

RESEARCH ARTICLE

Open Access

# The population genetics of wild chimpanzees in Cameroon and Nigeria suggests a positive role for selection in the evolution of chimpanzee subspecies

Matthew W Mitchell<sup>1,2†</sup>, Sabrina Locatelli<sup>2,3†</sup>, Lora Ghobrial<sup>2,4</sup>, Amy A Pokempner<sup>2,5</sup>, Paul R Sesink Clee<sup>1,2</sup>, Ekwoke E Abwe<sup>6,7</sup>, Aaron Nicholas<sup>8</sup>, Louis Nkembu<sup>9</sup>, Nicola M Anthony<sup>10</sup>, Bethan J Morgan<sup>6,7,11</sup>, Roger Fotso<sup>12</sup>, Martine Peeters<sup>3</sup>, Beatrice H Hahn<sup>13</sup> and Mary Katherine Gonder<sup>1,2\*</sup>

## Abstract

**Background:** Chimpanzees (*Pan troglodytes*) can be divided into four subspecies. Substantial phylogenetic evidence suggests that these subspecies can be grouped into two distinct lineages: a western African group that includes *P. t. verus* and *P. t. ellioti* and a central/eastern African group that includes *P. t. troglodytes* and *P. t. schweinfurthii*. The geographic division of these two lineages occurs in Cameroon, where the ranges of *P. t. ellioti* and *P. t. troglodytes* appear to converge at the Sanaga River. Remarkably, few population genetic studies have included wild chimpanzees from this region.

**Results:** We analyzed microsatellite genotypes of 187 wild, unrelated chimpanzees, and mitochondrial control region sequencing data from 604 chimpanzees. We found that chimpanzees in Cameroon and eastern Nigeria comprise at least two, and likely three populations. Both the mtDNA and microsatellite data suggest that there is a primary separation of *P. t. troglodytes* in southern Cameroon from *P. t. ellioti* north and west of the Sanaga River. These two populations split ~200-250 thousand years ago (kya), but have exchanged one migrant per generation since separating. In addition, *P. t. ellioti* consists of two populations that split from one another ~4 kya. One population is located in the rainforests of western Cameroon and eastern Nigeria, whereas the second population appears to be confined to a savannah-woodland mosaic in central Cameroon.

**Conclusions:** Our findings suggest that there are as many as three genetically distinct populations of chimpanzees in Cameroon and eastern Nigeria. *P. t. troglodytes* in southern Cameroon comprises one population that is separated from two populations of *P. t. ellioti* in western and central Cameroon, respectively. *P. t. ellioti* and *P. t. troglodytes* appear to be characterized by a pattern of isolation-with-migration, and thus, we propose that neutral processes alone can not explain the differentiation of *P. t. ellioti* and *P. t. troglodytes*.

\* Correspondence: gonder@drexel.edu

†Equal contributors

<sup>1</sup>Department of Biology, Drexel University, Philadelphia, Pennsylvania 19104, USA

<sup>2</sup>Department of Biological Sciences, University at Albany - State University of New York, Albany, New York 12222, USA

Full list of author information is available at the end of the article

## Background

Chimpanzees (*Pan troglodytes*) are still widely distributed across sub-Saharan Africa, and the species exploits a wide range of habitats including rainforests, ecotones and savannas [1-3]. Across this range they exhibit considerable genetic [4], behavioral [5] and ecological diversity [3]. Studies consisting mostly of samples from wild-born captive chimpanzees have given great insights regarding the phylogenetic history of this species, but genetic data from wild individuals remain sparse. Recent studies of captive wild-born chimpanzees include datasets such as complete mitochondrial (mt) genomes [6], genome-wide single nucleotide polymorphisms (SNPs) [7], and complete genome sequences [8]. The overall picture to emerge from these studies suggests that chimpanzees are divided into two geographically- and genetically-defined groups: a western African group that includes *P. t. verus* and *P. t. ellioti* and a central/eastern African group that includes *P. t. troglodytes* and *P. t. schweinfurthii* [8] (Figure 1). This classification is consistent with earlier studies using mtDNA control region sequence diversity from wild individuals from Nigeria and adjacent parts of Cameroon, which suggested that animals in this region form a genetically distinct population of chimpanzees [9,10], now widely recognized as *P. t. ellioti* [11]. These studies have also been important for understanding differences in subspecies population histories, and how these histories are connected to landscape and forest history. For example, the western and central/eastern groups appear to have split from one another very early in the history of this species. Analysis of complete genomes suggests that since their genetic and geographic separation, these groups have experienced markedly different demographic histories, and the subspecies within each group show different patterns of population growth and decline throughout their respective histories [8].

While the phylogenetic relationships among chimpanzee subspecies are now well-resolved, their population structure, migration patterns and patterns of population growth still remain unknown, particularly in areas where the ranges of subspecies converge. This is mostly due to a lack of fine-scale geographic sampling of wild chimpanzee populations around areas where subspecies overlap, which has revealed mechanisms of diversification in gorillas [12]. Analysis of mtDNA from wild chimpanzees from Nigeria and Cameroon has suggested that the Sanaga River delimits *P. t. ellioti* from *P. t. troglodytes* in southern Cameroon [9]. This split is old, but incomplete, as there is a zone in central Cameroon where the ranges of the mtDNA haplotypes that define these two subspecies overlap. These findings suggest that some introgression may have occurred between these two subspecies [4,10,13] or that a pattern of

isolation-by-distance (IBD) might be revealed by more fine scale population sampling across this region.

The Sanaga River has been proposed to delimit the distributions of several primate species and subspecies [12,14-18] including chimpanzees [4,19]. This region of Africa also contains diverse habitats. Cameroon lies at the intersection of two major rainforest biomes, the Guinean and Congolian rainforests [20] (Figure 1a). These two forests converge in central Cameroon and are connected by a zone of open woodland, savannah and riparian forest [20,21] which has been termed an 'ecotone' [22]. Therefore, denser geographic sampling and more comprehensive genetic data are necessary to tease apart the relative contributions of forest history and biogeographic boundaries in driving patterns of genetic variation in chimpanzees.

