bioRxiv preprint doi: https://doi.org/10.1101/814103; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Whole genome sequencing and the application of a SNP panel reveal primary evolutionary lineages and genomic diversity 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] Field Museum of Natural History, Integrative Research Center, Chicago, IL 60605 USA

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

bioRxiv preprint doi: https://doi.org/10.1101/814103; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

[13] University of Antwerp, Department of Biology, Evolutionary Ecology Group, Groenenborgerlaan 171, 2020 Antwerpen, Belgium

* Corresponding author:
Laura Bertola
Email: laura.bertola@gmail.com

Abstract

Previous phylogeographic studies of the lion (*Panthera leo*) have offered increased insight into the distribution of genetic diversity, as well as a revised taxonomy which now distinguishes between a northern (*Panthera leo leo*) and a southern (*Panthera leo melanochaita*) subspecies. However, existing whole range phylogeographic studies on lions focused on mitochondrial DNA and/or a limited set of microsatellites. The geographic extent of genetic lineages and their phylogenetic relationships remain uncertain, clouded by incomplete lineage sorting and sex-biased dispersal. In this study we present results of whole genome sequencing and subsequent variant calling in ten lions sampled throughout the geographic range. This resulted in the discovery of >150,000 Single Nucleotide Polymorphisms (SNPs), of which ~100,000 SNPs had sufficient coverage in at least half of the individuals. Phylogenetic analyses revealed the same basal split between northern and southern populations as had previously been reported using other genetic markers. Further, we designed a SNP panel, including 125 nuclear and 14 mitochondrial SNPs, which was subsequently tested on >200 lions from across their range. Results allow us to assign individuals to one of the four major clades: West & Central Africa, India, East Africa, or Southern Africa. This SNP panel can have important applications, not only for studying populations on a local geographic scale, but also for guiding conservation management of *ex situ* populations, and for tracing samples of unknown origin for forensic purposes. These genomic resources should not only contribute to our understanding of the

bioRxiv preprint doi: https://doi.org/10.1101/814103; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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), SNP discovery, genome sequencing, SNP panel, phylogeography, evolutionary history, genomic diversity, lion (*Panthera leo*)

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]. Apart from improved insight into patterns of biodiversity, this type of data can inform conservation efforts on how best to maintain this diversity [5, 6].

The lion (*Panthera leo*) has been the subject of several phylogeographic studies which have provided insights into the evolution and distribution of genetic diversity in the African populations (formerly subspecies *Panthera leo leo*) and its connection to the Asiatic population (formerly subspecies *Panthera leo persica*), located in the Gir forest, India. These studies included data from mitochondrial DNA (mtDNA) [7–16], autosomal DNA [12–14, 16] and subtype variation in lion Feline Immunodeficiency Virus (FIV_{Ple}) [13]. Studies using mitochondrial markers and/or complete mitogenomes, describe a basal dichotomy, consisting of a northern group that includes populations from West and Central Africa as well as the Asiatic population (formerly subspecies *P. leo persica*), and a southern group with populations from East and Southern Africa [7, 10, 12, 15]. Another study focused on populations in Tanzania, but also included three populations from West and Central Africa, and confirmed this distinction based on Single Nucleotide Polymorphisms (SNPs) [16]. The distinction of lions in West and Central Africa was further corroborated by autosomal microsatellite data [14]. However, because of limited variation in the remaining Asiatic population as a result of

bioRxiv preprint doi: https://doi.org/10.1101/814103; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

past bottlenecks [13, 17], it is impossible to reliably infer the evolutionary relationship between African and Asiatic populations from these type of data. Nevertheless, these newly described evolutionary relationships led to a review of the taxonomic split between the African and Asiatic subspecies, resulting in a revision of the taxonomy which now recognizes a northern subspecies (*Panthera leo leo*) and a southern subspecies (*Panthera leo melanochaita*) [18].

However, the approaches used in the studies mentioned above have a number of limitations. For example, mtDNA cannot identify admixture and represents only a single locus, potentially misrepresenting phylogenetic relationships due to the stochastic nature of the coalescent [19, 20]. In addition, sex-biased dispersal and gene flow will alter patterns derived from mitochondrial versus nuclear data. Female lions exhibit strong philopatry whereas male lions are capable dispersers [21, 22], indicating that phylogeographic patterns based on mtDNA in lions may overestimate divergence between populations. Inference of phylogenetic relationships from microsatellite data, on the other hand, is problematic due to their high variability and their mutation pattern. Finally, most studies mentioned above were limited by sparse coverage of populations, notably in West and Central Africa. In particular, FIV_{Ple} prevalence is geographically restricted. Additional data from genome-wide markers, covering more lion populations and different conservation units, are necessary to overcome these restrictions. This will improve our understanding of the spatial distribution of diversity in the lion, as well as help guide future conservation efforts.

Here, we describe the discovery and phylogenetic analysis of genome-wide SNPs based on whole genomes and complete mitogenomes of ten lions, providing a comprehensive overview of the intraspecific genomic diversity. We further developed a SNP panel consisting of a subset of the discovered SNPs, which was then used for genotyping >200 samples from 14 lion range states, representing almost the entire current distribution of the lion. This resolves phylogeographic breaks at a finer spatial resolution, and may serve as a reference dataset for future studies. Finally, we

bioRxiv preprint doi: https://doi.org/10.1101/814103; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

discuss the applications and future directions of high-throughput genotyping for wildlife research and conservation that, hopefully, will contribute to future studies on lion genomics.

Materials & Methods

Sampling

Blood or tissue samples of ten lions, representing the main phylogeographic groups as identified in previous studies [7, 10, 13–15] (Figure 1: map, Supplemental Table 1), were collected and preserved 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 included were either free-ranging lions or captive lions with proper documentation of their breeding history and with no known occurrences of hybridization between aforementioned lineages. A sample from a leopard (*Panthera pardus orientalis*, captive) was included as an outgroup. The tiger genome [23] was used as a reference for mapping of the lion and leopard reads. All samples were collected in full compliance with specific legally required permits (CITES and permits related to national legislation in the countries of origin). Details on laboratory protocols, sequencing, assembly, SNP calling, and quality control are given in Supplemental Information 1.

Whole genome data and complete mitogenomes

Identified SNPs were attributed to a chromosome following the genomic architecture in the tiger [23] (Supplemental Table 2). For the nuclear SNPs, we applied 5 levels of filtering, in which we included 1) all SNPs, 2) SNPs called in at least three samples, 3) SNPs called in at least five samples, 4) SNPs called in at least eight samples, and 5) SNPs called in all samples (Supplemental Table 2). Positions with coverage <3 were replaced by an ambiguous nucleotide. Full mitogenomes were recovered as described in Supplemental Information 1 and [15].

Phylogenetic analyses were performed on the full mitogenomes and on the concatenated SNP datasets with varying levels of missing data with MrBayes v.3.1.2 [24, 25] and Garli [26] using

bioRxiv preprint doi: https://doi.org/10.1101/814103; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

parameters as determined by MrModeltest2 (v.2.3) [27]. MrBayes and Garli were run for one million generations and five million generations respectively, using a GTR substitution model with rate variation across sites set to equal. In addition, we ran SVDquartets [28], assuming multi-locus unlinked single-site data with 100 bootstrap replicates as implemented in PAUP* 4.0a164 [29]. For the mitogenome data, the coalescent process in the model was disregarded. Nodes receiving >95% PP in Bayesian analysis (MrBayes) and/or 0.7 bootstrap support in Maximum Likelihood (ML) analysis (Garli) and SVDquartets are considered to have significant support. Individual ancestry coefficients were estimated using sparse nonnegative matrix factorization algorithms in sNMF, as implemented in the R package LEA [30, 31], exploring K=1-10, using 20 replicates and 4 values for the alpha regularization parameter (1, 10, 100 and 1,000).

