Whole genome sequencing and the application of a SNP panel reveal primary evolutionary lineages and genomic variation in the lion (*Panthera leo*)

L.D. Bertola[1,2,3,*], M. Vermaat[4,5], F. Lesilau[2,6], M. Chege[2,6], P.N. Tumenta[7,8], E.A. Sogbohossou[9], O.D. Schaap[3], H. Bauer[10], B.D. Patterson[11], P.A. White[12], H.H. de Iongh[2,13], J.F.J. Laros[4,5] and K. Vrieling[3]

[1] City University of New York, City College of New York, 160 Convent Avenue, NY 10031, New York, USA

[2] Leiden University, Institute of Environmental Sciences (CML), PO Box 9518, 2300 RA Leiden, The Netherlands

[3] Leiden University, Institute of Biology Leiden (IBL), PO Box 9505, 2300 RA Leiden, The Netherlands

[4] Department of Human Genetics, Leiden University Medical Center, 2300 RC Leiden, The Netherlands

[5] Leiden Genome Technology Center, Leiden University Medical Center, 2300 RC Leiden, The Netherlands

[6] Kenya Wildlife Service, Nairobi, Kenya

[7] Centre for Environment and Developmental Studies, Cameroon (CEDC)

[8] Regional Training Centre Specialized in Agriculture, Forest and Wood. University of Dschang, BP 138, Yaounde, Cameroon

[9] Laboratoire d'Ecologie Appliquée, Université d'Abomey-Calavi, 03 BP 294, Cotonou, Benin

[10] Wildlife Conservation Research Unit, Zoology, University of Oxford Recanati-Kaplan Centre, Tubney OX13 5QL, United Kingdom

[11] Negaunee Integrative Research Center, Field Museum of Natural History, Chicago, IL 60605 USA

[12] University of California, Center for Tropical Research, Institute of the Environment and Sustainability, Los Angeles, CA 90095-1496, USA

26 [13] University of Antwerp, Department of Biology, Evolutionary Ecology Group, Groenenborgerlaan
27 171, 2020 Antwerpen, Belgium
28
29 * Corresponding author:
30 Laura Bertola
31 Email: laura.bertola@gmail.com
32

33 **Abstract**
34 **Background**
35 Previous phylogeographic studies of the lion (*Panthera leo*) have improved our insight into the
36 distribution of genetic variation, as well as a revised taxonomy which now recognizes a northern
37 (*Panthera leo leo*) and a southern (*Panthera leo melanochaita*) subspecies. However, existing whole
38 range phylogeographic studies on lions either consist of very limited numbers of samples, or are
39 focused on mitochondrial DNA and/or a limited set of microsatellites. The geographic extent of
40 genetic lineages and their phylogenetic relationships remain uncertain, clouded by massive sampling
41 gaps, sex-biased dispersal and incomplete lineage sorting.
42 **Results**
43 In this study we present results of low depth whole genome sequencing and subsequent variant
44 calling in ten lions sampled throughout the geographic range, resulting in the discovery of >150,000
45 Single Nucleotide Polymorphisms (SNPs). Phylogenetic analyses revealed the same basal split
46 between northern and southern populations as well as four population clusters on a more local scale.
47 Further, we designed a SNP panel, including 125 autosomal and 14 mitochondrial SNPs, which was
48 tested on >200 lions from across their range. Results allow us to assign individuals to one of these
49 four major clades (West & Central Africa, India, East Africa, or Southern Africa) and delineate these
50 clades in more detail.

Conclusions

The results presented here, particularly the validated SNP panel, have important applications, not only for studying populations on a local geographic scale, but also for tracing samples of unknown origin for forensic purposes, and for guiding conservation management of *ex situ* populations. Thus, these genomic resources not only contribute to our understanding of the evolutionary history of the lion, but may also play a crucial role in conservation efforts aimed at protecting the species in its full diversity.

Keywords: Single Nucleotide Polymorphism (SNP), lion, genomics, genotyping, evolutionary history, genomic variation, wildlife, conservation, Africa, biodiversity

Background

Recent developments in next generation sequencing (NGS) techniques allow for the application of massive parallel sequencing to non-model organisms (1,2), such as the lion (*Panthera leo*). As a result, both the evolutionary history of a species and population histories can be reconstructed based on vastly expanded datasets (3,4). In addition to improved insight into the geographic distribution of genetic variation, this type of data can inform conservation efforts on how best to maintain this diversity (5,6). This is particularly relevant considering the dire situation of the lion populations in West and Central Africa, where downward populations trends are the strongest, and local extinctions have been reported in recent decades (7–11). The IUCN Red List currently classifies the lion as 'Vulnerable' across its range, but states that it would actually meet the criteria for 'Endangered' in East and Central Africa, with lions in West Africa being 'Critically Endangered' (12,13). Having a better understanding of intra-specific diversity in the lion can steer conservation efforts towards halting the loss of diversity.

76 The lion (*P. leo*) has been the subject of several phylogeographic studies which have provided insights
77 into the evolution and distribution of genetic variation in the African populations (formerly
78 subspecies *P. leo leo*) and its connection to the Indian population (formerly subspecies *P. leo persica*),
79 located in and around the Gir National Park, Gujarat state, India. Although this Africa-Asia split has
80 long been used to inform management (e.g. there is a separate studbook for Asiatic lions in zoos),
81 this taxonomic distinction has since been overhauled. Phylogeographic studies played an important
82 part in this, by providing improved understanding of the evolutionary history and relationships
83 between populations. These studies included data from mitochondrial DNA (mtDNA) (14–23),
84 autosomal DNA (18–20,22–24) and subtype variation in lion Feline Immunodeficiency Virus (FIV_{Ple})
85 (19). Studies using mitochondrial markers and/or complete mitogenomes, describe a basal
86 dichotomy, consisting of a northern group that includes populations from West and Central Africa as
87 well as the Indian population (formerly recognized as a distinct subspecies), and a southern group
88 with populations from East and Southern Africa (14,16,18,21).
89 Two studies have included Single Nucleotide Polymorphisms, although both have included only two
90 sampling localities in West and Central Africa and exact sampling localities were often unknown,
91 hence calling for improved sampling in the region (22,24). The distinction of lions in West and Central
92 Africa was further corroborated by autosomal microsatellite data (20). Reliable inference of the
93 evolutionary relationships between African and the Indian populations based on these data proved
94 to be a challenge, as the extremely low variation in the Indian population led to an unresolved
95 position in the distance tree (20).. Nevertheless, these newly described evolutionary relationships
96 are reflected in a revised taxonomy which now recognizes a northern subspecies (*P. leo leo*) and a
97 southern subspecies (*P. leo melanochaita*) (25).
98
99 However, each of the approaches used in the studies mentioned above have limitations. Although
100 mtDNA is regarded as a useful genetic marker for gaining insight into phylogeographic patterns,

101 partly because of its shorter coalescence time compared to nuclear markers, it may lead to 'over-
102 splitting', reflecting fully coalesced groups based on mtDNA data, but incomplete lineage sorting of
103 nuclear DNA (nuDNA) alleles. Moreover, mtDNA cannot identify admixture and represents only a
104 single locus, potentially misrepresenting phylogenetic relationships due to the stochastic nature of
105 the coalescent (26,27). In addition, sex-biased dispersal and gene flow will alter patterns derived
106 from mitochondrial versus nuclear data. Because female lions exhibit strong philopatry whereas male
107 lions are capable dispersers (28,29), phylogeographic patterns based on mtDNA in lions may
108 overestimate divergence between populations. Inference of phylogenetic relationships from
109 microsatellite data, on the other hand, is problematic due to their high variability and their mutation
110 pattern leading to homoplasy. Moreover microsatellite studies typically employ only a few dozen
111 markers which effectively represents <1% of the genome. Finally, most studies mentioned above
112 were limited by sparse coverage of populations, notably in West and Central Africa. In particular,
113 FIV_{Ple} prevalence is geographically restricted. Additional data from genome-wide markers, and
114 covering more lion populations, are necessary to overcome these shortcomings. This will improve
115 our understanding of the spatial distribution of variation in the lion, as well as help guide future
116 conservation efforts that seek to preserve the species' genetic diversity. As patterns of instra-specific
117 variation are often shared across co-distributed taxa (21,30), a deeper insight into these patterns of
118 lion variation will also be relevant for other species.
119
120 Here, we describe the discovery and phylogeographic analysis of genome-wide SNPs based on whole
121 genomes and complete mitogenomes of ten lions, providing an overview of the intraspecific genomic
122 variation. We further developed a SNP panel consisting of a subset (N=125) of the discovered SNPs,
123 which was then used for genotyping >200 samples from 14 countries, representing almost the entire
124 current distribution of the lion. This resolves phylogeographic breaks at a finer spatial resolution, and
125 may serve as a reference dataset for future studies. Finally, we discuss the applications and future

126 directions of high-throughput genotyping for wildlife research and conservation with
127 recommendations on how they may contribute to future studies on conservation genomics.
128
129 **Results**
130 **Sequencing**
131 The sequencing runs yielded a total of $6.5 \cdot 10^8$ reads for all samples combined, corresponding to an
132 average coverage of ~3.8x per lion (see Supplemental Table 1 for coverage of autosomes and
133 mitogenomes per individual). Following quality control, a total of $5.9 \cdot 10^8$ reads (94.4%) were retained
134 for subsequent alignment (Supplemental Table 1). More information on quality control is included in
135 Supplemental Information 1 (and associated Supplemental Figures 1-4). Filtering of variable positions
136 in lions with GATK yielded a total of 155,678 SNPs. Upon filtering for positions with coverage $\geq$3, we
137 retained 118,270 SNPs of which 98,952 SNPs were called in at least five individuals (see
138 Supplemental Table 2 for results for different levels of missing data). Missing data ranged from 90%
139 (Benin) to 10% (Kenya) for all variable positions. Results reported for downstream phylogenetic
140 analysis are based on the lion-specific variable positions which could be called in at least half of the
141 included samples (98,952 SNPs).
142
143 **Whole genome data and complete mitogenomes**
144 Phylogenetic analyses, based on 98,952 SNPs, show a well-supported dichotomy between the
145 northern populations (Benin, Cameroon, Demoratic Republic of the Congo (DRC), India) and the
146 southern ones (Somalia, Kenya, Zambia1, Zambia2, Republic of South Africa (RSA), Namibia) (Figure
147 1: left tree). Phylogenetic trees based on the mitochondrial genomes show the same basal
148 dichotomy, with the Indian population nested within the northern populations (Figure 1: right tree),
149 as was previously reported (21). However, it must be noted that the individual from RSA contains a
150 haplotype from Namibia, likely the result of historic translocation, as was previously described (21).

