

# The Evolutionary Dynamics of the Lion *Panthera leo* Revealed by Host and Viral Population Genomics

Agostinho Antunes<sup>1,2\*</sup>, Jennifer L. Troyer<sup>3</sup>, Melody E. Roelke<sup>3</sup>, Jill Pecon-Slaterry<sup>1</sup>, Craig Packer<sup>4</sup>, Christiaan Winterbach<sup>5</sup>, Hanlie Winterbach<sup>5</sup>, Graham Hemson<sup>6</sup>, Laurence Frank<sup>7</sup>, Philip Stander<sup>8a</sup>, Ludwig Siefert<sup>9</sup>, Margaret Driciru<sup>10</sup>, Paul J. Funston<sup>11</sup>, Kathy A. Alexander<sup>12ab</sup>, Katherine C. Prager<sup>13</sup>, Gus Mills<sup>14ac</sup>, David Wildt<sup>15ad</sup>, Mitch Bush<sup>15ad</sup>, Stephen J. O'Brien<sup>1</sup>, Warren E. Johnson<sup>1\*</sup>

**1** Laboratory of Genomic Diversity, National Cancer Institute, Frederick, Maryland, United States of America, **2** CIMAR, Centro Interdisciplinar de Investigação Marinha e Ambiental, Universidade do Porto, Porto, Portugal, **3** Laboratory of Genomic Diversity, SAIC-Frederick, Inc., NCI-Frederick, Frederick, Maryland, United States of America, **4** Department of Ecology, Evolution, and Behavior, University of Minnesota, St. Paul, Minnesota, United States of America, **5** Tau Consultants, Maun, Botswana, **6** Wildlife Conservation Research Unit, Tubney, Oxon, United Kingdom, **7** Laikipia Predator Project, Museum of Vertebrate Zoology, University of California Berkeley, Berkeley, California, United States of America, **8** Ministry of Environment and Tourism, Windhoek, Namibia, **9** Department of Wildlife and Animal Resources Management, Makerere University, Kampala, Uganda, **10** Uganda Wildlife Authority, Kamwokya, Kampala, Uganda, **11** Department of Nature Conservation, Tshwane University of Technology, Pretoria, South Africa, **12** Wildlife Veterinary Unit, Department of Wildlife and National Parks, Kasane, Botswana, **13** Department of Pathology, Microbiology, and Immunology, School of Veterinary Medicine, University of California Davis, Davis, California, United States of America, **14** SANParks, Endangered Wildlife Trust and Mammal Research Institute, University of Pretoria, Skukuza, South Africa, **15** Smithsonian's National Zoological Park, Conservation & Research Center, Front Royal, Virginia, United States of America

## Abstract

The lion *Panthera leo* is one of the world's most charismatic carnivores and is one of Africa's key predators. Here, we used a large dataset from 357 lions comprehending 1.13 megabases of sequence data and genotypes from 22 microsatellite loci to characterize its recent evolutionary history. Patterns of molecular genetic variation in multiple maternal (mtDNA), paternal (Y-chromosome), and biparental nuclear (nDNA) genetic markers were compared with patterns of sequence and subtype variation of the lion feline immunodeficiency virus (FIV<sub>Pl</sub>), a lentivirus analogous to human immunodeficiency virus (HIV). In spite of the ability of lions to disperse long distances, patterns of lion genetic diversity suggest substantial population subdivision (mtDNA  $\Phi_{ST}$  = 0.92; nDNA  $F_{ST}$  = 0.18), and reduced gene flow, which, along with large differences in seroprevalence of six distinct FIV<sub>Pl</sub> subtypes among lion populations, refute the hypothesis that African lions consist of a single panmictic population. Our results suggest that extant lion populations derive from several Pleistocene refugia in East and Southern Africa (~324,000–169,000 years ago), which expanded during the Late Pleistocene (~100,000 years ago) into Central and North Africa and into Asia. During the Pleistocene/Holocene transition (~14,000–7,000 years), another expansion occurred from southern refugia northwards towards East Africa, causing population interbreeding. In particular, lion and FIV<sub>Pl</sub> variation affirms that the large, well-studied lion population occupying the greater Serengeti Ecosystem is derived from three distinct populations that admixed recently.

**Citation:** Antunes A, Troyer JL, Roelke ME, Pecon-Slaterry J, Packer C, et al. (2008) The Evolutionary Dynamics of the Lion *Panthera leo* Revealed by Host and Viral Population Genomics. PLoS Genet 4(11): e1000251. doi:10.1371/journal.pgen.1000251

**Editor:** Arnaud Estoup, INRA, France

**Received:** July 27, 2007; **Accepted:** October 2, 2008; **Published:** November 7, 2008

This is an open-access article distributed under the terms of the Creative Commons Public Domain declaration which stipulates that, once placed in the public domain, this work may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose.

**Funding:** This research has been funded in part by federal funds from the National Cancer Institute, National Institutes of Health (N01-CO-12400), by the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research, by the National Science Foundation (No. 0343960) and by the Portuguese Foundation for Science and Technology (PTDC/BIA-BDE/69144/2006). AA received a grant from the Portuguese Foundation for Science and Technology (SFRH/BPD/5700/2001).

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: aantunes@ncifcrf.gov (AA); johnsonw@ncifcrf.gov (WEJ)

<sup>a</sup> Current address: Predator Conservation Trust, Windhoek, Namibia

<sup>b</sup> Current address: Department of Fisheries and Wildlife Sciences, Virginia Polytechnic Institute and State University, Blacksburg, Virginia, United States of America

<sup>c</sup> Current address: The Tony and Lisette Lewis Foundation, Kgalagadi Cheetah Project, Upington, South Africa

<sup>d</sup> Current address: Khao Kheow Open Zoo, Zoological Park Organization, Sriracha, Chonburi, Thailand

## Introduction

Lion fossils trace to the Late Pliocene in Eastern Africa and the Early Pleistocene in Eastern and Southern Africa coincident with the flourishing of grasslands ~2–1.5 million years ago [1,2]. By Mid Pleistocene (~500,000 years ago), lions occupied Europe and by the Late Pleistocene (~130,000–10,000 years ago) lions had the greatest intercontinental distribution for a large land mammal

(excluding man), ranging from Africa into Eurasia and the Americas [3]. Lions were extirpated from Europe 2,000 years ago and within the last 150 years from the Middle East and North Africa. Today, there are less than 50,000 free-ranging lions [4] that occur only in sub-Saharan Africa and the Gir Forest, India (Figure 1A).

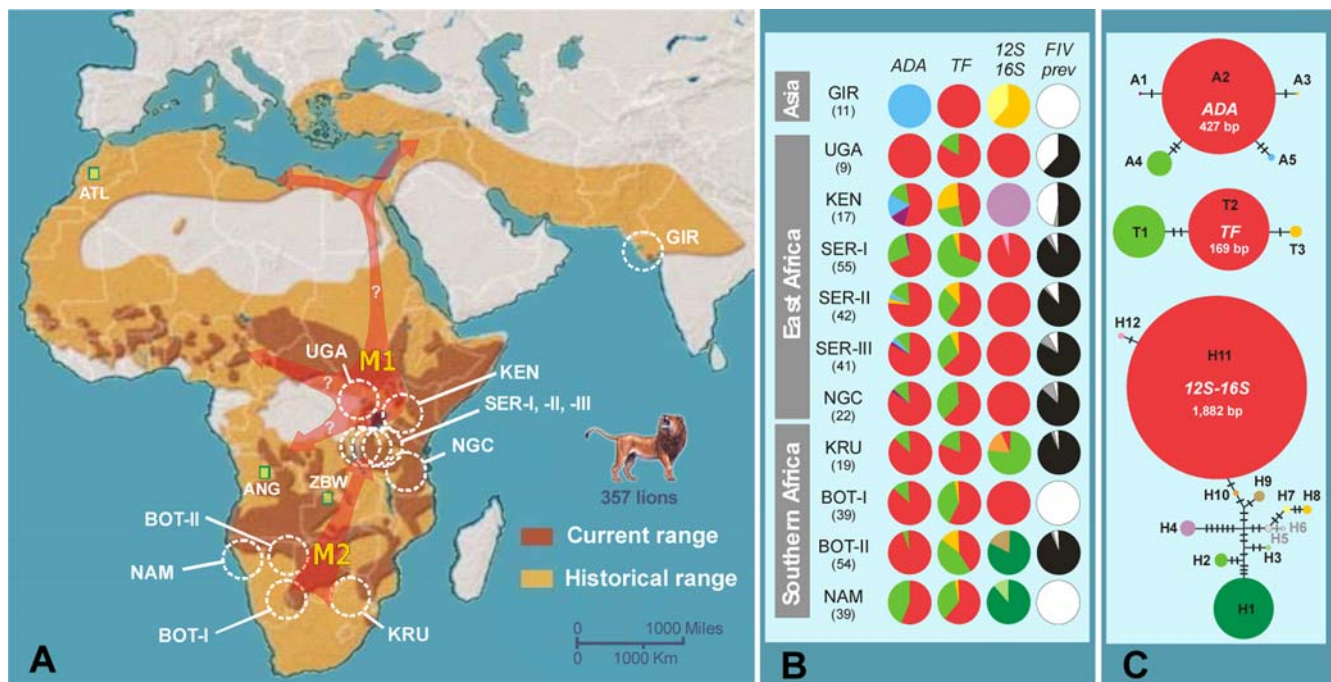
Understanding the broader aspects of lion evolutionary history has been hindered by a lack of comprehensive sampling and

## Author Summary

The lion *Panthera leo*, a formidable carnivore with a complex cooperative social system, has fascinated humanity since pre-historical times, inspiring hundreds of religious and cultural allusions. Here, we use a comprehensive sample of 357 individuals from most of the major lion populations in Africa and Asia. We assayed appropriately informative autosomal, Y-chromosome, and mitochondrial genetic markers, and assessed the prevalence and genetic variation of the lion-specific feline immunodeficiency virus (FIV<sub>Plc</sub>), a lentivirus analogous to human immunodeficiency virus (HIV) that causes AIDS-like immunodeficiency disease in domestic cats. We compare the large multigenic dataset from lions with patterns of genetic variation of the FIV<sub>Plc</sub> to characterize the population-genomic legacy of lions. We refute the hypothesis that African lions consist of a single panmictic population, highlighting the importance of preserving populations in decline rather than prioritizing larger-scale conservation efforts. Interestingly, lion and FIV<sub>Plc</sub> variation revealed evidence of unsuspected genetic diversity even in the well-studied lion population of the Serengeti Ecosystem, which consists of recently admixed animals derived from three distinct genetic groups.