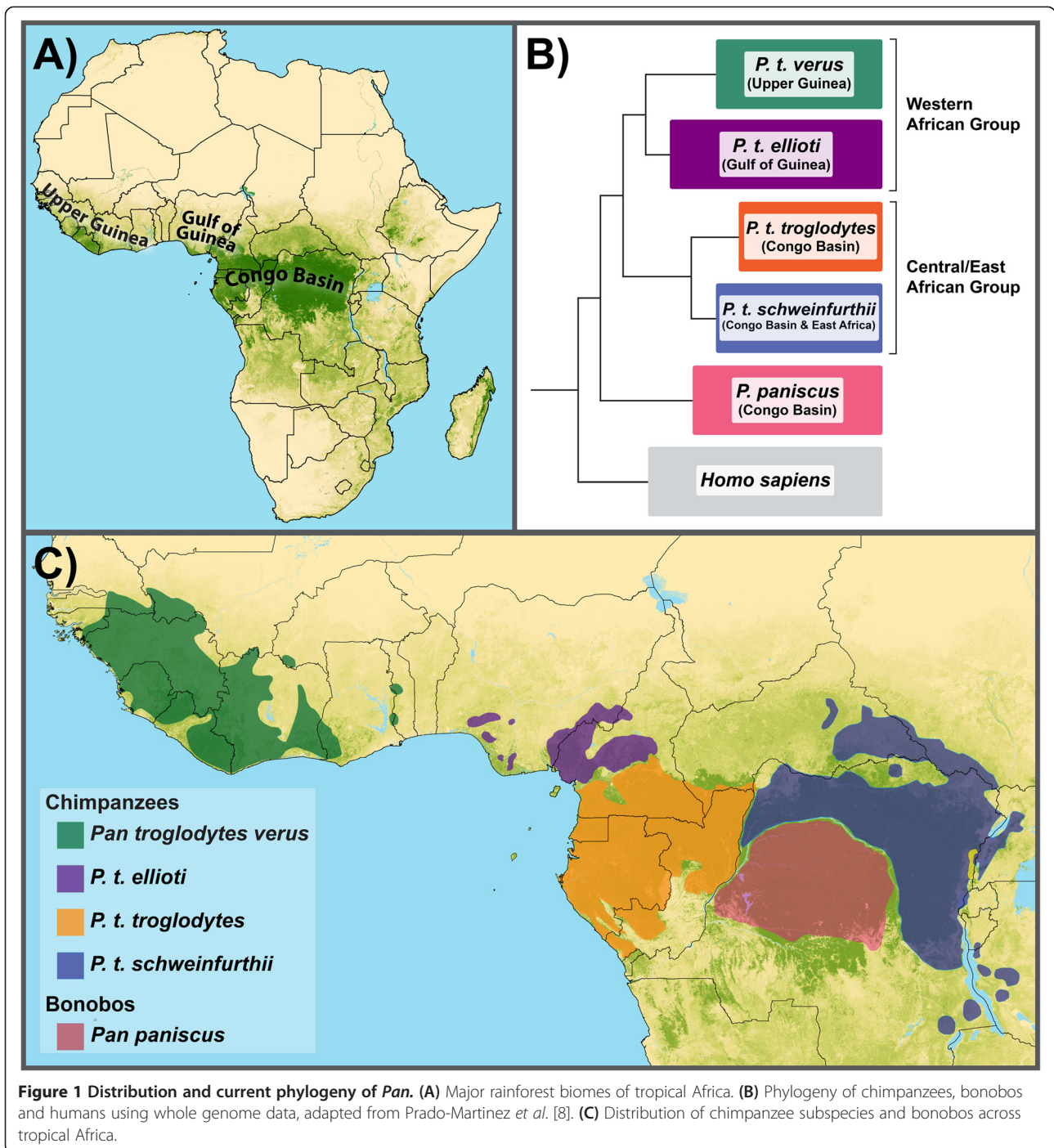
This study uses DNA extracted from fecal and hair samples to examine the population structure and genetic history of chimpanzees from Cameroon and eastern Nigeria at a fine geographic scale across this ecologically diverse region (Figure 2). The overall goals of the study were to use mtDNA haplotypes along with microsatellite genotype profiles of wild individuals to: (1) test between a variety of hypotheses for the presence and type of population structure (Table 1), including (i) panmixia, (ii) isolation-by-distance, (iii) population structure with complete isolation, and (iv) population structure with ongoing migration; (2) test the specific hypothesis that the Sanaga River has been important in delimiting the range of *P. t. ellioti* from neighboring *P. t. troglodytes*; and (3) compare the demographic histories of these two subspecies by inferring the time to their most recent common ancestor ( $T_{MRC A}$ ), their historical and current effective population sizes, and finally, the rates and directions of migration between them.

## Results and discussion

### Dataset preparation

We calculated pairwise estimates of relatedness for all pairs of genotyped individuals using the software program COANCESTRY [23] and removed duplicate individuals and those that were closely related to one another. A total of three individuals with a relatedness index value [24] above 0.75 were determined to be either identical individuals or highly related and were excluded from all further analyses. The analyses presented here include mtDNA sequencing data from 604 sequences from 35 sampling locations, and microsatellite data were included from 187 unrelated individuals from 28 sampling locations (Additional file 1).

In order to ensure that all of the genetic markers were selectively neutral, and therefore suitable for further analysis, the data were subjected to various neutrality tests. We found that all microsatellite loci met expectations of

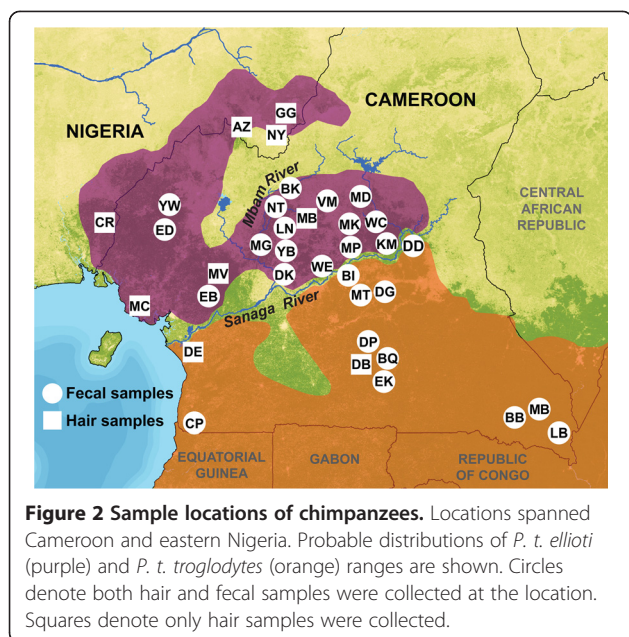


Hardy-Weinberg equilibrium and expectations of neutrality (Additional files 2 and 3). We further subjected the microsatellite loci to an outlier test based on observed heterozygosity and  $F_{ST}$  and all 21 loci fell within the acceptable range of neutrality (Additional file 4).

#### mtDNA diversity analysis

We constructed a median-joining mtDNA haplotype network (Figure 3a) and plotted a frequency distribution

of inferred haplotypes found at each sampling location across Cameroon and Nigeria (Figure 3b). The results were similar to observations from previous studies [10,13]. Specifically, there are two primary mtDNA haplogroups, each comprised of two distinct haplotypes, found across the study area. One haplogroup (Figure 3a light purple and dark purple) occurs only in individuals north of the Sanaga River (range of *P. t. ellioti*). The second haplogroup (Figure 3a light orange and dark orange)



occurs in *P. t. troglodytes* primarily south of the Sanaga. However, these two haplogroups overlap with each other in the ecotone in central Cameroon, north of the Sanaga River and east of the Mbam River. Interestingly, the *P. t. troglodytes*-like mtDNA sequences found in the ecotone belong to a single sub-type within haplogroup 2A. This suggests either a pattern of historic gene flow between the subspecies, which is consistent with other studies [10,13], or alternatively that there was a single migration event from south to north of the Sanaga followed by the proliferation of this haplotype exclusively within central Cameroon.

