is 3,256,841 km², of which
 871 2,778,309 km² is considered to be the current range. Within the current range, snow leopards are
 872 only definitively present in 32% of the area (889,059 km², definitive observation of snow leopards
 873 within the year leading up to the assessment (2007–2008)), probable in 8% (222,265 km², likely
 874 present based on habitat, prey, and connectivity and there is recent non-definitive information or
 875 definitive information of snow leopard presence within the last five years), and possible in 60%
 876 (1,666,985 km², possibility present due to habitat and connectivity to known populations, but no
 877 specific information about snow leopards in this area within the last five years) (32). B) Sample
 878 locations are indicated with different sized circles indicating the number of samples from each
 879 location. The basemap (98) indicates elevation and landscape features discussed in the text are
 880 labeled. Grey dashed ovals indicate the geographic distribution of the three subspecies
 881 suggested by Janecka et al. (24)— a) *P. u. irbis*, b) *P. u. uncia*, c) *P. u. uncioides*. In both maps,
 882 country boundaries available from the Snow Leopard Trust (99) are shown in dark grey. Not all
 883 country boundaries are in agreement, so a dotted line is used for India and a dashed line is used
 884 for China to make overlapping boundaries visible. Maps were created using ArcGIS software by
 885 ESRI.

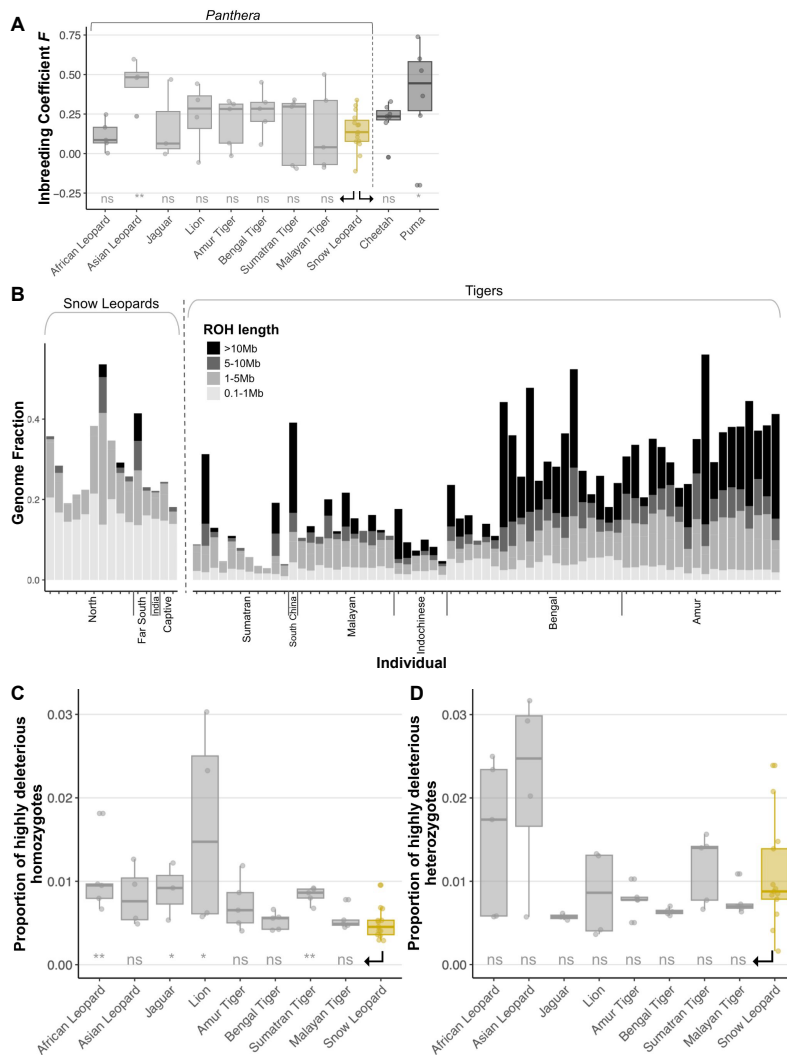


886
 887 **Figure 2. Principal components analysis (PCA), Admixture, and rare allele sharing.** A) PCA
 888 of genetic variation of 37 unrelated snow leopards using 1,448,657 SNPs. B) PCA after removing
 889 two outlier samples, India and sample U13 from Tajikistan. PCA axis labels include the percent
 890 variation explained by PC1 and PC2. C) Admixture results for ten independent runs for K=2 and
 891 K=3 distinct ancestry groups. The ancestry assignments shown for K=2 were supported by all ten
 892 iterations and the ancestry assignments shown for K=3 were supported by nine of the ten
 893 iterations (the alternate ancestry assignments supported by one run is shown in Supplementary
 894 Fig. 1B). D) Doubleton sharing between populations compared to null distributions under
 895 panmixia. We identified all doubletons where each minor allele occurred in a different individual.
 896 For each group, identified above each graph, we then calculated the fraction of doubletons
 897 occurring in an individual of that group that were shared with individuals of each other group.
 898 Observed values are shown in color and null distributions are shown in gray. We made null
 899 distributions by randomly shuffling population assignment among the samples and recalculating
 900 doubleton sharing 10,000 times. P-values for comparisons between observed data and the null
 901 distribution using the Wilcoxon Rank Sum Test are shown. The lower and upper edges of the
 902 boxes correspond to the first and third quartiles and the whiskers extend to the lowest/highest
 903 value that is no further than 1.5*IQR (inter-quartile range) from the box. Points falling further than
 904 1.5*IQR from the box are plotted individually. In all analyses, we have removed one member of
 905 each related pair and in the case of doubleton sharing, we have downsampled populations to n=6
 906 for each group.



907

908 **Figure 3. Genome-wide heterozygosity across all big cats and demographic history of**
 909 **snow leopards.** A) Comparison of genome-wide heterozygosity across all big cat species. We
 910 used publicly available data to call SNPs for every big cat species using GATK and calculated
 911 observed heterozygosity using VCFtools. In the case of leopard and tiger, we called SNPs
 912 separately for genetically distinct groups. We calculated snow leopard heterozygosity from SNPs
 913 called using the same pipeline that was used for SNP calling in all the other big cat species. Only
 914 snow leopard samples with a depth of 8X or higher are included (n=15). The lower and upper
 915 edges of the boxes correspond to the first and third quartiles and the whiskers extend to the
 916 lowest/highest value that is no further than 1.5*IQR (inter-quartile range) from the box. Points
 917 falling further than 1.5*IQR from the box are plotted individually. Stars below each boxplot
 918 indicate the p-value when comparing each species/subspecies to snow leopards using the
 919 Wilcoxon rank-sum test ($p < 0.01^{**}$ and $p < 0.001^{***}$). B) Reconstruction of effective population
 920 sizes using PSMC with a mutation rate of 3.62×10^{-9} per site per generation and a generation time
 921 of five years (26). Thirty bootstraps are shown for each sample in thinner, fainter lines of the
 922 same color. The timing of the Last Glacial Maximum (LGM) in the mountains of Asia (~40,000-
 923 100,000 years ago) (56) and the average depth of coverage of each sample used is indicated.



924
 925
 926
 927
 928
 929
 930
 931
 932
 933
 934
 935
 936
 937
 938
 939
 940
 941
 942

Figure 4. Inbreeding estimates across all big cats, ROH estimates for snow leopards and tigers, and highly deleterious load across *Panthera*. A) Comparison of inbreeding coefficient F measured by method of moments across all big cat species. We used publicly available data to call SNPs for every big cat species using GATK and calculated the inbreeding coefficient F using VCFtools. Only snow leopard samples with 8X coverage or greater are included (n=15). B) The fraction of the genome in ROH of each size bin (0.1-1Mb, 1-5Mb, 5-10Mb, and >10Mb) for snow leopards and tigers. All ROH estimates were calculated using GARLIC. We pulled tiger ROH data from Armstrong et al. (54). Each bar is one individual. Snow leopard samples are ordered by group as indicated along the x-axis and tigers are ordered by subspecies. All ROH estimates, for snow leopard and tiger, only include samples with greater than 8X coverage. C) Comparison of the proportion of homozygous derived protein-coding SNPs that are highly deleterious across *Panthera*. D) Comparison of the proportion of heterozygous derived protein-coding SNPs that are highly deleterious across *Panthera*. In all boxplots, each point represents one individual. The lower and upper edges of the boxes correspond to the first and third quartiles and the whiskers extend to the lowest/highest value that is no further than 1.5*IRQ (inter-quartile range) from the box. Points falling further than 1.5*IRQ from the box are plotted individually. The p-value when comparing each species/subspecies to snow leopards using the Wilcoxon rank-sum test is indicated below each boxplot (ns = non-significant, $p < 0.05^*$, $p < 0.01^{**}$).

1 **Supplementary Materials for:**

2 **Extreme in Every Way: Exceedingly Low Genetic Diversity in Snow Leopards Due to**
3 **Persistently Small Population Size**

4
5 Katherine A. Solari^{1&}, Simon Morgan¹, Andrey D. Poyarkov², Byron Weckworth³, Gustaf
6 Samelius^{4,5}, Koustubh Sharma^{4,6}, Stephane Ostrowski⁷, Uma Ramakrishnan⁸, Zairbek
7 Kubanychbekov⁹, Shannon Kachel³, Örjan Johansson^{4,10}, Purevjav Lkhagvajav¹¹, Heather
8 Hemmingmoore¹⁰, Dmitry Y. Alexandrov², Munkhtsog Bayaraa¹², Alexey Grachev¹³, Miroslav P.
9 Korablev², Jose A. Hernandez-Blanco², Bariushaa Munkhtsog¹⁴, Barry Rosenbaum¹⁵,
10 Viatcheslav V. Rozhnov², Ali Madad Rajabi¹⁶, Hafizullah Noori¹⁶, Kulbhushansingh
11 Suryawanshi^{3,17,18}, Ellie E. Armstrong^{1,19*}, Dmitri A. Petrov^{1,20,21*&}

12
13 **Supplementary Methods:**

14 **Species ID of Biosample SAMN17432540:**

15 Biosample SAMN17432540, SRA accession SRR13500277, is identified as *Panthera*
16 *uncia* on NCBI and in the following publications: (1, 2). We downloaded publicly available whole-
17 genome sequencing (WGS) data from NCBI for two individuals from each species in the
18 *Panthera* clade - lion (ERR4139877, ERR4139880), tiger (SRR13647584, SRR13647599),
19 jaguar (SRR4444359, SRR11097154), leopard (Asian - ERR5671313, ERR5671317,
20 SRR5382750; African - ERR5671315, ERR5671314) and snow leopard (SRR13500277,
21 SRR836372). We used BWA-MEM (3) to map all of this data, as well as WGS data from five
22 additional snow leopard samples that we generated, to the snow leopard reference genome (4).
23 Mapped data was sorted and indexed using SAMtools (5). We then used ANGSD (6) followed
24 by ngstools (7) to generate a PCA. A total of 40,205,810 variable sites were identified in
25 ANGSD. All of the *Panthera* species group clearly in the PCA except for biosample
26 SAMN17432540 which groups with Asian leopards and not snow leopards (Supplementary
27 Figure 7).