appropriately informative genetic markers [5–9], which in species of modern felids requires large, multigenic data sets due to its generally rapid and very recent speciation [10,11]. Nevertheless, the unique social ecology of lions [12–14] and the fact that lions have experienced well-documented infectious disease outbreaks, including canine distemper virus, feline parvovirus, calicivirus, coronavirus, and lion feline immunodeficiency virus (FIV<sub>Plc</sub>) [15–18] provide a good opportunity to study lion evolutionary history using both host and virus genetic information. Indeed, population genetics of transmitted pathogens can accurately reflect the demographic history of their hosts [19,20]. Unlike other of the 36 cat species, lions have a cooperative social system (prides of 2–18 adult females and 1–9 males) and their populations can have high frequencies of FIV<sub>Plc</sub>, a lentivirus analogous to human immunodeficiency virus (HIV), which causes AIDS-like immunodeficiency disease in domestic cats. FIV<sub>Plc</sub> is a retrovirus that integrates into the host genome and is transmitted by cell-to-cell contact, which in felids occurs during mating, fighting and mother-to-offspring interactions. Thus, viral dissemination is a function in part of the frequency of contact between infected and naïve lions within and among populations. The virus is quite genetically diverse in lions [15,18], offering a unique marker for assessing ongoing lion demographic processes.

To unravel lion population demographic history we used a large multigenic dataset. Distinct sets of markers may not necessarily



**Figure 1. Geographic location of the lion samples and the variability of host and viral genetic markers among lion populations.** (A) Historical and current geographic distribution of lion, *Panthera leo*. A three-letter code pointing to a white dotted circle represents the geographic location of the 11 lion populations determined by Bayesian analyses [22] and factorial correspondence analyses [23] of the genetic distinctiveness of 357 lion samples (see text): GIR, Gir Forest, India; UGA, Uganda (Queen Elizabeth National Park); KEN, Kenya (Laikipia), SER, Serengeti National Park, Tanzania; NGC, Ngorongoro Crater, Tanzania; KRU, Kruger National Park, South Africa; BOT-I, southern Botswana and Kalahari, South Africa; BOT-II, northern Botswana; and NAM, Namibia. Green squares represent captive individual samples to explore the relationship of lions from more isolated/endangered/depleted areas: ATL, Morocco Atlas lions ( $n=4$ ); ANG, Angola ( $n=2$ ); and ZBW, Zimbabwe ( $n=1$ ). Deduced historical expansions (M1 and M2) are represented by red arrows (see text). (B) Haplotype frequencies observed in the 11 lion populations for nDNA (ADA and TF), and mtDNA (12S–16S) genes, paralleled with the FIV<sub>Plc</sub> serum-prevalence frequencies (black – sero-positive; gray – indeterminate; white – sero-negative). Population sample sizes are indicated within parenthesis. (C) Statistical parsimony networks of lion ADA, TF, and 12S–16S haplotypes. Circle size is proportional to the haplotype frequency and crossbars represent the number of step mutations connecting haplotypes. The mtDNA haplotypes H5 and H6 are shaded gray as they were detected only in the individual samples from ANG, ATL, and ZBW, which do not group in unique population clusters (see text).

doi:10.1371/journal.pgen.1000251.g001

yield similar inferences of population history, as coalescent times vary as a function of their pattern of inheritance [21]. There is also a large variance in coalescent times across loci sharing a common pattern of inheritance especially in complex demographical histories (Table 1). However, the accurate interpretation of the differences among loci can provide a more resolved and coherent population history, affording more-nuanced insights on past demographic processes, levels of admixture, taxonomic issues, and on the most appropriate steps for effective conservation and management of remaining populations.

The goal of this study was to assess the evolutionary history of lion by (1) characterizing lion population structure relative to patterns of FIV<sub>Ple</sub> genetic variation, (2) detect signatures of migration using both host and viral population genomics, and (3) reconstruct lion demographic history and discuss its implication for lion conservation. We assess genetic variation from 357 lions from most of its current distribution, including mitochondrial (mtDNA; *12S–16S*, 1,882 bp), nuclear (nDNA) Y-chromosome (*SRY*, 1,322 bp) and autosomal (*ADA*, 427 bp; *TF*, 169 bp) sequences, and 22 microsatellites markers. We further document patterns of FIV<sub>Ple</sub> variation in lions (FIV<sub>Ple</sub> *pol*-RT gene, up to 520 bp).

## Results/Discussion

### Population Structure of Lion

Genetic analyses of 357 lions from throughout the extant species range showed that paternally inherited nDNA (*SRY*) and maternally inherited (mtDNA) sequence variation was generally low (only one paternal *SRY*-haplotype and 12 mtDNA haplotypes;  $\pi = 0.0066$ ) (Figure 1; Figure S1; Tables S1 and S2). The most common mtDNA haplotype H11 was ubiquitous in Uganda/Tanzania and parts of Botswana/South Africa, H1 was common in Southern Africa, and H7 and H8 were unique to Asian lions. The autosomal nDNA sequences showed fairly distinct patterns of variation (Figure 1; Figure S1). Of the five *ADA* haplotypes, A2 was the most common and most-widely distributed. The other four haplotypes, which are derived and much less common, included one (A5) that was fixed in Asian lions. The three *TF* haplotypes were more widely and evenly distributed.

Levels of population subdivision among lions were assessed using microsatellite and sequencing data. Eleven groups were

identified using Bayesian analyses [22] and three-dimensional factorial correspondence analyses [23] (Figure 2; Table S3). Most clusters represented geographically circumscribed populations: Namibia (NAM), Kruger National Park (KRU), Ngorongoro Crater (NGC), Kenya (KEN), Uganda (UGA), and Gir (GIR). Two distinct clusters were found in Botswana, BOT-I that included lions from southern Botswana and Kalahari (South Africa) ( $F_k = 0.24$ ) and BOT-II found exclusively in northern Botswana ( $F_k = 0.18$ ). Surprisingly, three distinct clusters were found in a single geographical locale (approximately 60×40 km square) in the large panmycotic population of the Serengeti National Park (SER-I/SER-II/SER-III) ( $F_k = 0.18, 0.21$ , and  $0.15$ , respectively).

Two captive lions from Angola (ANG), one from Zimbabwe (ZBW) and four Morocco Zoo Atlas lions (ATL; presently extinct from the wild) (Figure 1A) were included in the analyses to explore the relationship of lions from more isolated, endangered, or depleted areas. ANG and ZBW lions were assigned to BOT-II ( $q = 0.90$  and  $0.87$ ; 90% CI:  $0.47–1.00$ ) and KRU ( $q = 0.85$ ; 90% CI:  $0.52–1.00$ ) (Bayesian analyses [22]) populations, respectively, as expected based on their geographical proximity. However, these lions differed from BOT-II and KRU by up to 8 mtDNA mutations, sharing haplotypes with the ATL lions (H5 in ANG and H6 in ZBW) (Figure 1B and 1C). The ATL lions did not group in a unique cluster.

Both nDNA and mtDNA pairwise genetic distances among the 11 lion populations showed a significant relationship with geographic distance ( $R^2 = 0.75$ ; Mantel's test,  $P = 0.0097$ ; and  $R^2 = 0.15$ ; Mantel's test,  $P = 0.0369$ ; respectively) (Figure 3). The significant positive and monotonic correlation across all the scatterplot pairwise comparisons for the nDNA markers (bi-parental) was consistent with isolation-by-distance across the sampled region. However, the correlation between nDNA  $F_{ST}$  and geographic distance considerably decreased when the Asian GIR population was removed ( $R^2 = 0.19$ ; Mantel's test,  $P = 0.0065$ ) suggesting that caution should be taken in interpreting the pattern of isolation-by-distance in lions. We further compared linearized  $F_{ST}$  estimates [24] plotted both against the geographic distance (model assuming habitat to be arrayed in an infinite one-dimensional lattice) and the log geographic distance (model assuming an infinite two-dimensional lattice). The broad distribution of lions might suggest *a priori* that a two-dimensional isolation-by-distance model would provide the best fit for the nDNA data

**Table 1.** Expected and observed coalescent times for the different markers studied in lions according to their pattern of inheritance.

DNA region	Pattern of inheritance	Expected coalescence time in generations	Size (bp)	Number of haplotypes	$F_{ST}$	Divergence Ppa/Ple (H)	Substitution rate (site/yr)	TMRCA (Myr ago)
mtDNA	Maternal haploid	$2N_{eff} \approx N_{ef}$						
<i>12S and 16S</i>			1,882	12	0.92	0.0161	$8.1 \times 10^{-8}$	0.324
Y-Chromosomal DNA	Paternal haploid	$2N_{efm} \approx N_{ef}$						
<i>SRY-3'</i>			1,322	1	—	0.0005	$0.25 \times 10^{-8}$	—
Autosomal DNA	Bisexual diploid	$4N_{ef}$						
<i>ADA</i>			427	5	0.20	0.0023	$1.2 \times 10^{-8}$	4
<i>TF</i>			169*	3	0.09	0.0030*	$1.5 \times 10^{-8}$	4.8
microsatellites			na	na	0.18	na	$2.1 \times 10^{-3}$	0.027

TMRCA - Time to the most recent common ancestral haplotype.

$N_{ef}$  - long-term inbreeding effective size of the population;  $N_{eff}$  - inbreeding effective size of females; and  $N_{efm}$  - inbreeding effective size of males.