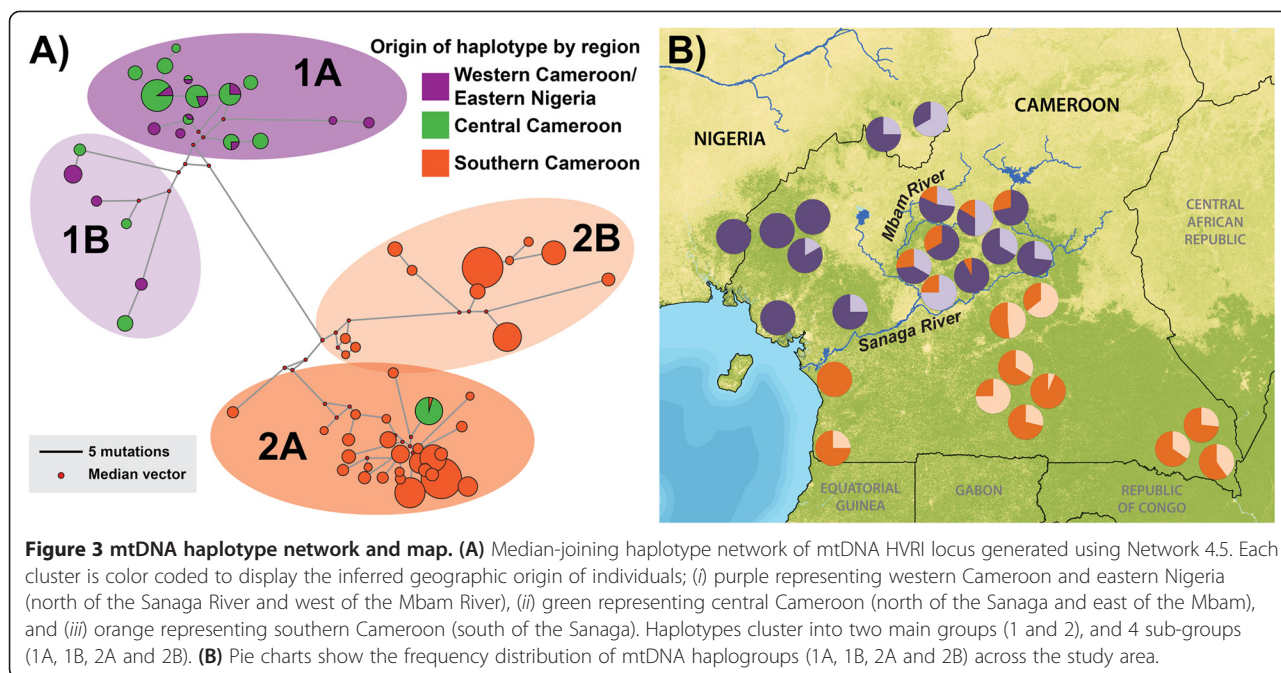
Overall, the results of the mtDNA haplotype analysis suggest that panmixia does not describe the population structure of chimpanzees. There is clear evidence that there are two main mtDNA haplogroups: a group in eastern Nigeria and western Cameroon and a second group in southern Cameroon. The geographic distributions of these

groups converge with one another in central Cameroon, with an abrupt transition between them along the Sanaga River. Overlap between these groups appears to be confined to the ecotone of central Cameroon, which also coincides with the confluence of the Sanaga with the Mbam River—another river of biogeographic significance for primates [16]. It is also worthy to note that *P. t. ellioti* in central Cameroon are more diverse than those in western Cameroon/eastern Nigeria (Figure 3b). Lower haplotype diversity in western Cameroon and eastern Nigeria, which has been proposed to be an area of putative Pleistocene refugia [25,26], may either be the result of local extinctions of haplotypes, or a diversification of mtDNA haplotypes as the result of increased habitat variation in central Cameroon.

We grouped results of the Analysis of Molecular Variance (AMOVA) (Table 2) for mtDNA sequences according to variation: (i) within populations ( $\Phi_{ST}$ ); (ii) among populations in groups ( $\Phi_{SC}$ ); and (iii) among groups north or south of the Sanaga River ( $\Phi_{CT}$ ), the major population partition indicated in previous studies [9,10]. Dividing the haplotypes by their origins north versus south of the Sanaga River accounted for 52.41% of the variation among groups, whereas 43.22% of the variation occurred within sampled populations within these groups. This population pattern is markedly different from the partitioning of genetic variation found in *P. t. schweinfurthii* [27] where less than 2% of the genetic variation were reported to have accounted for the differences between *P. t. schweinfurthii* groups that were separated from one another by more than a 600 km straight-line distance (Table 2), whereas the groups included in this AMOVA are separated by no more than 30 km at their points of closest sampling north versus south of the Sanaga River (Figure 2). Similarly, results for a spatial AMOVA (SAMOVA) [28] that assumed two populations ( $k = 2$ ) were present across the study area confirmed the geographic grouping of sampling locations north and south of the Sanaga River. The division of

**Table 1 Hypotheses and predictions**

Models of population history	Patterns of genetic differentiation		Demographic history
	mtDNA	Microsatellites	
<b>Panmixia</b>	One cluster of haplotypes spanning entire study region; Majority of genetic variation found between individuals	No evidence of discernable population structure into $k > 2$ demes	All individuals equally divergent from one another
<b>Isolation-by-distance</b>	Positive linear relationship between genetic and geographic distance	Positive linear relationship between genetic and geographic distance	Populations closer to one another share a more recent relationship than those more distant for one another
<b>Population structure with complete isolation</b>	Haplotypes cluster into discrete and geographically distinct groups	Individuals cluster into discrete, geographically distinct groups	No evidence of recent migration between populations, but ancestral polymorphism may still be present
<b>Population structure with ongoing migration</b>	Haplotypes cluster into geographically distinct groups with overlap	Individuals cluster into geographically distinct groups with some individuals showing evidence of admixture	Evidence of recent migration between populations that is distinguishable from ancestral polymorphism



haplotypes according to geographically homogeneous groups, accounted for 49.37%, and 45.56% of the variation within sample locations within these groups (Additional file 5). This pattern was statistically significant and recapitulated the results of the AMOVA (Table 2).

We generated two sets of mismatch distributions to examine whether the sampled populations had experienced recent bottlenecks, population expansions or demographic stability (Additional files 6 and 7). Individuals were grouped into either two or three separate groups, as identified by microsatellite cluster analysis. None of the mismatch distributions were found to be significantly different from the null model, which indicates demographic stability of all populations included in the analyses (Additional files 6 and 7) [29].

**Microsatellite genotype analysis**

Mantel tests of 21 microsatellite loci (Figure 4) revealed a significant ( $p < 0.05$ ) pattern of isolation-by-distance across the study area as a whole. However, the relationship between genetic and geographic distance is weak. The  $R^2$  value is low ( $R^2 = 0.0186$ ) and overall accounts for less than 2% of the total genetic variation present across the study area. Because isolation-by-distance only weakly explains the population structure of chimpanzees, we next examined the number and distribution of populations across the study area.