28

29

30 **Variant filtering:**

31 Variant calls received from Gencove were first filtered based on mappability across the
32 genome. We indexed the reference genome and generated mappability scores using GenMap
33 v1.3.0 (8) with flags '-K 30' and '-E 2'. GenMap mappability scores represent the uniqueness of
34 k-mers (k-mer size given by flag -K) for each position in the genome while allowing for a certain
35 number of mismatches (given by the flag -E) where a mappability score of one at a position
36 indicates that the k-mer at that position occurs only once with up to E mismatches. We sorted
37 the reference genome and mappability BED file using BEDtools (9) and removed SNPs falling in
38 a region of the genome with a mappability score less than one using BEDtools *intersect* with the
39 flags '-v', '-header', and '-sorted' (leaving 63,199,070 sites). We then used VCFtools (10) to
40 remove indels using the option '--remove-indels' (leaving 51,199,725 SNPs). We removed
41 individuals with less than 60% of the genome represented (samples U02, U12, and U20) using
42 the VCFtools option '--remove-ind'. Most SNPs were removed at this step because sample U02
43 had over 1.3 million unique singletons (Supplementary Table 1) and over 44 million unique
44 doubletons, likely due to some unknown contamination. We then used VCFtools to remove sites
45 that no longer had any variation using flag '--non-ref-ac-any 1' (leaving 2,379,069 SNPs), SNPs
46 that did not fall on putative autosomes (leaving 2,213,174 SNPs), and non-biallelic SNPs using
47 the flags '--min-alleles 2 --max-alleles 2' (leaving 2,146,294 SNPs). We also split scaffold 22 at
48 base pair 67,650,000 into chromosomes E1 and F1, respectively, to remedy the misassembly
49 described in Armstrong et al.(4). Next, with bam files as input, we used GATK v4.1.4.1 to add
50 the following annotations to our filtered VCF file using *VariantAnnotator*: QD (QualByDepth -
51 variant confidence normalized by unfiltered depth of variant samples), FS (FisherStrand - strand
52 bias estimated using Fisher's exact test), MQ (RMSMappingQuality - root mean square of the
53 mapping quality of reads across all samples), MQRankSum (MappingQualityRankSumTest -
54 rank sum test for mapping qualities of reference versus alternate reads), and
55 ReadPosRankSum (ReadPosRankSumTest - rank sum test for relative positioning of reference

56 versus alternate alleles within reads). After adding these annotations, we used GATK
57 *VariantFiltration* to filter SNPs using the following flag: '--filter-expression "QD < 2.0 || FS > 60.0
58 || MQ < 40.0 || MQRankSum < -12.5 || ReadPosRankSum < -8.0"'. We then used VCFtools to
59 remove SNPs that did not pass the GATK *VariantFiltration* step using option '--remove-filtered-
60 all' (leaving 2,065,453 SNPs). Lastly, we removed SNPs missing data in more than 10% of the
61 individuals using flag '--max-missing 0.9' (leaving a final SNP set of 1,591,978). We calculated
62 the number of singletons and private doubletons using the VCFtools flag '--singletons' and
63 removed one of the captive samples (10x_SDzoo) due to an excess of singletons
64 (Supplementary Table 1) which we suspected to be due to the 10X library prep method unique
65 to this sample.

66

67 **Relatedness assessment:**

68 We estimated relatedness among samples using SNPrelate (11) in R (12). We divided
69 the samples into three genetically distinct groups based on preliminary PCA and Admixture
70 analyses—Mongolia and Russia, Kyrgyzstan and captive, and all samples south of Kyrgyzstan.
71 We conducted preliminary PCA and Admixture analyses using the same methods described
72 below but in this case included samples that were later removed due to relatedness. We ran
73 each group through SNPrelate separately. We used the function *snpGDSBED2GDS* to convert
74 BED files to GDS (Genomic Data Structures) files. We used *snpGDSLDPurging* to prune the
75 input data using the following flags - 'method="corr", slide.max.n=50, ld.threshold=0.3, maf =
76 0.05, autosome.only = F'. We then calculated kinship coefficients using the function
77 *snpGDSIBDMLE* which calculates IBD coefficients for individual pairs using Maximum Likelihood
78 Estimation. A kinship coefficient of 0.5 indicates a duplicate sample, 0.25 indicates a first-order
79 related pair (parent-offspring or full sibling), and 0.125 indicates a second-order related pair
80 (half-siblings or grandparent-grandchild). All sample pairs with non-zero kinship coefficients are
81 listed in Supplementary Table 2.

82

83 **Principal component analyses:**

84 We conducted principal component analyses (PCAs) on the dataset after filtering for first
85 and second degree relatives (N=33, 1,448,657 SNPs) using PLINK2 (13). First, we generated
86 BED files from VCF files using the flags '--make-bed' and '--allow-extra-chr' and allele
87 frequencies were computed using the flags '--freq' and '--allow-extra-chr'. Then, we ran the PCA
88 by inputting the resultant BED file with flag '--bfile' and frequency file with flag '--read-freq' and
89 using the flags '--allow-extra-chr' and '--pca 38'. We visualized PCA output using ggplot2 in R.

90

91 **Admixture analysis:**

92 We conducted Admixture analyses on the dataset after filtering for first and second
93 degree relatives. Since these analyses can be sensitive to having a small sample size for a
94 genetically distinct group (14), we ran Admixture analyzes both with and without the India
95 sample (15, 16). We used VCFtools to convert VCF files into PLINK format files. We then used
96 PLINK1.9 to recode alleles numerically using the flags '--allow-extra-chr --recode12'. We then
97 ran Admixture (17) for K=1-10 ten times using the flag '-s time' so that a different random seed
98 was used for each run based on the time. We visualized all runs using Clumpak (18).

99 We also conducted PCA and Admixture analyses on the dataset after removing captive
100 samples in order to assess the impact of captive samples on population structure results
101 (Supplementary Fig. 2) as we know these individuals have at least somewhat mixed ancestry
102 and have been separated from the wild population for many decades.

103

104 **Phylogenetic tree construction:**

105 We constructed a phylogenetic tree as an additional way to visualize genetic groupings
106 among the samples using the same data set used for PCA and Admixture analyses. We
107 included two tiger samples (SRR13647578 and SRR13500748) to serve as an outgroup. We

108 first mapped the tiger WGS data to the snow leopard reference genome using BWA-MEM. We
109 then used the function *vcf2bed* from bedops (19) with the flags '-snvs' and '-d' to convert our
110 filtered snow leopard VCF file to a BED file. We then used BCFtools *mpileup* with the flags '-A -
111 a AD,DP -R' followed by BCFtools *call* with the flag '-m' to generate a VCF file of SNP calls for
112 just the snow leopard SNP sites from the tiger data. We then merged the tiger calls to our snow
113 leopard VCF file using *vcf-merge* from VCFtools. We used *vcf2phylip* (20) to convert the merged
114 VCF to phylip format and constructed a maximum likelihood phylogeny using IQtree (21) with
115 the flags '-st DNA -m GTR+ASC -T 10 -B 1000'. We visualized the tree using FigTree v1.4.4
116 (22).

117

118 **Quantifying population structure:**

119 We further characterized population divides identified in Admixture and PCA analyses by
120 calculating the number of shared versus private SNPs among groups, pairwise F_{ST} , and the rate
121 of rare variant sharing among groups. We excluded captive samples from these analyses as
122 well as the sample from India as the PCA indicates that this sample is genetically distinct from
123 all groups.

124 We did not filter for relatedness when calculating shared versus private SNPs among the
125 groups; however, we did downsample as necessary to have an equal number of samples in
126 each group for all comparisons (n=15 for the K=2 north vs. south comparison and n=7 for the
127 K=3 North-Kyrgyzstan-FarSouth comparison). We used BCFtools (23) to create and index VCF
128 files with only the individuals from each group using the *view* and *index* function, respectively.
129 We then used BCFtools *isec* to calculate how many SNPs were shared and private among the
130 groups.

131 Before calculating F_{ST} , we removed one individual from each first or second degree
132 related pair. We also downsampled to have an equal number of samples in each group (n=13
133 for the K=2 north vs. south comparison and n=6 for the K=3 North-Kyrgyzstan-FarSouth

134 comparison). We then used VCFtools to calculate the Weir and Cockerham weighted F_{ST}
135 estimate (24) between each group using the flag '--weir-fst-pop'. In order to assess the
136 significance of F_{ST} estimates at K=3, we conducted 10,000 permutations by shuffling the
137 population assignments and recalculating F_{ST} to generate a null distribution.

138 To assess doubleton sharing among groups, we excluded one individual for each related
139 pair and downsampled all groups so that each had an equal number of individuals (N=6 for
140 K=3). We then used VCFtools to retain only SNPs with a minor allele count of two (doubletons)
141 using the flags '--mac 2 --max-mac 2'. We removed any SNPs where the doubletons were
142 present in the same individual and identified allele sharing among samples using the flag '--
143 genome full' in PLINK. We created a null distribution under panmixia by calculating doubleton
144 sharing between populations after randomly shuffling population assignments among the
145 samples 10,000 times. We compared fractions of observed doubleton sharing to the null
146 distribution using the Wilcoxon Rank Sum Test in R.