151 MrBayes, Garli and SVDquartets resulted largely in the same topology, although the three methods

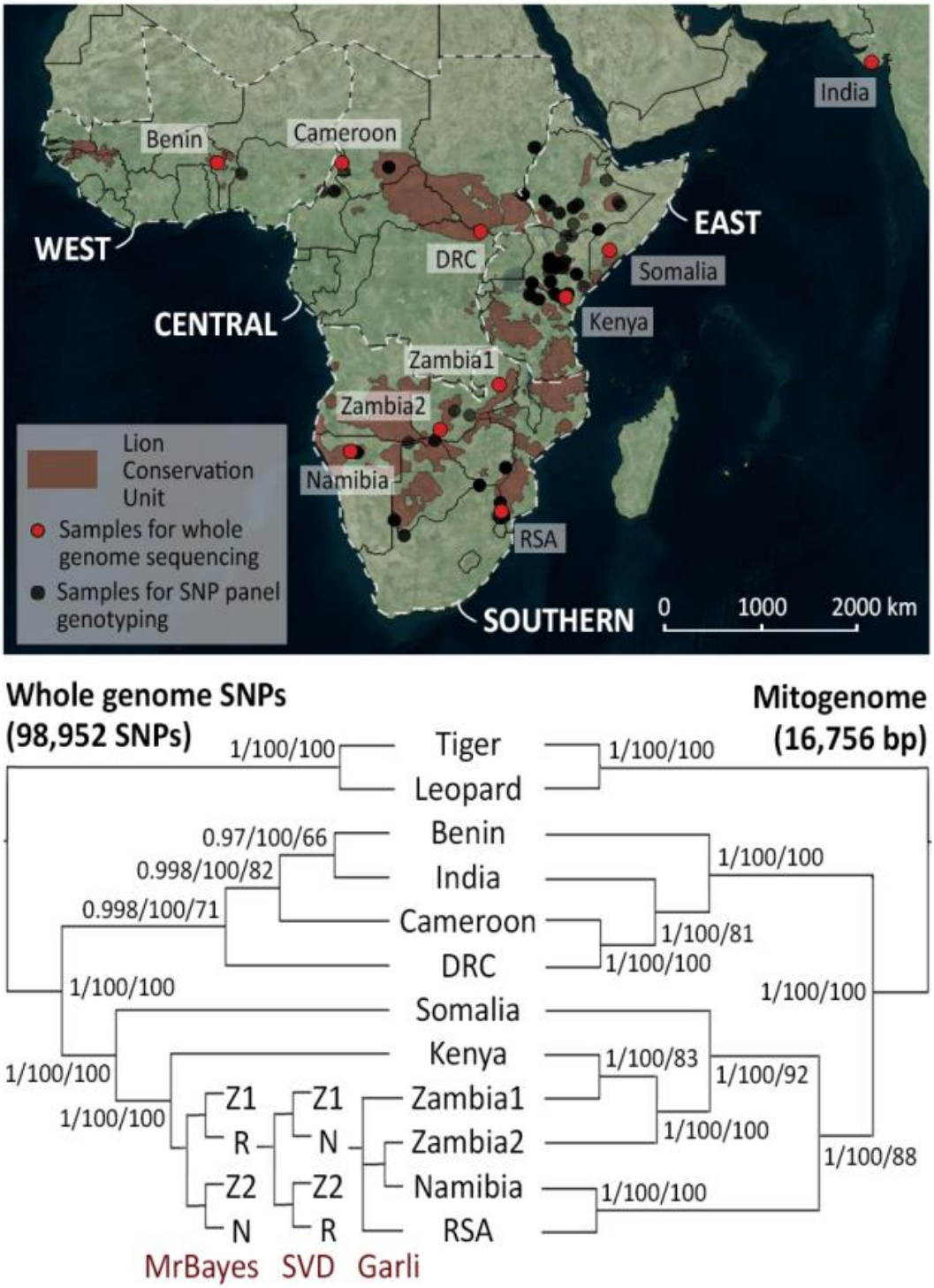
152 do not agree on the relationships of Zambia1, Zambia2, RSA and Namibia based on the autosomal

153 SNPs. A dendrogram based on genotype likelihoods shows the same basal dichotomy, including a

154 split within the Southern subspecies (Supplemental Information 2, Supplemental Figure 5).

155

156



157 **Figure 1. Distribution of lion sampling localities and inferred phylogenetic relationships between populations.** Map
158 indicating sampling locations of lions for whole genome sequencing (red) and SNP panel genotyping (black) (top). Red
159 shading indicates Lion Conservation Units, and white delineation defines the regions (West, Central, East, and
160 Southern), sensu the lion conservation strategies (93). Phylogenetic trees, based on 98,952 nuclear SNPs (left) and
161 16,756 bp mitogenomes (right) for 10 lions which were subjected to whole genome sequencing (bottom). Support
162 values indicate posterior probabilities (MrBayes), bootstrap support from SVDquartets and bootstrap support from
163 Garli. Topologies are indicated per method in the southern branch of the autosomal SNP tree, and each split is
164 maximally supported, with the exception of the Zambia2+Namibia branch in the Garli tree, which received a
165 bootstrap support of 96. Z1: Zambia1, Z2: Zambia2, N: Namibia, R: RSA.
166
167 PCA based on genotype likelihoods and called genotypes show a comparable pattern, with a strong
168 geographic signal (Supplemental Information 2, Supplemental Figure 6). Clustering of the ten
169 individuals using NGSadmix and sNMF, identifies the same basal split into a northern and a southern
170 cluster (Figure 2A, sNMF results included in Supplemental Information 2, Supplemental Figure 7)).
171 Based on the NGSadmix results, both Somalia and Kenya show signatures of admixture between the
172 two subspecies. This is in line with results from the abbababa function in ANGSD. In the sNMF results,
173 we observe a similar pattern for Somalia, but not for Kenya. This is further corroborated by the D-
174 statistic (using GATK genotype calls) which resulted in significant admixture only for Somalia (but not
175 for Kenya or DRC).
176
177