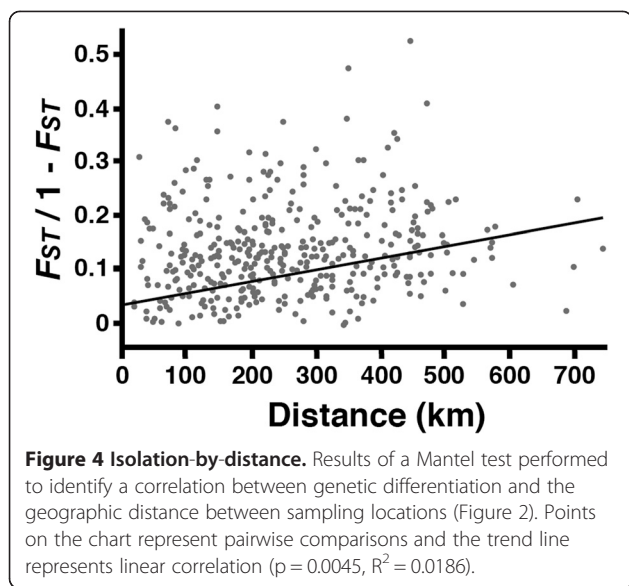
A Principal Components Analysis (PCA) was carried out to examine how many different gene pools were present across the study region. The PCA shown in Figure 5 classified the first four eigenvectors as significant ( $p < 0.05$ ). The

**Table 2 Analysis of Molecular Variance (AMOVA) for mtDNA HVRI**

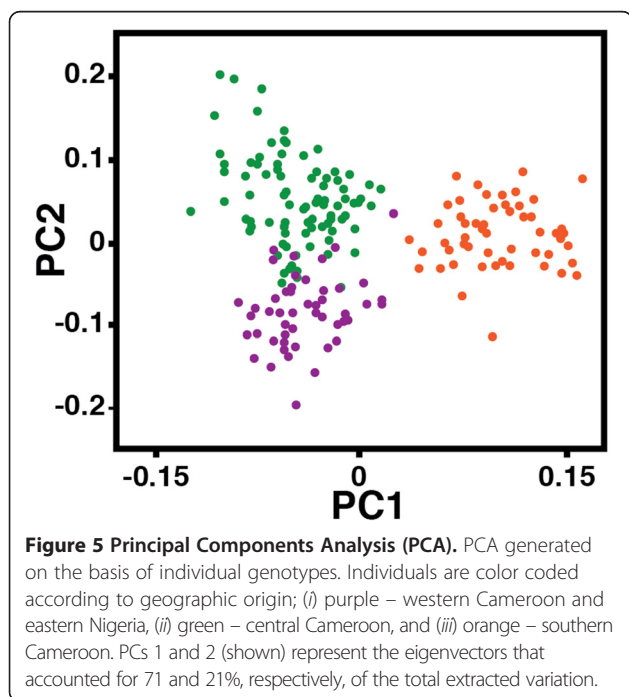
Partition	Fixation indices	Variance components	Percentage of variation	Significance
North ( <i>P. t. ellioti</i> ) versus south of Sanaga ( <i>P. t. troglodytes</i> ) <sup>a</sup>				
Among groups ( $\Phi_{CT}$ )	00.52	10.09	52.41	$p < 0.05$
Among populations in groups ( $\Phi_{SC}$ )	00.56	00.84	04.37	$p < 0.05$
Within populations ( $\Phi_{ST}$ )	00.09	08.32	43.22	$p < 0.05$
DRC versus eastern forests (both <i>P. t. schweinfurthii</i> ) <sup>b</sup>				
Among groups ( $\Phi_{CT}$ )	00.00	00.00	00.00	Not significant
Among populations in groups ( $\Phi_{SC}$ )	00.10	00.05	10.23	$p < 0.05$
Within populations ( $\Phi_{ST}$ )	00.10	00.45	89.77	$p < 0.05$

<sup>a</sup>AMOVA results of data from this study.

<sup>b</sup>AMOVA analysis reported by Goldberg and Ruvolo [27]. The samples included eastern chimpanzees (*P. t. schweinfurthii*) divided into two groups of samples recovered from chimpanzees in northeastern Democratic Republic of Congo (DRC) and ‘eastern forests’ that includes sampling locations in Rwanda, Tanzania and Uganda. These two groupings are separated from one another by more than 600km straight-line distance.



first Principal Component (PC 1) separated chimpanzees north and south of the Sanaga River (*P. t. ellioti* and *P. t. troglodytes*), and accounted for 71% of the total extracted variation in the dataset. PC 2 accounted for 21% of the total extracted variation and separated *P. t. ellioti* chimpanzees into two groups: one cluster of individuals mostly from the forests of western Cameroon (further designated as: *P. t. ellioti* (Rainforest), purple) and another cluster of individuals from the ecotone in central Cameroon (further designated as: *P. t. ellioti* (Ecotone), green). PC 3 and PC4 accounted for 5% and 3% respectively, and separated small