147

148 **Heterozygosity in other big cats:**

149 We calculated heterozygosity in all big cat species using publicly available data in order
150 to see how snow leopard heterozygosity levels compared to other big cat species. We included
151 all other species in the genus *Panthera* (leopard, lion, tiger, jaguar) as well as cheetah and
152 puma. Accession numbers of publicly available WGS data and reference genomes used in this
153 analysis are listed in Supplementary Table 3 and 4, respectively. We mapped all FASTQ data to
154 the corresponding reference genome using BWA-MEM (3). We sorted and indexed resulting
155 bam files using SAMtools (5). We added read groups and marked duplicates using picard (25).
156 We calculated depth of coverage and breadth of coverage for each sample using SAMtools. We
157 then used GATK (26) *HaplotypeCaller* to generate a GVCF file for each sample and GATK
158 *CombineGVCFs* to combine the GVCF files for every species (or subgroup in the case of tigers

159 and leopards). Lastly, we used GATK *GenotypeGVCFs* to create a final VCF file for every
160 group.

161 In order to limit the need for excessive computing power, we downsampled some
162 samples with over 30X coverage using SAMtools *view* as indicated in Supplementary Table 3.
163 The leopard samples required a larger amount of computing power than the other species to
164 call haplotypes, so we split these samples into one BAM file per contig using BAMtools *split* and
165 ran each contig through GATK *HaplotypeCaller* and *CombineGVCFs* in parallel. We also found
166 it necessary to split the jaguar and snow leopard data into intervals to combine GVCF files using
167 GATK *GenomicsDBImport* followed by *CombineGVCFs*. In both cases, we concatenated
168 resultant VCF files, one per contig or interval, using BCFtools *concat*.

169 We filtered each reference genome for mappability using genmapv1.3.0 to first index
170 and then calculate mappability using flags '-K 30' and '-E 2'. We then filtered each VCF to only
171 include nucleotides with a mappability score of one. To do this, we sorted both VCF files and
172 mappability BED files using BEDtools and retained only SNPs falling in regions with a
173 mappability score of one in each VCF using BEDtools *intersect*. We then filtered each VCF
174 using GATK *VariantFiltration* with the following flag: '--filter-expression "QD < 2.0 || FS > 60.0 ||
175 MQ < 40.0 || MQRankSum < -12.5 || ReadPosRankSum < -8.0"'. Then, we filtered VCF files to
176 remove indels and non-biallelic SNPs using VCFtools flags '--remove-indels --min-alleles 2 --
177 max-alleles 2'.

178 We calculated the mappable length of each genome using mappability BED files to
179 calculate the number of nucleotides with a mappability score of one. Using the filtered VCF for
180 each species or group, we calculated observed homozygosity for each sample using VCFtools
181 (10) with the flag '--het'. In R, we calculated the number of heterozygous sites by subtracting the
182 number of observed homozygous sites column ((O)HOM) from the total number of sites column
183 (NSITES). We then calculated heterozygosity by dividing the number of heterozygous sites by

184 the length of the genome consisting of putative autosomes with a mappability score of one. We
185 used ggplot2 in R to create box plots of heterozygosity results.

186

187 **Genetic load across *Panthera*:**

188 We characterized the prevalence of deleterious derived mutations in all five *Panthera*
189 species in order to assess the genetic load in each species. First, we used SnpEff (27) to build a
190 new database for each species using publicly available gene annotations. In the case of snow
191 leopard, leopard, and jaguar, the genomes used for SNP calling were annotated
192 (Supplementary Table 4), so these same reference genomes were used for SnpEff database
193 creation. In the case of tiger and lion, the reference genome used for SNP calling was not
194 annotated, so a different reference genome was used for SnpEff database creation—Genbank
195 accession GCA_018350195 and GCA_018350215.1 for tiger and lion, respectively.

196 For tiger and lion, we then had to project our SNP calls onto the annotated genomes
197 used to make the SnpEff database. To do this, we first aligned the reference genome that we
198 used for SNP calling to the annotated reference genome for each species using minimap2 (28)
199 with the flag '-c'. We then used paf2chain (29) to convert the minimap2 output to a chain file. We
200 used picard (30) to make a dictionary from the annotated reference genome using
201 *CreateSequenceDictionary* and then used *LiftoverVcf* (31) from picard to project our SNP calls
202 onto the annotated reference genome using the chain file.

203 Once having a SnpEff database for each species, as well as a VCF for each
204 species/subspecies made using the same annotated reference genome used to make the
205 SnpEff database, we used SnpEff to annotate each VCF file using the flag '-classic'.

206 In order to identify which SNPs were derived in each species, we followed the protocol
207 outlined in (32). For example, for snow leopards, we first mapped all of the other *Panthera*
208 reference genomes to the snow leopard reference genome using the

209 Pipeline_input_alignment_GERP.sh script from (32). We then used htsbox (33) *pileup* to create
210 a consensus sequence for each pairwise genome alignment. We then split each consensus
211 alignment up into scaffolds and combined species to make a multispecies alignment for each
212 scaffold. We then ran genomic evolutionary rate profiling (GERP) (34) on each scaffold as
213 outlined in (32) using the species tree from (35). We then used the Derived_alleles_extract.R
214 script from (32) to use the GERP output to identify derived SNPs in the snow leopard VCF. This
215 process was repeated for each *Panthera* species.

216 Assessing load in this way requires the calling of derived variants through the alignment
217 of reference genomes for species in the clade of interest. In order to be able to use the same
218 genome alignments for derived variant calling in all species, providing greater consistency and
219 comparability among species, we excluded non-*Panthera* species (cheetah and puma) from this
220 analysis.

221 Once having a list of derived SNPs for each sample, we focused only on SNPs in protein
222 coding transcripts by only considering SNPs that had a functional class annotation from SnpEff.
223 We then counted the number of these derived protein-coding SNPs that were classified as
224 “High” effect by SnpEff—defined as having a highly disruptive impact on the protein, including
225 loss of function and stop-gained mutations, as well as SNPs classified as “Moderate” effect by
226 SnpEff—defined as non-disruptive variants that might change protein effectiveness, including
227 missense mutations. We counted the number of homozygous and heterozygous SNPs in each
228 effect category using a custom python script. We quantified the highly deleterious homozygous
229 load of each individual by dividing the number of highly deleterious homozygous SNPs by the
230 total number of derived protein-coding homozygous SNPs observed in that individual. Highly
231 deleterious heterozygous load and moderately deleterious homozygous and heterozygous load
232 were calculated in the same way. We tested if values observed in snow leopards were
233 significantly different from those observed in other species/subspecies using the Wilcoxon rank

234 sum test in R (12). Scripts, as well as additional detail on steps taken to calculate load in each
235 species, are available on our project's github (https://github.com/ksolari/SL_WGS).

236 Comparing load estimates between studies is not advisable since output can be
237 impacted by numerous factors such as SNP filtering, species trees used for derived state
238 identification, etc. That said, we note that the proportions of moderately deleterious mutations
239 observed in our data (Supplementary Fig. 9) are higher than those observed in some other
240 animal studies (36), but are indeed on par with others (37–39). These high proportions indicate
241 that many of the SNPs identified as moderately deleterious by SnpEff in our dataset are likely
242 extremely weakly deleterious to the point of being effectively neutral.