SNP panel data

In order to obtain better insight into the geographic locations of phylogeographic breaks, we made a selection of SNPs for inclusion in a SNP panel for genotyping more samples. We used the following criteria for the selection process: 1) minimum coverage of 20 for all lion samples combined, across 50 bp upstream and downstream of the SNP position, 2) maximum of one variable position in these 50 bp flanking regions, 3) high quality mapping of the flanking regions as identified by eye using IGV Genome Browser [32, 33], 4) SNPs evenly spread across all chromosomes, as implied by the chromosomal architecture in tiger (max. one SNP per scaffold), 5) preferably both homozygotes and heterozygote genotype present among the ten genotyped lions, otherwise both alleles identified. In addition, we included a total of 14 mitochondrial SNPs selected to represent each major branch within the mitochondrial phylogenetic tree (as described in [15]), as well as the distinction between the northern and the southern subspecies is represented by two SNPs. The selected mtDNA SNPs had already been assessed in a wide range of populations [15], making it more likely that the selected SNPs are diagnostic throughout the lion range. Finally, we ensured that the SNPs were not located in any of the nuclear copies (numts) which are known to exist in cats [34–38]. After test runs and

bioRxiv preprint doi: https://doi.org/10.1101/814103; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

quality control (see Supplemental Information 1), we retained 125 nuclear 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).

Resulting nuclear genotypes were analysed with STRUCTURE [39], using correlated allele frequencies and running the program for 5,000,000 generations, discarding the first 500,000 generation 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 [40]; CLUMPP [41] was used before generating the barplots of population assignment. To reduce the effect of missing data on the assignment results, STRUCTURE was then repeated including only the samples with >75% of the SNPs successfully called (N=171). MtDNA SNPs were used to assign a specific haplotype to each individual, matching with previously described lineages (see delineation of haplogroups in [15]): West Africa, Central Africa, North East Africa, East/Southern Africa, South West Africa, and India.

A Principal Component Analysis (PCA) was performed in Genalex [42]. PCA was repeated, only retaining samples with >75% of the SNPs successfully called (171 individuals), and excluding the Asiatic population.

Patterns of connective zones and barriers were investigated using Estimating Effective Migration Surfaces (EEMS) [43] for all individuals genotyped with the SNP panel. Three independent runs were performed, using 10 million generations and discarding the first 5 million generations as burn-in. We followed the author´s suggestions for tuning the proposal variances, using 0.1 and 1 for mSeedsProposalS2 and mEffctProposalS2 respectively, and 1.5 and 0.015 for qSeedsProposalS2 and qEffctProposalS2 respectively.

bioRxiv preprint doi: https://doi.org/10.1101/814103; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Heterozygosity

The level of heterozygosity was assessed for each individual for which the whole genome had been sequenced, ignoring ambiguous nucleotides (i.e. positions which were not scored due to insufficient quality or coverage). SNP panel results were included if several individuals from the same population had been included. Results were then compared to known levels of heterozygosity based on earlier studies using microsatellite data from the same populations [12, 14].

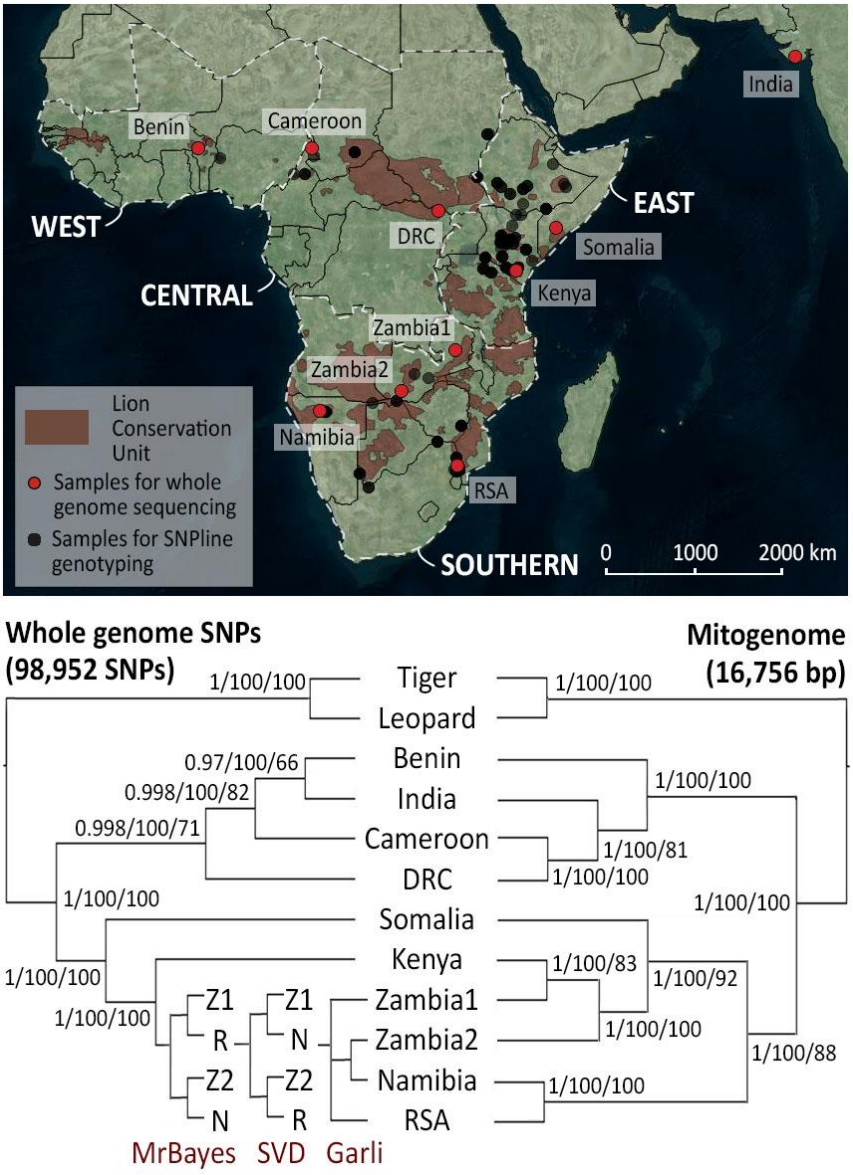

Figure 1. Map indicating sampling locations of lion samples for whole genome sequencing and SNP panel genotyping (top). Red shading indicates Lion Conservation Units, and white delineation defines the regions (West, Central, East, and Southern), sensu the lion conservation strategies [74]. Phylogenetic trees, based on 98,952 nuclear SNPs (left) and 16,756 bp mitogenomes (right) for 10 lions which were subjected to whole genome sequencing (bottom). Support values indicate posterior probabilities (MrBayes), bootstrap support from SVDquartets and bootstrap support from Garli. All methods agreed on overall topology, with the exception of the southern branch in the nuDNA SNP tree. Topologies are indicated per method, and each split is maximally supported, with the exception of the Zambia2+Namibia branch in the Garli tree, which received a bootstrap support of 96. Z1: Zambia1, Z2: Zambia2, N: Namibia, R: RSA.