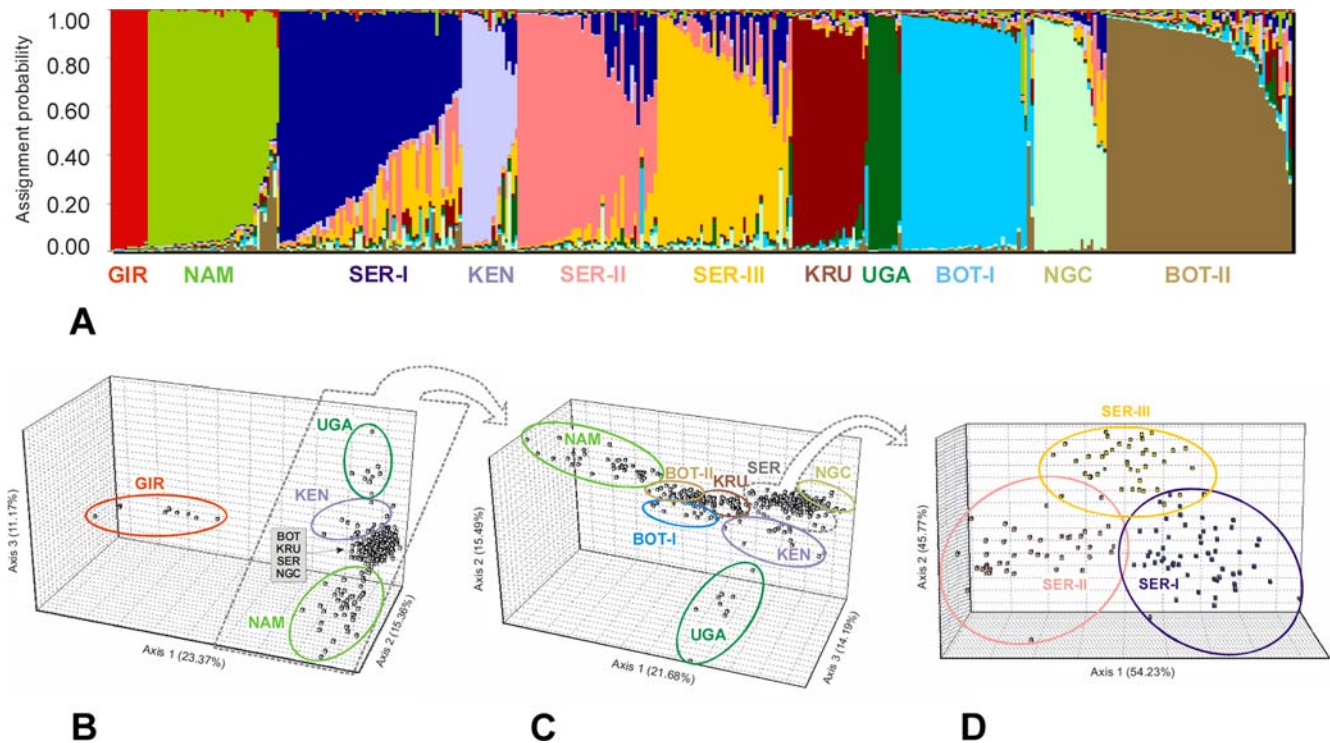
The approximations to the expected coalescence time are made under the assumption that  $N_{eff} = N_{efm} = N_{ef}/2$ .

\*The *TF* gene segment used to estimate the divergence Ppa/Ple was 450 bp long.

na - not applicable.

doi:10.1371/journal.pgen.1000251.t001





**Figure 2. Population structure analyses in lions.** (A) Bayesian population assignment test [22] of the 357 lions using 24 nDNA loci (*ADA*, *TF*, and 22 microsatellites) and mtDNA data, and considering  $K=11$  (11 populations). (B) Three-dimensional factorial correspondence analysis [23] (FCA) based on the 24 nDNA loci genotypes in the 357 lions. Axis 1, 2, and 3 represent 49.90% of the genetic variation observed. (C) FCA representation excluding the GIR lions. Axis 1, 2, and 3 represent 51.35% of the genetic variation observed. (D) FCA representation considering only the SER lions supportive of a three distinct population clusters subdivision (SER-I, SER-II, and SER-III). doi:10.1371/journal.pgen.1000251.g002

( $R^2 = 0.25$ ; Mantel's test,  $P = 0.0022$ ), but instead the one-dimensional isolation-by-distance model performed better ( $R^2 = 0.71$ ; Mantel's test,  $P = 0.0476$ ) (Figure S2).

The pattern observed for the mtDNA (maternal) was more complex. While there was a significant relationship between mtDNA  $F_{ST}$  and geographic distance, there was an inconsistent pattern across broader geographic distances (Figure 3). This is partly due to the fixation or near fixation of haplotype H11 in six populations and the fixation of a very divergent haplotype H4 in KEN population (Figure 1B and 1C). The removal of the KEN population considerably increased the correlation between mtDNA  $F_{ST}$  and geographic distance ( $R^2 = 0.27$ ; Mantel's test,  $P = 0.0035$ ). Thus, the null hypothesis of regional equilibrium for mtDNA across the entire sampled region is rejected despite the possibility that isolation-by-distance may occur regionally.

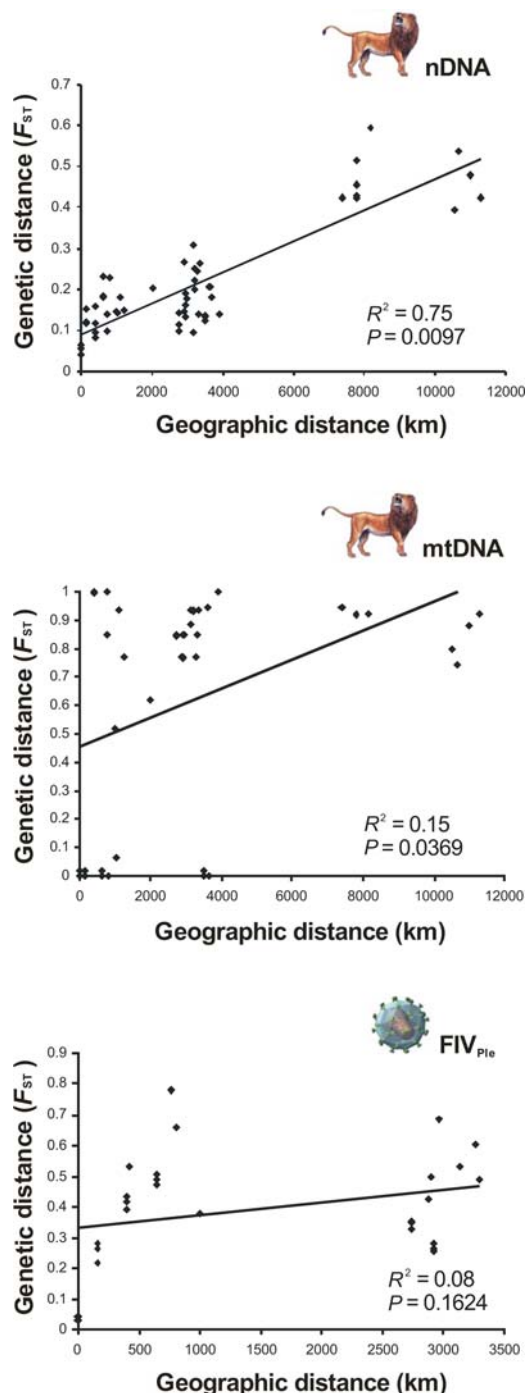
These contrasting nDNA and mtDNA results may be indicative of differences in dispersal patterns between males and females, which would be consistent with evidence that females are more philopatric than males. Alternatively, selection for matrilineally transmitted traits upon which neutral mtDNA alleles hitchhike is possible, given the low values of nucleotide diversity of the mtDNA ( $\pi = 0.0066$ ). A similar process has been suggested in whales ( $\pi = 0.0007$ ) [25] and African savannah elephants ( $\pi = 0.0200$ ) [26], where both species have female philopatry and like lions, a matriarchal social structure. However, genetic drift tends to overwhelm selection in small isolated populations, predominantly affecting haploid elements due to its lower effective population size (Table 1). Therefore, we suggest that the contrasting results obtained for nDNA and mtDNA are more likely further evidence that lion populations underwent severe bottlenecks. The highly

structured lion matrilineal groups comprise four monophyletic mtDNA haplo-groups (Figure 4A; Figure S3). Lineage I consisted of a divergent haplotype H4 from KEN, lineage II was observed in most Southern Africa populations, lineage III was widely distributed from Central and Northern Africa to Asia, and lineage IV occurred in Southern and Eastern Africa.

#### Population Structure of FIV<sub>Ple</sub>

Seroprevalence studies indicate that FIV<sub>Ple</sub> is endemic in eight of the 11 populations but absent from the Asian GIR lions in India and in Namibia and southern Botswana/Kalahari regions (NAM/BOT-I) in Southwest Africa (Figure 1B). Phylogenetic analysis of the conserved *pol*-RT region in 117 FIV-infected lions depicted monophyletic lineages [15,18] that affirm six distinct subtypes (A–F) that are distributed geographically in three distinct patterns (Figure 4B; Figure S4). First, multiple subtypes may circulate within the same population as exemplified by subtypes A, B and C all ubiquitous within the three Serengeti lion populations (SER-I, SER-II and SER-III) and subtypes A and E within lions of Botswana (BOT-II) (Figure 4B and 4C and Figure S4). Second, unique FIV<sub>Ple</sub> subtypes may be restricted to one location as subtype F in Kenya, subtype E in Botswana (BOT-II), subtype C in Serengeti, and subtype D in Krugar Park (Figure 4B and 4C and Figure S4). Third, intra-subtype strains cluster geographically, as shown by distinct clades within subtype A that were restricted to lions within Krugar Park, Botswana and Serengeti and within subtype B that corresponded to Uganda, Serengeti and Ngorongoro Crater lions (Figure 4B and 4C and Figure S4).