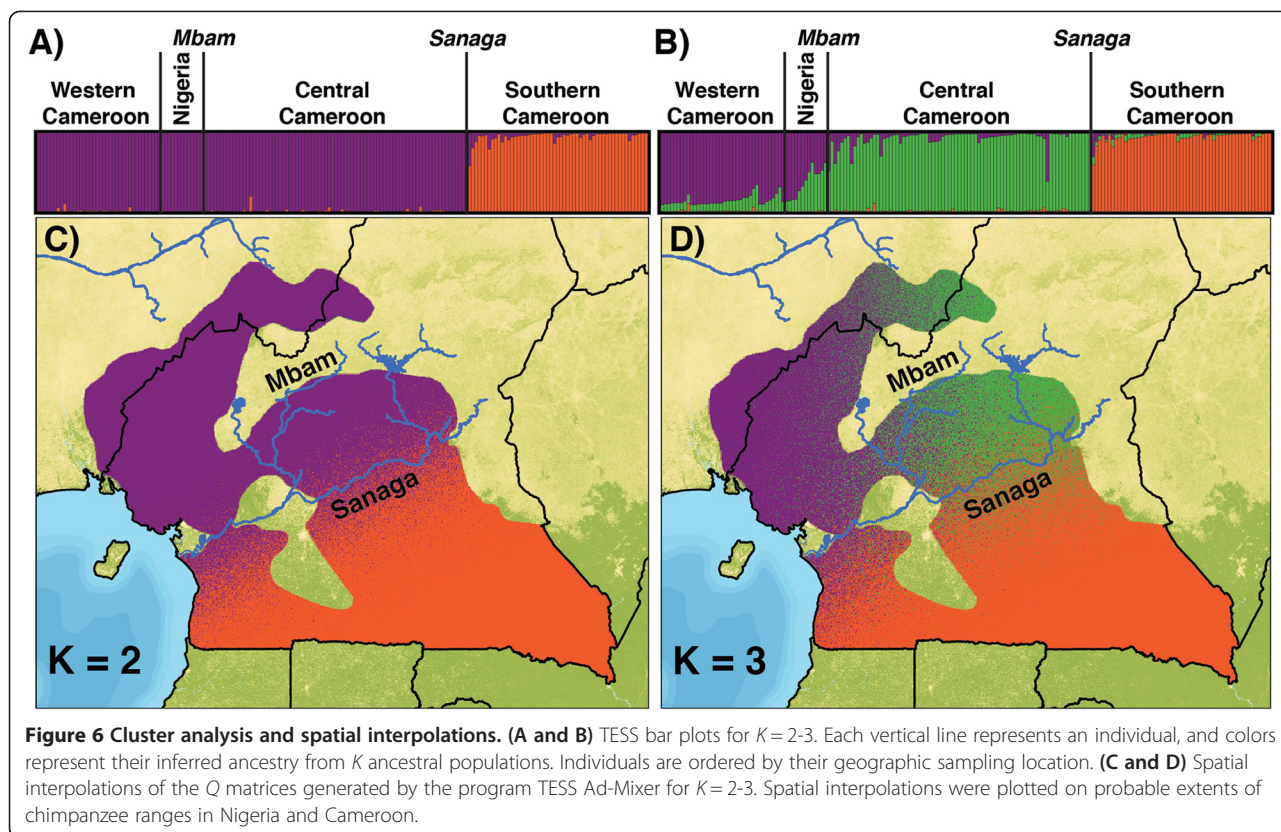


clusters of individuals from the same sampling locations from the rest of the dataset. These final PCs were sample location specific and are possibly the result of sampling bias (i.e. sampling of related individuals below our relatedness threshold).

Results of the cluster analysis in TESS [30] are shown in Figures 6a and b. We used several methods to infer the maximum number of populations,  $K$ , represented by this dataset. Deviance Information Criteria (DIC) [31] values suggest that  $K_{MAX}$  is 2 (Additional file 8a). In contrast, the *post hoc* statistic,  $\Delta K$  [32], suggests a  $K_{MAX}$  of 3 (Additional file 8b). Overall, these results suggest that there is a primary separation along the banks of the Sanaga River. *P. t. ellioti*, (purple) is found north and west of the river, whereas *P. t. troglodytes* (orange) occurs south of the Sanaga (Figure 6a). There is additional evidence of a population subdivision within *P. t. ellioti*: a population located in the forests of western Cameroon (*P. t. ellioti* (Rainforest), purple) and another population located in the ecotone in central Cameroon (*P. t. ellioti* (Ecotone), green) (Figure 6b).

The spatial interpolations generated by TESS Ad-Mixer [33] show the separation of these two populations (*P. t. ellioti*, purple and *P. t. troglodytes*, orange) corresponds to banks of the Sanaga River (Figure 6c). Setting  $k = 3$ , which was supported by the  $\Delta K$  statistic and the PCA, provided evidence of further subdivision within *P. t. ellioti*. Setting  $k = 2$  and  $k = 3$  using TESS (Figure 6) distinguished the same populations as PC1 and PC2 (Figure 5). The TESS Ad-Mixer [33] results show that this subdivision exists exclusively north of the Sanaga River, and the east–west division of populations (purple and green) coincides with the Mbam River, although this division is much less pronounced than the division of *P. t. ellioti* and *P. t. troglodytes* at the Sanaga River (Figure 6d), and there is predicted overlap between these two groups in Nigeria, north of the Mbam.

Values for three measures of genetic differentiation,  $D^2$  [34],  $R_{ST}$  [35], and  $\delta\mu^2$  [36], are shown in Additional file 9 between the three populations identified in the cluster analyses (Figures 5 and 6). All three measures were correlated with one another;  $D^2$  and  $R_{ST}$ ,  $r^2 = 0.94$ ;  $D^2$  and  $\delta\mu^2$ ,  $r^2 = 0.58$ ; and  $R_{ST}$  and  $\delta\mu^2$ ,  $r^2 = 0.81$ ; and all pairwise differences were significantly different from null expectations of the data (Additional file 9). Genetic differentiation was highest between populations of chimpanzee across the Sanaga River (*P. t. ellioti* versus *P. t. troglodytes*) and lowest between populations of chimpanzee north of the Sanaga River (*P. t. ellioti* (Rainforest) versus *P. t. ellioti* (Ecotone)). Additional file 9 shows the estimated number of migrants exchanged between populations per generation ( $2Nm$ ) as calculated using in the Arlequin version 3.5 software package [37]. Between 1 and 2 migrants per generation are exchanged between populations located



north vs. south of the Sanaga River, and approximately 11 migrants exchanged between the two populations found north of the Sanaga River in western and central Cameroon, respectively.

Allele richness and number of private alleles by population, corrected for unequal population size, is shown in Additional files 10a and 10b. Allele richness varied considerably between populations, with *P. t. troglodytes* displaying almost twice as many distinct alleles per locus compared to *P. t. ellioti* (Rainforest) and *P. t. ellioti* (Ecotone). However, the mean number of private alleles per locus did not vary considerably between regions, although we observed a similar pattern, with *P. t. troglodytes* exhibiting the greatest

number of private alleles, when corrected for unequal sample size. Additional file 10c shows the mean number of shared private alleles between the three combinations of population pairs, when corrected for equal population size. Interestingly, the highest number of shared private alleles occurs between *P. t. ellioti* (Rainforest) and *P. t. troglodytes*, with the other two population pairs exhibiting similar numbers of shared alleles.

**Population history**

We pooled the mtDNA and microsatellite data together for analysis in IMa [38], which allowed for the estimation of several population parameters (Table 3), including

**Table 3 Summary of demographic parameters for population pairs**

Comparison			Migrants per generation*		Population Size*			Population Divergence*	
Population 1	x	Population 2	Into Pop. 1	Into Pop. 2	Ne Pop. 1	Ne Pop. 2	Na	T <sub>MRC</sub> A (kya)	
<i>P. t. ellioti</i>	x	<i>P. t.</i>	<i>MLE</i>	1.10	1.15	2,140	3,365	3,002	200,979
(Rainforest)		<i>troglodytes</i>	95% CI	0.52 - 2.37	0.32 - 2.53	1,748 - 2,696	2,720 - 4,091	222 - 15,294	180,157 - 334,096
<i>P. t. ellioti</i>	x	<i>P. t.</i>	<i>MLE</i>	1.58	0.79	2,633	9,084	7,208	254,553
(Ecotone)		<i>troglodytes</i>	95% CI	1.03 - 2.34	0.20 - 1.71	1,984 - 3,282	7,549 - 10,961	1,493 - 45,421	202,149 - 328,715
<i>P. t. ellioti</i>	x	<i>P. t. ellioti</i>	<i>MLE</i>	2.49	4.39	2,368	860	5,920	4,219
(Rainforest)		(Ecotone)	95% CI	0.62 - 13.73	1.84 - 8.84	1,466 - 3,077	658 - 1,231	4,200 - 12,465	2,298 - 67,636