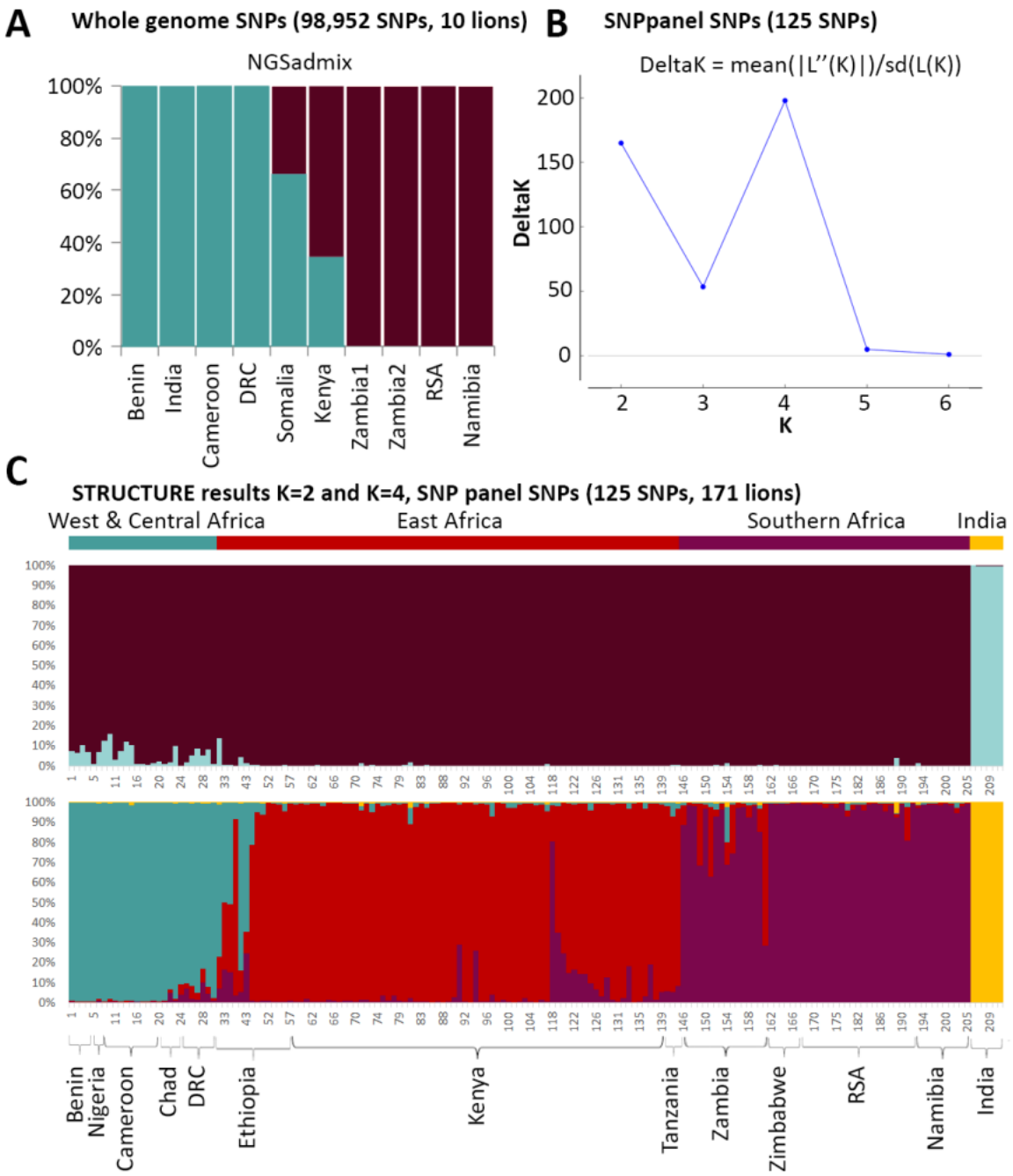


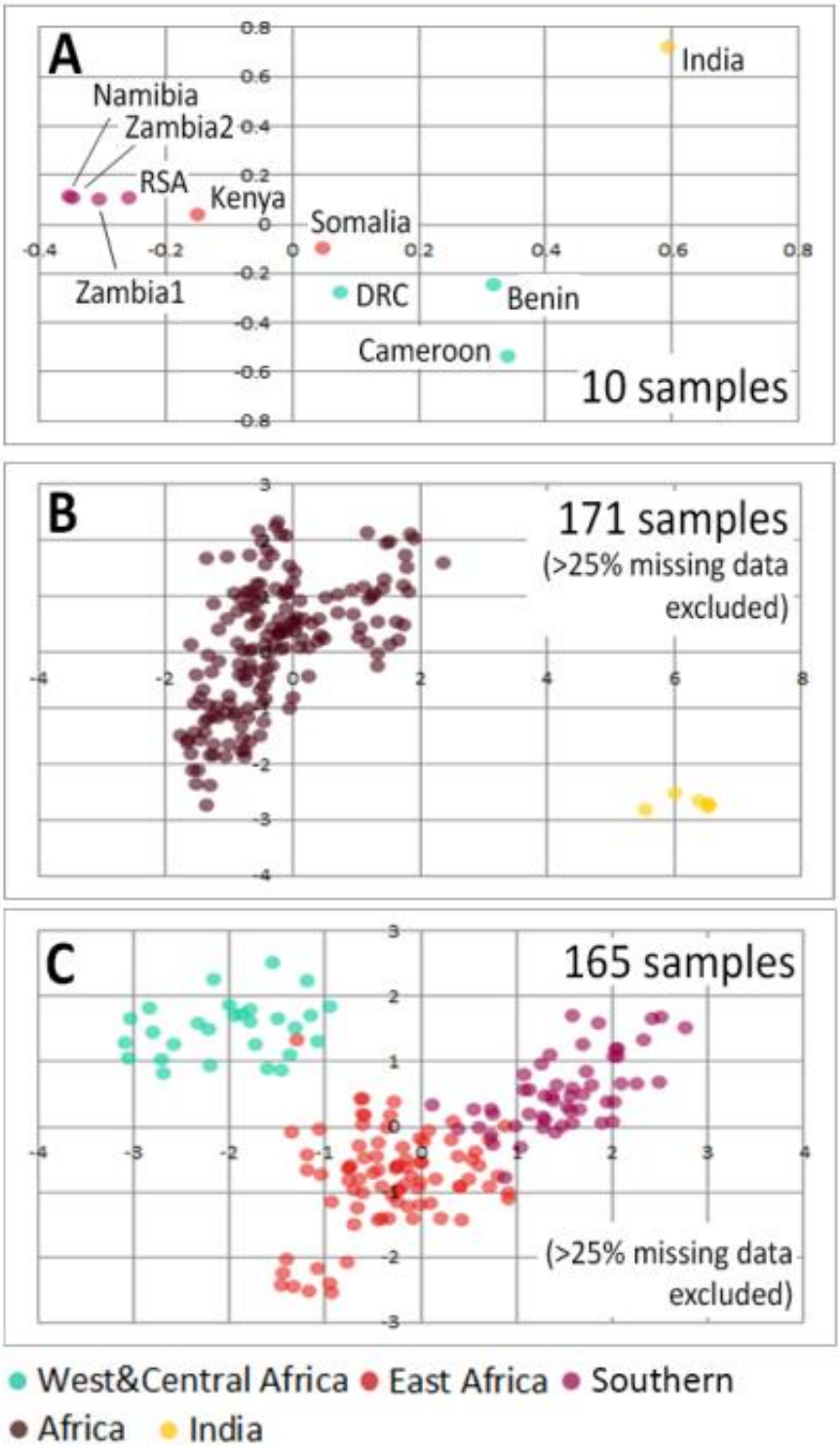
Figure 2. Assessment of population structure using NGSadmix and STRUCTURE. Results of NGSadmix (K=2) based on SNP genotype likelihoods from 10 lions (A)**.** Result of STRUCTURE Harvester for a STRUCTURE run of 125 SNPs genotyped (B). Assignment values based on the STRUCTURE run for K=2 (upper bar plot) and K=4 (lower bar plot) for 171 lions (i.e. excluding samples with >25% missing data; see Supplemental Figure 6 for results for all 211 lions) (C). Sample numbers correspond with numbers in Supplemental Table 4.

184 **SNP panel data**

185 A total of 211 samples from across the entire range of the lion were genotyped for 125 autosomal
186 SNPs and 14 mitochondrial SNPs. Missing data for the autosomal SNPs ranged from to 0 to 115 with a
187 median value of 11; for the mitochondrial data, these ranged from 0 to 10 with a median value of 0
188 (Supplemental Table 4, Supplemental Figure 8, included in Supplemental Information 2). As a
189 confirmation of our initial SNP calling, the individuals which were subjected to whole genome
190 sequencing have also been included in genotyping through the SNP panel. No contradictory
191 genotypes were found, corroborating the results from our initial SNP calling. STRUCTURE suggests an
192 optimal number of four clusters (Figure 2B) corresponding to West & Central Africa, India, East
193 Africa, and Southern Africa (Figure 2C: lower bar plot). A lower peak can be detected for K=2, driven
194 by India's very low genetic diversity resulting in the formation of a strong separate cluster (Figure 2C:
195 upper bar plot). Because missing data can affect the results of a STRUCTURE run, we present here the
196 results excluding individuals with >25% missing data (N=171). STRUCTURE plots including the entire
197 lion dataset (N=211) are available in Supplemental Figure 9 (included in Supplemental Information 2).
198 Assignment values to clusters and to mitochondrial haplogroups are reported in Supplemental Table
199 5. In the PCA, African and Indian lions are distinguished as two clouds (Figure 3B). Removing the
200 Indian population reveals more structure within African lions (Figure 3C), resulting in two nearly
201 distinct clouds representing the northern (West & Central Africa) and the southern (East and
202 Southern Africa) subspecies. PCA results based on whole genome genotype likelihoods are included
203 as Figure 3A, showing a congruent geographic differentiation compared to the SNP panel SNPs. PCA
204 results for the entire datasets (i.e. including samples with >25% missing data) are presented in
205 Supplemental Information 2, Supplemental Figure 10.

206

207



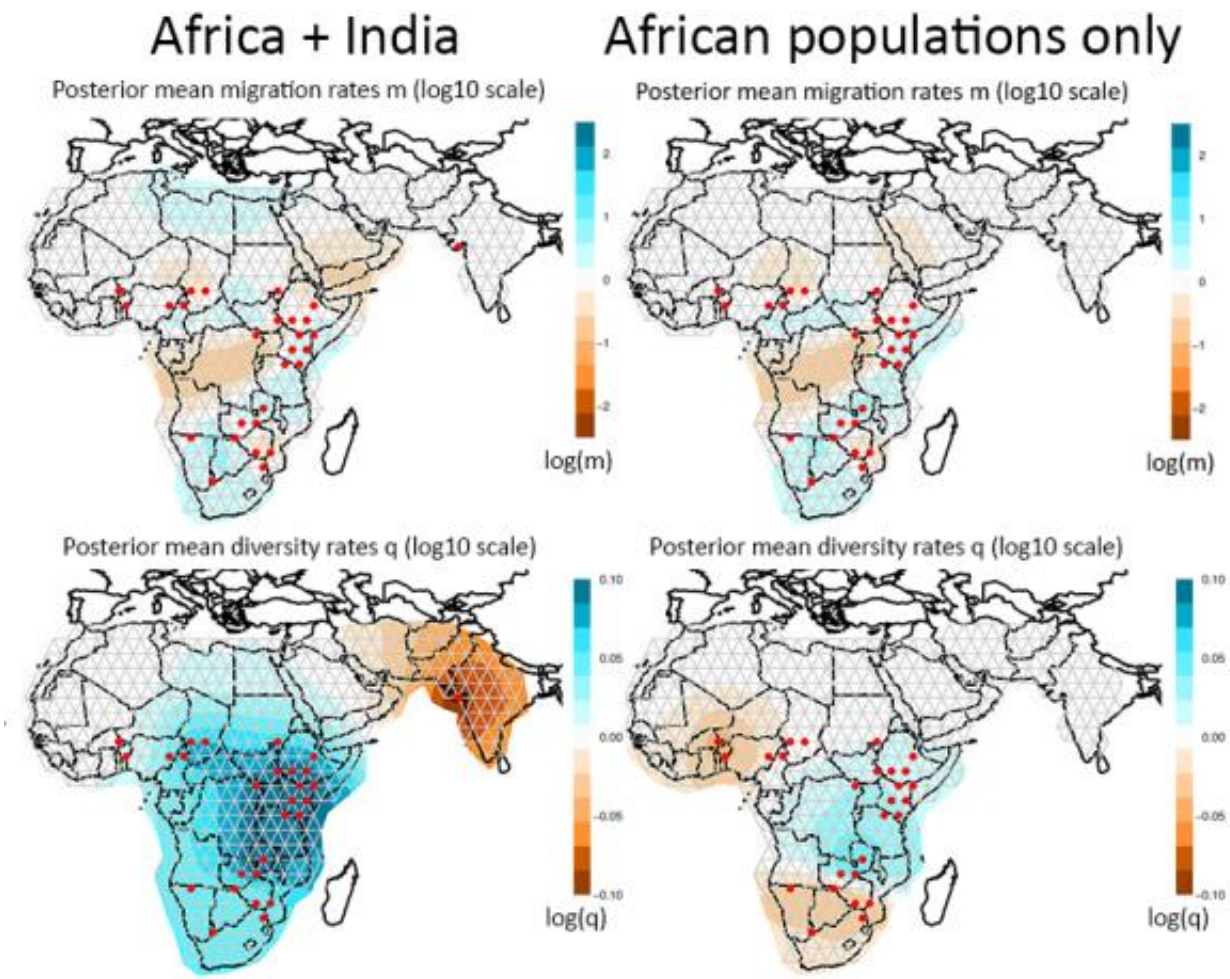
208

209 **Figure 3. PCA plots based on whole genome genotype likelihoods and SNP panel SNPs.** Plots are based on genotype

210 likelihoods of 10 lions (A), 125 SNP panel SNPs in 171 lion samples (B), and after exclusion of India (C). Colours

211 indicate region of origin.