Not unexpectedly, FIV<sub>Ple</sub> pairwise genetic distances, represented as population  $F_{ST}$  among the eight lion FIV-infected



**Figure 3. Genetic differentiation of host and viral genetic markers with geographic distance.** Regression of lion pairwise  $F_{ST}$  (nDNA and mtDNA) and FIV<sub>Ple</sub> (pol-RT) on geographic distance. doi:10.1371/journal.pgen.1000251.g003

populations, were not significantly correlated with geographic distance ( $R^2 = 0.08$ ; Mantel's test,  $P = 0.165$ ) (Figure 3), and affirms that patterns of viral dissemination do not conform to a strict isolation-by-distance model. Rather, the two distinct clusters observed (Figure 3) reflect the complex distribution of FIV<sub>Ple</sub> among African lions. Indeed, despite the low geographic distance within East-African lion populations, the FIV<sub>Ple</sub> genetic divergence showed a broader range in  $F_{ST}$  (0.03 to 0.79 for most of first cluster; Figure 3). By contrast, approximately half of the range in

$F_{ST}$  (0.26 to 0.69 for the second cluster; Figure 3) was observed among East and Southern Africa in spite of its large geographic separation. In contrast with the patterns observed in lions, linearized  $F_{ST}$  estimates [24] for FIV<sub>Ple</sub> were better correlated with log geographic distance (two-dimensional lattice model) ( $R^2 = 0.15$ ) than with geographic distance (one-dimensional model) ( $R^2 = 0.02$ ), although in both cases the Mantel's test was not significant ( $P > 0.2474$ ) (Figure S2).

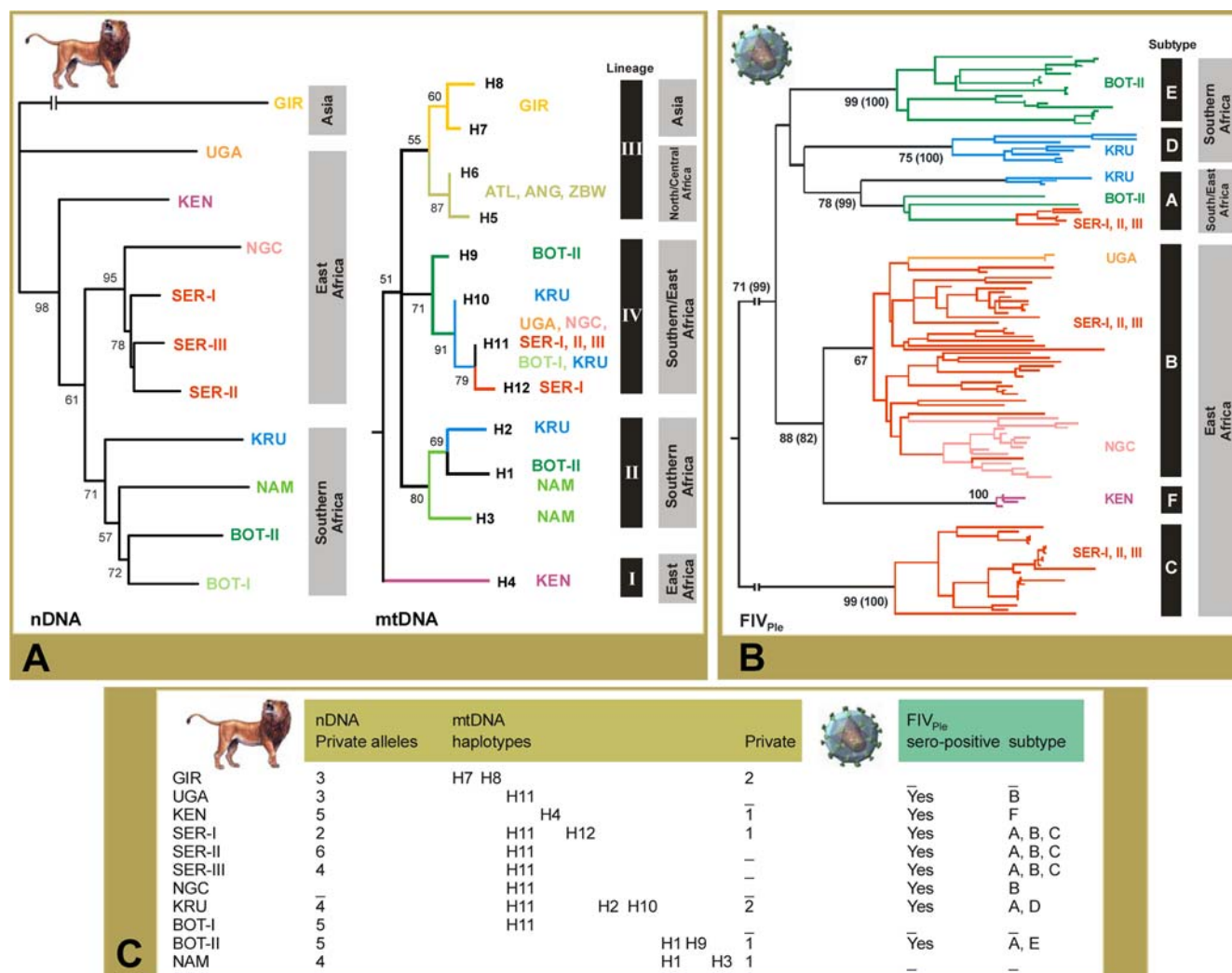
### Natural History of Lions as Inferred from Lion and FIV<sub>Ple</sub> Markers

The mtDNA coalescence dating suggested that the East African lineage I (KEN haplotype H4) had an old origin of ~324,000 years (95% CI: 145,000–502,000). Extant East African populations (KEN/NGC/SER-I/SER-II/SER-III) also showed a slightly significant higher nDNA allelic richness and genetic diversity (Table S4) relative to populations to the south (KRU/NAM/BOT-I/BOT-II) and north (UGA/GIR) ( $A = 2.43$ , 2.39, and 1.62,  $P = 0.021$ ;  $H_O = 0.64$ , 0.62, and 0.34,  $P = 0.019$ ; respectively). Moreover, the FIV<sub>Ple</sub> subtype diversity was higher in East African clades (exhibiting four out of the six known viral-strains), including the most divergent FIV<sub>Ple</sub> subtype C (Figure 4B and 4C). These genetic data from lions and FIV<sub>Ple</sub> is consistent with the older origin of extant East African lions, which is further supported by the oldest lion fossils discovered in East Africa [1].

Relative to East Africa, Southern lions have a slightly more recent mtDNA coalescence. Lineage II, found in NAM, BOT-II and KRU has an estimated coalescence of 169,000 years (95% CI: 34,000–304,000) and the more widespread lineage IV found in the Southern populations of BOT-I, BOT-II and KRU as well as the Eastern populations of SER (I, II, and III), NGC and UGA, coalesces ~101,000 years ago (95% CI: 11,000–191,000). However, the similar levels of nDNA genetic diversity, the occurrence of an exclusively Southern mtDNA lineage II and highly divergent FIV<sub>Ple</sub> subtypes, FIV<sub>Ple</sub> subtype D found only in KRU and subtype E exclusive to BOT-II, suggests that both East and Southern Africa were important refugia for lions during the Pleistocene. Therefore, the co-occurrence of divergent mtDNA haplotypes (6 to 10 mutations; Figure 1B and 1C) in southern populations may be the consequence of further isolation within refugia during colder climatic periods. Contemporary fragmentation of lion populations could further explain the results of nested-clade phylogeographical analysis (NCPA [27]) (Figure S5), which inferred restricted gene flow with isolation-by-distance between mtDNA haplotypes H9 (BOT-II) and H10 (KRU) ( $\chi^2 = 10.00$ ,  $P = 0.0200$ ), between haplotypes H1 (BOT-II/NAM) and H2 (KRU) ( $\chi^2 = 71.00$ ,  $P \leq 0.0001$ ), and between haplotypes H9–H10 (BOT-II/KRU) and haplotypes H11–H12 (BOT-I/KRU/SER/NGC/UGA) ( $\chi^2 = 187.83$ ,  $P \leq 0.0001$ ).

Further isolation within refugia (sub-refugia) may also have occurred in East Africa. This is suggested by the distinctive mtDNA haplotype H4 and the unique FIV<sub>Ple</sub> subtype F found in the Kenya population, which may have resulted from reduced gene flow across the Rift valley, a scenario that has been suggested for several bovid and carnivore populations (see [28] and references therein).

The best example of concordance between host genome markers and viral transmission patterns is observed in the Serengeti National Park in Tanzania. Our previous findings described markedly high levels of FIV<sub>Ple</sub> subtype A, B and C circulating within the Serengeti lion population to such an extent that 43% of the lions sampled were multiply-infected with two or three subtypes [15,18] and were hypothesized to represent recent admixture of three formerly separated populations. Such result is



**Figure 4. Evolutionary relationships of the host and viral genetic markers among lion populations.** (A) Unrooted neighbour-joining (NJ) tree from nDNA genotypes of 24 loci (*ADA*, *TF*, and 22 microsatellites) in the 11 lion populations (left), and rooted NJ tree for the distinct mtDNA (725–16S, 1,882 bp) haplotypes in lion (right). The distinct mtDNA lineages were labelled I to IV. Bootstrap support (BPS) values >50 are indicated. (B) NJ tree of the 301 bp FIV<sub>Ple</sub> *pol*-RT sequences. The distinct FIV<sub>Ple</sub> subtypes were labelled A to F. BPS values are placed at each branchpoint and in parenthesis are the BPS values obtained for a tree established with 520 bp of FIV<sub>Ple</sub> *pol*-RT sequence for a representative subset of individuals. (C) Distinctiveness of host and viral molecular genetics in lion populations. doi:10.1371/journal.pgen.1000251.g004

confirmed here by lion genomic markers (Figure 2). Further, although lions within the Serengeti can be assigned to one of three populations (SER-I, SER-II or SER-III) by host genomic markers, FIV<sub>Ple</sub> subtypes are distributed ubiquitously in all three, characteristic of rapid horizontal retroviral transmission subsequent to host population admixture. The possible isolating mechanism remains to be elucidated as there is no apparent barrier to gene flow in this ecosystem.