\*Demographic parameters were inferred in IMa [38]. These parameters were scaled assuming an mtDNA mutation rate of  $1.64 \times 10^{-7}$  [89], an intermediate microsatellite mutation rate of  $7.75 \times 10^{-5}$  [87], and 20-year generation time [90]. Demographic estimates assuming different microsatellite mutation rates are shown in Additional file 11.

descendant ( $N_1$  and  $N_2$ ) and ancestral ( $N_A$ ) population sizes, time of divergence and migration rates between each of the three populations recovered by the cluster analysis (Figure 6) and PCA (Figure 5). All three population comparisons yielded well-resolved posterior probability distributions for each population parameter (Additional files 11, 12 and 13). Both *P. t. ellioti* (Rainforest) and *P. t. ellioti* (Ecotone) last shared a common ancestor with *P. t. troglodytes* approximately 200 to 250 kya, which is consistent with other studies using similar data [39]. The two *P. t. ellioti* populations last shared a common ancestor with each other much more recently, about 4 kya. The results of these analyses also included estimates of ancestral and descendant effective population sizes for all pairs of populations. Estimates of effective population size for the present *P. t. ellioti* (Rainforest) population were consistent across runs, showing an effective population size of approximately 2,000 individuals. There was considerably more variation detected between comparisons for effective population sizes of *P. t. ellioti* (Ecotone) and *P. t. troglodytes*. In *P. t. ellioti* (Ecotone), effective population size was estimated to be between approximately 1,000 and 2,500 individuals. The current effective population size of *P. t. troglodytes* in southern Cameroon, may range from 3,000 to 10,000 individuals. The estimates for ancestral population size show that *P. t. ellioti* north of the Sanaga River were approximately 6,000 individuals, and that the ancestral populations of chimpanzees likely ranged between approximately 3,000 and 7,000 individuals. The number of migrants per generation was also estimated for all three pairs of populations. Between 0.79 and 1.58 migrants per generation have been exchanged between populations across the Sanaga River, since the initial separation of *P. t. troglodytes* and *P. t. ellioti*. Higher levels of migration were found between the *P. t. ellioti* populations (Table 3 and Additional file 7).

Overall the results of the IMA suggest very clearly that the gene pools of *P. t. ellioti* and *P. t. troglodytes* are mostly, but not completely isolated from one another (Figures 3, 5 and 6). We conclude that these two subspecies are characterized by an isolation-with-migration population history model [40] for several reasons. The results of the IMA analysis shows that the *P. t. ellioti* populations last shared a common ancestor with *P. t. troglodytes* approximately 200–250 kya, and since this separation they have exchanged approximately one migrant per generation. Moderate levels of gene flow ( $\geq 1$  migrant/generation) between populations should prevent divergence, under a model that assumes that drift occupies the dominant role in population differentiation [40]. Population divergence in the presence of gene flow is often cited as evidence that local adaptation is driving the separation of two or more populations [41–43]. However, the observed value of migration between *P. t. ellioti* and *P. t. troglodytes* ( $2 Nm \sim 1$ ) is a

minimal threshold [40], making it difficult to fully distinguish whether allopatric speciation or local adaptation has occupied the dominant role in driving this separation.

Demographic histories can be influenced by climatic histories. The climate of Africa has been affected by oscillations in global temperature and species assemblages in tropical Africa are often influenced by these changes, particularly the Last Glacial Maximum (LGM) [25]. It is also important to understand demographic history in the context of climatic histories. Interestingly, the dates of divergence between *P. t. ellioti* and *P. t. troglodytes* predate the LGM, but do coincide with a pronounced glacial episode approximately 250kya [26], with evidence that sand dunes extended all the way to the Niger Delta [25].

It may be possible that these fluctuations in local climate may have influenced the Sanaga River, but little is known regarding its historical course and size. Rivers are known to dramatically change in course and size over time, thus affecting their ability to act as dispersal barriers [44]. Thus, it is possible that chimpanzee migration across the Sanaga could have occurred in bursts when river size was minimal, but we are unable to directly test this hypothesis, given the nature of the genetic data used in this study.

## Conclusions

This study represents the most comprehensive genetic dataset available for wild chimpanzees from Cameroon, which is important because it is the only known area where the distributions of two chimpanzee subspecies overlap. Microsatellite genotype profiles from 187 unrelated individuals and mtDNA haplotypes of 607 individuals were used to test between a variety of hypotheses for the presence and type of population structure of chimpanzees across the study area, including (i) panmixia, (ii) isolation-by-distance, (iii) population structure with complete isolation, and (iv) population structure with ongoing migration. Overall, the results suggest that *P. t. ellioti* and *P. t. troglodytes* represent genetically distinct populations in Cameroon, confirming results from previous studies of captive chimpanzee with inferred origins [6–8,39] and previous analyses of mtDNA haplotypes of wild individuals [9,10]. Surprisingly, we found additional evidence that *P. t. ellioti* consists of two genetically- and geographically-distinct populations. One population is located in forested regions of eastern Nigeria and western Cameroon and the second is located in central Cameroon, in a savanna-woodland mosaic that occurs between the Mbam and Sanaga Rivers. The results of the IMA analysis show that the *P. t. ellioti* populations and *P. t. troglodytes* diverged from one another 200–250 kya, and have experience moderately high ( $2Nm \sim 1$ ) gene flow since (Table 3). This, coupled with the overlap of these populations in the mtDNA haplotype network analysis (Figure 3) and microsatellite cluster analyses (Figures 5 and 6) call in question



the role of the Sanaga River in acting as a biogeographic barrier that has separated these chimpanzee subspecies.

Several riverine barriers have been proposed as biogeographic boundaries for primates in the region [4,9,12,14-18,45], but the role of habitat variation in driving primate speciation remains unknown. For instance, in addition to being important for chimpanzees, the Sanaga River has been proposed to influence the distribution of several pairs of primates, including *Mandrillus leucophaeus*/*M. sphinx*, *Cercopithecus erythrotis*/*C. cephus*, *C. nictitans martini*/*C. n. nictitans*, and *C. pogonias pogonias*/*C. p. grayi* [4,14-16,18]. These pairs of primates all occupy vastly different habitats and niches [14,15]. This observation suggests that other factors along with, or instead of, the Sanaga River may be important in separating the distribution of these species, subspecies and populations across the region. It is well documented that non-riverine barriers, such as the Dahomey Gap, also separate taxa, and likely separates *P. t. verus* from *P. t. ellioti* (although the role of the Niger River is unclear) [10,44]. Other factors have also been proposed to shape African tropical biodiversity in the region [12,46-49]. Thus, it is important that we reevaluate the role that the allopatric speciation, as driven by the Sanaga River, may have had in governing distributions of rainforest taxa, particularly primates.