243

244

245

246

247

248

249

250

251

252

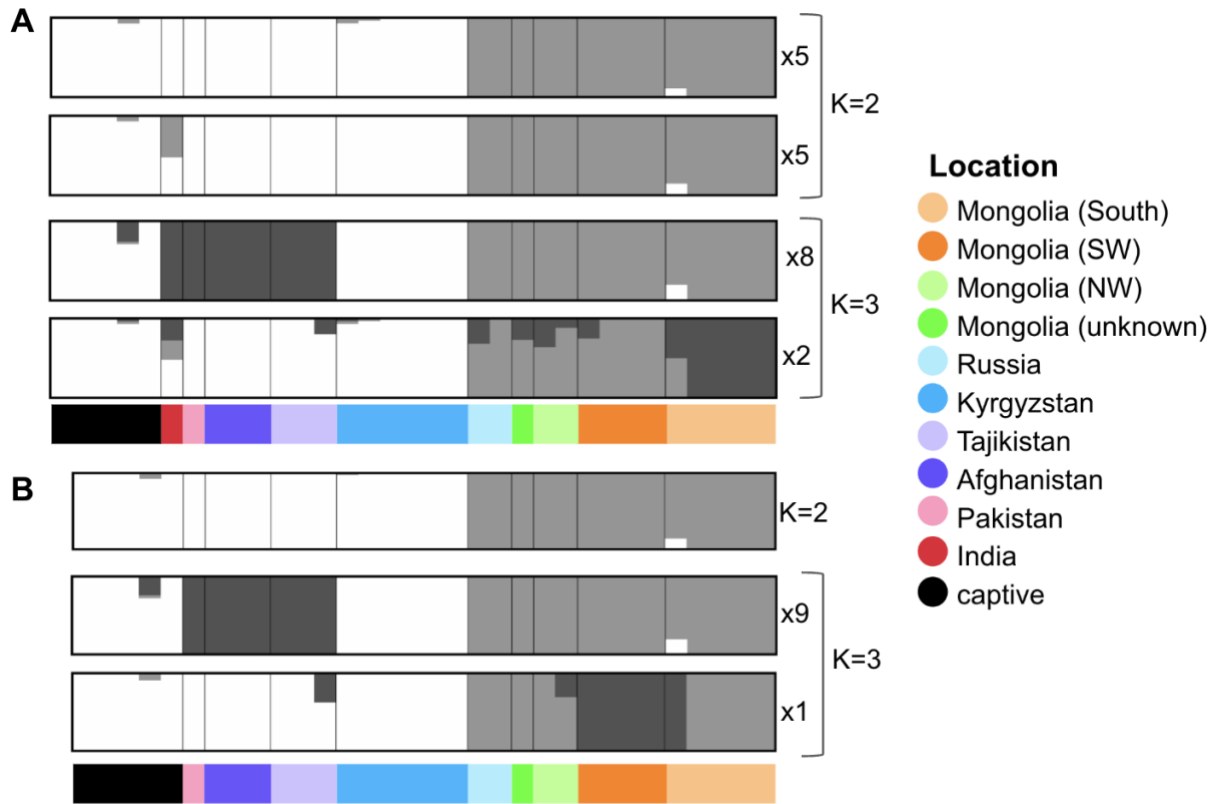
253

254

255

256

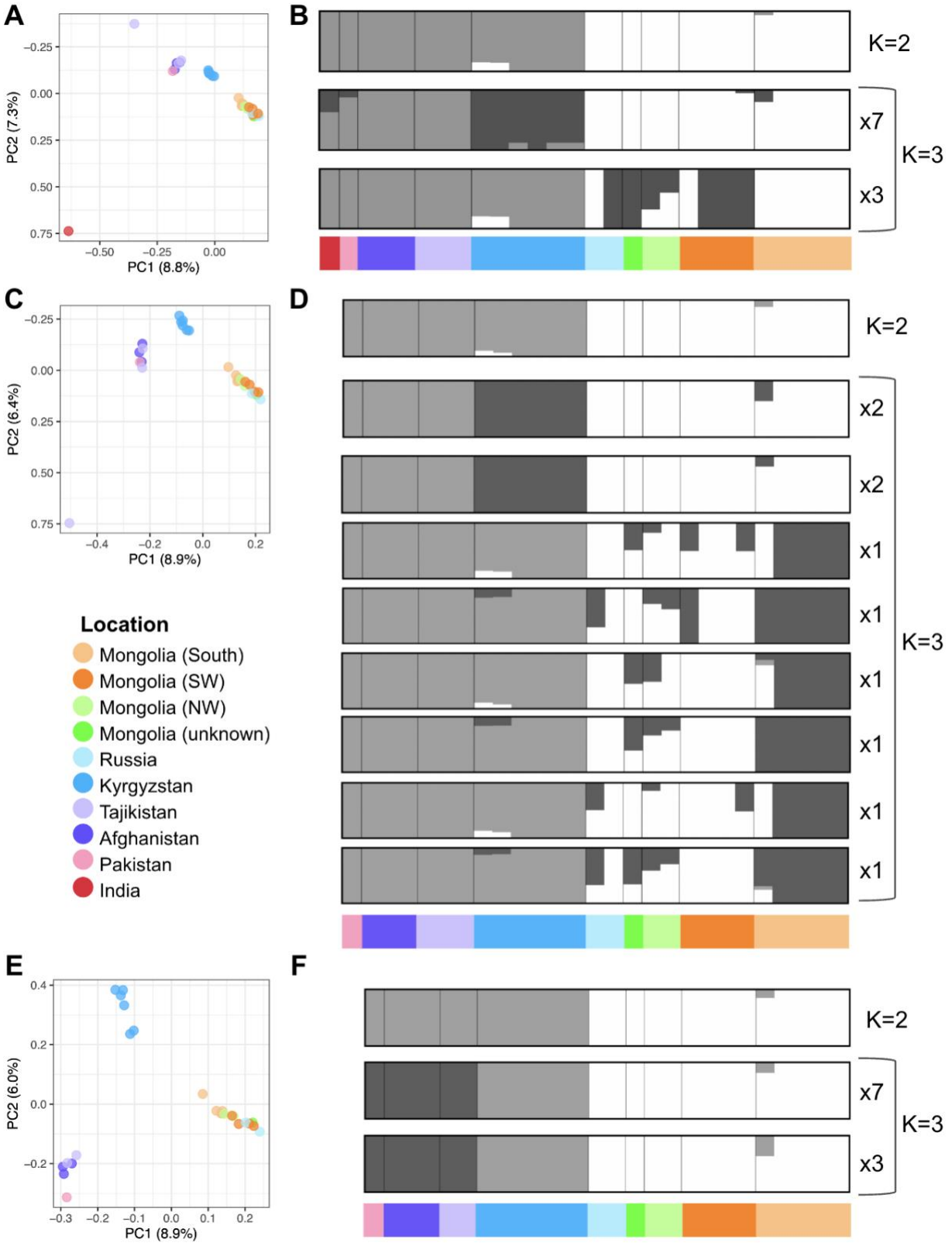
257 **Supplementary Figures:**



258

259 **Supplementary Figure 1. Admixture results with and without the India sample. A)**

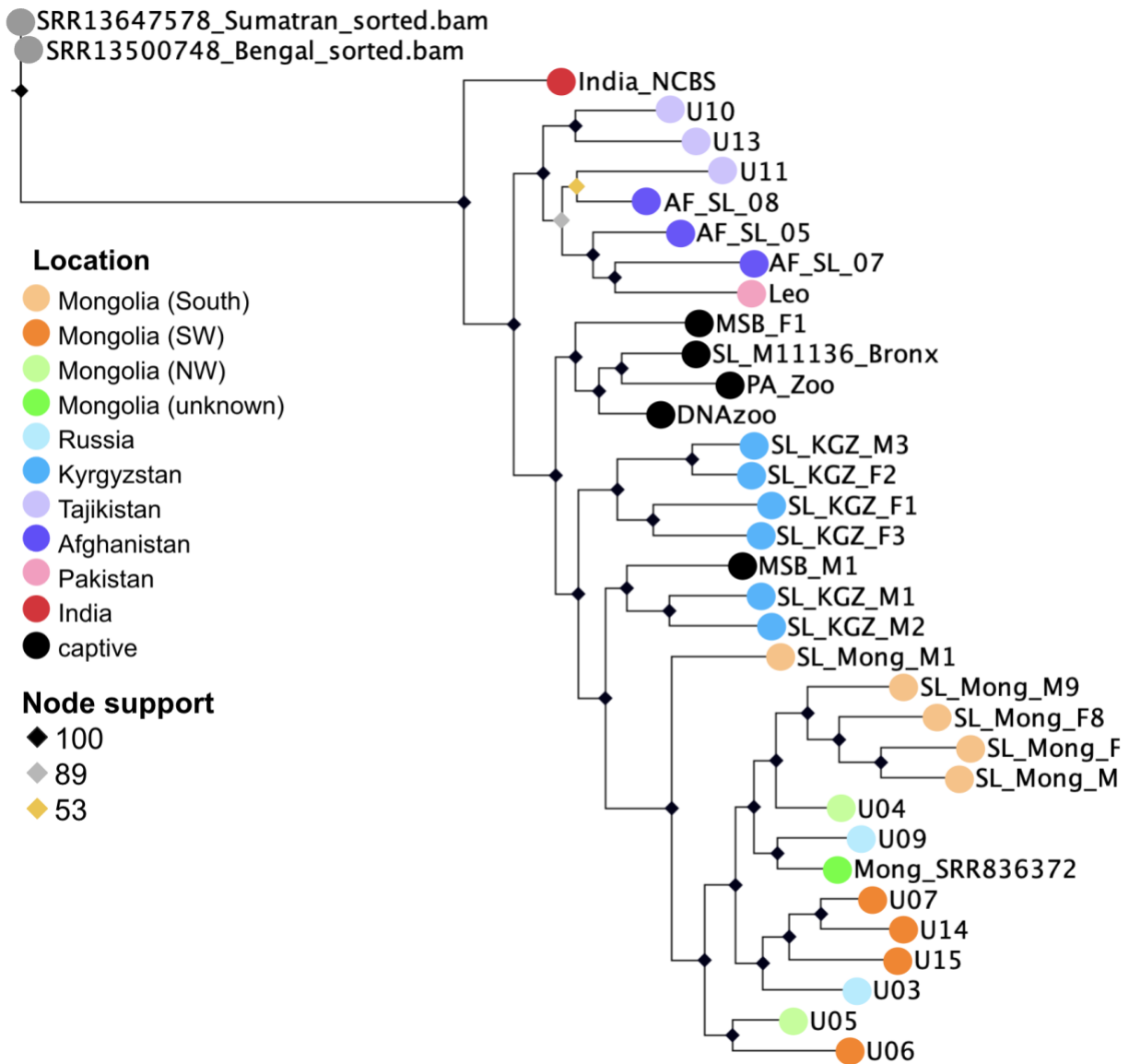
260 Admixture results for ten independent runs for K=2 and K=3 distinct ancestry groups with all
 261 non-related individuals included (n=37). The number to the right of each bar plot indicates the
 262 number of iterations supporting each. B) Admixture results for ten independent runs for K=2 and
 263 K=3 distinct ancestry groups with the India sample excluded. The ancestry assignments shown
 264 for K=2 were supported by all ten iterations and the ancestry assignments shown for K=3 were
 265 supported by nine (top) and one (bottom) of the ten iterations. The top two admixture plots
 266 shown in B are also included in Figure 2 of the main manuscript.



267

268 **Supplementary Figure 2. Admixture and PCA results with captive samples removed. A)**