### Genomic Signatures Left by Migration

Based on patterns of genetic diversity and phylogenetic analysis of lion nDNA/mtDNA and FIV<sub>Ple</sub> markers, we propose a scenario of a period of refugia/isolation in the Late Pleistocene followed by two major lion expansions across Africa and Asia. The first expansion, supported by the mtDNA NCPA [27] ( $\chi^2 = 690.00$ ,  $P \leq 0.0001$ ; Figure S5), was a long-distance colonization of mtDNA lineage-III (GIR/ATL/ANG/ZBW) around 118,000 years ago (95% CI: 28,000–208,000), with subsequent fragmentation of haplotypes

H5–H6 into Central and North Africa and haplotypes H7–H8 into West Asia (M1- Figure 1A). Support for this initial expansion is also found in nDNA. The *ADA* haplotype A5 fixed in GIR in also present in KEN, SER-II, and SER-III, suggesting that lions likely colonized West Asia from the East Africa refugia (Figure 1B). Such an expansion may have been favored by the start of a warmer and less arid period in Africa 130,000–70,000 years ago [29]. This “out-of-Africa event” would have occurred much later than the initial lion expansion through Eurasia based on fossils (~500,000 years ago) [3]. It is likely that multiple lion expansions occurred in the Pleistocene, as occurred with humans [21].

A second, more recent lion expansion probably occurred at the Pleistocene/Holocene transition, this one from Southern Africa toward East Africa (M2- Figure 1A, Figure 3). This is reflected in the mtDNA lineage IV, where haplotypes present in Southern lions are basal (older) to those found in the East. Overall, mtDNA population nucleotide diversity decreases from Southern to East Africa (Figure 1B and 1C), a finding supported by pairwise

mismatch analysis [30] (raggedness,  $r = 0.086$ ;  $P < 0.001$ ). The fixation of mtDNA haplotype H11 in BOT-I (otherwise fixed only in East Africa populations) suggests that the colonizing lions expanded northwards from the Kalahari Desert, which included bush, woodland and savannah habitats during the climatic fluctuations of the Pleistocene [31]. This expansion would have occurred relatively recently as the single rare tip mtDNA haplotype H12, found only in SER-I, is derived from the interior widespread haplotype H11 (~14,000–7,000 years; given one mtDNA substitution every 7,000 years; Table 1). This expansion is also supported by FIV<sub>Ple</sub> subtype A where haplotypes present in Southern lions (KRU and BOT-I) are basal to those found in the East (SER-I, SER-II and SER-III) and a decrease of nucleotide-diversity of this FIV<sub>Ple</sub> subtype is observed from Southern ( $\pi = 0.15$ ) to Eastern Africa ( $\pi = 0.03$ ) (Figure 3B). Interestingly, a similar northward colonization process from Southern Africa has been suggested for some of the lion preys, namely the impala, greater kudu, and wildebeest [32,33].

### Utility of Population Genomic Datasets

If we had restricted our inferences to mtDNA, we might have concluded that East African lion populations, which are fixed or nearly fixed for haplotype H11, went extinct during the Pleistocene/Holocene transition (similar to the well known mega-fauna extinctions of the Late Pleistocene [34]) and were then colonized by Southern populations. However, our population genomics data better fit a scenario of lion population expansion and interbreeding rather than simple replacement. First, genetic diversity and allelic richness at nDNA are slightly higher in East Africa populations relatively to those in Southern Africa. This is contrary to the expected pattern of population expansion in which there is usually a progressive decline in genetic diversity and allelic richness. Second, SER lions carry two diverse FIV<sub>Ple</sub> subtypes found only in East Africa (B–C), and not only FIV<sub>Ple</sub> subtype A, which was presumably introduced in East Africa coincidentally with the mtDNA expansion event northwards from South. Third, the East African FIV<sub>Ple</sub> subtype B found in UGA/SER-I/SER-II/SER-III/NGC showed evidence of a population expansion (raggedness,  $r = 0.004$ ;  $P < 0.01$ ;  $F_s = -20.37$ ;  $P < 0.00001$ ) and the highest nucleotide diversity observed within FIV<sub>Ple</sub> subtypes ( $\pi = 0.09$ ). Four, the FIV<sub>Ple</sub> subtype diversity is higher in East African clades (four out of the six viral strains).

The utility of FIV<sub>Ple</sub> *pol*-RT as a marker of lion population structure and natural history is that it can be informative on a contemporaneous time scale, though it may be less useful at capturing more ancient demographic events. The extreme divergence among FIV<sub>Ple</sub> subtypes, considered with high sero-prevalence in eight of the 11 lion populations, and combined with patterns of geographic concordance, support the hypothesis that FIV<sub>Ple</sub> is not a recent emergence within modern lions [35]. Populations that harbor one private FIV<sub>Ple</sub> subtype such as KEN (subtype F), BOT-II (subtype E), and KRU (subtype D) must have been sufficiently isolated for enough time for the virus to evolve into unique subtypes, a result corroborated by the high nDNA and mtDNA genetic structure present in these lion populations (Figure 4). Thus, it is possible that the initial emergence of FIV<sub>Ple</sub> pre-dates the Late-Pleistocene expansions of contemporary lion populations [36], but present day distributions are more useful indicators of very recent host population dynamics, a result also observed with FIV<sub>Pco</sub> in a panmictic population of pumas in western North America [19].

### Conservation Implications

Accurate interpretation of past and contemporary population demographic scenarios is a primary goal for the effective

conservation of endangered species. In this study, we found substantial population subdivision, reduced gene flow, and large differences in FIV<sub>Ple</sub> sequence and sero-prevalence among lion populations, as well as evidence of historic secondary contact between populations (Figure 3C; Table S4 to S9). The very low population level of mtDNA nucleotide diversity, the number of haplotypes private to a single population (Figure 1), and probably also the lack of *SRY* genetic variation across all male lions (haplotype S1,  $n = 183$ ) suggests that lion numbers diminished considerably following the Late Pleistocene. The last century reduction in lion distribution further eroded its genomic diversity, and microsatellite variation suggested recent population bottlenecks in seven out of the 11 populations (standardized differences test,  $P < 0.05$ ; Table S5) [37].

Although we did not explicitly try to address the adequacy of lion subspecies designations (currently only one African subspecies is widely recognized) [38,39], we provided strong evidence that there is no evidence of substantial genetic exchange of matrilineal among existing populations as the AMOVA [40] within-population component was uniformly high in all distinct subdivision scenarios ( $\Phi_{ST} \approx 0.920$ ;  $P < 0.0001$ ; three-six groups; Table S6). Similarly, significant population structure was detected from nDNA ( $F_{ST} = 0.18$ ), with low levels of admixture evident from Bayesian analysis [22] ( $\alpha = 0.033$ ). Therefore, employing a bottom-up perspective that prioritizes populations, rather than large-scale units (e.g. all African lions), might preserve and maintain lion diversity and evolutionary processes most efficiently [41].

## Material and Methods

### Study Site, Sampling, and Molecular Genetic Analyses of Lions

A total of 357 individuals were obtained across most of the lion range in Africa and Asia (Figure 1A; Table S1). Genetic variation among lion specimens was assessed using maternal (*12S* and *16S* genes), paternal (*SRY* gene) and bi-parental autosomal (22 microsatellite loci, and the *ADA* and *TF* genes) markers (GenBank accession numbers: FJ151632–FJ151652). Analyses of mtDNA in *Panthera* species are complicated by the presence of a 12.5 kb mtDNA integration into chromosome F2 [42]. Accordingly, mtDNA specific primers were designed for the *12S* and *16S* genes (Table S2) and we used long-range PCR amplification. We designed primers to amplify segments of the *ADA* (exon 10 and intron 10) and the *TF* (intron 3) genes (Table S2), two of the most variable protein loci in lion populations [5], localized on the domestic cat *Felis catus* chromosome A3p and C2q, respectively. The Y-chromosome *SRY*-3'UTR gene was also amplified [43].

PCR products were amplified from 50 ng of genomic DNA in a 25  $\mu$ L reaction system containing 1.5 mM  $MgCl_2$ , 1.0 mM dNTPs, 0.25 units of AmpliTaq Gold DNA polymerase (Applied Biosystems), and 1 $\times$  PCR buffer II; the amplification protocol was: denaturation 10 min at 95°C, a touch-down cycle of 95°C for 30 s, 52°C for 60 s decreased by 1°C in the next cycle for 10 cycles, 72°C for 120 s, then 35 amplification cycles of 95°C for 30 s, 52°C for 60 s, and 72°C for 120 s, followed by an extension of 10 min at 72°C. PCR products were sequenced on an ABI 377. Sequences were aligned and cleaned using SEQUENCHER (Gene Codes).

Twenty two polymorphic microsatellite loci (20 dinucleotide repeats: FCA006, FCA008, FCA014, FCA069, FCA077, FCA085, FCA091, FCA098, FCA105, FCA126, FCA129, FCA139, FCA205, FCA208, FCA211, FCA224, FCA229, FCA230, FCA247, and FCA281; and two tetranucleotide repeats:



FCA391 and FCA441) were amplified [44]. Microsatellites were scored in an ABI 377 and analyzed using GENESCAN 2.1 and GENOTYPER 2.5. These loci are located on 11 of the 19 *F. catus* chromosomes, occurring in different linkage groups or at least 12 centimorgans apart [44,45].

### Sero-Prevalence and Molecular Genetics of FIV<sub>Ple</sub>

Western blots using domestic cat and lion FIV as antigen were performed as previously described [46,47]. The supernatant from virus-infected cells was centrifuged at 200 g for 10 min at 5°C. The resultant supernatant was centrifuged at 150,000 g at 4°C for 2 hours. Pelleted viral proteins were resuspended in 1/20<sup>th</sup> of the original volume and total protein content was assayed using the Biorad Protein Assay. Twenty mg of viral protein were run on 4–20% Tris-Glycine gels and transferred to PDVF membranes (BioRad). Membrane strips were exposed 2–12 h to a 1:25 or 1:200 dilution of serum or plasma. After washing, samples were labeled with goat anti-cat HRP or phosphate conjugated antibody (KPL laboratories) at a 1:2000 dilution, washed, and incubated in ECL Western Blotting detection reagents (Amersham Biosciences) for 2 min, then exposed to Lumi-Film Chemiluminescent Detection Film (Boehringer Mannheim) or incubated in BCIP/NBP phosphatase substrate (KPL laboratories) for 15 min [46–48]. Results were visualized and scored manually based on the presence and intensity of antibody binding to the *p24 gag* capsid protein.