212 EEMS infers the existence of corridors and barriers for dispersal and gene flow from the spatial decay
213 of genetic similarity. Results show that the Central African rainforest is highlighted as a barrier
214 (indicated with orange shading, Figure 4: upper row), whether or not the Indian population is
215 included in the analysis. Barriers are further identified between East and Southern Africa, and across
216 the Arabian peninsula. Diversity indices, which reflect the expected genetic dissimilarities of
217 individuals sampled from the same deme, illustrate that the Indian population has a much lower
218 genetic diversity than the African populations (Figure 4: bottom left). After repeating the analysis
219 with only African populations, low genetic diversity is detected in West Africa and Southern Africa
220 (Figure 4: bottom right).
221



222

Figure 4. EEMS results based on 211 lions, including India (left), and 205 lions, excluding India (right). The upper row shows posterior mean migration rates, the bottom row shows posterior mean diversity rates. Orange colours indicate low values, blue colours indicate high values.

Heterozygosity

Comparisons of levels of heterozygosity across datasets and to previously published data (18,20) (Supplemental Table 6) show that full genome observed heterozygosity is highly positively correlated with population-level observed heterozygosity derived from SNP panel SNPs and microsatellites (Supplemental Table 6: scatter plot and bar plot). The Indian population consistently shows a low observed heterozygosity for both SNPs and microsatellites (Supplemental Table 6: bar plot). However, for the estimates based on whole genome data, levels of heterozygosity are likely to be underestimated in the samples with low coverage (notably Benin).

Discussion

Whole genome data, complete mitogenomes and phylogeographic inference

Although SNPs can be powerful markers, ascertainment bias as a result of the study's design is a concern (31). To reflect the full diversity of the species, it is necessary to base the SNP discovery, for SNPs which can subsequently used in a SNP panel, on samples representing all lineages. Different types of markers used in lion phylogeographic studies show largely congruent results with some local discrepancies (e.g. widespread East/Southern Africa haplogroup not recovered from nuDNA data (20) and admixture in the Kruger area (21,32)). Together, they provide useful criteria for selecting populations to be subjected to whole genome sequencing, as was done in this study

247 Based on the whole genome data, we explored both genotype likelihoods, derived from ANGSD and
248 genotype calls derived from GATK, as well as different levels of missing data, balancing the number
249 of SNPs and the number of samples with an accepted call at a given position. As a higher number of
250 SNPs represents a denser sampling of the coalescent (e.g. see (33)), presented trees are based on
251 SNPs which were present in at least 50% of the samples. Increasing the number of SNPs (and
252 therefore also the amount of missing data), did not change the topology or the support of the
253 phylogenetic trees. Simulation studies and studies using empirical data have shown that
254 concatenated SNP data are able to produce reliable trees reflecting the true topology, as long as
255 enough genes are sampled (34,35). The underlying assumption is that there is enough phylogenetic
256 signal in the data, and that discordant coalescent histories will effectively cancel each other out
257 when all histories are considered together. The conclusion that our concatenated phylogenetic trees
258 (MrBayes and Garli) produce reliable topologies is supported by the fact that observed patterns are
259 congruent with results of SVDquartets, which does assume multi-locus unlinked data. An exception is
260 the more fine-scale topology for Southern Africa where patterns of differentiation are likely to be
261 affected by continuous gene flow and incomplete lineage sorting. Further, described topologies are
262 in line with previously described mitochondrial trees (14,21) with similar discordances as found in
263 microsatellite datasets (20). Notably, the wide-spread haplogroup labeled as East/Southern Africa
264 (21), stretching from Kenya to Namibia, is not recovered from microsatellite (20) or SNP data (this
265 study). The previously mentioned dichotomy between the northern subspecies, *P. leo leo*, and the
266 southern subspecies *P. leo melanochaita*, is also supported by the assignment values of NGSadmix
267 and sNMF, with possible admixture where both subspecies overlap.
268
269 **SNP panel data**
270 Based on the ten whole genomes, we generated a SNP panel of 125 autosomal and 14 mitochondrial
271 SNPs which allows cost-effective genotyping of larger numbers of samples.

272 STRUCTURE results from the SNP panel data from lions across their natural range identify four
273 clusters: West & Central Africa, India (the two clades of the northern subspecies, *P. leo leo*), East
274 Africa, Southern Africa (the two clades of the southern subspecies, *P. leo melanochaita*) (Figure 2C).
275 Main regions of admixture are Ethiopia and Zambia, as is visible in the STRUCTURE plots (Figure 2C),
276 and was also found using microsatellite data (18,20) and mtDNA (36). It is well known that
277 STRUCTURE is sensitive to groups of closely related individuals, such as siblings, family groups, or in
278 our case, an inbred population (37,38) and that identification of ancestral populations may be an
279 over-interpretation of the data, depending on demographic histories (39). Here, the West and
280 Central African populations show an ancestral relationship to the Indian cluster, while also
281 harbouring higher genetic diversity, leading to similarity to the southern subspecies cluster. The PCA
282 results further illustrate that variation between African populations is masked by the extremely low
283 diversity in the Indian population, likely the result of isolation and small population size due to
284 multiple bottlenecks and subsequent genetic drift (also see results on heterozygosity, Supplemental
285 Table 6). The variation between African populations only becomes apparent when exploring a higher
286 number of clusuters (i.e., K=4; Figure 2C, bottom plot) or excluding the Indian populations (Figure
287 3C). After excluding the Indian population in the PCA, three overlapping clouds are apparent
288 corresponding to West & Central Africa, East Africa and Southern Africa (Figure 3C). The individual
289 which has a color code from East Africa (red) but falls in the West and Central Africa cluster (green)
290 originates from Ethiopia, where both subspecies are known to overlap and admixture has been
291 previously described (20,21). The distinction between East and Southern Africa seems to be more
292 gradual in PCA space, which is in line with the widespread haplogroup that occurs throughout almost
293 the entire region.
294
295 **Evolutionary history and connectivity**

296 In order to put the phylogenetic patterns of the lion into an evolutionary perspective, it is worthwhile
297 to explore current and historical barriers to lion dispersal. Paleoclimatic data show that cyclical
298 contraction and expansion of vegetation zones, e.g. rain forest and desert, may have acted as
299 temporal barriers to lion dispersal (21). The discrete genetic lineages recognizable in the mtDNA are
300 likely to be the result of the restriction of suitable lion habitat to a number of refugia (21). The
301 pattern found in mtDNA data of the lion is congruent with that of other African savanna species
302 (21,30,40) and predicted refugial areas based on climate models (41). Faster coalescence times of
303 mtDNA may have led to reciprocally monophyletic mtDNA clades in the lion, while isolation in refugia
304 may not have lasted long enough for coalescence in autosomal markers (20,21). In addition, dispersal
305 in lions is male-biased (28,29), which may explain the more discrete phylogeographic pattern found
306 in mtDNA data. This is reflected by the fact that we do not retrieve a North East Africa cluster or a
307 South West Africa cluster based on autosomal SNPs, even though they are represented by diverged
308 mitochondrial haplotypes. Interestingly, the discrepancy in population structure between autosomal
309 and mitochondrial markers in East and Southern Africa, where a wide-spread mitochondrial
310 haplogroup occurs from Kenya to Namibia and autosomal markers suggest a phylogeographic break
311 around Zambia and Mozambique, is identical to the discrepancy found in giraffe (42). Current
312 barriers for gene flow seem to be mainly represented by the recent population disjunction in North
313 Africa/Middle East and the longstanding barrier representing the Central African rain forest, as is also
314 inferred by EEMS (Figure 4). Although the Rift valley, stretching from Ethiopia in the north to Malawi
315 and Mozambique in the south, has been mentioned as a potential barrier for gene flow in the lion
316 (Dubach *et al.* 2005; Barnett *et al.* 2006a; Barnett *et al.* 2006b; Bertola *et al.* 2011) and gene flow
317 may be reduced in that region, co-occurrence of strongly diverged haplotypes and admixture
318 detected in microsatellite and SNP data in Ethiopia indicates that the Rift valley does not represent
319 an impenetrable barrier for lion dispersal (20,21).
320

321 **Applications for wildlife research and conservation**

322 The design of a SNP panel and a reference dataset of >200 lions from 14 lion range states has
323 important applications for wildlife research and conservation. First, it allows us to distinguish
324 phylogeographic breaks and overlap between lineages which can be used to study the evolutionary
325 history of the species in more detail, e.g. at the national level. This may have important conservation
326 implications. For example, genomic data can be used to prioritize lion management at the population
327 level, including providing key information for decisions regarding translocations of individuals within
328 and among lion range states. Based on CITES documentation, >1000 live lions have been imported
329 into lion range states since the 1980s with (future) potential to interbreed with resident lions
330 (Bertola *et al.*, in review). Human-mediated movement of wildlife, regarding lions and other species,
331 only rarely take genetic considerations into account. Providing a baseline of the distribution of
332 genetic variation and a tool to screen potential candidates for translocations, can help support these
333 initiatives with scientific data. To assess the suitability of this SNP panel on a smaller geographic
334 scale, we are currently genotyping >200 lion samples from across Kenya. The aim is to provide a
335 robust tool for wildlife managers seeking genetic support for management decisions. Secondly, it
336 enables tracing of lion samples of unknown origin, such as material confiscated from illegal trade
337 chains or through anti-poaching efforts. Illegal trade in wildlife products is currently estimated to be
338 the fifth largest illegal industry globally, and a major concern for conservationists (44–46). Part of the
339 trade is thought to cater to domestic markets, but growing evidence suggests an increase in illegal
340 shipments to international markets (47). Genetic toolkits can contribute to combatting wildlife
341 crimes by identifying confiscated material, source populations and tracking trade routes (32,48–51).
342 For lions, a tool that assists law enforcement intelligence has already been designed using mtDNA
343 haplotypes (lionlocalizer.org), and it offers an ideal opportunity to include autosomal SNPs in future
344 updates of this tool. Thirdly, a SNP panel can be used to guide breeding efforts for ex situ
345 conservation (32,52). Recently, our SNP panel was used to genotype a batch of lion samples from