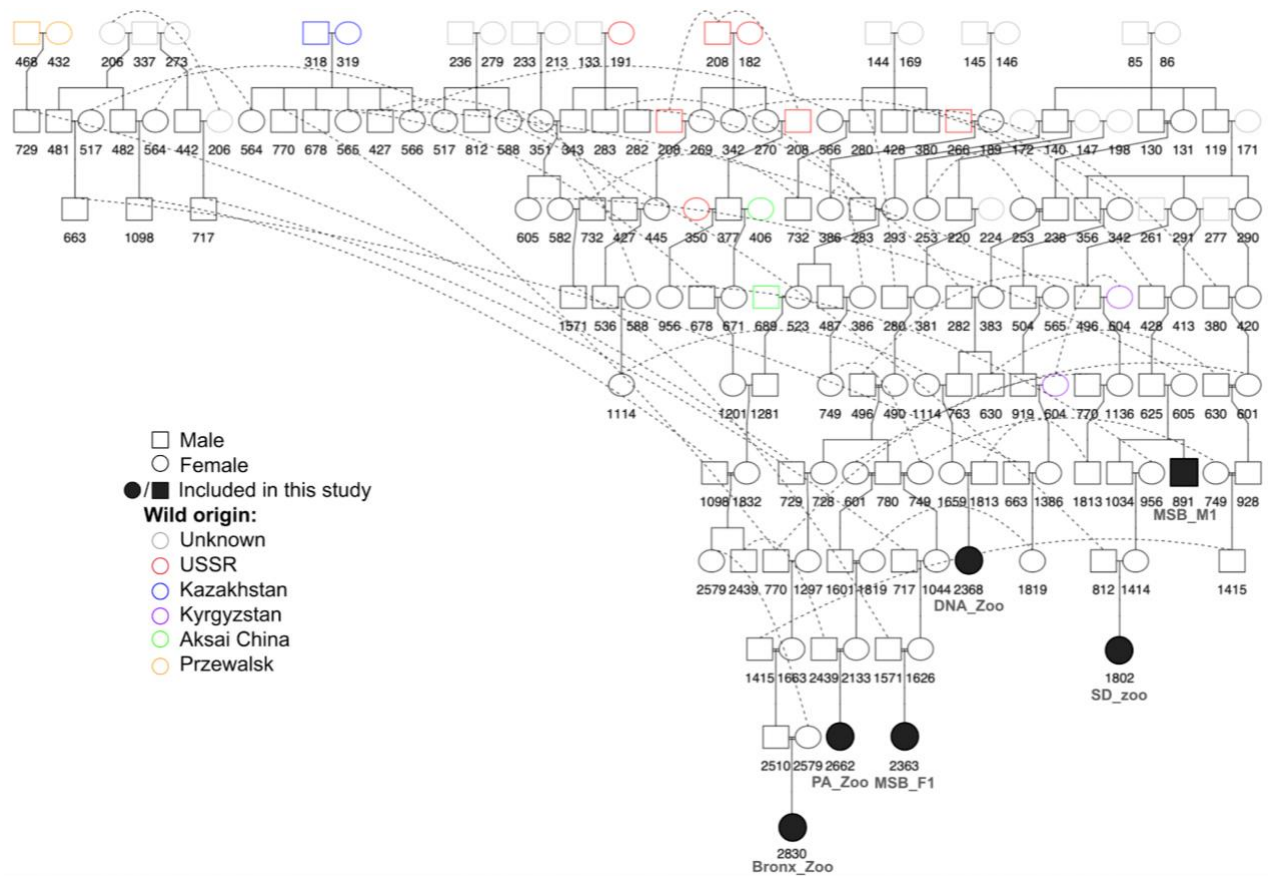
269 PCA of genetic variation for all non-related wild-origin samples. B) Admixture results for all non-
270 related wild-origin samples. C) PCA of genetic variation for all non-related wild-origin samples
271 excluding India. D) Admixture results for all non-related wild-origin samples excluding India. E)
272 PCA of genetic variation for all non-related wild-origin samples excluding India and one
273 Tajikistan sample, U13. F) Admixture results for all non-related wild-origin samples excluding
274 India and one Tajikistan sample, U13. PCA axis labels include the percent variation explained
275 by PC1 and PC2. All Admixture plots show ten independent runs for K=3 and K=2. In all cases,
276 the ancestry assignments shown for K=2 are supported by all ten iterations. The number of
277 iterations supporting each K=3 Admixture plot is indicated to the right of each.
278



279

280 **Supplementary Figure 3. Snow leopard phylogeny.** Maximum likelihood phylogeny using
 281 genome-wide SNP data from unrelated snow leopard samples with two tiger samples serving as
 282 an outgroup. 1000 bootstraps were run and node support is indicated by different colored
 283 diamonds.

284



285

286 **Supplementary Figure 4. Pedigree of captive snow leopards included in this study.** This

287 pedigree was constructed using data from studbooks and visualized using the Kinship2 package

288 in R. The origin of wild founders are indicated with different colors as identified in the legend. All

289 individuals shown with a black outline of a circle (female) or square (male) are captive-born.

290 Studbook numbers are shown under each individual. Individuals included in this study are

291 shown with solid black shapes and sample names are provided under studbook numbers.

292

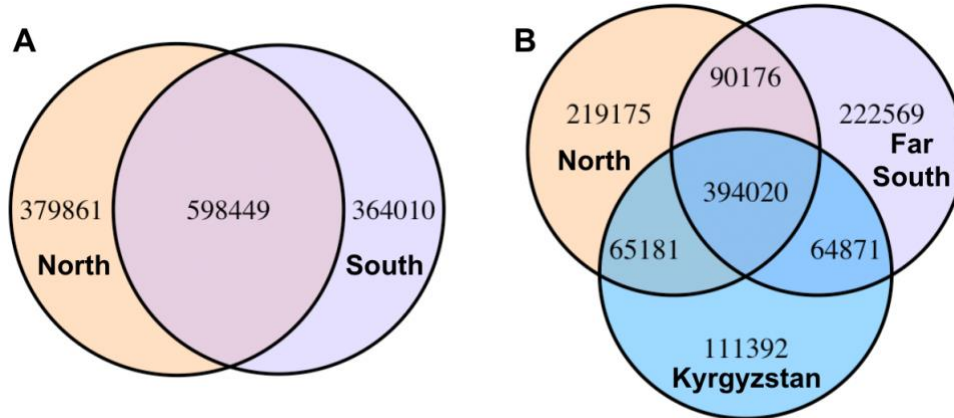
293

294

295

296

297



298

299 **Supplementary Figure 5. Shared and private SNPs among sample locations.** A) Comparing
 300 North (N=15) versus South (N=15) as identified by Admixture at K=2 (1,342,320 SNPs total). B)
 301 Comparing North, Kyrgyzstan, and Far South (all subsampled to n=7) as identified by Admixture
 302 at K=3 (1,167,384 SNPs total). Captive individuals and the Indian sample are not included in
 303 these analyses.

304

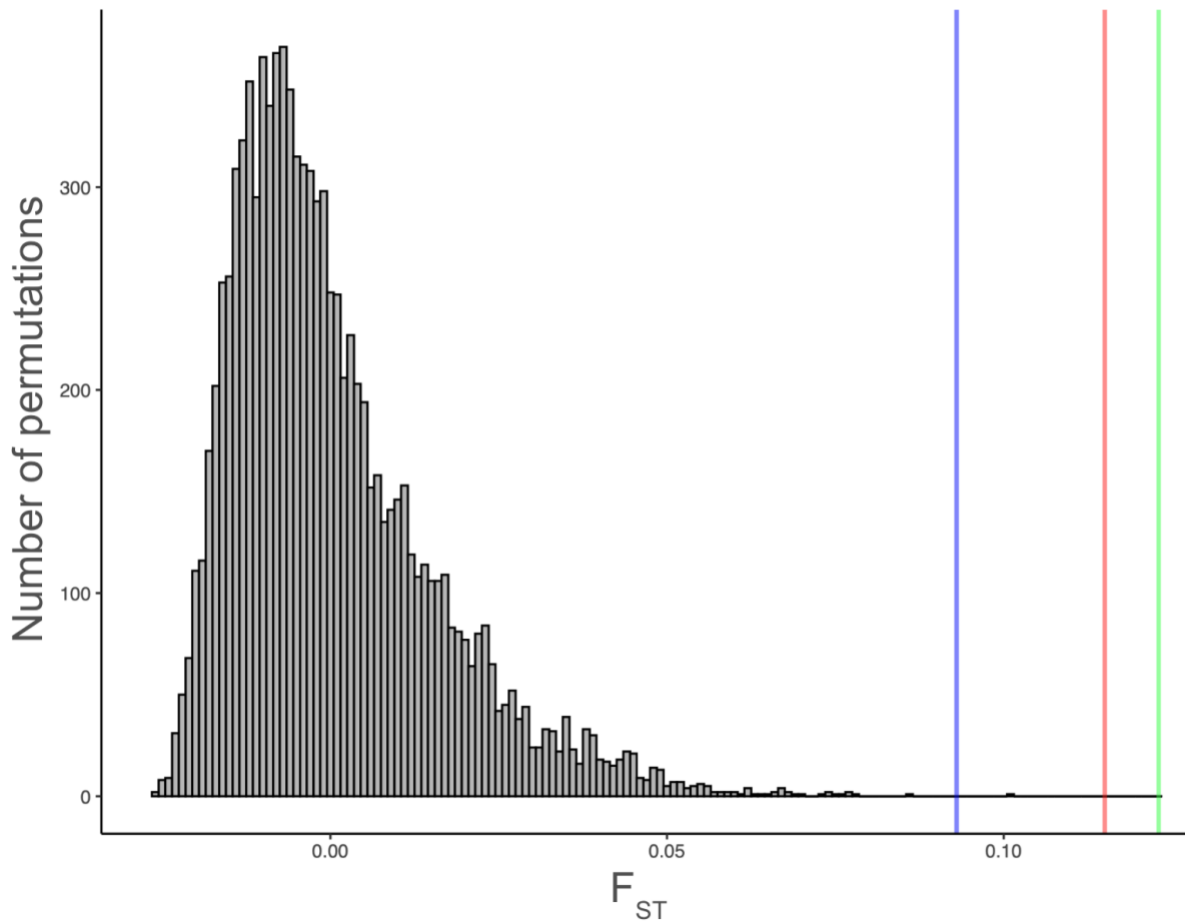
305

306

307

308

309



310

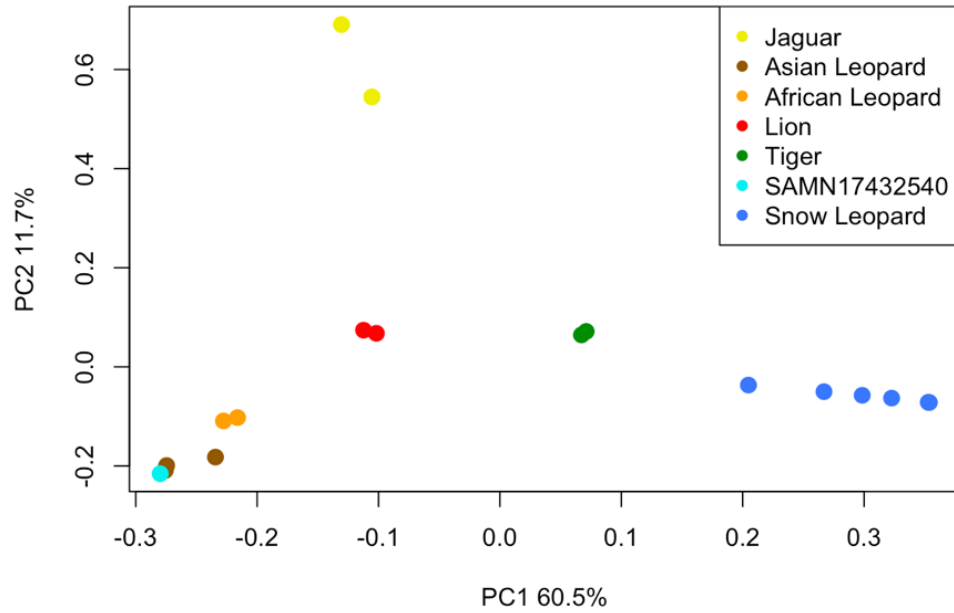
311 **Supplementary Figure 6. Comparison of observed F_{ST} values to the null distribution.** The
 312 null distribution of weighted pairwise F_{ST} generated by randomly permuting individual location
 313 assignments 10,000 times is shown in grey. Observed F_{ST} between the North and Far South
 314 (0.123) is shown with a vertical green line. Observed F_{ST} between North and Kyrgyzstan (0.115)
 315 is shown with a vertical red line. Observed F_{ST} between Kyrgyzstan (0.093) and the Far South is
 316 shown with a vertical blue line.