bioRxiv preprint doi: https://doi.org/10.1101/814103; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Results

Sequencing

The sequencing runs yielded a total of $6.5 \cdot 10^8$ reads. Following quality control, a total of $5.9 \cdot 10^8$ reads (94.4%) were retained for subsequent alignment (Supplemental Table 1). Filtering of variable positions in lions yielded a total of 155,678 SNPs. Upon filtering for positions which had reliable callings (i.e. coverage ≥ 3), we retained 118,270 SNPs of which 98,952 SNPs were called in at least five individuals (see Supplemental Table 2 for results for different levels of missing data). Missing data ranged from 90% (Benin) to 10% (Kenya) for all variable positions. Results reported for downstream phylogenetic analysis are based on the lion-specific variable positions which could be called in at least half of the included samples (98,952 SNPs).

Whole genome data and complete mitogenomes

Phylogenetic analyses, based on 98,952 SNPs, show a well-supported dichotomy between the northern populations (Benin, Cameroon, DRC, India) and the southern ones (Somalia, Kenya, Zambia1, Zambia2, Republic of South Africa (RSA), Namibia) (Figure 1: left tree). Phylogenetic trees based on the mitochondrial genomes show the same basal dichotomy, with the Asiatic population nested within Central Africa (Figure 1: right tree), as was previously reported [15]. However, it must be noted that the individual from RSA contains a haplotype from Namibia, likely the result of historic translocation, as was previously described [15]. MrBayes, Garli and SVDquartets resulted largely in the same topology, although the three methods do not agree on the relationships of Zambia1, Zambia2, RSA and Namibia based on the nuclear SNPs. Clustering of these ten individuals, as inferred by sNMF, identifies the same basal split into a northern and a southern cluster (Figure 2: map and inserted barplot).

bioRxiv preprint doi: https://doi.org/10.1101/814103; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

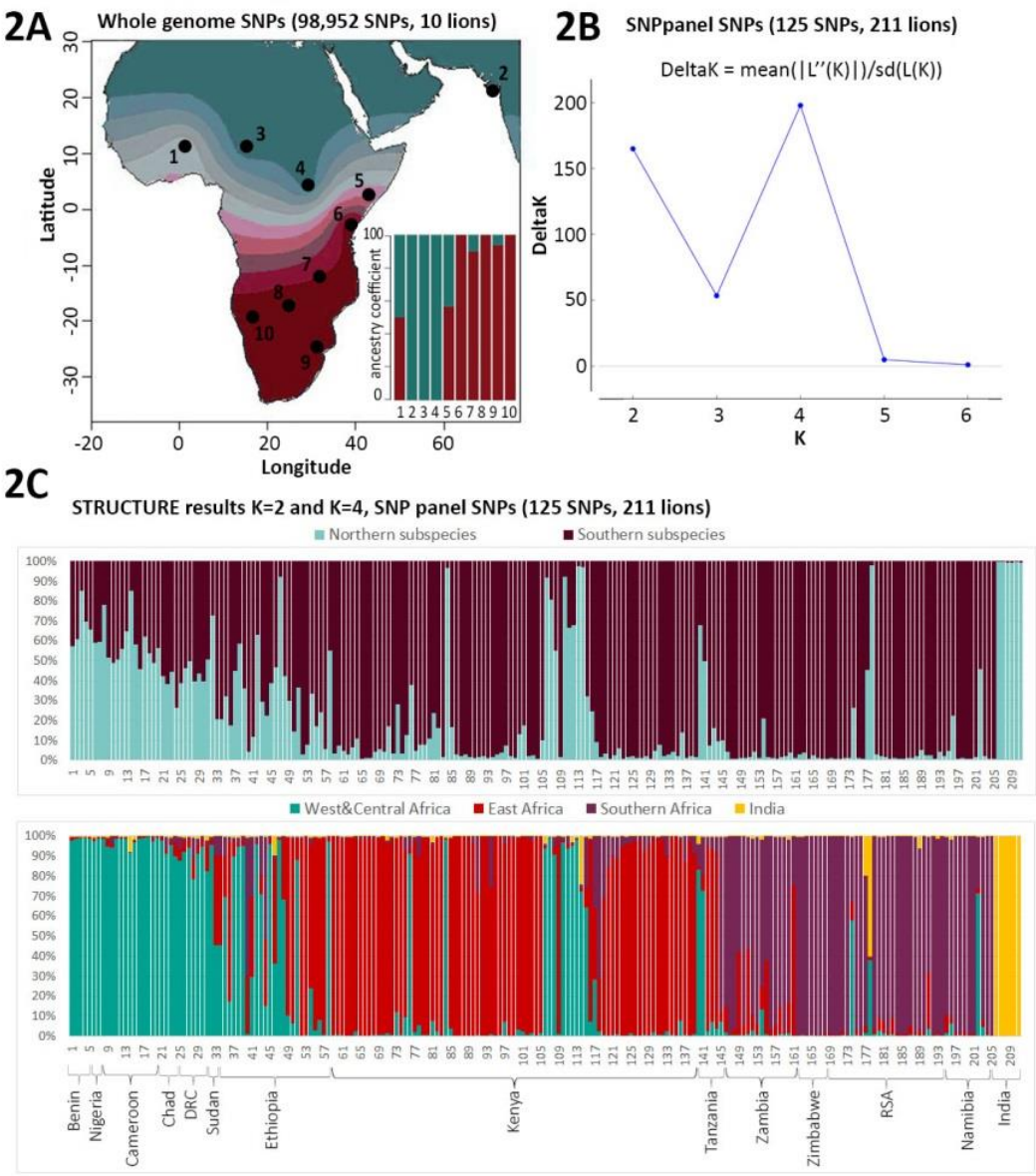

Figure 2. Results of sNMF (K=2) based on 98,952 nuclear SNPs in 10 lions plotted on a map, with the assignment values indicated in the inserted barplot (top, left). Result of STRUCTURE Harvester for a STRUCTURE run of 125 SNPs genotyped in 211 lions (top, right). Assignment values based on the STRUCTURE run for K=2 (upper barplot) and K=4 (lower barplot) for all 211 lions. Colours indicate region of origin. Sample numbers correspond with numbers in Supplemental Table 3.

SNP panel data

A total of 211 samples from across the entire range of the lion were genotyped for 125 nuclear SNPs and 14 mitochondrial SNPs. Missing data for the nuclear SNPs ranged from to 0 to 115 with a median value of 11; for the mitochondrial data, these ranged from 0 to 10 with a median value of 0 (Supplemental Table 3, Supplemental Figure 1). STRUCTURE suggests an optimal number of four clusters (Figure 2: graph) corresponding to West & Central Africa, India, East Africa, and Southern Africa (Figure 2: lower barplot). A lower peak can be detected for K=2, distinguishing roughly

bioRxiv preprint doi: https://doi.org/10.1101/814103; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

between the northern and the southern subspecies with India forming a strong separate cluster (Figure 2: upper barplot). Assignment values to clusters and to mitochondrial haplogroups are reported in Supplemental Table 4. In the PCA, African and Asiatic lions are distinguished as two clouds (A and 3C). Removing the Asiatic population reveals more structure within African lions (Figure 3B), which becomes increasingly apparent when only including samples with >75% of their SNPs successfully called (N=171) are included (Figure 3D). This results in two nearly distinct clouds representing the northern (West and Central Africa) and the southern (East and Southern Africa) subspecies.