Cameroon lies at the intersection of two major rainforest biomes. The Congolian Rainforest extends northward into southern Cameroon from central Africa, and the Guinean Rainforest extends southward into eastern Nigeria and western Cameroon from western Africa (Figure 1a). These two biomes converge in central Cameroon at the location of a pronounced ecotone [20], which is composed of open woodland, savannah and riparian forest [20,21]. Recent studies in this area that have combined genetic, morphological and environmental data have found that this ecotone appears to drive evolutionary diversification in insects [50], reptiles [51] and birds [22,52]. This is especially interesting, given that neutral processes are unlikely to explain the population history of chimpanzees in the region, as they seem to follow an isolation-with-migration model. The results of the mtDNA haplotype analysis shows an overlap of the two major haplogroups (Figure 3), and the results of the cluster analysis show a distinct deme of chimpanzee in this ecotone (Figure 6d). However, the loci included in this study meet expectations of neutral evolution, and thus, do not allow for us to draw additional conclusions regarding how ecological variation may drive diversification in chimpanzees. These observations underscore the importance of testing alternative hypotheses with loci that are neutral *and* under selection.

Even though the migration rate between *P. t. ellioti* and *P. t. troglodytes* ( $2Nm \sim 1$ ) is a minimal threshold to propose that selection has driven the separation of these populations, there is mounting evidence that allopatric

speciation alone cannot account for the divergence of these populations. A complimentary study of the ecological niche differentiation in these populations shows that the demes of chimpanzees found in this study also occupy significantly different niches [53], suggesting that environmental variation contributes to driving the differentiation of these populations. The alternative hypothesis, that speciation in chimpanzees is driven by allopatric speciation, would be explained by an absence of niche divergence among these three demes [54,55]. Furthermore, in another complimentary study we found significant associations between genetic differentiation in these three demes and environmental variation across the study area [56], which led us to conclude that these populations likely follow a pattern of isolation-by-environment [57], a relationship between populations that arises as a result of local adaptation to different environments. Taken together, these findings are consistent with the observation that chimpanzees in Cameroon and Nigeria may be adapted to their local environments, and that this variation has contributed to the genetic differentiation of chimpanzee subspecies.

A broader understanding about the role of local adaptation in chimpanzees and other taxa may provide important clues regarding why this region of Africa contains such a high proportion of the Earth's biodiversity. Future studies that more closely examine the history of the Sanaga River and the role of historic climatic variation in shaping chimpanzee genetic diversity are crucial for building this understanding. Additionally, a better understanding of how chimpanzee social structures and dispersal patterns are shaped by habitat variation, and how this may contribute to regional genetic diversity, is of vital importance to unravelling the forces that have helped make this region such an incredibly diverse place on Earth.

## Methods

### Dataset preparation

We gathered fecal ( $n = 247$ ) and hair samples ( $n = 223$ ) from wild-living, non-habituated chimpanzee populations from 36 remote regions spanning eastern Nigeria through southern Cameroon (Figure 2 and Additional file 1). Most samples originating from the regions of Belgique (BQ), Boumba Bek (BB), Diang (DG), Dja Biosphere (DB), Duomo Pierre (DP), Ekom (EK), Lobeke (LB), Mambele (MB) and Minta (MT) were collected between 2004 and 2006 by the team of BHH and MP. Hair and tissue samples from Nigeria and Cameroon were collected during a series of studies from 1994 – 2010 by the team of MKG. We recruited a team of ~9 field assistants for each mission and base camps were established in the vicinity of forests where chimpanzee presence was reported by local hunters and villagers. We initially walked transects, hunters' paths or elephants paths in search of chimpanzees presence (e.g. night nests, foot prints and vocalizations). The Campo

Ma'an (CP), Cross River (CR), Deng Deng (DD), Diang (DG), Dja Biosphere (DB), Douala-Edea (DE), Gashaka Gumti (GG), Mbam et Djerem (MD) and Mount Cameroon (MC) sites are located in National Parks or forests reserves, whereas the remaining field sites are located in unprotected forests. Fecal samples were identified to be of likely chimpanzee origin by experienced trackers and/or by the researchers. Fecal samples (15–20 g) were placed into 30 or 50 mL tubes and mixed with equal amounts of RNA-later<sup>®</sup> (Ambion, Austin, TX). Hair samples were collected directly from abandoned sleeping nests and a minimum of three hairs per nest were stored into glassine envelopes, and kept dry in silica gel. Fecal samples were inspected to estimate their likely time of deposition, whereas night nests age was estimated according to a leaf decay index [58]. Time, date, location, longitude, latitude and name of collector were also recorded. Fecal and hair samples were generally kept at ambient temperature for no longer than 2 weeks and subsequently stored at  $-20^{\circ}\text{C}$  once back in Yaoundé, Cameroon. Samples were shipped to the United States at ambient temperature, then stored at  $-20^{\circ}\text{C}$  upon receipt. All samples were transported from Cameroon to the United States in full compliance with Convention of International Trade in Endangered Species of Wild Fauna and Flora (CITES) and Center for Disease Control (CDC) export and import regulations. This research was carried out with IACUC approval from the University at Albany – State University of New York.

We extracted fecal DNA using the QIAamp Stool DNA Mini kits (Qiagen, Valencia, CA) following well-established protocols [59]. We briefly resuspended 1.5 mL of fecal RNA-later<sup>®</sup> mixture in stool lysis buffer and clarified them by centrifugation. We treated the supernatants with an InhibitEx tablet, subjected them to proteinase K digestion, and passed them through a DNA binding column. Bound DNA was eluted in 100  $\mu\text{L}$  elution buffer. We extracted DNA from hair samples using a chelating resin protocol [60] followed by filtration using Microcon 100 columns (Millipore, Billerica, MA) to concentrate DNA extracts. We conducted the preparation of the DNA quantification standards, reagents and reactions according to the Quantifiler<sup>®</sup> DNA Manufacturer's protocol [61]. We analyzed the data using a 7500 System v1.2.3 software as an 'Absolute Quantification (standard curve)' assay with the settings recommended in the Quantifiler<sup>®</sup> Kit User's Manual [61].

We used primers L15997 (5'-CACCATAGCACCCA AAGCT-3') and H16498 (5'-CCTGAAGTAGGAACCA GATG-3') to amplify a 460 – 500-bp mtDNA fragment spanning the hypervariable D-loop region (HVRI) in all samples that were newly collected for this study, using methods described in previous studies [10,13,59]. We assembled and aligned the resulting sequences with SEQMAN DNASTAR software (Lasergene, DNASTAR, Inc., Madison, WI), along with georeferenced sequences from

previous studies [9,10,59,62]. We deposited all newly generated sequences from this study in GenBank under accession numbers KM401682-KM401815.