317

318

319

320

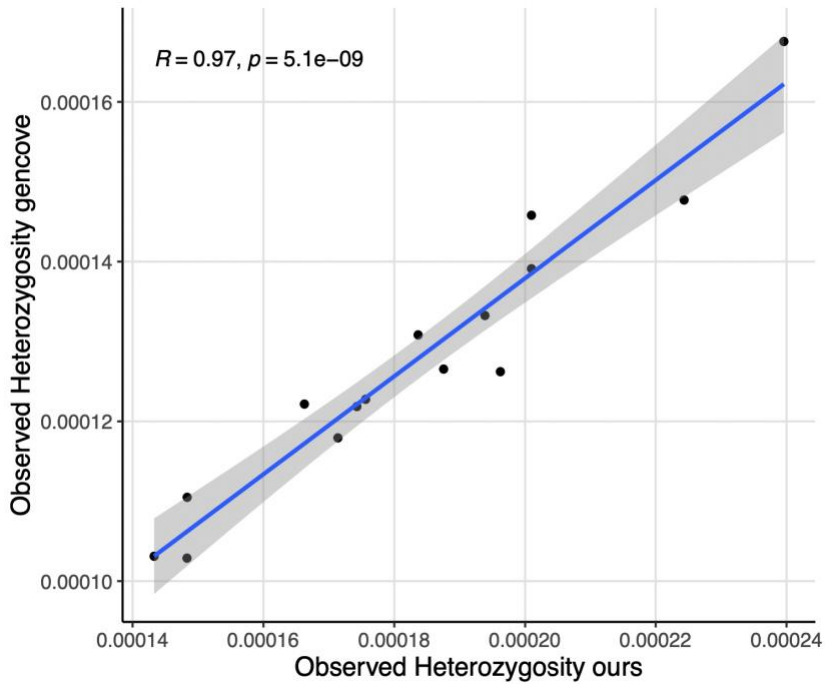


321

322 **Supplementary Figure 7. Principal component analysis showing the species ID of**
323 **Biosample SAMN17432540.** The PCA is based on whole-genome sequencing data. PC1 and
324 PC2 are visualized and the percent variance explained by each principal component is
325 indicated.

326

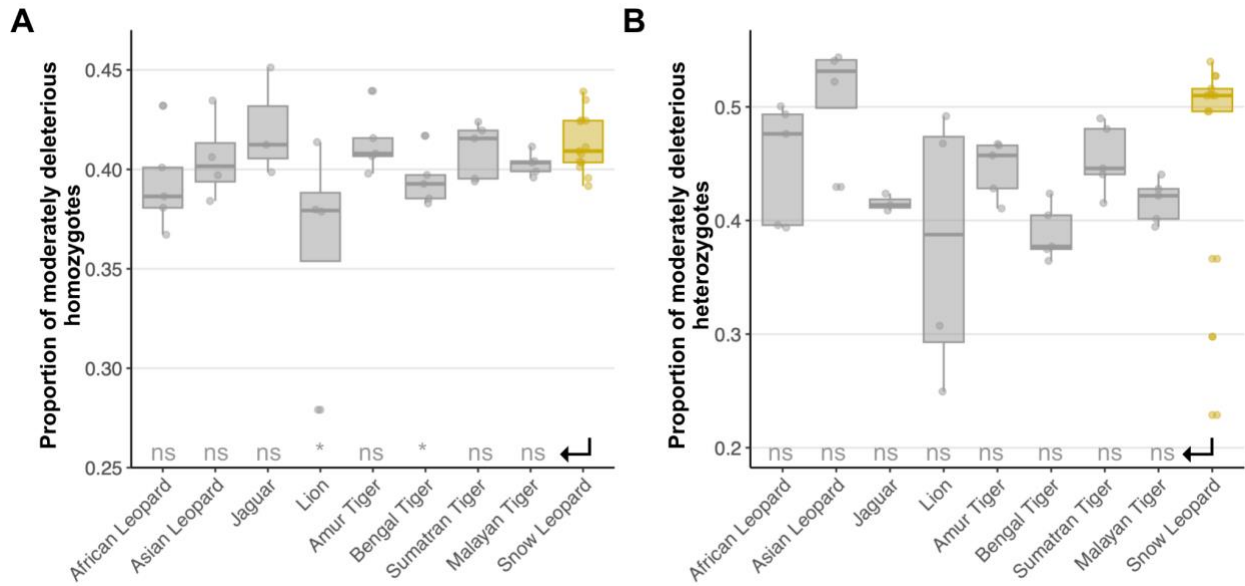
327



328

329 **Supplementary Figure 8. Comparison of observed heterozygosity calculated from**
330 **different implementations of GATK to call SNPs.** Observed heterozygosity calculated using
331 SNPs called with our pipeline are shown on the x-axis and heterozygosity calculated using
332 SNPs called by Gencove are shown on the y-axis. Pearson correlation coefficient calculated
333 using the ggpubr package in R is indicated.

334



335

336 **Supplementary Figure 9. Moderately deleterious load across *Panthera*.** A) Comparison of

337 the proportion of moderately deleterious mutations in the homozygous state across *Panthera*. B)

338 Comparison of the proportion of moderately deleterious mutations in the heterozygous state

339 across *Panthera*. In all boxplots, each point represents one individual. The lower and upper

340 edges of the boxes correspond to the first and third quartiles and the whiskers extend to the

341 lowest/highest value that is no further than 1.5*IRQ (inter-quartile range) from the box. Points

342 falling further than 1.5*IRQ from the box are plotted individually. The p-value when comparing

343 each species/subspecies to snow leopards using the Wilcoxon rank-sum test is indicated below

344 each boxplot (ns = non-significant, $p < 0.05^*$).

345 **Supplementary Tables:**
 346 **Supplementary Table 1. Sample Information.**

Sample ID	Country	Source	Group assignment at K=2	Group assignment at K=3	Depth	Breadth	# of singletons	Stud book #	Lat	Long	Collection year	Sex	NCBI Accession	DNA extraction method	Library preparation method
SD_zoo*	captive	(4)			33.65	99.1	81952	1802				F	SAMN13911153	Qiagen MagAttract kit	10× Genomics Chromium library preparation
DNA_Zoo	captive	DNA zoo			28.37	99.03	12253	2368				F	SAMN15801464		
MSB_F1	captive	current study			5.37	95.72	4090	2363				F	SAMN38638818	Qiagen DNeasy Blood and Tissue kit	Nextera DNA Library Prep
MSB_M1	captive	current study			5.37	96.08	4491	891				M	SAMN38638819	Qiagen DNeasy Blood and Tissue kit	Nextera DNA Library Prep
PA_Zoo	captive	current study			4.66	94.84	4036	2662				F	SAMN38638820	Qiagen DNeasy Blood and Tissue kit	Nextera DNA Library Prep
Bronx_Zoo	captive	current study			8.12	98.11	11983	2830				F	SAMN38638821	Qiagen DNeasy Blood and Tissue kit	Nextera DNA Library Prep
Mong_F8	Mongolia (South)	current study	North	North	5.7	95.68	3059		43.199	100.436	2012	F	SAMN38638822	Qiagen DNeasy Blood and Tissue kit	Nextera DNA Library Prep
Mong_F9	Mongolia (South)	current study	North	North	5.37	96.16	2900		43.272	100.707	2012	F	SAMN38638823	Qiagen DNeasy Blood and Tissue kit	Nextera DNA Library Prep
Mong_M1	Mongolia (South)	current study	North	North	6.22	96.37	7574		43.134	100.550	2008	M	SAMN38638824	Qiagen DNeasy Blood and Tissue kit	Nextera DNA Library Prep
Mong_M10	Mongolia (South)	current study	North	North	5.3	95.85	2834		43.183	100.437	2012	M	SAMN38638825	Qiagen DNeasy Blood and Tissue kit	Nextera DNA Library Prep
Mong_M9	Mongolia (South)	current study	North	North	6.88	97.07	3634		43.151	100.694	2011	M	SAMN38638826	Qiagen DNeasy Blood and Tissue kit	Nextera DNA Library Prep

U06	Mongolia (SW)	current study	North	North	10.72	95.61	17607		47.796	90.872	2015	M	SAMN38638827	Isogen Diatom DNA kit	NEBNext Ultra II DNA Library Prep
U07	Mongolia (SW)	current study	North	North	11.65	98.6	5082		47.569	92.587	2015	F	SAMN38638828	Qiagen QIAmp Fast DNA Tissue Kit	NEBNext Ultra II DNA Library Prep
U08	Mongolia (SW)	current study	North	North	9.19	98.43	2627		47.561	92.591	2015	M	SAMN38638829	Qiagen QIAmp Fast DNA Tissue Kit	NEBNext Ultra II DNA Library Prep
U14	Mongolia (SW)	current study	North	North	10.86	98.84	1648		46.673	93.441	2017	M	SAMN38638830	Qiagen QIAmp Fast DNA Tissue Kit	NEBNext Ultra II DNA Library Prep
U15	Mongolia (SW)	current study	North	North	4.71	90.35	3044		46.445	93.463	2016	F	SAMN38638831	Isogen Diatom DNA kit	NEBNext Ultra II DNA Library Prep
U20*	Mongolia (SW)	current study	North	North	1.37	46.59	10663		47.906	90.872	2019		SAMN38638832	CTAB (40)	NEBNext Ultra II DNA Library Prep
U04	Mongolia (NW)	current study	North	North	9.57	98.56	4474		50.294	91.068	2014	F	SAMN38638833	CTAB (40)	NEBNext Ultra II DNA Library Prep
U05	Mongolia (NW)	current study	North	North	16.79	99.03	4099		50.286	91.119	2015	M	SAMN38638834	Qiagen QIAmp Fast DNA Tissue Kit	NEBNext Ultra II DNA Library Prep
Mong_SRR836372	Mongolia (unknown)	(41)	North	North	28.87	98.86	14363					F	SAMN02086968		
U01	Russia	current study	North	North	9.59	98.12	10550		51.300	89.787	2009		SAMN38638835	CTAB (40)	NEBNext Ultra II DNA Library Prep
U02*	Russia	current study	North	North	5.14	53.73	1322527		51.847	92.068	2013	M	SAMN38638836	CTAB (40)	NEBNext Ultra II DNA Library Prep
U03	Russia	current study	North	North	9.96	97.74	21605		50.737	86.900	2015	F	SAMN38638837	Qiagen QIAmp Fast DNA Tissue Kit	NEBNext Ultra II DNA Library Prep