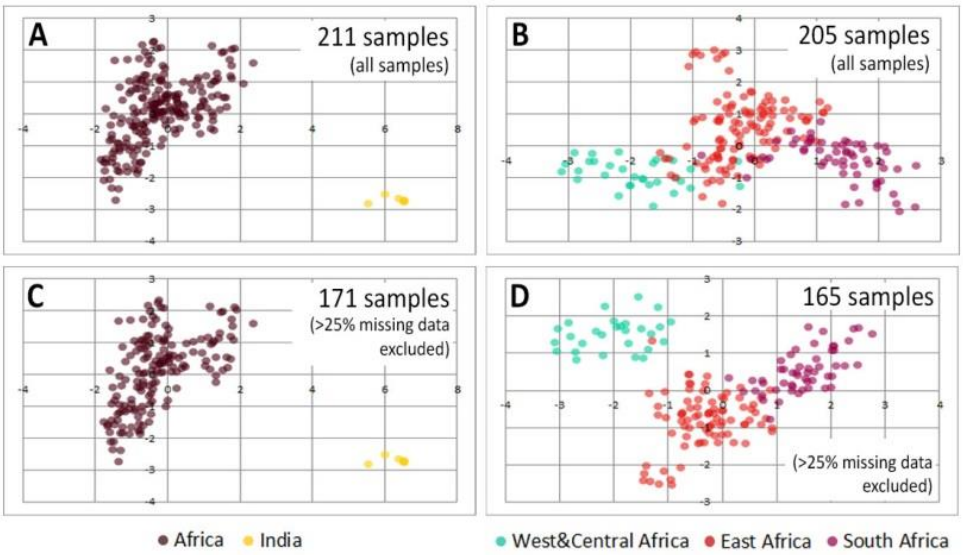

Figure 3. PCA plots based on 125 SNPs in 211 lion samples (A), after exclusion of the Asiatic population (B), and PCA plots based on 125 SNPs in samples with >75% SNPs successfully called (N=171) (C), and after exclusion of the Asiatic population (D). Colours indicate region of origin.

EEMS infers corridors and barriers for dispersal and gene flow from spatial the decay of genetic similarity. Results show that the Central African rainforest is highlighted as a barrier (indicated with orange shading, Figure 4: upper row) , whether or not the Asiatic population is included in the analysis. Barriers are further identified between East and Southern Africa, and across the Arabian peninsula. Diversity indices illustrate that the Asiatic population has a much lower genetic diversity than the African populations (Figure 4: bottom left). After repeating the analysis with only African populations, low genetic diversity is detected in West Africa and Southern Africa (Figure 4: bottom right).

bioRxiv preprint doi: https://doi.org/10.1101/814103; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

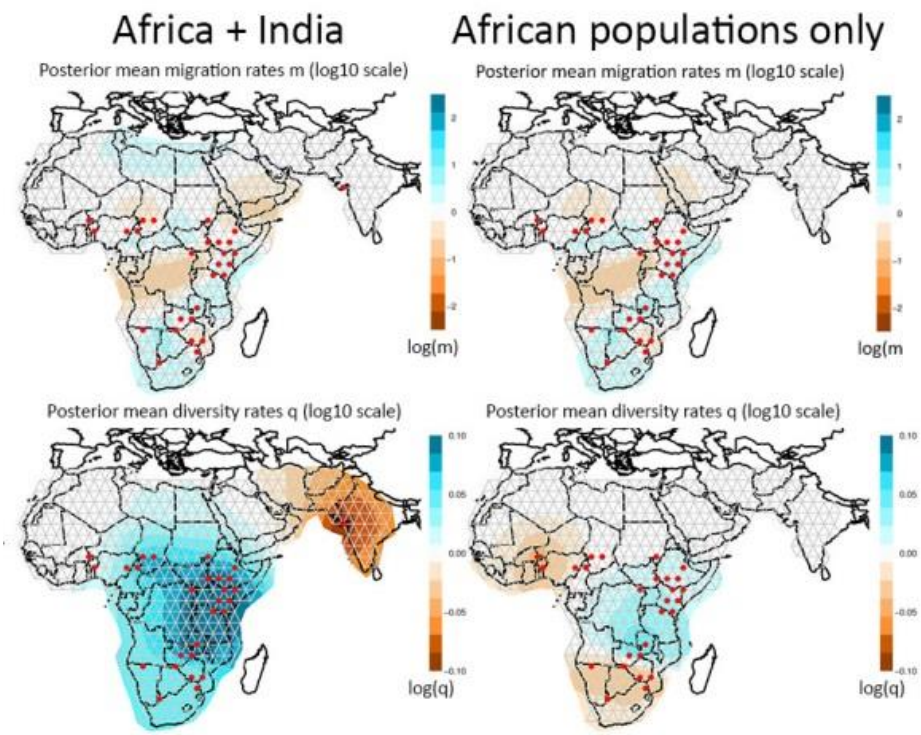

Figure 4. EEMS results based on 125 SNPs in 211 lions samples, including the Asiatic population (left panels), and 205 samples, excluding the Asiatic population (right panels). The upper row shows posterior mean migration rates, the bottom row shows posterior mean diversity rates. Orange colours indicate low values, blue colours indicate high values.

Heterozygosity

Comparisons of levels of heterozygosity to previously published data [12, 14] (Supplemental Table 5) show that ranking between SNP data and microsatellite data is roughly congruent. 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 and RSA).

Discussion

This study is the first to use whole genome sequencing in an attempt to describe the full genomic diversity over the lion's range. Phylogenetic analyses show strong congruence with previously published patterns, confirming the basal dichotomy separating a northern and a southern subspecies in the lion. The use of a SNP panel provides further insight into phylogeographic patterns at a higher spatial resolution.

bioRxiv preprint doi: https://doi.org/10.1101/814103; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Whole genome data and complete mitogenomes

Although genome-wide SNPs can be a powerful marker, ascertainment bias as a result of the study's design is a concern [44]. To reflect the full diversity of the species, it is necessary to base the SNP discovery on samples representing all lineages. MtDNA has proven to be a useful genetic marker for gaining insight into phylogeographic patterns, partly because of its shorter coalescence time compared to nuclear markers. However, this may lead to 'oversplitting', reflecting fully coalesced groups based on mtDNA data, but incomplete lineage sorting of nuclear DNA (nuDNA) alleles. 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 [14], admixture in Kruger area [15, 45]). Together, they provide useful criteria for selecting populations to be subjected to whole genome sequencing.