346 institutions linked to the European Association for Zoos and Aquaria (EAZA). Based on the resulting
347 information, managers are currently deciding how many and which lineages to include in future
348 breeding efforts. In particular, given the dire situation for lions in West and Central Africa, and a
349 severe under-representation of this lineage in captivity, the use of SNP data can be helpful to
350 efficiently identify and conserve diversity *ex situ* and prevent inbreeding.
351
352 **Future perspectives**
353 With the increase of genomic data alongside computational and technological developments, the
354 field of conservation genomics is undergoing a major transition. Demographic histories and patterns
355 of gene flow can be inferred in greater detail (53,54). Searching genomes for signatures of adaptation
356 and deleterious alleles has been highlighted as a powerful tool for gauging the sensitivity of
357 populations in a changing environment (55–59). New developments allowing SNP genotyping from
358 poor quality (e.g., non-invasively collected) samples will further contribute to the applicability of SNP
359 genotyping in the field (60). And, finally, the development of mobile, hand-held sequencing devices
360 shows great promises for rapid, real-time identification of samples (61). Such tools represent
361 tremendous potential for training and capacity building (62) especially for biodiverse countries with
362 limited facilities to process samples, thereby making the field of wildlife research and conservation
363 more democratized and inclusive.
364
365 **Conclusions**
366 This study is the first to report phylogeographic relationships between lion populations throughout
367 their entire range based on ~150,000 SNPs derived from whole genome sequencing. We present a
368 SNP panel, containing 125 autosomal and 14 mitochondrial SNPs which has been validated on >200
369 individuals from 14 lion range states, spanning most of the lion's range. The results reveal a detailed
370 population structure and confirm a basal distinction between a northern and a southern subspecies

371 that supports the recent revision of the lion's taxonomy. We highlight several applications for the
372 genomic resources we present here, and the SNP panel in particular. The samples on which the panel
373 was tested can serve as a reference database for future research and conservation efforts.
374
375 **Methods**
376 **Sampling**
377 Blood or tissue samples of ten lions, representing the main phylogeographic groups as identified in
378 previous studies (14,16,19–21) (Figure 1: map, Supplemental Table 1), were collected and preserved
379 in a buffer solution (0.15 M NaCl, 0.05 M Tris-HCl, 0.001 M EDTA, pH = 7.5) at -20°C. All individuals
380 included were either free-ranging lions or captive lions with proper documentation of their breeding
381 history that included no known occurrences of hybridization between aforementioned lineages. A
382 sample from an Amur leopard (*P. pardus orientalis*, captive) was included as an outgroup. The Amur
383 tiger (*P. tigris altaica*) genome (63) was used as a reference for mapping of the lion and leopard
384 reads. All samples were collected in full compliance with specific legally required permits (CITES and
385 permits related to national legislation in the countries of origin). Details on laboratory protocols,
386 sequencing, mapping to the reference genome, SNP calling, and quality control are given in
387 Supplemental Information 1 (including Supplemental Figures 1-3).
388
389 **SNP calling and complete mitogenomes**
390 After mapping of the reads to the reference genome (tiger reference genome from (63) and lion
391 mitogenome from (21)), autosomal SNPs were called using two methods: 1) we used ANGSD (64) to
392 derive genotype likelihoods, and 2) we called and filtered SNPs using GATK (65) and VCFtools (66)
393 (details in Supplemental Information 1). Because these results stem from low coverage genomes,
394 and acknowledging the potential confounding effect of missing data, we explored different levels of
395 missing data applying the –minInd flag in ANGSD. For the SNPs derived from GATK we replaced

396 positions with coverage <3 by an ambiguous nucleotide, and we applied 5 levels of filtering as
397 follows: 1) all SNPs, 2) SNPs called in at least three samples, 3) SNPs called in at least five samples, 4)
398 SNPs called in at least eight samples, and 5) SNPs called in all samples (Supplemental Table 2).
399 Identified SNPs were attributed to a chromosome following the genomic architecture in the tiger (63)
400 (Supplemental Table 2). Full mitogenomes were recovered by mapping all reads to a mitogenome
401 reference generated with long range PCR and a known numt sequence, allowing us to differentiate
402 between reads of mitochondrial and nuclear origin. This procedure is described in more detail in
403 Supplemental Information 1 and (21).
404 Phylogenetic analyses were performed on the full mitogenomes and on the concatenated SNP
405 datasets with varying levels of missing data with MrBayes v.3.1.2 (67,68) and Garli (69) using
406 parameters as determined by MrModeltest2 (v.2.3) (70). MrBayes and Garli were run for one million
407 generations and five million generations respectively, using a GTR substitution model with rate
408 variation across sites set to equal. In addition, we ran SVDquartets (71), assuming multi-locus
409 unlinked single-site data with 100 bootstrap replicates as implemented in PAUP* 4.0a164 (72). For
410 the mitogenome data, the coalescent process in the model was disregarded. Nodes receiving >95%
411 Posterior Probability (PP) in Bayesian analysis (MrBayes) and 0.7 bootstrap support in Maximum
412 Likelihood (ML) analysis (Garli) and SVDquartets are considered to have significant support, as is
413 common practice. A dendrogram was created on the genotype likelihoods derived from ANGSD,
414 using the hclust algorithm. Genotype likelihoods (ANGSD) and genotype calls (GATK) were further
415 subjected to Principal Component Analysis (PCA), using PCAngsd (73) and the PCA function from the
416 Adegenet R package (74), respectively. Individual ancestry coefficients were estimated using
417 NGSadmix (75) on the genotype likelihoods. For GATK genotypes, we used sparse non-negative
418 matrix factorization algorithms in sNMF, as implemented in the R package LEA (76,77), exploring K=1-
419 10, using 20 replicates and 4 values for the alpha regularization parameter (1, 10, 100 and 1,000). To
420 explore putative admixture, we used the abbababa function in ANGSD and we calculated D-statistics

421 (ABBA-BABA test) on the GATK genotypes using CalcD from the evobiR package (78) with 1000
422 bootstraps. We used Cameroon and Zambia1 as the ingroup as they are clearly assigned to *P. leo leo*
423 and *P. leo melanochaita* respectively, and tiger as the outgroup.
424
425 **SNP panel data**
426 In order to obtain better insight into the geographic locations of phylogeographic breaks, we made a
427 selection of SNPs for inclusion in a SNP panel for genotyping more samples. We used the following
428 criteria for the selection process: 1) minimum coverage of 20 for all lion samples combined, across 50
429 bp upstream and downstream of the SNP position, 2) maximum of one variable position in these 50
430 bp flanking regions, 3) high quality mapping of the flanking regions as identified by eye using IGV
431 Genome Browser (79,80), 4) at least eight individuals represented at each selected site, 5) SNPs
432 evenly spread across all chromosomes, as implied by the chromosomal architecture in tiger (max.
433 one SNP per scaffold), 6) preferably both homozygotes and heterozygote genotype present among
434 the ten genotyped lions (but sites for which each individual was scored as a heterozygote were
435 excluded, as these may represent gene duplications). In addition, we included a total of 14
436 mitochondrial SNPs selected to represent each major branch within the mitochondrial phylogenetic
437 tree (as described in (21)), as well as two diagnostic SNPs that represent the distinction between the
438 northern and the southern subspecies. These mtDNA SNPs had already been assessed in a wide
439 range of populations (21), making it more likely that the selected SNPs are diagnostic throughout the
440 lion range. Finally, we ensured that the mitochondrial SNPs were not located in any of the nuclear
441 copies (numts) which are known to exist in cats (81–85) by aligning our mitogenome sequences with
442 known numt sequences. Coordinates of each of the SNPs and its associated chromosome, both for
443 the tiger genome (63) and for a nearly-chromosome level lion genome (86) (lion genome was not
444 available when the original mapping was done) are reported in Supplemental Table 3 and visualized
445 in Supplemental Figure 4 (included in Supplemental Information 1). After test runs and quality

control (see Supplemental Information 1), we retained 125 autosomal SNPs and 14 mtDNA SNPs which were used for genotyping 211 lions of known origin from 14 lion range states representing the entire geographic range of the species (Figure 1: map). There is a special focus on the region where the ranges of the northern and southern subspecies overlap i.e., Ethiopia and Kenya. This region is known to harbor haplotypes from strongly diverged lineages (21), and microsatellite data have suggested admixture (20). SNP genotyping was done at the SNP genotyping facility at Leiden University using allele-specific primers and KASP technology (LGC Genomics).

Resulting autosomal genotypes were analysed with STRUCTURE (87) using correlated allele frequencies and running the program for 5 million generations, discarding the first 500,000 generations as burn-in. Five replicates were run for K=1 to K=7. The optimal number of K was assessed by using the DeltaK method as implemented in STRUCTURE Harvester (88,89); CLUMPP (90) was used before generating the bar plots of population assignment. To reduce the effect of missing data on the assignment results, we ran STRUCTURE using the complete dataset (N=211) and a dataset excluding samples with >25% missing data (N=171). MtDNA SNPs were used to assign a specific haplotype to each individual, matching with previously described lineages (see delineation of haplogroups in (21)): West Africa, Central Africa, North East Africa, East/Southern Africa, South West Africa, and India.

A PCA was performed in Genalex (91), using the entire dataset (N=211) and the reduced dataset excluding samples with >25% missing data (N=177). Because the PCA is disproportionately affected by the very low genetic diversity of Indian lions, we repeatedthe analysis excluding this population, thereby allowing a more detailed picture from the African populations to emerge.