U09	Russia	current study	North	North	9.61	98.45	10453		50.784	89.740	2013		SAMN38638838	CTAB (40)	NEBNext Ultra II DNA Library Prep
KGZ_F1	Kyrgyzstan	current study	South	Kyrgyzstan	5.06	93.87	3963		41.945	78.576	2015	F	SAMN38638839	Qiagen DNeasy Blood and Tissue kit	Nextera DNA Library Prep
KGZ_F2	Kyrgyzstan	current study	South	Kyrgyzstan	4.64	93.4	3443		41.977	78.537	2016	F	SAMN38638840	Qiagen DNeasy Blood and Tissue kit	Nextera DNA Library Prep
KGZ_F3	Kyrgyzstan	current study	South	Kyrgyzstan	3.73	89.93	3603		41.910	78.591	2017	F	SAMN38638841	Qiagen DNeasy Blood and Tissue kit	Nextera DNA Library Prep
KGZ_F4	Kyrgyzstan	current study	South	Kyrgyzstan	4.48	93.07	2654		41.909	78.590	2017	F	SAMN38638842	Qiagen DNeasy Blood and Tissue kit	Nextera DNA Library Prep
KGZ_M1	Kyrgyzstan	current study	South	Kyrgyzstan	6.14	95.87	5193		41.873	78.579	2016	M	SAMN38638843	Qiagen DNeasy Blood and Tissue kit	Nextera DNA Library Prep
KGZ_M2	Kyrgyzstan	current study	South	Kyrgyzstan	4.78	93.47	4115		41.909	78.590	2016	M	SAMN38638844	Qiagen DNeasy Blood and Tissue kit	Nextera DNA Library Prep
KGZ_M3	Kyrgyzstan	current study	South	Kyrgyzstan	6.27	96.17	5150		41.876	78.578	2016	M	SAMN38638845	Qiagen DNeasy Blood and Tissue kit	Nextera DNA Library Prep
U10	Tajikistan	current study	South	Far South	6.56	88.07	7407		37.875	73.691	2016	F	SAMN38638846	Isogen Diatom DNA kit	NEBNext Ultra II DNA Library Prep
U11	Tajikistan	current study	South	Far South	7.3	73.44	4185		37.875	73.692	2016		SAMN38638847	Isogen Diatom DNA kit	NEBNext Ultra II DNA Library Prep
U12*	Tajikistan	current study	South	Far South	1	13.62	1319		37.835	74.076	2018		SAMN38638848	CTAB (40)	NEBNext Ultra II DNA Library Prep
U13	Tajikistan	current study	South	Far South	6.71	95.15	22190		38.455	74.372	2018		SAMN38638849	CTAB (40)	NEBNext Ultra II DNA Library Prep
AF_05	Afghanistan	current study	South	Far South	7.66	98.5	10554		38.301	71.177	2019	F	SAMN38638850	Macherey-Nagel Nucleospin Genomic DNA from tissue	Illumina DNA Prep

AF_06	Afghanistan	current study	South	Far South	3	85.4	6779		38.200	70.933	2020	F	SAMN38638851	Macherey-Nagel Nucleospin Genomic DNA from tissue	Illumina DNA Prep
AF_07	Afghanistan	current study	South	Far South	3.4	87.73	8081		36.948	72.973	2010		SAMN38638852	Macherey-Nagel Nucleospin Genomic DNA from tissue	Illumina DNA Prep
AF_08	Afghanistan	current study	South	Far South	9.25	98.83	10463		36.963	72.902	2011		SAMN38638853	Macherey-Nagel Nucleospin Genomic DNA from tissue	Illumina DNA Prep
Leo	Pakistan (wild born, currently captive)	current study	South	Far South	9.06	98.13	8484	2630	36.249	74.045	2021	M	SAMN38638854	Qiagen DNeasy Blood and Tissue kit	Nextera DNA Library Prep
India_NCBS	India	National Centre for Biological Sciences-TIFR			12.01	98.88	49158		32.300	78.010	2012		PRJNA1051290		

347 *Samples that were removed from all analyses due to sequencing quality issues.

348
349
350

Supplementary Table 2. Kinship coefficients calculated using SNPrelate for all pairs with non-zero coefficients.

Group	Individual 1	Individual 2	Kinship
North	U09	U01*	0.3350
Kyrgyzstan_captive	SL_KGZ_F1	SL_KGZ_F4*	0.1482
Far South	AF_SL_07	AF_SL_06*	0.1354
North	U14	U08*	0.1242
Kyrgyzstan_captive	SL_KGZ_F2	SL_KGZ_M3	0.0699
North	U08	U07	0.0641
North	SL_Mong_M10	SL_Mong_F9	0.0628
North	U14	U07	0.0623
North	SL_Mong_M10	SL_Mong_F8	0.0488
North	Mong_SRR836372	U04	0.0375
Kyrgyzstan_captive	SL_KGZ_F3	SL_KGZ_F4	0.0343
Kyrgyzstan_captive	SL_KGZ_M2	SL_KGZ_M1	0.0334
North	SL_Mong_F9	SL_Mong_F8	0.0313
North	U15	U07	0.0226
North	SL_Mong_M10	SL_Mong_M9	0.0219
North	Mong_SRR836372	U05	0.0154
North	SL_Mong_F9	SL_Mong_M9	0.0153
Kyrgyzstan_captive	PA_Zoo	SL_M11136_Bronx	0.0143
Kyrgyzstan_captive	SL_KGZ_F3	SL_KGZ_F1	0.0139
Kyrgyzstan_captive	DNAzoo	SL_M11136_Bronx	0.0096
North	U03	U07	0.0085
North	U05	U04	0.0075
North	SL_Mong_F8	SL_Mong_M9	0.0072
Kyrgyzstan_captive	MSB_F1	MSB_M1	0.0060
North	U06	SL_Mong_M1	0.0047
North	SL_Mong_M1	U06	0.0047
North	U15	U14	0.0046
North	U03	U04	0.0044
North	U03	U06	0.0040
Kyrgyzstan_captive	SL_KGZ_F2	SL_KGZ_F3	0.0028
North	U08	U05	0.0023
North	U15	U06	0.0019
Kyrgyzstan_captive	SL_M11136_Bronx	MSB_F1	0.0017

351 *Samples removed from analyses potentially impacted by the presence of related individuals as
352 indicated in the main text.

353
354
355
356

357
358

Supplementary Table 3. Published whole-genome sequencing data used for heterozygosity comparisons among all big cat species.

Species/group	sample	depth	breadth	Observed Heterozygosity
Amur Tiger	SRR13647652	5.81	96.27	0.00038
Amur Tiger	SRR13647654	7.70	98.68	0.00041
Amur Tiger	SRS8209282	17.46	99.45	0.00058
Amur Tiger	SRS8209284	13.89	99.32	0.00063
Amur Tiger	SRS8209286	8.67	99.10	0.00044
Bengal Tiger	SRR13500746	8.45	99.32	0.00066
Bengal Tiger	SRR13500748	15.01	99.41	0.00087
Bengal Tiger	SRR13500752	21.20	99.49	0.00063
Bengal Tiger	SRR13500756	18.19	99.53	0.00051
Bengal Tiger	SRR13500762	8.95	99.37	0.00073
Malayan Tiger	SRR13647621	6.20	97.63	0.00040
Malayan Tiger	SRR13647627	6.14	97.10	0.00052
Malayan Tiger	SRS8209294	16.35	99.43	0.00080
Malayan Tiger	SRS8209296	32.27	99.56	0.00091
Malayan Tiger	SRS8209300	22.27	99.53	0.00090
Sumatran Tiger	SRR13647577	6.35	97.52	0.00035
Sumatran Tiger	SRR13647578	6.32	97.50	0.00033
Sumatran Tiger	SRR13647591	5.67	96.56	0.00031
Sumatran Tiger	SRS8209305	31.23	99.52	0.00057
Sumatran Tiger	SRS8209313	22.27	99.54	0.00058
Lion	SRR10009886	37.66	99.15	0.0010
Lion	SRR13242484	21.22	98.98	0.00075
Lion	SRR836361	29.38	98.77	0.00064
Lion	SRR836370	24.16	98.70	0.00054
Jaguar	SRR11097154	22.75*	99.00	0.00089
Jaguar	SRR14572000	21.67*	98.66	0.00050
Jaguar	SRR4444360	22.42*	98.52	0.00094
African Leopard	ERR5671299	19.50	99.23	0.0021

African Leopard	ERR5671307	16.42	98.98	0.0023
African Leopard	ERR5671308	18.04*	99.35	0.0021
African Leopard	ERR5671314	20.22*	99.85	0.0017
African Leopard	ERR5671315	22.64	99.69	0.0019
Asian Leopard	ERR5671303	12.55	81.12	0.00061
Asian Leopard	ERR5671311	18.14*	99.30	0.00055
Asian Leopard	ERR5671313	22.08*	99.48	0.00071
Asian Leopard	ERR5671323	7.87	96.00	0.00099
Cheetah	SRR2737540	5.64	97.58	0.00032
Cheetah	SRR2737541	7.12	98.23	0.00036
Cheetah	SRR2737542	7.76	98.19	0.00037
Cheetah	SRR2737543	6.64	98.10	0.00037
Cheetah	SRR2737544	6.29	97.81	0.00034
Cheetah	SRR2737545	7.52	98.13	0.00038
Cheetah	SRR9951918	24.29	99.55	0.00051
Puma	BR406 (SRR7542886, SRR7542887, SRR7542888)	32.49	94.10	0.0018
Puma	CYP47 (SRR7664677, SRR7664678)	42.56	94.18	0.00040
Puma	EVG21 (SRR7660678, SRR7660679)	51.92	94.23	0.00066
Puma	SC29 (SRR7537344, SRR7537345)	29.22	94.15	0.00072
Puma	SMM22 (SRR7543017, SRR7543018)	34.81	94.18	0.00061
Puma	YNP198 (SRR7610940, SRR7610941)	39.03	94.19	0.00097