Based on the whole genome data, we explored different levels of missing data, balancing the number of SNPs and the number of samples with an accepted call at a given position. As a higher number of SNPs represents a denser sampling of the coalescent (e.g. see [46]), presented trees are based on SNPs which were present in at least 50% of the samples. Increasing the number of SNPs (and therefore also the amount of missing data), did not change the topology or support of the phylogenetic trees. Simulation studies and studies on empirical data have shown that concatenated SNP data are able to produce reliable trees reflecting the true topology, as long as enough genes are sampled [47, 48]. The underlying assumption is that there is enough phylogenetic signal in the data, and that discordant coalescent histories will effectively cancel each other out when all histories are considered together. The conclusion that our concatenated phylogenetic trees (MrBayes and Garli) produce reliable topologies is supported by the fact that observed patterns are congruent with results of SVDquartets, which does assume multi-locus unlinked data. An exception is the topology for Southern Africa where patterns of differentiation are likely to be affected by continuous gene flow and incomplete lineage sorting. Further, described topologies are in line with previously

bioRxiv preprint doi: https://doi.org/10.1101/814103; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

described mitochondrial trees [7, 15] with similar discordances as found in microsatellite datasets [14]. Notably, the wide-spread haplogroup labeled as East/Southern Africa (Supplemental Table 4, [15]), stretching from Kenya to Namibia, is not recovered from microsatellite or SNP data [14] (this study). The previously mentioned dichotomy between the northern subspecies, *Panthera leo leo*, and the southern subspecies *Panthera leo melanochaita*, is also supported by the assignment values of sNMF. The putative hybrid character of the individual from Benin is likely the result from the substantial amount of missing data, and does not reflect its evolutionary origin or admixture.

SNP panel data

The whole genome sequencing of ten individuals allowed identification of variable positions in the lion. Based on the ten whole genomes, we generated a SNP panel of 125 nuclear and 14 mitochondrial SNPs which allows cost-effective genotyping of larger numbers of samples. As part of quality control, i.e. to ensure reliable SNP callings, we explored different concentrations of the starting amount of DNA and the effect of using a Whole Genome Amplification (WGA) kit. In addition, we included a number of samples which were known to be of lower quality DNA (older, degraded samples). We observed an increase in missing values for samples of lower quality or quantity of DNA, however, we never observed changes in the called genotype (allelic drop-out). This illustrates that the SNP panel is producing reliable results, although the sensitivity of downstream analyses with high numbers of missing values needs to be taken into account.

The SNP panel was tested on 211 samples from across the range of the lion, with 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 [15], and microsatellite data have suggested admixture [14]. Our results assign all tested lions to one of the four identified clusters: West & Central Africa, India (the two clades of the northern subspecies, *Panthera leo leo*), East Africa or Southern Africa (the two clades of the southern subspecies, *Panthera*

bioRxiv preprint doi: https://doi.org/10.1101/814103; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

leo melanochaita). Main regions of admixture are Ethiopia and Zambia, which was also found using microsatellite data [12, 14]. The high assignment values to West and Central Africa within the Kenyan population are likely to be an artefact of missing data in these individuals; this pattern disappears after repeating the run with only individuals with >75% called SNP data (Supplemental Figure 2). However, as lions from the northern part of Kenya are thought to represent a different evolutionary lineage, some admixture between lineages is plausible and merits further research. It must be noted that when K=2 the result is a split that resembles the distinction between the northern and the southern subspecies, however, due to the very low diversity in the Asiatic population, assignment values are driven to ~1 in the Asiatic lions. Samples from West and Central Africa are largely assigned to the northern cluster, however they also show assignment to the cluster representing the southern subspecies. We wish to emphasize that this does not indicate region-wide hybridization between the two subspecies, nor does it imply that the populations with the highest assignment values constitute ancestral populations. Rather, this result is driven by the extremely low genetic diversity in the Asiatic population, comparable to the Africa/Asia split which is found based on STRUCTURE analysis or PCA of microsatellite data [14]. It is well known that STRUCTURE is sensitive to groups of closely related individuals, such as siblings, family groups, or in our case, an inbred population [49, 50] and that identification of ancestral populations may be an over-interpretation of the data, depending on demographic histories [51]. Also in this case, clustering of West and Central African lions with Asiatic lion decreases after running STRUCTURE with only individuals with >75% SNPs called (Supplemental Figure 2).

Comparing the assignments based on STRUCTURE and the assignments based on the mitochondrial SNPs show a largely congruent pattern. Six individuals show an unexpected haplotype (i.e. a haplotype not previously documented from this country): Ethiopia13 – West African haplotype, Ethiopia26 – East/Southern African haplotype, Kenya49 – Central African haplotype, Tanzania1 – North East African haplotype, RSA10 – North East African haplotype, Namibia2 – North East African

bioRxiv preprint doi: https://doi.org/10.1101/814103; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

haplotype. However, it must be noted that all of these individuals have >35% missing nuDNA data (with Kenya49, Tanzania1, RSA10 and Namibia2 even missing >50% nuDNA data). In addition, Ethiopia26, RSA10 and Namibia2 also have >20% missing data in the mitochondrial SNPs, and therefore results should be interpreted with caution.

PCA illustrates the strong clustering of the Asiatic population, likely the result of low diversity from multiple bottlenecks (see above). After excluding the Asiatic population, three overlapping clouds are apparent, roughly indicating West & Central Africa, East Africa and Southern Africa. Including only samples with >75% of the SNPs successfully called, a clearer divide is visible between the northern and southern subspecies in the PCA (Figure 3D). The individual which has a color code from East Africa (red) but falls in the West and Central Africa cluster (green) in the analysis originates from Ethiopia, where both subspecies are known to overlap and admixture has been described before [14, 15]. The distinction between East and Southern Africa seems to be more gradual in PCA space, which is in line with the widespread haplogroup which occurs throughout almost the entire region.

Barriers to dispersal and evolutionary history

In order to put the phylogenetic patterns of the lion into an evolutionary perspective, it is worthwhile to explore current and historical barriers to lion dispersal. Paleoclimatic data show that cyclical contraction and expansion of vegetation zones, e.g. rain forest and desert, may have represented temporal barriers for lion dispersal [15]. The discrete genetic lineages recognizable in the mtDNA are likely to be the result of the restriction of suitable lion habitat to a number of refugia [15]. The pattern found in mtDNA data of the lion is congruent with that of other African savanna species [15, 52, 53] and predicted refugial areas based on climate models [54]. Faster coalescence times of mtDNA may have led to reciprocally monophyletic mtDNA clades in the lion, while isolation in refugia may not have lasted long enough for coalescence in autosomal markers [14, 15]. In addition, dispersal in lions is male-biased [21, 22], which may explain the more discrete phylogeorgaphic

bioRxiv preprint doi: https://doi.org/10.1101/814103; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

pattern found in mtDNA data. This is reflected by the fact that we do not retrieve a North East Africa cluster or a South West Africa cluster based on nuclear SNPs, even though they are represented by diverged mitochondrial haplotypes. Interestingly, the discrepancy in population structure between nuclear and mitochondrial markers in East and Southern Africa, where a wide-spread mitochondrial haplogroup occurs from Kenya to Namibia and nuclear markers suggest a phylogeographic break around Zambia and Mozambique, is identical to the discrepancy found in giraffe [55]. Current barriers for gene flow seem to be mainly presented by the recent population disjunction in North Africa/Middle East and the longstanding barrier representing the Central African rain forest. Although the Rift valley has been mentioned as a potential barrier for gene flow in the lion [7–9, 11] and gene flow may be reduced in that region, co-occurrence of strongly diverged haplotypes and admixture detected in microsatellite and SNP data in Ethiopia indicates that the Rift valley does not pose a complete barrier for lion dispersal [14, 15].