Nested PCR amplification of partial FIV<sub>Ple</sub> *pol*-RT was performed [18,46]. Briefly, first round PCR reactions used 100 ng of genomic DNA, 2.5 mM MgCl<sub>2</sub> and an annealing temperature of 52°C. Second round PCR reactions used identical conditions with 1–5 µl of first-round product as template. All PCR products were sequenced as described above for lion genetic analyses (GenBank accession numbers: AY549217–AY552683; AY878208–AY878235; FJ225347–FJ225382).

### Statistical Analyses

We used the GENETIX 4.02 [49], GENEPOP 3.3 [50], MICROSAT [51], and DNASP 4.10 [52] to calculate the following descriptive statistics: (i) percentage of polymorphic loci ( $P_{95}$ ), number of alleles per locus ( $A$ ), observed and expected heterozygosity ( $H_E$  and  $H_O$ ), and number of unique alleles ( $A_U$ ); (ii) assess deviations from HWE; (iii) estimate the coefficient of differentiation ( $F_{ST}$ ), and (iv) nucleotide ( $\pi$ ) and haplotype ( $h$ ) diversity.

We tested the hypothesis that all loci are evolving under neutrality for both the lion and the FIV<sub>Ple</sub> data. For frequency data, we used the method described by Beaumont and Nichols [53] and implemented in FDist (<http://www.rubic.rdg.ac.uk/~mab/software.html>). The  $F_{ST}$  values estimated from microsatellite loci plotted against heterozygosity showed that all values fall within the expected 95% confidence limit and consequently no outlier locus were identified. For sequence data (lion nDNA/mtDNA and FIV<sub>Ple</sub> *pol*-RT), we ruled out any significant evidence for genetic hitchhiking and background selection by assessing Fu and Li's  $D^*$  and  $F^*$  tests [54] and Fu's  $F_S$  statistics [55].

A Bayesian clustering method implemented in the program STRUCTURE [22] was used to infer number of populations and assign individual lions to populations based on multilocus genotype (microsatellites) and sequence data (*ADA*, *TF*, and mtDNA genes) and without incorporating sample origin. For haploid mtDNA data, each observed haplotype was coded with a unique integer (e.g. 100, 110) for the first allele and missing data for the second (STRUCTURE [22] analyses with or without the mtDNA data were essentially identical). For  $K$  population clusters, the program estimates the probability of the data,  $\Pr(X|K)$ , and the probability

of individual membership in each cluster using a Markov chain Monte Carlo method under the assumption of Hardy-Weinberg equilibrium (HWE) within each cluster. Initial testing of the HWE in each of the populations defined by the geographic origin of sampling revealed no significant deviation from HW expectations with the exception of SER and BOT population (later subdivided by STRUCTURE [22] in 3 and 2 clusters, respectively; such deviations from HW expectations were interpreted as evidence of further population structuring). We conducted six independent runs with  $K=1-20$  to guide an empirical estimate of the number of identifiable populations, assuming an admixture model with correlated allele frequencies and with burn-in and replication values set at 30,000 and 10<sup>6</sup>, respectively. STRUCTURE also estimates allele frequencies of a hypothetical ancestral population and an alpha value that measures admixed individuals in the data set. The assignment of admixed individuals to populations using STRUCTURE [22] has been considered in subsequent population analyses. For each population cluster  $k$ , the program estimates  $F_k$ , a quantity analogous to Wright's  $F_{ST}$ , but describing the degree of genetic differentiation of population  $k$  from the ancestral population.

Patterns of gene flow and divergence among populations were described using a variety of tests. First, to visualize subtle relationships among individual autosomal genotypes, three-dimensional factorial correspondence analyses [23] (FCA) were performed in GENETIX [49], which graphically projects the individuals on the factor space defined by the similarity of their allelic states. Second, neighbor-joining (NJ) analyses implementing the Cavalli-Sforza & Edwards' chord genetic distance [56] ( $D_{CE}$ ) were estimated in PHYLIP 3.6 [57], and the tree topology support was assessed by 100 bootstraps. Third, the difference in average  $H_O$  and  $A$  was compared among population groups using a two-sided test in FSTAT 2.9.3.2 [58], which allows to assess the significance of the statistic  $OS_x$  using 1,000 randomizations. Four, the equilibrium between drift and gene flow was tested using a regression of pairwise  $F_{ST}$  on geographic distance matrix among all populations for host nDNA(microsatellites)/mtDNA and FIV<sub>Ple</sub> data. A Mantel test [59] was used to estimate the 95% upper probability for each matrix correlation. Assuming a stepping stone model of migration where gene flow is more likely between adjacent populations, one can reject the null hypothesis that populations in a region are at equilibrium if (1) there is a non-significant association between genetic and geographic distances, and/or (2) a scatterplot of the genetic and geographic distances fails to reveal a positive and monotonic relationship over all distance values of a region [60]. We also evaluated linearized  $F_{ST}$  [i.e.  $F_{ST}/(1-F_{ST})$ ] [24] among populations. We tested two competing models of isolation-by-distance, one assuming the habitat to be arrayed in an infinite one-dimensional lattice and another assuming an infinite two-dimensional lattice. Both models showed that genetic differentiation increased with raw and log-transformed Euclidean distances, respectively [24]. We determined the confidence interval value of the slope of the regression for the nDNA data using a non parametric ABC bootstrap [61] in GENEPOP 4.0 [62].

The demographic history of populations was compared using a variety of estimators based on the coalescence theory. First, signatures of old demographic population expansion were investigated for mtDNA and FIV<sub>Ple</sub> *pol*-RT haplotypes using pairwise mismatch distributions [63] in DNASP [52]. The goodness-of-fit of the observed data to a simulated model of expansion was tested with the raggedness ( $r$ ) index [64].

Second, the occurrence of recent bottlenecks was evaluated for microsatellite data using the method of Cornuet & Luikart [37] in



BOTTLENECK [65] and using 10,000 iterations. This approach, which exploits the fact that rare alleles are generally lost first through genetic drift after reduction in population size, employs the standardize differences test, which is the most appropriate and powerful when using 20 or more polymorphic loci [37]. Tests were carried out using the stepwise mutation model (SMM), which is a conservative mutation model for the detection of bottleneck signatures with microsatellites [66].

Third, to discriminate between recurrent gene flow and historical events we used the nested-clade phylogeographical analysis [27,67] (NCPA) for the mtDNA data. When the null-hypothesis of no correlation between genealogy and geography is rejected, biological inferences are drawn using a priori criteria. The NCPA started with the estimation of a 95% statistical parsimony [68] mtDNA network in TCS 1.20 [69]. Tree ambiguities were further resolved using a coalescence criteria [70]. The network was converted into a series of nested branches (clades) [71,72], which were then tested against their geographical locations through a permutational contingency analysis in GEODis 2.2 [73]. The inferences obtained were also corroborated with the automated implementation of the NCPA in ANECA [74]. To address potential weaknesses in some aspects of the NCPA analysis [75,76], we further validated the NCPA inferences with independent methods for detecting restricted gene-flow/isolation-by-distance (using matrix correlation of pairwise  $F_{ST}$  and geographic distance) and population expansion (using pairwise mismatch distributions).

Four, to test the significance of the total mtDNA genetic variance, we conducted hierarchical analyses of molecular variance [40] (AMOVA) using ARLEQUIN 2.0 [77]. Total genetic variation was partitioned to compare the amount of difference among population groups, among populations within each group, and within populations.

Phylogenetic relationships among mtDNA and FIV<sub>Ple</sub> *pol*-RT sequences were assessed using Minimum evolution (ME), Maximum parsimony (MP), and Maximum likelihood (ML) approaches implemented in PAUP [78]. The ME analysis for mtDNA consisted of NJ trees constructed from Kimura two-parameter distances followed by a branch-swapping procedure and for FIV<sub>Ple</sub> data employed the same parameter estimates as were used in the ML analysis. The MP analysis was conducted using a heuristic search, with random additions of taxa and tree-bisection-reconnection branch swapping. The ML analysis was done after selecting the best evolutionary model fitting the data using MODELTEST 3.7 [79]. Tree topologies reliability was assessed by 100 bootstraps. For the FIV<sub>Ple</sub> data, the reliability of the tree topology was further assessed through additional analyses using 520 bp of FIV<sub>Ple</sub> *pol*-RT sequences in a representative subset of individuals.

The time to the most recent common ancestor (TMRCA) for the *ADA* and *TF* haplotypes was estimated following Takahata et al. [80], where we calculate the ratio of the average nucleotide differences within the lion sample to one-half the average nucleotide difference between leopards (*P. pardus*) and lions and multiplying the ratio by an estimate of the divergence time between lions and leopards (2 million years based on undisputed lion fossils in Africa) [81,82]. The mtDNA TMRCA was estimated with a linearized tree method in LINTREE [83] and using the equation  $H = 2\mu T$ , where  $H$  was the branch height (correlated to the average pairwise distance among haplotypes),  $\mu$  the substitution rate, and  $T$  the divergence time. Leopard and snow leopard (*P. uncia*) sequences were used as outgroups. Inference of the TMRCA for microsatellite loci followed Driscoll et al. [6] where the estimate of microsatellite variance in average allele repeat-size was used as a surrogate for evolutionary time based on the rate of

allele range reconstitution subsequent to a severe founder effect. Microsatellite allele variance has been shown to be a reliable estimator for microsatellite evolution and demographic inference in felid species [6].