470 Patterns of connective zones and barriers were investigated using Estimated Effective Migration
471 Surfaces (EEMS) (92) for all individuals genotyped with the SNP panel. Three independent runs were
472 performed, using 10 million generations and discarding the first 5 million generations as burn-in. We
473 followed the authors' suggestions for tuning the proposal variances, using 0.1 and 1 for
474 mSeedsProposalS2 and mEffctProposalS2 respectively, and 1.5 and 0.015 for qSeedsProposalS2 and
475 qEffctProposalS2 respectively.
476
477 **Heterozygosity**
478 The level of observed heterozygosity was assessed for each individual for which the whole genome
479 had been sequenced, expressed as the proportion of heterozygote positions compared to the total
480 number of SNPs, excluding positions with ambiguous nucleotides (i.e. positions which were not
481 scored due to insufficient quality or coverage). This was done on the individual level for all SNPs, and
482 on the populations level for all SNP panel SNPs which were successfully called in five or more
483 individuals from the same population. Results were then compared to known levels of heterozygosity
484 based on earlier studies using microsatellite data from the same populations (18,20).
485
486 **List of abbreviations**
487 CITES – Convention on International Trade in Endangered Species of Wild Fauna and Flora
488 DRC – Demoratic Republic of the Congo
489 EAZA – European Association of Zoos and Aquaria
490 FIV_{Ple} – Feline Immunodeficiency Virus
491 IUCN – International Union for Conservation of Nature
492 ML – Maximum Likelihood
493 mtDNA – mitochondrial DNA
494 NGS – Next Generation Sequencing

NP – National Park

nuDNA – nuclear DNA

PCA – Principal Component Analysis

RSA – Republic of South Africa

SNP – Single Nucleotide Polymorphism

Declarations

Ethics approval and consent to participate

Not applicable. All blood and tissue samples had been collected in the context of previous studies.

Consent for publication

Not applicable

Availability of data and materials

The datasets generated and analysed during the current study are available through NCBI under Bioproject ID … (accession numbers …-…) (fastq format), and included within the article and its supplementary information files (SNP panel genotypes).

Competing interests

The authors declare that they have no competing interests.

Funding

The investigations were supported by the Division for Earth and Life Sciences (ALW) with financial aid from the Netherlands Organization for Scientific Research (NWO) (project no. 820.01.002).

Authors' contributions

L.D.B. performed analyses and wrote the manuscript, M.V. performed and J.F.J.L. supervised bioinformatics analyses, O.D.S. performed SNP genotyping, F.L., M.C., P.N.T., E.A.S., H.B., B.D.P. and P.A.W. supplied material and assisted with useful advice, H.H.d.I. and K.V. supervised analyses. All authors contributed to writing the manuscript.

Acknowledgements

Samples were kindly provided by P. Henschel, Yohanna Saidu and the Nigerian National Park Service (Nigeria), S. Adam, R. Buij and B. Croes (Cameroon), N. Vanherle (CURESS, Chad), ICCN, Garamba NP (DRC), NABU (Ethiopia, Kafa BR), T. Jirmo (Kenya), South Africa National Parks (SANParks) (RSA), S. Miller, R. Groom, Save Valley Conservancy and DeBeers Venetia-Limpopo Nature Reserve/The Diamond Route (RSA and Zimbabwe), O. Aschenborn (Namibia), C.A.Driscoll (India), Safaripark Beekse Bergen (Hilvarenbeek, The Netherlands) (Somalia), Ouwehands Dierenpark (Rhenen, The Netherlands) (RSA), and Planckendael (Muizen, Belgium) (Leopard). We further thank N. Schidlo, R. Hennevelt, H. Buermans, Y. Ariyurek, and S. Greve-Onderwater for assisting in processing of the samples and BioIT for bioinformatics support.

References

1. Ellegren H. Genome sequencing and population genomics in non-model organisms. Trends Ecol Evol [Internet]. 2014 Jan [cited 2014 Jul 10];29(1):51–63. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24139972

2. Ekblom R, Galindo J. Applications of next generation sequencing in molecular ecology of non-model organisms. Heredity (Edinb) [Internet]. 2011 Jul [cited 2014 Jul 10];107(1):1–15. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3186121&tool=pmcentrez&rendertype=abstract

3. Shafer ABA, Wolf JBW, Alves PC, Bergström L, Bruford MW, Brännström I, et al. Genomics and the challenging translation into conservation practice. Trends Ecol Evol [Internet]. 2015 Dec [cited 2014 Dec 19];30(2):78–87. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0169534714002511

544 4. Stephens ZD, Lee SY, Faghri F, Campbell RH, Zhai C, Efron MJ, et al. Big Data: Astronomical or
545 Genomical? PLOS Biol [Internet]. 2015 Jul 7 [cited 2015 Jul 8];13(7):e1002195. Available from:
546 http://dx.plos.org/10.1371/journal.pbio.1002195

547 5. Hendricks S, Anderson EC, Antao T, Bernatchez L, Forester BR, Garner B, et al. Recent
548 advances in conservation and population genomics data analysis. Evol Appl. 2018;11(8):1197–
549 211.

550 6. Supple MA, Shapiro B. Conservation of biodiversity in the genomics era. Genome Biol.
551 2018;19(1):1–12.

552 7. Craigie ID, Baillie JEM, Balmford A, Carbone C, Collen B, Green RE, et al. Large mammal
553 population declines in Africa's protected areas. Biol Conserv [Internet]. 2010 Sep [cited 2011
554 Jul 25];143(9):2221–8. Available from:
555 http://linkinghub.elsevier.com/retrieve/pii/S0006320710002739

556 8. Bauer H, Chapron G, Nowell K, Henschel P, Funston P, Hunter LT, et al. Lion (Panthera leo)
557 populations are declining rapidly across Africa, except in intesively managed areas. Proc Natl
558 Acad Sci. 2015;1–6.

559 9. Riggio J, Jacobson A, Dollar L, Bauer H, Becker M, Dickman A, et al. The size of savannah
560 Africa: a lion's (Panthera leo) view. Biodivers Conserv [Internet]. 2012 Dec 2 [cited 2012 Dec
561 3];22(1):17–35. Available from: http://www.springerlink.com/index/10.1007/s10531-012-
562 0381-4

563 10. Henschel P, Azani D, Burton C, Malanda GUY, Saidu Y, Sam M, et al. Lion status updates from
564 five range countries in West and Central Africa. CATnews. 2010;52:34–9.

565 11. Brugière D, Chardonnet B, Scholte P. Large-scale extinction of large carnivores (lion Panthera
566 leo , cheetah Acinonyx jubatus and wild dog Lycaon pictus) in protected areas of West and
567 Central Africa. Trop Conserv Sci. 2015;8(2):513–27.

568 12. Bauer H, Packer C, Funston PF, Henschel P, Nowell K. The IUCN Red List of Threatened Species
569 2015: Panthera leo. 2015.

570 13. Henschel P, Coad L, Burton C, Chataigner B, Dunn A, MacDonald D, et al. The Lion in West
571 Africa Is Critically Endangered. Hayward M, editor. PLoS One [Internet]. 2014 Jan 8 [cited 2014
572 Jan 9];9(1):e83500. Available from: http://dx.plos.org/10.1371/journal.pone.0083500

573 14. Bertola LD, van Hooft WF, Vrieling K, Uit de Weerd DR, York DS, Bauer H, et al. Genetic
574 diversity, evolutionary history and implications for conservation of the lion (Panthera leo) in
575 West and Central Africa. J Biogeogr [Internet]. 2011 Jul 31 [cited 2011 Sep 22];38(7):1356–67.
576 Available from: http://doi.wiley.com/10.1111/j.1365-2699.2011.02500.x

577 15. Barnett R, Yamaguchi N, Barnes I, Cooper A. The origin, current diversity and future
578 conservation of the modern lion (Panthera leo). Proc R Soc B Biol Sci [Internet]. 2006 Sep
579 7;273(1598):2119–25. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16901830

580 16. Barnett R, Yamaguchi N, Shapiro B, Ho SY, Barnes I, Sabin R, et al. Revealing the maternal
581 demographic history of Panthera leo using ancient DNA and a spatially explicit genealogical
582 analysis. BMC Evol Biol [Internet]. 2014;14(1):70. Available from:
583 http://www.biomedcentral.com/1471-2148/14/70

584 17. Dubach J, Patterson BD, Briggs MB, Venzke K, Flamand J, Stander P, et al. Molecular genetic
585 variation across the southern and eastern geographic ranges of the African lion, Panthera leo.
586 Conserv Genet [Internet]. 2005 Jan;6(1):15–24. Available from:
587 http://www.springerlink.com/index/10.1007/s10592-004-7729-6

588 18. Dubach JM, Briggs MB, White PA, Ament BA, Patterson BD. Genetic perspectives on “Lion
589 Conservation Units” in Eastern and Southern Africa. Conserv Genet [Internet]. 2013 Feb 3
590 [cited 2013 Feb 6];1942(Guggisberg 1961). Available from:
591 http://link.springer.com/10.1007/s10592-013-0453-3

592 19. Antunes A, Troyer JL, Roelke ME, Pecon-Slattery J, Packer C, Winterbach C, et al. The
593 evolutionary dynamics of the lion Panthera leo revealed by host and viral population
594 genomics. PLoS Genet [Internet]. 2008 Nov;4(11):e1000251. Available from:
595 http://www.ncbi.nlm.nih.gov/pubmed/18989457

596 20. Bertola LD, Tensen L, van Hooft P, White P a, Driscoll C a, Henschel P, et al. Autosomal and
597 mtDNA Markers Affirm the Distinctiveness of Lions in West and Central Africa. PLoS One
598 [Internet]. 2015 Jan [cited 2015 Oct 19];10(10):e0137975. Available from:
599 http://www.ncbi.nlm.nih.gov/pubmed/26466139

600 21. Bertola LD, Jongbloed H, Van Der Gaag KJ, De Knijff P, Yamaguchi N, Hooghiemstra H, et al.
601 Phylogeographic Patterns in Africa and High Resolution Delineation of Genetic Clades in the
602 Lion (Panthera leo). Sci Rep [Internet]. 2016;6(August):1–11. Available from:
603 http://dx.doi.org/10.1038/srep30807

604 22. Smitz N, Jouvenet O, Ligate FA, Crosmary W-G, Ikanda D, Chardonnet P, et al. A genome-wide
605 data assessment of the African lion (Panthera leo) population genetic structure in Tanzania.
606 PlosOne. 2018;1–24.