Applications for wildlife research and conservation

The design of a SNP panel and a reference dataset of >200 lions from 14 lion range states has important applications for wildlife research and conservation. First, it allows us to distinguish phylogeographic breaks and overlap between lineages which can be used to study the evolutionary history of the species in more detail, e.g. at the national level. This may have important conservation implications and genomic data can be used to inform both population management and translocations. Secondly, it enables tracing of lion samples of unknown origin, such as confiscated material from illegal trade or poaching. Illegal trade in wildlife products is currently estimated to be the fifth largest illegal industry globally, and a major concern for conservationists [56–58]. Part of the trade is thought to cater to domestic markets, but growing evidence suggests an increase in illegal shipments to international markets [59]. Genetic toolkits can contribute to combatting wildlife crimes by identifying confiscated material, source populations and tracking trade routes [45, 60–63]. Thirdly, a SNP panel can be used to guide breeding efforts for ex situ conservation [45, 64]. Recently,

bioRxiv preprint doi: https://doi.org/10.1101/814103; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

this SNP panel was used to genotype a batch of lion samples from institutions linked to the European Association for Zoos and Aquaria (EAZA). Based on the resulting information, managers are currently deciding how many and which lineages to include in future breeding efforts.

Future perspectives

With the increase of genomic data and computational and technological developments, the field of conservation genomics is undergoing a major transition. Demographic histories and patterns of gene flow can be inferred in greater detail [65, 66]. Searching genomes for signatures of adaptation has been highlighted as a powerful tool for gauging the sensitivity of populations in a changing environment [67–70]. New developments allowing SNP genotyping from poor quality (e.g. non-invasively collected) samples, will further contribute to the applicability of SNP genotyping in the field [71]. Finally, the development of mobile, hand-held sequencing devices shows great promises for rapid, real-time identification of samples [72]. Such tools represent tremendous potential for training and capacity building [73], especially for biodiverse countries with limited facilities to process samples, thereby making the field of wildlife research and conservation more democratized and inclusive.

This study is the first to report phylogeographic relationships between lion populations throughout their entire range based on ~100,000 SNPs derived from whole genome sequencing. We present a SNP panel, containing 125 nuclear and 14 mitochondrial SNPs which has been tested on >200 individuals from 14 lion range states, spanning most of the lion's range. The results confirm a basal distinction between a northern and a southern subspecies that supports the recent revision of the lion's taxonomy. The samples on which the panel was tested can serve as a reference database for future research and conservation efforts.

bioRxiv preprint doi: https://doi.org/10.1101/814103; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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.

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).

References

1. Ellegren H: **Genome sequencing and population genomics in non-model organisms.** *Trends Ecol Evol* 2014, **29**:51–63.

2. Ekblom R, Galindo J: **Applications of next generation sequencing in molecular ecology of non-model organisms.** *Heredity (Edinb)* 2011, **107**:1–15.

3. Shafer ABA, Wolf JBW, Alves PC, Bergström L, Bruford MW, Brännström I, Colling G, Dalén L, De

bioRxiv preprint doi: https://doi.org/10.1101/814103; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Meester L, Ekblom R, Fawcett KD, Fior S, Hajibabaei M, Hill JA, Hoezel AR, Höglund J, Jensen EL, Krause J, Kristensen TN, Krützen M, McKay JK, Norman AJ, Ogden R, Österling EM, Ouborg NJ, Piccolo J, Popović D, Primmer CR, Reed FA, Roumet M, et al.: **Genomics and the challenging translation into conservation practice**. *Trends Ecol Evol* 2015, **30**:78–87.

4. Stephens ZD, Lee SY, Faghri F, Campbell RH, Zhai C, Efron MJ, Iyer R, Schatz MC, Sinha S, Robinson GE: **Big Data: Astronomical or Genomical?** *PLOS Biol* 2015, **13**:e1002195.

5. Hendricks S, Anderson EC, Antao T, Bernatchez L, Forester BR, Garner B, Hand BK, Hohenlohe PA, Kardos M, Koop B, Sethuraman A, Waples RS, Luikart G: **Recent advances in conservation and population genomics data analysis**. *Evol Appl* 2018, **11**:1197–1211.

6. Supple MA, Shapiro B: **Conservation of biodiversity in the genomics era**. *Genome Biol* 2018, **19**:1–12.

7. Bertola LD, van Hooft WF, Vrieling K, Uit de Weerd DR, York DS, Bauer H, Prins HHT, Funston PJ, Udo de Haes HA, Leirs H, van Haeringen WA, Sogbohossou E, Tumenta PN, Iongh HH de, De Iongh HH: **Genetic diversity, evolutionary history and implications for conservation of the lion (Panthera leo) in West and Central Africa**. *J Biogeogr* 2011, **38**:1356–1367.

8. Barnett R, Yamaguchi N, Barnes I, Cooper A: **The origin, current diversity and future conservation of the modern lion (Panthera leo).** *Proc R Soc B Biol Sci* 2006, **273**:2119–2125.

9. Barnett R, Yamaguchi N, Barnes I, Cooper A: **Lost populations and preserving genetic diversity in the lion Panthera leo: Implications for its ex situ conservation**. *Conserv Genet* 2006, **7**:507–514.

10. Barnett R, Yamaguchi N, Shapiro B, Ho SY, Barnes I, Sabin R, Werdelin L, Cuisin J, Larson G: **Revealing the maternal demographic history of Panthera leo using ancient DNA and a spatially explicit genealogical analysis**. *BMC Evol Biol* 2014, **14**:70.

11. Dubach J, Patterson BD, Briggs MB, Venzke K, Flamand J, Stander P, Scheepers L, Kays RW: **Molecular genetic variation across the southern and eastern geographic ranges of the African lion, Panthera leo**. *Conserv Genet* 2005, **6**:15–24.

bioRxiv preprint doi: https://doi.org/10.1101/814103; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

12. Dubach JM, Briggs MB, White PA, Ament BA, Patterson BD: **Genetic perspectives on "Lion Conservation Units" in Eastern and Southern Africa**. *Conserv Genet* 2013, **1942**(Guggisberg 1961).

13. Antunes A, Troyer JL, Roelke ME, Pecon-Slattery J, Packer C, Winterbach C, Winterbach H, Hemson G, Frank L, Stander P, Siefert L, Driciru M, Funston PJ, Alexander KA, Prager KC, Mills G, Wildt D, Bush M, O'Brien SJ, Johnson WE: **The evolutionary dynamics of the lion Panthera leo revealed by host and viral population genomics.** *PLoS Genet* 2008, **4**:e1000251.

14. Bertola LD, Tensen L, van Hooft P, White P a, Driscoll C a, Henschel P, Caragiulo A, Dias-Freedman I, Sogbohossou E a, Tumenta PN, Jirmo TH, de Snoo GR, de Iongh HH, Vrieling K: **Autosomal and mtDNA Markers Affirm the Distinctiveness of Lions in West and Central Africa.** *PLoS One* 2015, **10**:e0137975.

15. Bertola LD, Jongbloed H, Van Der Gaag KJ, De Knijff P, Yamaguchi N, Hooghiemstra H, Bauer H, Henschel P, White PA, Driscoll CA, Tende T, Ottosson U, Saidu Y, Vrieling K, De Iongh HH: **Phylogeographic Patterns in Africa and High Resolution Delineation of Genetic Clades in the Lion (Panthera leo)**. *Sci Rep* 2016, **6**(August):1–11.