## Supporting Information

**Figure S1** Genetic variation of *12S–16S* (mtDNA) and *ADA* and *TF* (nDNA) genes in lions. (A) Haplotypes and variable sites for the *12S–16S* mtDNA region surveyed in lions (total length 1,882 bp). Position 1 corresponds to position 1441 of the domestic cat (*Felis catus*) mtDNA genome (GenBank U20753). The “-” represents a gap and “.” matches the nucleotide in the first sequence. Shading indicates a fixed difference in the mtDNA lineage. (B) Haplotypes and variable sites for the *ADA* gene segment surveyed in lions (total length 427 bp). (C) Haplotypes and variable sites for the *TF* gene segment surveyed in lions (total length 427 bp).  
Found at: doi:10.1371/journal.pgen.1000251.s001 (0.12 MB PDF)

**Figure S2** Linearized genetic differentiation of host and viral genetic markers with geographic distance. Regression of linearized  $F_{ST}$  estimates [24] for lion (nDNA and mtDNA) and FIV<sub>Ple</sub> (*pol*-RT) genetic data plotted both against the geographic distance (model assuming habitat to be arrayed in an infinite one-dimensional lattice; one-dimension isolation-by-distance [IBD]) and the log geographic distance (model assuming an infinite two-dimensional lattice; two-dimension isolation-by-distance) on geographic distance.  
Found at: doi:10.1371/journal.pgen.1000251.s002 (0.13 MB PDF)

**Figure S3** Phylogenetic relationships of the *12S–16S* mtDNA lion haplotypes. Neighbour-joining tree of the 1,882 bp *12S–16S* mtDNA sequences. Bootstrap values are placed at each branchpoint for the minimum evolution/maximum parsimony/maximum likelihood analyses, respectively (ME/MP/ML). Outgroups: Ppa – leopard, *Panthera pardus*; Pun – snow-leopard, *Panthera uncia*. The symbol (•) represents nodes with bootstrap support <50 or an inferred polytomy in the bootstrap 50% majority-rule consensus tree.  
Found at: doi:10.1371/journal.pgen.1000251.s003 (0.10 MB PDF)

**Figure S4** Phylogenetic relationships of the FIV<sub>Ple</sub> *pol*-RT sequences. Neighbour-joining tree of the 301 bp FIV<sub>Ple</sub> *pol*-RT sequences. The distinct FIV<sub>Ple</sub> subtypes were labelled A to F. Bootstrap (BPS) values are placed at each branchpoint (ME/MP/ML) and in parenthesis are the BPS values obtained for a tree established with 520 bp of FIV<sub>Ple</sub> *pol*-RT sequence for a representative subset of individuals.  
Found at: doi:10.1371/journal.pgen.1000251.s004 (0.13 MB PDF)

**Figure S5** Nested design and summary results of the nested clade phylogeographic analysis (NCPA) for lion mtDNA data. (A) Nested design of the mtDNA haplotype network used for the NCPA. (B) Summary results of the NCPA. RGF/IBD - Restricted gene flow/isolation by distance. LDC/FR – long distance colonization/fragmentation.  
Found at: doi:10.1371/journal.pgen.1000251.s005 (0.10 MB PDF)

**Table S1** List of the lion samples used in this study.  
Found at: doi:10.1371/journal.pgen.1000251.s006 (0.11 MB PDF)

**Table S2** Primers used to amplify the mtDNA (*12S–16S*) and nDNA (*ADA* and *TF*) portions surveyed in this study.  
Found at: doi:10.1371/journal.pgen.1000251.s007 (0.06 MB PDF)

**Table S3** Structure cluster assignment results of 357 lions based on nDNA (*ADA*, *TF*, and 22 microsatellites) and mtDNA markers. Burn-in and replication values set at 30,000 and 1,000,000, respectively.  
Found at: doi:10.1371/journal.pgen.1000251.s008 (0.06 MB PDF)

**Table S4** Gene diversity and frequency values in lion populations.

Found at: doi:10.1371/journal.pgen.1000251.s009 (0.07 MB PDF)

**Table S5** Bottleneck analysis in lion populations using the standardized differences test and the stepwise mutation model (SMM).

Found at: doi:10.1371/journal.pgen.1000251.s010 (0.05 MB PDF)

**Table S6** Results of the hierarchical AMOVA in lions for four different geographical scenarios.

Found at: doi:10.1371/journal.pgen.1000251.s011 (0.06 MB PDF)

**Table S7** Lion population pairwise  $F_{ST}$  estimates. Below the diagonal mtDNA data (*I2S–I6S*) and above the diagonal microsatellite data (22 loci).

Found at: doi:10.1371/journal.pgen.1000251.s012 (0.06 MB PDF)

**Table S8** Taxon specific unique nDNA alleles in lion populations (FCA-microsatellites and ADA locus).

Found at: doi:10.1371/journal.pgen.1000251.s013 (0.06 MB PDF)

## References

- Werdelin L, Lewis ME (2005) Plio-Pleistocene Carnivora of eastern Africa: species richness and turnover patterns. *Zool J Linn Soc* 144: 121–144.
- Petter G (1973) Carnivores Pléistocènes du Ravin d'Olduvai (Tanzanie). In Leakey LSB, Savage RJG, Coryndon SC, eds. *Fossil vertebrates of Africa*. London: Academic Press 3: 43–100.
- Hemmer H (1974) Untersuchungen zur Stammesgeschichte der Pantherkatzen (Pantherinae) Teil 3. Zur Artgeschichte de Löwen *Panthera (Panthera) leo* (Linnaeus, 1758). *Veröffentlichungen der Zoologischen Staatssammlung* 17: 167–280.
- Bauer H, Nowell K (2006) *Panthera leo*. In: IUCN 2006, 2006 IUCN Red List of Threatened Species (<http://www.iucnredlist.org/search/details.php/15951/all>).
- O'Brien SJ, Martenson JS, Packer C, Herbst L, De Vos V, Joslin P, Ott-Joslin J, Wildt DE, Bush M (1987) Biochemical genetic variation in geographic isolates of African and Asiatic lions. *National Geographic Research* 3: 114–124.
- Driscoll CA, Menotti-Raymond M, Nelson G, Goldstein D, O'Brien SJ (2002) Genomic microsatellites as evolutionary chronometers: a test in wild cats. *Genome Res* 12: 414–423.
- Burger J, Rosendahl W, Loreille O, Hemmer H, Eriksson T, et al. (2004) Molecular phylogeny of the extinct cave lion *Panthera leo spelaea*. *Mol Phylogenet Evol* 30: 841–849.
- Dubach J, Patterson B, Briggs M, Venzke K, Flamand J, et al. (2005) Molecular genetic variation across the southern and eastern geographic ranges of the African lion, *Panthera leo*. *Cons Genet* 6: 15–24.
- Barnett R, Yamaguchi N, Barnes I, Cooper A (2006) Lost populations and preserving genetic diversity in the lion *Panthera leo*: implications for its *ex situ* conservation. *Cons Genet* 7: 507–514.
- Luo SJ, Kim JH, Johnson WE, van der Walt J, Martenson J, et al. (2004) Phylogeography and genetic ancestry of tigers (*Panthera tigris*). *PLoS Biol* 2(12): e442.
- Johnson WE, Eizirik E, Pecon-Slatery J, Murphy WJ, Antunes A, et al. (2006) The late Miocene radiation of modern Felidae: a genetic assessment. *Science* 311: 73–77.
- Packer C, Gilbert DA, Pusey AE, O'Brien SJ (1991) A molecular genetic analysis of kinship and cooperation in Africa lions. *Nature* 351: 562–565.
- Packer C, Hilborn R, Mosser A, Kissui B, Borner M, et al. (2005) Ecological change, group territoriality, and population dynamics in Serengeti lions. *Science* 307: 390–393.
- Spong G, Stone J, Creel S, Björklund M (2002) Genetic structure of lions (*Panthera leo* L.) in the Selous Game Reserve: implications for the evolution of sociality. *J Evol Biol* 15: 945–953.
- Brown EW, Yuhki N, Packer C, O'Brien SJ (1994) A lion lentivirus related to feline immunodeficiency virus: epidemiologic and phylogenetic aspects. *J Virol* 68: 5953–5968.
- Roelke-Parker ME, Munson L, Packer C, Kock R, Cleaveland S, et al. (1996) A canine distemper virus epidemic in Serengeti lions (*Panthera leo*). *Nature* 379: 441–445.
- Packer C, Altizer S, Appel M, Brown E, Martenson J, et al. (1999) Viruses of the Serengeti: patterns of infection and mortality in African lions. *J Anim Ecol* 68: 1161–1178.
- Troyer JL, Pecon-Slatery J, Roelke ME, Black L, Packer C, et al. (2004) Patterns of feline immunodeficiency virus multiple infection and genome divergence in a free-ranging population of African lions. *J Virol* 78: 3777–3791.
- Bick R, Drummond AJ, Poss M (2006) A virus reveals population structure and recent demographic history of its carnivore host. *Science* 311: 538–541.
- Falush D, Wirth T, Linz B, Pritchard JK, Stephens M, et al. (2003) Traces of human migrations in *Helicobacter pylori* populations. *Science* 299: 1582–1585.
- Templeton AR (2002) Out of Africa again and again. *Nature* 416: 45–51.
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155: 945–959.
- Benzécri JP (1982) L'analyse des données, vol II. L'analyse des correspondances. Paris: Dunod.
- Rousset F (1997) Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics* 145: 1219–1228.
- Whitehead H (1998) Cultural selection and genetic diversity in matrilineal whales. *Science* 282: 1708–1711.
- Nyakaana S, Arctander P, Siegmund HR (2002) Population structure of the African savannah elephant inferred from mitochondrial control region sequences and nuclear microsatellite loci. *Heredity* 89: 90–98.
- Templeton AR, Routman E, Phillips CA (1995) Separating population structure from population history: a cladistic analysis of the geographical distribution of mitochondrial DNA haplotypes in the tiger salamander, *Ambystoma tigrinum*. *Genetics* 140: 767–782.
- Pitra C, Hansen AJ, Lieckfeldt D, Arctander P (2002) An exceptional case of historical outbreeding in African sable antelope populations. *Mol Ecol* 11: 1197–208.
- van Andel TH, Tzedakis PC (1996) Palaeolithic landscapes of Europe and environs, 150,000–25,000 years ago: an overview. *Quat Sci Rev* 15: 481–500.
- Rogers AR, Harpending H (1992) Population growth makes waves in the distribution of pairwise genetic differences. *Mol Biol Evol* 9: 552–569.
- Partridge T (1993) Warming phases in Southern Africa during the last 150,000 years; an overview. *Palaeogeography, Palaeoclimatology, Palaeoecology* 101: 237–244.
- Arctander P, Johansen C, Coutellec-Vreto MA (1999) Phylogeography of three closely related African bovids (tribe Alcelaphini). *Mol Biol Evol* 16: 1724–39.
- Nersting LG, Arctander P (2001) Phylogeography and conservation of impala and greater kudu. *Mol Ecol* 10: 711–719.
- Barnosky AD, Koch PL, Feranec RS, Wing SL, Shabel AB (2004) Assessing the causes of late Pleistocene extinctions on the continents. *Science* 306: 70–75.
- Carpenter MA, Brown EW, Culver M, Johnson WE, Pecon-Slatery J, et al. (1996) Genetic and phylogenetic divergence of feline immunodeficiency virus in the puma (*Puma concolor*). *J Virol* 70: 6682–6693.
- Pecon-Slatery J, Troyer JL, Johnson WE, O'Brien SJ (2008) Evolution of Feline Immunodeficiency Virus in Felidae: Implications For Human Health And Wildlife Ecology. *Vet Immunol Immunopathol* 123: 32–44.
- Cornuet JM, Luikart G (1996) Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics* 144: 2001–2014.
- Neff N (1983) *The Big Cats: the Paintings of Guy Cheleach*. New York: Abrams.
- Nowak RM, Paradiso JL (1983) *Walker's Mammals of the World*, 4<sup>th</sup> ed. Baltimore: Johns Hopkins University Press.
- Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131: 479–491.
- Antunes A, Faria R, Weiss S, Alexandrino P (2001) Complex evolutionary history in the brown trout: insights on the recognition of conservation units. *Cons Genet* 2: 337–347.
- Kim JH, Antunes A, Luo SJ, Menninger J, Nash WG, et al. (2006) Evolutionary analysis of a large mtDNA translocation (*numt*) into the nuclear genome of the *Panthera* genus species. *Gene* 366: 292–302.
- King V, Goodfellow PN, Pearks Wilkerson AJ, Johnson WE, O'Brien SJ, et al. (2007) Evolution of the male determining gene SRY within the cat family Felidae. *Genetics* 175: 1855–1867.