607 23. Curry CJ, Davis BW, Bertola LD, White PA, Murphy WJ. Spatiotemporal Genetic Diversity of
608 Lions Reveals the Influence of Habitat Fragmentation Across Africa. Mol Biol Evol. 2020;1–27.

609 24. Manuel M De, Barnett R, Sandoval-Velasco M, Yamaguchi N, Vieira FG, Mendoza LZ, et al. The
610 evolutionary history of extinct and living lions. Proc Natl Acad Sci. 2020;

611 25. Kitchener A, Breitenmoser-Würsten C, Eizirik E, Gentry A, Werdelin L, Wilting A, et al. A
612 revised taxonomy of the Felidae. Cat News. 2017;80.

613 26. Zink RM, Barrowclough GF. Mitochondrial DNA under siege in avian phylogeography. Mol Ecol
614 [Internet]. 2008 May [cited 2010 Jul 14];17(9):2107–21. Available from:
615 http://www.ncbi.nlm.nih.gov/pubmed/18397219

616 27. Edwards S, Bensch S. Looking forwards or looking backwards in avian phylogeography? A
617 comment on Zink and Barrowclough 2008. Mol Ecol [Internet]. 2009 Jul;18(14):2930–3;
618 discussion 2934-6. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19552688

619 28. Pusey AE, Packer C, Erhoff-Mulder MB. The Evolution of Sex-biased Dispersal in Lions.
620 Behaviour. 1987;101(4):275–310.

621 29. Spong GF, Creel S. Deriving dispersal distances from genetic data. Proc Biol Sci [Internet]. 2001
622 Dec 22 [cited 2011 Feb 3];268(1485):2571–4. Available from:
623 http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1088917&tool=pmcentrez&rend
624 ertype=abstract

625 30. Lorenzen ED, Heller R, Siegismund HR. Comparative phylogeography of African savannah
626 ungulates. Mol Ecol [Internet]. 2012 Aug [cited 2014 Feb 20];21(15):3656–70. Available from:
627 http://www.ncbi.nlm.nih.gov/pubmed/22702960

628 31. Lachance J, Tishkoff SA. SNP ascertainment bias in population genetic analyses: Why it is
629 important, and how to correct it. Bioessays. 2014;35(9):780–6.

630 32. Bertola LD. Genetic Diversity in the Lion (Panthera leo (Linnaeus 1758)): Unravelling the Past
631 and Prospects for the Future. Leiden University; 2015.

632 33. Tripp EA, Tsai YHE, Zhuang Y, Dexter KG. RADseq dataset with 90% missing data fully resolves
633 recent radiation of Petalidium (Acanthaceae) in the ultra-arid deserts of Namibia. Ecol Evol.
634 2017;7(19):7920–36.

635 34. Tonini J, Moore A, Stern D, Shcheglovitova M, Ort?? G. Concatenation and species tree
636 methods exhibit statistically indistinguishable accuracy under a range of simulated conditions.
637 PLoS Curr. 2015;7(TREE OF LIFE).

638 35. Xi Z, Liu L, Davis CC. The impact of missing data on species tree estimation. Mol Biol Evol.
639 2016;33(3):838–60.

640 36. Curry CJ, White PA, Derr JN. Mitochondrial Haplotype Diversity in Zambian Lions: Bridging a
641 Gap in the Biogeography of an Iconic Species. PlosOne. 2015;1–14.

642 37. Waples RS, Anderson EC. Purging putative siblings from population genetic data sets: A
643 cautionary view. Mol Ecol. 2017;1211–24.

644 38. Rodríguez-Ramilo ST, Wang J. The effect of close relatives on unsupervised Bayesian clustering
645 algorithms in population genetic structure analysis. Mol Ecol Resour. 2012;12(5):873–84.

646 39. Lawson DJ, van Dorp L, Falush D. A tutorial on how not to over-interpret STRUCTURE and
647 ADMIXTURE bar plots. Nat Commun [Internet]. 2018;9(1):1–11. Available from:
648 http://dx.doi.org/10.1038/s41467-018-05257-7

649 40. Hewitt GM. The structure of biodiversity - insights from molecular phylogeography. Front Zool
650 [Internet]. 2004 Oct 26 [cited 2014 Feb 26];1(1):4. Available from:
651 http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=544936&tool=pmcentrez&rende
652 rtype=abstract

653 41. Levinsky I, Araújo MB, Nogués-Bravo D, Haywood AM, Valdes PJ, Rahbek C. Climate envelope
654 models suggest spatio-temporal co-occurrence of refugia of African birds and mammals. Glob
655 Ecol Biogeogr [Internet]. 2013 Mar 17 [cited 2014 Jun 7];22(3):351–63. Available from:
656 http://doi.wiley.com/10.1111/geb.12045

657 42. Fennessy J, Bidon T, Reuss F, Vamberger M, Fritz U, Janke Correspondence A, et al. Multi-locus
658 Analyses Reveal Four Giraffe Species Instead of One. Curr Biol [Internet]. 2016;26:1–7.
659 Available from: http://dx.doi.org/10.1016/j.cub.2016.07.036

660 43. Barnett R, Yamaguchi N, Barnes I, Cooper A. Lost populations and preserving genetic diversity
661 in the lion Panthera leo: Implications for its ex situ conservation. Conserv Genet [Internet].
662 2006 Mar 1;7(4):507–14. Available from:
663 http://www.springerlink.com/index/10.1007/s10592-005-9062-0

664 44. Nijman V, Morcatty T, Smith JH, Atoussi S, Shepherd CR, Siriwat P, et al. Illegal wildlife trade–
665 surveying open animal markets and online platforms to understand the poaching of wild cats.
666 Biodiversity [Internet]. 2019;20(1):58–61. Available from:
667 https://doi.org/10.1080/14888386.2019.1568915

668 45. 't Sas-Rolfes M, Challender DWS, Hinsley A, Veríssimo D, Milner-Gulland EJ. Illegal Wildlife
669 Trade: Patterns, Processes, and Governance. Annu Rev Environ Resour. 2019;44(1):1–28.

670 46. Biggs D, Cooney R, Roe D, Dublin HT, Allan JR, Challender DWS, et al. Developing a theory of
671 change for a community-based response to illegal wildlife trade. Conserv Biol. 2017;31(1):5–
672 12.

673 47. Williams VL, Loveridge AJ, Newton DJ, Macdonald DW. A roaring trade? The legal trade in
674 Panthera leo bones from Africa to East-Southeast Asia. PLoS One. 2017;12(10):1–22.

675 48. Wasser S, Poole J, Lee P, Lindsay K, Dobson A, Hart J, et al. Elephants, Ivory, and Trade.
676 Science (80-). 2010;327(March):1331–2.

677 49. Wasser SK, Brown L, Mailand C, Mondol S, Clark W, Laurie C, et al. Genetic assignment of large
678 seizures of elephant ivory reveals major poaching hotspots. Science (80-). 2015;(June):1–7.

679 50. Wasser SK, Joseph Clark W, Drori O, Stephen Kisamo E, Mailand C, Mutayoba B, et al.
680 Combating the illegal trade in African elephant ivory with DNA forensics. Conserv Biol
681 [Internet]. 2008 Aug [cited 2010 Aug 30];22(4):1065–71. Available from:
682 http://www.ncbi.nlm.nih.gov/pubmed/18786100

683 51. Mondol S, Sridhar V, Yadav P, Gubbi S, Ramakrishnan U. Tracing the Geographic Origin of
684 Traded Leopard Body Parts in the Indian Subcontinent with DNA-Based Assignment Tests.
685 Conserv Biol [Internet]. 2014 Nov 5 [cited 2014 Nov 11];00(0):1–9. Available from:
686 http://www.ncbi.nlm.nih.gov/pubmed/25376464

687 52. Henry P, Miquelle D, Sugimoto T, McCullough DR, Caccone A, Russello M a. In situ population

688 structure and ex situ representation of the endangered Amur tiger. Mol Ecol [Internet]. 2009
689 Aug [cited 2013 Oct 30];18(15):3173–84. Available from:
690 http://www.ncbi.nlm.nih.gov/pubmed/19555412

691 53. Kamm JA, Terhorst J, Durbin R, Song YS. Efficiently inferring the demographic history of many
692 populations with allele count data. bioRxiv [Internet]. 2018;287268. Available from:
693 https://www.biorxiv.org/content/early/2018/03/23/287268

694 54. Elleouet JS, Aitken SN. Exploring Approximate Bayesian Computation for inferring recent
695 demographic history with genomic markers in nonmodel species. Mol Ecol Resour.
696 2018;18(3):525–40.

697 55. Harrisson KA, Pavlova A, Telonis-Scott M, Sunnucks P. Using genomics to characterize
698 evolutionary potential for conservation of wild populations. Evol Appl [Internet]. 2014 Mar 14
699 [cited 2014 Sep 5];7(9):1008–25. Available from: http://doi.wiley.com/10.1111/eva.12149

700 56. Funk WC, Mckay JK, Hohenlohe PA, Allendorf FW. Harnessing genomics for delineating
701 conservation units. Trends Ecol Evol [Internet]. 2012 Sep [cited 2014 Jul 10];27(9):489–96.
702 Available from: http://dx.doi.org/10.1016/j.tree.2012.05.012

703 57. Funk W, Lovich R, Hohenlohe P, Hofman C, Morrison S, Sillett T, et al. Adaptive divergence
704 despite strong genetic drift: genomic analysis of the evolutionary mechanisms causing genetic
705 differentiation in the island fox (Urocyon littoralis). Mol Ecol. 2016;25(10):2176–94.

706 58. Razgour O, Forester B, Taggart JB, Bekaert M, Juste J, Ibáñez C, et al. Considering adaptive
707 genetic variation in climate change vulnerability assessment reduces species range loss
708 projections. Proc Natl Acad Sci U S A. 2019;116(21):10418–23.

709 59. Henn BM, Botigué LR, Bustamante CD, Clark AG, Gravel S. Estimating the mutation load in
710 human genomes. Nat Rev Genet. 2015;16(6):333–43.

711 60. Natesh M, Taylor RW, Truelove NK, Hadly EA, Palumbi SR, Petrov DA, et al. Empowering

712 conservation practice with efficient and economical genotyping from poor quality samples.
713 Methods Ecol Evol. 2019;2019(August 2018):1–7.