16. Smitz N, Jouvenet O, Ligate FA, Crosmary W-G, Ikanda D, Chardonnet P, Fusari A, Meganck K, Gillet F, Melletti M, Michaux J: **A genome-wide data assessment of the African lion (Panthera leo) population genetic structure in Tanzania**. *PlosOne* 2018:1–24.

17. Driscoll CA, Menotti-Raymond M, Nelson G, Goldstein D, O'Brien SJ: **Genomic microsatellites as evolutionary chronometers: a test in wild cats.** *Genome Res* 2002, **12**:414–423.

18. Kitchener A, Breitenmoser-Würsten C, Eizirik E, Gentry A, Werdelin L, Wilting A, Yamaguch N: **A revised taxonomy of the Felidae**. *Cat News* 2017:80.

19. Zink RM, Barrowclough GF: **Mitochondrial DNA under siege in avian phylogeography.** *Mol Ecol* 2008, **17**:2107–2121.

20. Edwards S, Bensch S: **Looking forwards or looking backwards in avian phylogeography? A comment on Zink and Barrowclough 2008.** *Mol Ecol* 2009, **18**:2930-3-6.

bioRxiv preprint doi: https://doi.org/10.1101/814103; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

21. Pusey AE, Packer C, Erhoff-Mulder MB: **The Evolution of Sex-biased Dispersal in Lions**. *Behaviour* 1987, **101**:275–310.

22. Spong GF, Creel S: **Deriving dispersal distances from genetic data.** *Proc Biol Sci* 2001, **268**:2571–2574.

23. Cho YS, Hu L, Hou H, Lee H, Xu J, Kwon S, Oh S, Kim H, Jho S, Kim S, Shin Y, Kim BC, Kim H, Kim C, Turner JA, Luo S, Johnson WE, Koepfli K, Schmidt-Küntzel A, Marker L, Harper C, Miller SM, Jacobs W, Bertola LD, Kim TH, Lee S, Zhou Q, Jung H, Xu X, Gadhvi P, et al.: **The tiger genome and comparative analysis with lion and snow leopard genomes**. *Nat Commun* 2013, **4**.

24. Huelsenbeck JP, Ronquist F: **MRBAYES: Bayesian inference of phylogenetic trees.** *Bioinformatics* 2001, **17**:754–5.

25. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard M a, Huelsenbeck JP: **MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space.** *Syst Biol* 2012, **61**:539–42.

26. Zwickl DJ: **Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion**. The University of Texas at Austin; 2006.

27. Nylander JAA: **MrModeltest v2.3**. 2004:Program distributed by the author.

28. Chifman J, Kubatko L: **Quartet inference from SNP data under the coalescent model**. *Bioinformatics* 2014, **30**:3317–3324.

29. Swofford DL: **PAUP*: Phylogenetic analysis using parsimony (*and other methods), version 4b10.** 2002.

30. Frichot E, François O: **LEA: An R package for landscape and ecological association studies**. *Methods Ecol Evol* 2015, **6**:925–929.

31. Frichot E, Mathieu F, Trouillon T, Bouchard G, François O: **Fast and efficient estimation of individual ancestry coefficients**. *Genetics* 2014, **196**:973–983.

bioRxiv preprint doi: https://doi.org/10.1101/814103; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

32. Robinson JT, THorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov J: **Integrative Genomic Viewer**. *Nat Biotechnol* 2013, **29**:24–26.

33. Thorvaldsdóttir H, Robinson JT, Mesirov JP: **Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration**. *Brief Bioinform* 2013, **14**:178–192.

34. Lopez J V., Cevario S, O'Brien SJ: **Complete nucleotide sequences of the domestic cat (Felis catus) mitochondrial genome and a transposed mtDNA tandem repeat (Numt) in the nuclear genome.** *Genomics* 1996, **33**:229–46.

35. Lopez J V, Yuhki N, Masuda R, Modi W, Brien SJO: **Numt , a Recent Transfer and Tandem Amplification of Mitochondrial DNA to the Nuclear G e n o m e of the Domestic Cat**. *J Mol Evol* 1994, **39**:174–190.

36. Antunes A, Ramos MJ: **Discovery of a large number of previously unrecognized mitochondrial pseudogenes in fish genomes**. *Genomics* 2005, **86**:708–717.

37. Jae-Heup K, Eizirik E, O'Brien SJ, Johnson WE: **Structure and patterns of sequence variation in the mitochondrial DNA control region of the great cats.** *Mitochondrion* 2001, **1**:279–292.

38. Kim J-H, Antunes A, Luoa S-J, Menninger J, Nash WG, O'Brien SJ, Johnson WE: **Evolutionary analysis of a large mtDNA translocation (numt) into the nuclear genome of the Panthera genus species**. *Brain, Behav Immun* 2006, **22**:629–629.

39. Pritchard JK, Stephens M, Donnelly P: **Inference of population structure using multilocus genotype data.** *Genetics* 2000, **155**:945–59.

40. Evanno G, Regnaut S, Goudet J: **Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study.** *Mol Ecol* 2005, **14**:2611–20.

41. Jakobsson M, Rosenberg NA: **CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure.** *Bioinformatics* 2007, **23**:1801–6.

bioRxiv preprint doi: https://doi.org/10.1101/814103; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

42. Peakall R, Smouse PE: **GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research--an update.** *Bioinformatics* 2012, **28**:2537–9.

43. Petkova D, Novembre J, Stephens M: **Visualizing spatial population structure with estimated effective migration surfaces**. *Nat Genet* 2016, **48**:94–100.

44. Lachance J, Tishkoff SA: **SNP ascertainment bias in population genetic analyses: Why it is important, and how to correct it**. *Bioessays* 2014, **35**:780–786.

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

46. Tripp EA, Tsai YHE, Zhuang Y, Dexter KG: **RADseq dataset with 90% missing data fully resolves recent radiation of Petalidium (Acanthaceae) in the ultra-arid deserts of Namibia**. *Ecol Evol* 2017, **7**:7920–7936.

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

48. Xi Z, Liu L, Davis CC: **The impact of missing data on species tree estimation**. *Mol Biol Evol* 2016, **33**:838–860.

49. Waples RS, Anderson EC: **Purging putative siblings from population genetic data sets: A cautionary view**. *Mol Ecol* 2017:1211–1224.

50. Rodríguez-Ramilo ST, Wang J: **The effect of close relatives on unsupervised Bayesian clustering algorithms in population genetic structure analysis**. *Mol Ecol Resour* 2012, **12**:873–884.

51. Lawson DJ, van Dorp L, Falush D: **A tutorial on how not to over-interpret STRUCTURE and ADMIXTURE bar plots**. *Nat Commun* 2018, **9**:1–11.

52. Hewitt GM: **The structure of biodiversity - insights from molecular phylogeography.** *Front Zool* 2004, **1**:4.

bioRxiv preprint doi: https://doi.org/10.1101/814103; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

53. Lorenzen ED, Heller R, Siegismund HR: **Comparative phylogeography of African savannah ungulates.** *Mol Ecol* 2012, **21**:3656–70.