**Table S9** Summary statistics for FIV<sub>plc</sub> data.

Found at: doi:10.1371/journal.pgen.1000251.s014 (0.00 MB PDF)

## Acknowledgments

Tissues were obtained in full compliance with specific Federal Fish and Wildlife Permits. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government. We would like to thank A. Beja-Pereira, M. Branco and J. Martenson for suggestions and technical assistance. Comments made by the Associate Editor A. Estoup and three anonymous referees improved a previous version of this manuscript.

## Author Contributions

Conceived and designed the experiments: AA JLT SJO WEJ. Performed the experiments: AA JLT MER WEJ. Analyzed the data: AA JLT MER SJO WEJ. Contributed reagents/materials/analysis tools: AA JLT MER JPS CP CW HW GH LF PS LS MD PJF KAA KCP GM DW MB SJO WEJ. Wrote the paper: AA JLT JPS CP SJO WEJ.

44. Menotti-Raymond M, David VA, Lyons LA, Schaffer AA, Tomlin JF, et al. (1999) A genetic linkage map of microsatellites in the domestic cat (*Felis catus*). *Genomics* 57: 9–23.
45. Menotti-Raymond M, David VA, Roelke ME, Chen ZQ, Menotti KA, et al. (2003) Second-generation integrated genetic linkage/radiation hybrid maps of the domestic cat (*Felis catus*). *J Hered* 94: 95–106.
46. Troyer JL, Pecon-Slattery J, Roelke ME, Johnson W, VandeWoude S, et al. (2005) Seroprevalence and genomic divergence of circulating strains of feline immunodeficiency virus among Felidae and Hyaenidae species. *J Virol* 79: 8282–8294.
47. VandeWoude S, O'Brien SJ, Langelier K, Hardy WD, Pecon-Slattery J, et al. (1997) Growth of lion and puma lentiviruses in domestic cat cells and comparisons with FIV. *Virology* 233: 185–192.
48. Brown EW, Miththapala S, O'Brien SJ (1993) Prevalence of exposure to feline immunodeficiency virus is exotic felid species. *J Zool Wildl Med* 20: 265–272.
49. Belkir K, Borsa P, Goudet J, Chikhi L, Bonhomme F (1999) Genetix, logiciel sous Windows TM pour la génétique des populations. Montpellier (France): Laboratoire Génome et Populations, CNRS UPR 9060, Université de Montpellier II.
50. Raymond M, Rousset F (1995) Population genetics software for exact tests and ecumenicism. *J Hered* 86: 248–249.
51. Minch E, Ruiz-Linares A, Goldstein D, Feldman M, Cavalli-Sforza LL (1995) MICROSAT Version 1.4d: a computer program for calculating various statistics on microsatellite allele data.
52. Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* 19: 2496–2497.
53. Beaumont MA, Nichols RA (1996) Evaluating loci for use in the genetic analysis of population structure. *Proc R Soc Lond B* 263: 1619–1626.
54. Fu Y-X, Li W-H (1993) Statistical tests of neutrality of mutations. *Genetics* 133: 693–709.
55. Fu Y-X (1997) Statistical tests of neutrality against population growth, hitchhiking and background selection. *Genetics* 147: 915–925.
56. Cavalli-Sforza LL, Edwards AWF (1967) Phylogenetic analysis: models and estimation procedures. *Evolution* 32: 550–570.
57. Felsenstein J (1993) PHYLIP (Phylogeny inference package). Version 3.5c. Seattle: Distributed by the author, Department of Genetics, University of Washington.
58. Goudet J (1995) FSTAT. Vers. 1.2. A computer program to calculate *F*-statistics. *J Hered* 86: 485–486.
59. Mantel N (1967) The detection of disease clustering and a generalized regression approach. *Cancer Res* 27: 209–220.
60. Hutchison DW, Templeton AR (1999) Correlation of pairwise genetic and geographic distance measures: inferring the relative influences of gene flow and drift on the distribution of genetic variability. *Evolution* 53: 1898–1914.
61. Leblois R, Estoup A, Rousset F (2003) Influence of mutational and sampling factors on the estimation of demographic parameters in a “continuous” population under isolation by distance. *Mol Biol Evol* 20: 491–502.
62. Rousset F (2008) Genepop'007: a complete re-implementation of the Genepop software for Windows and Linux. *Mol Ecol Res* 8: 103–106.
63. Rogers AR, Harpending H (1992) Population growth makes waves in the distribution of pairwise genetic differences. *Mol Biol Evol* 9: 552–569.
64. Harpending HC (1994) Signature of ancient population growth in a low-resolution mitochondrial DNA mismatch distribution. *Hum Biol* 66: 591–600.
65. Piry S, Luikart G, Cornuet JM (1999) BOTTLENECK: a computer program for detecting recent effective population size reductions from allele frequency data. *J Hered* 90: 502–503.
66. Leblois R, Estoup A, Streiff R (2006) Genetics of recent habitat contraction and reduction in population size: does isolation by distance matter? *Mol Ecol* 15: 3601–3615.
67. Templeton AR (1998) Nested clade analysis of phylogeographic data: Testing hypotheses about gene flow and population history. *Mol Ecol* 7: 381–397.
68. Templeton AR, Crandall KA, Sing CF (1992) A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. *Genetics* 132: 619–633.
69. Clement M, Posada D, Crandall KA (2000) TCS: a computer program to estimate gene genealogies. *Mol Ecol* 9: 1657–1660.
70. Crandall KA, Templeton AR (1993) Empirical tests of some predictions from coalescent theory with applications to intraspecific phylogeny reconstruction. *Genetics* 134: 959–969.
71. Templeton AR, Boerwinkle E, Sing CF (1987) A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping. I. Basic theory and an analysis of alcohol dehydrogenase activity in *Drosophila*. *Genetics* 117: 343–351.
72. Templeton AR, Sing CF (1993) A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping. IV. Nested analyses with cladogram uncertainty and recombination. *Genetics* 134: 659–669.
73. Posada D, Crandall KA, Templeton AR (2000) GeoDis: a program for the cladistic nested analysis of the geographical distribution of genetic haplotypes. *Mol Ecol* 9: 487–488.
74. Panchal M (2007) The automation of Nested Clade Phylogeographic Analysis. *Bioinformatics* 23: 509–10.
75. Knowles LL, Maddison WP (2002) Statistical phylogeography. *Mol Ecol* 11: 2623–2635.
76. Panchal M, Beaumont MA (2007) The automation and evaluation of nested clade phylogeographic analysis. *Evolution* 61: 1466–1480.
77. Schneider S, Kueffer JM, Roessli D, Excoffier L (1997) Arlequin Version 1.1. Switzerland: Genetics and Biometry Laboratory, Dept. of Anthropology, University of Geneva.
78. Swofford DL (2001) PAUP\* Phylogenetic Analysis Using Parsimony and Other Methods' Computer Program. Sunderland, MA: Sinauer.
79. Posada D, Crandall KA (1998) MODELTEST: Testing the model of DNA substitution. *Bioinformatics* 14: 817–818.
80. Takahata N, Lee SH, Satta Y (2001) Testing multiregionality of modern human origins. *Mol Biol Evol* 18: 172–183.
81. Petter G (1973) Carnivores Pléistocènes du Ravin d'Olduvai (Tanzanie). In Leakey LSB, Savage RJG, Coryndon SC, eds. Fossil vertebrates of Africa. London: Academic Press 3: 43–100.
82. Werdelin L, Lewis ME (2005) Plio-Pleistocene Carnivora of eastern Africa: species richness and turnover patterns. *Zool J Linn Soc* 144: 121–144.
83. Takezaki N, Rzhetsky A, Nei M (1995) Phylogenetic test of the molecular clock and linearized trees. *Mol Biol Evol* 12: 823–833.