714 61. Johri S, Solanki J, Cantu VA, Fellows SR, Edwards RA, Moreno I, et al. 'Genome skimming' with
715 the MinION hand-held sequencer identifies CITES-listed shark species in India's exports
716 market. Sci Rep [Internet]. 2019;9(1):1–13. Available from: http://dx.doi.org/10.1038/s41598-
717 019-40940-9

718 62. Pomerantz A, Peñafiel N, Arteaga A, Bustamante L, Pichardo F, Coloma LA, et al. Real-time
719 DNA barcoding in a rainforest using nanopore sequencing: Opportunities for rapid biodiversity
720 assessments and local capacity building. Gigascience. 2018;7(4):1–14.

721 63. Cho YS, Hu L, Hou H, Lee H, Xu J, Kwon S, et al. The tiger genome and comparative analysis
722 with lion and snow leopard genomes. Nat Commun. 2013;4(2433).

723 64. Korneliussen TS, Albrechtsen A, Nielsen R. ANGSD: Analysis of Next Generation Sequencing
724 Data. BMC Bioinformatics. 2014;15(1):1–13.

725 65. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome
726 Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing
727 data. Genome Res. 2010;20:1297–303.

728 66. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call
729 format and VCFtools. Bioinformatics. 2011 Aug;27(15):2156–8.

730 67. Huelsenbeck JP, Ronquist F. MRBAYES: Bayesian inference of phylogenetic trees.
731 Bioinformatics [Internet]. 2001 Aug;17(8):754–5. Available from:
732 http://www.ncbi.nlm.nih.gov/pubmed/11524383

733 68. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, et al. MrBayes 3.2:
734 efficient Bayesian phylogenetic inference and model choice across a large model space. Syst
735 Biol [Internet]. 2012 May [cited 2013 Feb 28];61(3):539–42. Available from:

736 http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3329765&tool=pmcentrez&rend
737 ertype=abstract

738 69. Zwickl DJ. Genetic algorithm approaches for the phylogenetic analysis of large biological
739 sequence datasets under the maximum likelihood criterion. The University of Texas at Austin;
740 2006.

741 70. Nylander JAA. MrModeltest v2.3. Uppsala, Sweden: Evolutionary Biology Centre, Uppsala
742 University; 2004. p. Program distributed by the author.

743 71. Chifman J, Kubatko L. Quartet inference from SNP data under the coalescent model.
744 Bioinformatics. 2014;30(23):3317–24.

745 72. Swofford DL. PAUP*: Phylogenetic analysis using parsimony (*and other methods), version
746 4b10. Sinauer, Sunderland, Massachusetts, USA.; 2002.

747 73. Meisner J, Albrechtsen A. Inferring population structure and admixture proportions in low-
748 depth NGS data. Genetics. 2018;210(2):719–31.

749 74. Jombart T. Adegenet: A R package for the multivariate analysis of genetic markers.
750 Bioinformatics. 2008;24(11):1403–5.

751 75. Skotte L, Korneliussen TS, Albrechtsen A. Estimating individual admixture proportions from
752 next generation sequencing data. Genetics. 2013;195(3):693–702.

753 76. Frichot E, François O. LEA: An R package for landscape and ecological association studies.
754 Methods Ecol Evol. 2015;6(8):925–9.

755 77. Frichot E, Mathieu F, Trouillon T, Bouchard G, François O. Fast and efficient estimation of
756 individual ancestry coefficients. Genetics. 2014;196(4):973–83.

757 78. Durand EY, Patterson N, Reich D, Slatkin M. Testing for ancient admixture between closely
758 related populations. Mol Biol Evol [Internet]. 2011 Aug [cited 2014 May 25];28(8):2239–52.

759 Available from:
760 http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3144383&tool=pmcentrez&rend
761 ertype=abstract

762 79. Robinson JT, THorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative
763 Genomic Viewer. Nat Biotechnol [Internet]. 2013;29(1):24–6. Available from:
764 https://www.broadinstitute.org/igv/node/250

765 80. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): High-
766 performance genomics data visualization and exploration. Brief Bioinform. 2013;14(2):178–
767 92.

768 81. Lopez J V., Cevario S, O'Brien SJ. Complete nucleotide sequences of the domestic cat (Felis
769 catus) mitochondrial genome and a transposed mtDNA tandem repeat (Numt) in the nuclear
770 genome. Genomics [Internet]. 1996 Apr 15;33(2):229–46. Available from:
771 http://www.ncbi.nlm.nih.gov/pubmed/8660972

772 82. Lopez J V, Yuhki N, Masuda R, Modi W, Brien SJO. Numt , a Recent Transfer and Tandem
773 Amplification of Mitochondrial DNA to the Nuclear G e n o m e of the Domestic Cat. J Mol
774 Evol. 1994;39:174–90.

775 83. Antunes A, Ramos MJ. Discovery of a large number of previously unrecognized mitochondrial
776 pseudogenes in fish genomes. Genomics. 2005;86(6):708–17.

777 84. Jae-Heup K, Eizirik E, O'Brien SJ, Johnson WE. Structure and patterns of sequence variation in
778 the mitochondrial DNA control region of the great cats. Mitochondrion [Internet]. 2001
779 Oct;1(3):279–92. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16120284

780 85. Kim J-H, Antunes A, Luoa S-J, Menninger J, Nash WG, O'Brien SJ, et al. Evolutionary analysis of
781 a large mtDNA translocation (numt) into the nuclear genome of the Panthera genus species.
782 Brain, Behav Immun. 2006 Jul;22(5):629–629.

783 86. Armstrong EE, Taylor RW, Miller DE, Kaelin C, Barsh G, Hadly EA, et al. Long live the king:
784 chromosome-level assembly of the lion (Panthera leo) using linked-read, Hi-C, and long read
785 data. BMC Biol [Internet]. 2020;18(3):705483. Available from:
786 https://www.biorxiv.org/content/10.1101/705483v1.abstract

787 87. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus
788 genotype data. Genetics [Internet]. 2000 Jun;155(2):945–59. Available from:
789 http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1461096&tool=pmcentrez&rend
790 ertype=abstract

791 88. Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the
792 software STRUCTURE: a simulation study. Mol Ecol [Internet]. 2005 Jul [cited 2012 Mar
793 9];14(8):2611–20. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15969739

794 89. Earl DA, VonHoldt BM. STRUCTURE HARVESTER: a website and program for visualizing
795 STRUCTURE output and implementing the Evanno method. Conserv Genet Resour [Internet].
796 2012 Oct 13 [cited 2014 Jul 9];4(2):359–61. Available from:
797 http://link.springer.com/10.1007/s12686-011-9548-7

798 90. Jakobsson M, Rosenberg NA. CLUMPP: a cluster matching and permutation program for
799 dealing with label switching and multimodality in analysis of population structure.
800 Bioinformatics [Internet]. 2007 Jul 15 [cited 2014 Jan 10];23(14):1801–6. Available from:
801 http://www.ncbi.nlm.nih.gov/pubmed/17485429

802 91. Peakall R, Smouse PE. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for
803 teaching and research--an update. Bioinformatics [Internet]. 2012 Oct 1 [cited 2013 Dec
804 11];28(19):2537–9. Available from:
805 http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3463245&tool=pmcentrez&rend
806 ertype=abstract

807 92. Petkova D, Novembre J, Stephens M. Visualizing spatial population structure with estimated
808 effective migration surfaces. Nat Genet. 2016;48(1):94–100.

809 93. IUCN SSC Cat Specialist Group. Conservation strategy for the lion in West and Central Africa.
810 Gland, Switzerland: IUCN; 2006. 1–60 p.

811

812 **Figures**

813 **Figure 1. Distribution of lion sampling localities and inferred phylogenetic relationships between**
814 **populations.** Map indicating sampling locations of lions for whole genome sequencing (red) and SNP
815 panel genotyping (black) (top). Red shading indicates Lion Conservation Units, and white delineation
816 defines the regions (West, Central, East, and Southern), sensu the lion conservation strategies (93).
817 Phylogenetic trees, based on 98,952 nuclear SNPs (left) and 16,756 bp mitogenomes (right) for 10
818 lions which were subjected to whole genome sequencing (bottom). Support values indicate posterior
819 probabilities (MrBayes), bootstrap support from SVDquartets and bootstrap support from Garli.
820 Topologies are indicated per method in the southern branch of the autosomal SNP tree, and each
821 split is maximally supported, with the exception of the Zambia2+Namibia branch in the Garli tree,
822 which received a bootstrap support of 96. Z1: Zambia1, Z2: Zambia2, N: Namibia, R: RSA.

823

824 **Figure 2. Assessment of population structure using NGSadmix and STRUCTURE.** Results of
825 NGSadmix (K=2) based on SNP genotype likelihoods from 10 lions (A)**.** Result of STRUCTURE
826 Harvester for a STRUCTURE run of 125 SNPs genotyped (B). Assignment values based on the
827 STRUCTURE run for K=2 (upper bar plot) and K=4 (lower bar plot) for 171 lions (i.e. excluding samples
828 with >25% missing data; see Supplemental Figure 6 for results for all 211 lions) (C). Sample numbers
829 correspond with numbers in Supplemental Table 4.

830

831 **Figure 3. PCA plots based on whole genome genotype likelihoods and SNP panel SNPs.** Plots are
832 based on genotype likelihoods of 10 lions (A), 125 SNP panel SNPs in 171 lion samples (B), and after
833 exclusion of India (C). Colours indicate region of origin.

834

835 **Figure 4. EEMS results based on 211 lions, including India (left), and 205 lions, excluding India**

836 **(right).** The upper row shows posterior mean migration rates, the bottom row shows posterior mean

837 diversity rates. Orange colours indicate low values, blue colours indicate high values.

838

839 **Supplemental Files**

840 **Supplemental Information S1. Details on laboratory protocols, sequencing, mapping, SNP calling**

841 **and quality control** (including Supplemental Figures S1-S4)

842

843 **Supplemental Information S2. Validation of phylogenetic inferences and populations structure**

844 **from whole genome sequencing and SNP panel data** (including Supplemental Figures S5-S10)

845

846 **Supplemental Tables S1-S8 (see captions in Excel)**