54. Levinsky I, Araújo MB, Nogués-Bravo D, Haywood AM, Valdes PJ, Rahbek C: **Climate envelope models suggest spatio-temporal co-occurrence of refugia of African birds and mammals**. *Glob Ecol Biogeogr* 2013, **22**:351–363.

55. Fennessy J, Bidon T, Reuss F, Vamberger M, Fritz U, Janke Correspondence A, Kumar V, Elkan P, Nilsson MA, Janke A: **Multi-locus Analyses Reveal Four Giraffe Species Instead of One**. *Curr Biol* 2016, **26**:1–7.

56. Nijman V, Morcatty T, Smith JH, Atoussi S, Shepherd CR, Siriwat P, Nekaris KAI, Bergin D: **Illegal wildlife trade–surveying open animal markets and online platforms to understand the poaching of wild cats**. *Biodiversity* 2019, **20**:58–61.

57. ‘t Sas-Rolfes M, Challender DWS, Hinsley A, Veríssimo D, Milner-Gulland EJ: **Illegal Wildlife Trade: Patterns, Processes, and Governance**. *Annu Rev Environ Resour* 2019, **44**:1–28.

58. Biggs D, Cooney R, Roe D, Dublin HT, Allan JR, Challender DWS, Skinner D: **Developing a theory of change for a community-based response to illegal wildlife trade**. *Conserv Biol* 2017, **31**:5–12.

59. Williams VL, Loveridge AJ, Newton DJ, Macdonald DW: **A roaring trade? The legal trade in Panthera leo bones from Africa to East-Southeast Asia**. *PLoS One* 2017, **12**:1–22.

60. Wasser S, Poole J, Lee P, Lindsay K, Dobson A, Hart J, Douglas-hamilton I, Wittemyer G, Granli P, Morgan B, Gunn J, Alberts S, Beyers R, Chiyo P, Croze H, Estes R, Joram P, Kikoti A, Kingdon J, King L, Macdonald D: **Elephants, Ivory, and Trade**. *Science (80-)* 2010, **327**(March):1331–1332.

61. Wasser SK, Brown L, Mailand C, Mondol S, Clark W, Laurie C, Weir BS: **Genetic assignment of large seizures of elephant ivory reveals major poaching hotspots**. *Science (80-)* 2015(June):1–7.

62. Wasser SK, Joseph Clark W, Drori O, Stephen Kisamo E, Mailand C, Mutayoba B, Stephens M: **Combating the illegal trade in African elephant ivory with DNA forensics.** *Conserv Biol* 2008,

bioRxiv preprint doi: https://doi.org/10.1101/814103; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

22:1065–1071.

63. Mondol S, Sridhar V, Yadav P, Gubbi S, Ramakrishnan U: **Tracing the Geographic Origin of Traded Leopard Body Parts in the Indian Subcontinent with DNA-Based Assignment Tests.** *Conserv Biol* 2014, **0**:1–9.

64. Henry P, Miquelle D, Sugimoto T, McCullough DR, Caccone A, Russello M a: **In situ population structure and ex situ representation of the endangered Amur tiger.** *Mol Ecol* 2009, **18**:3173–84.

65. Kamm JA, Terhorst J, Durbin R, Song YS: **Efficiently inferring the demographic history of many populations with allele count data**. *bioRxiv* 2018:287268.

66. Elleouet JS, Aitken SN: **Exploring Approximate Bayesian Computation for inferring recent demographic history with genomic markers in nonmodel species**. *Mol Ecol Resour* 2018, **18**:525–540.

67. Harrisson KA, Pavlova A, Telonis-Scott M, Sunnucks P: **Using genomics to characterize evolutionary potential for conservation of wild populations**. *Evol Appl* 2014, **7**:1008–25.

68. Funk WC, Mckay JK, Hohenlohe PA, Allendorf FW: **Harnessing genomics for delineating conservation units**. *Trends Ecol Evol* 2012, **27**:489–496.

69. Funk W, Lovich R, Hohenlohe P, Hofman C, Morrison S, Sillett T, Ghalambor C, Maldonado J, Rick T, Day M, Polato N, Fitzpatrick S, Coonan T, Crooks K, Dillon A, Garcelon D, King J, Boser C, Gould N, Andelt W: **Adaptive divergence despite strong genetic drift: genomic analysis of the evolutionary mechanisms causing genetic differentiation in the island fox (Urocyon littoralis)**. *Mol Ecol* 2016, **25**:2176–2194.

70. Razgour O, Forester B, Taggart JB, Bekaert M, Juste J, Ibáñez C, Puechmaille SJ, Novella-Fernandez R, Alberdi A, Manel S: **Considering adaptive genetic variation in climate change vulnerability assessment reduces species range loss projections**. *Proc Natl Acad Sci U S A* 2019, **116**:10418–10423.

bioRxiv preprint doi: https://doi.org/10.1101/814103; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

71. Natesh M, Taylor RW, Truelove NK, Hadly EA, Palumbi SR, Petrov DA, Ramakrishnan U: **Empowering conservation practice with efficient and economical genotyping from poor quality samples**. *Methods Ecol Evol* 2019, **2019**(August 2018):1–7.

72. Johri S, Solanki J, Cantu VA, Fellows SR, Edwards RA, Moreno I, Vyas A, Dinsdale EA: **"Genome skimming" with the MinION hand-held sequencer identifies CITES-listed shark species in India's exports market**. *Sci Rep* 2019, **9**:1–13.

73. Pomerantz A, Peñafiel N, Arteaga A, Bustamante L, Pichardo F, Coloma LA, Barrio-Amorós CL, Salazar-Valenzuela D, Prost S: **Real-time DNA barcoding in a rainforest using nanopore sequencing: Opportunities for rapid biodiversity assessments and local capacity building**. *Gigascience* 2018, **7**:1–14.

74. IUCN SSC Cat Specialist Group: *Conservation Strategy for the Lion in West and Central Africa*. Gland, Switzerland: IUCN; 2006(February).

bioRxiv preprint doi: https://doi.org/10.1101/814103; this version posted October 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Supplemental Files

Supplemental Information 1. Details on laboratory protocols, sequencing, assembly, SNP calling and quality control

Supplemental Figure 1. Distribution of missing data in 211 lions which were genotyped using 125 nuclear and 14 mitochondrial SNPs.

Supplemental Figure 2. Barplots indicating assignment values for K=2 and K=4 from a STRUCTURE run, including only samples with >75% of the nuclear SNPs successfully called (171 samples and 125 nuclear SNPs).

Supplemental Figure 3. Coverage plots for putative Y-chromosomal scaffolds from one leopard and ten lions. Coverage shown for five scaffolds that had been identified as having an Y-chromosomal origin in cat (*Felis catus*).

Supplemental Figure 4. Read quality derived from the first Illumina run, containing one leopard and two lion samples. Drop in quality scores for (A) Leopard, (B) Benin and (C) Kenya and quality scores after hard clipping of reads after 30 bp for (D) Leopard, (E) Benin and (F) Kenya.

Supplemental Figure 5. GC content distribution for two lion samples showing signs of bacterial contamination. GC content of raw reads of (A) Benin and (D) RSA, (B+E) reads filtered against main contaminants and (C+F) reads after filtering and aligned against the reference genome of the tiger.

Supplemental Tables 1-7 (see captions in Excel)