359 In all cases, the whole-genome sequencing data was mapped to the reference genome for that
360 species (Supplementary Table 4), SNPs were called using GATK, and heterozygosity was
361 estimated using VCFtools. *Bam files that were downsampled. The depth after downsampling is
362 shown.
363
364
365
366
367

368 **Supplementary Table 4. Genome mappability calculations for all big cat reference**
 369 **genomes used in heterozygosity analyses.**

Species	Reference genome	Total bp	unmappable bp	mappable bp
Lion	GCA_008795835.1	2387958570	441792892	1946165678
Puma	GCF_003327715.1	2314358962	401315200	1913043762
Jaguar	GCA_028533385.1	2468400402	544548340	1923852062
Tiger	GCA_021130815.1	2405965077	460360448	1945604629
Cheetah	GCF_003709585.1	2373450562	428515259	1944935303
Leopard	GCA_024362965.1	2441143628	494445445	1946698183
Snow leopard	GCF_023721935.1	2205089538	386922644	1818166894

370

KGZ_F3	Kyrgyzstan	3.73	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X				
KGZ_F4	Kyrgyzstan	4.48											X	X							
KGZ_M1	Kyrgyzstan	6.14	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X				
KGZ_M2	Kyrgyzstan	4.78	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X				
KGZ_M3	Kyrgyzstan	6.27	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X				
U10	Tajikistan	6.56	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X				
U11	Tajikistan	7.3	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X				
U13	Tajikistan	6.71	X	X		X	X		X	X	X		X	X	X	X	X				
AF_05	Afghanistan	7.66	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X				
AF_06	Afghanistan	3											X								
AF_07	Afghanistan	3.4	X	X	X	X	X	X	X	X	X	X	X	X	X						
AF_08	Afghanistan	9.25	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Leo	Pakistan (wild born, currently captive)	9.06	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
India_NCB S	India	12.01	X			X			X			X						X	X	X	X

373 Note that samples U02, U12, U20 and SD_zoo were removed from all analyses for sample quality issues described in the methods.
374 Xs indicate samples that were included in each analysis. In all analyses where downsampling was required to have an equal number
375 of samples in each group, samples included are colored by group membership. Downsampling was done by first removing
376 geographically redundant samples followed by the lowest coverage samples.
377
378

379 **Supplementary Citations:**

- 380 1. C. Zhou, *et al.*, Characterization of Olfactory Receptor Repertoires in the Endangered Snow
381 Leopard Based on the Chromosome-Level Genome. *DNA Cell Biol.* **40**, 293–302 (2021).
- 382 2. C. Zhou, *et al.*, Characterization of microsatellites in the endangered snow leopard based
383 on the chromosome-level genome. *Mamm. Res.* **66**, 385–398 (2021).
- 384 3. H. Li, Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.
385 *arXiv [q-bio.GN]* (2013).
- 386 4. E. E. Armstrong, *et al.*, Genome report: chromosome-level draft assemblies of the snow
387 leopard, African leopard, and tiger (*Panthera uncia*, *Panthera pardus pardus*, and *Panthera*
388 *tigris*). *G3* **12** (2022).
- 389 5. H. Li, *et al.*, The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–
390 2079 (2009).
- 391 6. T. S. Korneliussen, A. Albrechtsen, R. Nielsen, ANGSD: Analysis of Next Generation
392 Sequencing Data. *BMC Bioinformatics* **15**, 356 (2014).
- 393 7. M. Fumagalli, F. G. Vieira, T. Linderoth, R. Nielsen, ngsTools: methods for population
394 genetics analyses from next-generation sequencing data. *Bioinformatics* **30**, 1486–1487
395 (2014).
- 396 8. C. Pockrandt, M. Alzamel, C. S. Iliopoulos, K. Reinert, GenMap: Fast and Exact
397 Computation of Genome Mappability. *Cold Spring Harbor Laboratory* 611160 (2019).
- 398 9. A. R. Quinlan, I. M. Hall, BEDTools: a flexible suite of utilities for comparing genomic
399 features. *Bioinformatics* **26**, 841–842 (2010).
- 400 10. P. Danecek, *et al.*, The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158
401 (2011).
- 402 11. X. Zheng, *et al.*, A high-performance computing toolset for relatedness and principal
403 component analysis of SNP data. *Bioinformatics* **28**, 3326–3328 (2012).
- 404 12. R Core Team, *R: A language and environment for statistical computing* (2020).
- 405 13. S. Purcell, *et al.*, PLINK: a tool set for whole-genome association and population-based
406 linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
- 407 14. A. G. Nazareno, J. B. Bemmels, C. W. Dick, L. G. Lohmann, Minimum sample sizes for
408 population genomics: an empirical study from an Amazonian plant species. *Mol. Ecol.*
409 *Resour.* **17**, 1136–1147 (2017).
- 410 15. K. S. Toyama, P.-A. Crochet, R. Leblois, Sampling schemes and drift can bias admixture
411 proportions inferred by structure. *Mol. Ecol. Resour.* **20**, 1769–1785 (2020).
- 412 16. S. J. Puechmaille, The program structure does not reliably recover the correct population
413 structure when sampling is uneven: subsampling and new estimators alleviate the problem.
414 *Mol. Ecol. Resour.* **16**, 608–627 (2016).

- 415 17. D. H. Alexander, J. Novembre, K. Lange, Fast model-based estimation of ancestry in
416 unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).
- 417 18. N. M. Kopelman, J. Mayzel, M. Jakobsson, N. A. Rosenberg, I. Mayrose, Clumpak: a
418 program for identifying clustering modes and packaging population structure inferences
419 across K. *Mol. Ecol. Resour.* **15**, 1179–1191 (2015).
- 420 19. S. Neph, *et al.*, BEDOPS: high-performance genomic feature operations. *Bioinformatics* **28**,
421 1919–1920 (2012).
- 422 20. Creators Edgardo M. Ortiz1 Show affiliations 1. Technical University of Munich, *vcf2phylop*
423 *v2.0: convert a VCF matrix into several matrix formats for phylogenetic analysis*.
- 424 21. L.-T. Nguyen, H. A. Schmidt, A. von Haeseler, B. Q. Minh, IQ-TREE: a fast and effective
425 stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**,
426 268–274 (2015).
- 427 22. FigTree. Available at: <http://tree.bio.ed.ac.uk/software/figtree/> [Accessed 30 May 2024].
- 428 23. H. Li, A statistical framework for SNP calling, mutation discovery, association mapping and
429 population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987–
430 2993 (2011).
- 431 24. B. S. Weir, C. C. Cockerham, Estimating F-Statistics for the Analysis of Population
432 Structure. *Evolution* **38**, 1358–1370 (1984).
- 433 25. Picard. Available at: <http://broadinstitute.github.io/picard/> [Accessed 26 November 2023].
- 434 26. G. A. Van der Auwera, B. D. O'Connor, *Genomics in the Cloud: Using Docker, GATK, and*
435 *WDL in Terra* (“O’Reilly Media, Inc.,” 2020).
- 436 27. P. Cingolani, *et al.*, A program for annotating and predicting the effects of single nucleotide
437 polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118;
438 iso-2; iso-3. *Fly (Austin)* **6**, 80–92 (2012).
- 439 28. H. Li, Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–
440 3100 (2018).
- 441 29. A. Guarracino, *AndreaGuarracino/paf2chain: v0.1.1* (Zenodo, 2024).
- 442 30. Picard. Available at: <http://broadinstitute.github.io/picard> [Accessed 13 August 2025].
- 443 31. A. S. Hinrichs, *et al.*, The UCSC Genome Browser Database: update 2006. *Nucleic Acids*
444 *Res* **34**, D590–8 (2006).
- 445 32. R. S. Taylor, M. Manseau, P. Liu, P. J. Wilson, Protocol for genetic load analysis in caribou
446 using a modified genomic evolutionary rate profiling. *STAR Protoc* **6**, 103789 (2025).
- 447 33. P. Danecek, *et al.*, Twelve years of SAMtools and BCFtools. *Gigascience* **10** (2021).
- 448 34. E. V. Davydov, *et al.*, Identifying a high fraction of the human genome to be under selective
449 constraint using GERP++. *PLoS Comput Biol* **6**, e1001025 (2010).

- 450 35. H. V. Figueiró, *et al.*, Genome-wide signatures of complex introgression and adaptive
451 evolution in the big cats. *Sci Adv* **3**, e1700299 (2017).
- 452 36. G. Femerling, *et al.*, Genetic load and adaptive potential of a recovered avian species that
453 narrowly avoided extinction. *Mol. Biol. Evol.* **40** (2023).
- 454 37. R. S. Taylor, *et al.*, High genetic load without purging in caribou, a diverse species at risk.
455 *Curr. Biol.* **34**, 1234–1246.e7 (2024).
- 456 38. T. Lan, *et al.*, Enhancing inbreeding estimation and global conservation insights through
457 chromosome-level assemblies of the Chinese and Malayan pangolin. *Gigascience* **14**
458 (2025).
- 459 39. N. Dussex, *et al.*, Adaptation to the High-Arctic island environment despite long-term
460 reduced genetic variation in Svalbard reindeer. *iScience* **26**, 107811 (2023).
- 461 40. J. J. Doyle, J. L. Doyle, A rapid DNA isolation procedure for small quantities of fresh leaf
462 tissue. *Phytochemical bulletin* (1987).
- 463 41. Y. S. Cho, *et al.*, The tiger genome and comparative analysis with lion and snow leopard
464 genomes. *Nat. Commun.* **4**, 2433 (2013).
- 465