We used twenty-one autosomal microsatellite loci to produce genotype profiles from both hair and fecal samples. Additional file 14 lists information about the markers including the primers flanking the selected regions and the fluorescent dye set (Applied Biosystems, Foster City, CA) chosen. These markers were originally developed for use in humans [63], but have been shown to be highly informative for use in chimpanzee population genetics studies [39,64]. We also included the Amelogenin locus to determine the sex of the individual sampled. We divided these 22 loci into 4 multiplex PCR reactions that were performed using the Quiagen Multiplex PCR Kit (Qiagen, Valencia, CA) in Eppendorf Mastercycler (Eppendorf, Westbury, NY). We used 0.5–1 ng DNA, along with Q-Solution (included in the kit). We adopted the following PCR conditions: an initialization at  $95^{\circ}\text{C}$  for 15 min, followed by 40 cycles including a denaturation step at  $95^{\circ}\text{C}$  for 30 sec, an annealing temperature of  $60^{\circ}\text{C}$  for 1 min, an elongation at  $72^{\circ}\text{C}$  for 1 min and 30 sec, and a final extension at  $72^{\circ}\text{C}$  for 10 min. For samples that failed to produce reliable genotype profiles, we increased the amount of DNA template to up to 8  $\mu\text{L}$  per reaction, regardless of the DNA quantitation results. Many of the hair samples from the Nigerian locations had been typed previously for some of the loci selected, and were adjusted to the differences in base pair sizes due to apparatus and protocol discrepancies [13,65]. All PCR reactions included negative control samples for quality assurance. We analyzed each multiplex PCR product on an ABI 3130 capillary array genetic analyzer (Applied Biosystems, Foster City, CA). We determined fragment sizes against a Genescan 600 Liz size standard (Applied Biosystems, Foster City, CA), and allele sizes using the Genemapper ID version 2.7 software (Applied Biosystems, Foster City, CA). We scored alleles between three and 6 times to avoid problems associated with allelic dropout, which frequently occurs when genotyping low-yield DNA samples [66]. Samples that did not include at least 15 (70%) or more loci after multiple attempts at PCR fragment amplification were excluded from this study.

We subjected all 21 autosomal microsatellite loci to outlier tests using LOSITAN [67]. LOSITAN is a Java based selection detection platform, based on the  $F_{ST}$  outlier methods [68]. We ran 1,000,000 simulations of the data while (i) assuming a stepwise mutation model, (ii) assuming three populations, (iii) forcing a correct mean  $F_{ST}$  and (iv) calculating a "neutral" mean  $F_{ST}$ . We also subjected the 21 autosomal microsatellites, in addition to the mtDNA HVRI sequences, to an exact test of Hardy-Weinberg equilibrium [69] using the Arlequin version 3.5 software package [37]. We tested deviations from Hardy-Weinberg

equilibrium for each locus against 1,000,000 random permutations of the data.

We calculated pairwise estimates of relatedness for all samples that produced reliable microsatellite genotypes using the software program COANCESTRY [23]. COANCESTRY implements three inbreeding estimators and seven relatedness estimators [24,70-76] to estimate relatedness between individuals and inbreeding coefficients from multi-locus genotype data. We tested 95% confidence intervals for relatedness and inbreeding estimates for all genotyped individuals against 1000 bootstrap permutations of the data. Samples estimated to come from the same individual or close relatives, with a relatedness index [24] above 0.75, were excluded in order to remove all pairs of individuals down to first cousins from all further data analyses.

#### mtDNA diversity analysis

We generated haplotype networks for HVRI mtDNA sequences via the median-joining algorithm of Network 4.5 (<http://www.fluxus-engineering.com>). Because it allows for reticulation, the median-joining approach for inferring haplotype relationships is appropriate for the analyses of mtDNA control region sequences, which exhibits high levels of homoplasy in humans [77,78]. We identified hypermutable sites by post-processing using the Steiner maximum parsimony algorithm within Network 4.5 and were excluded from the network analyses.

We performed an AMOVA using Arlequin version 3.5 software package [37]. We tested population differentiation values against 10,000 random permutations of the data. For the AMOVA, we grouped individuals according to their origin either north or south of the Sanaga River. In addition to the AMOVA, we ran a spatial AMOVA (SAMOVA) at  $k=2$ , to confirm the groupings used in the AMOVA, as well as to test how geospatial partitioning of the sampled populations affected population differentiation [28].

We generated mismatch distributions by computing distribution of the number of pairwise differences between mtDNA haplotypes using Arlequin version 3.5 software package [37]. We generated distributions by grouping individuals into either two or three separate groups (as identified by microsatellite cluster analysis), using 100 bootstrap replicates.

#### Microsatellite genotype analysis

We examined the relationship between genetic differentiation and geographic distance by carrying out partial Mantel tests on the microsatellite data using the Arlequin version 3.5 software package [37]. Pairwise  $F_{ST}$  values were generated between all sampling locations and genetic differentiation was plotted against straight-line, geographic distance. We fit the data with a linear regression, calculated a correlation coefficient and tested for statistical significance.

We used the EIGENSOFT software package [79] to perform PCA on individual genotypes to identify individuals that could be grouped into significantly different populations. We converted microsatellite data into a false SNP format by scoring the presence or absence of each of  $n-1$  alleles (where  $n$  is the number of alleles in the sample) using a script in MATLAB (The MathWorks, Natick, MA) described previously [39]. We processed this file in SmartPCA, which produced eigenvectors and eigenvalues. We conducted this analysis blindly to *a priori* population labels for individuals in the dataset. We tested the statistical significance of each eigenvector using Tracy-Widom statistics. Each significant eigenvector recovered by this PCA approach separates the samples in such a way that the first and subsequent eigenvectors distinguish, in order, the most to least differentiated populations in the sample [79]. All analyses using EIGENSOFT were performed blinded to *a priori* population labels.

We examined population structure and individual ancestry using a Bayesian clustering approach implemented in the TESS (version 2.3) software package [30]. This program estimates the shared population history of individuals based on their genotypes and geographic origin under a model that assumes Hardy-Weinberg equilibrium and linkage equilibrium, thereby making no *a priori* assumptions regarding population classifications. TESS estimates individual proportions of ancestry into  $K$  clusters, where  $K$  is specified for the program in advance across independent runs and corresponds to the number of putative ancestral populations. The program then assigns admixture estimates for each individual ( $Q$ ) from each inferred ancestral population cluster. It also produces posterior predictive values of  $Q$  for geographic areas located in between and around the individual data points.

TESS runs were performed: (i) with a model that allows individuals to have ancestry in multiple populations (CAR model [80]); (ii) with correlated allele frequencies; and (iii) blinded to *a priori* population labels. Runs were performed with a burn-in step of 500,000 Markov Chain Monte Carlo (MCMC) iterations and 1,000,000 MCMC iterations. Fifty runs each for  $K=2$  through  $K=8$  were performed for all datasets. We processed TESS outputs with CLUMPP [81] and a G-statistic >99% was used to assign groups of runs to a common clustering pattern. CLUMPP outputs for each  $K$  value were plotted with DISTRUCT [82]. We used a combination of methods to infer a maximum number of chimpanzee populations ( $K_{MAX}$ ) including, (i) the  $K$  value that had the highest average DIC [31], (ii) high stability of clustering patterns between runs, (iii) the  $K_{MAX}$  value at which  $K_{MAX} + 1$  no longer split the cluster distinguished by  $K_{MAX}$  [83], (iv) correspondence between maximum PPD values from TESS runs and significant eigenvectors recovered by PCA, and (v) calculating an *ad hoc* statistic,  $\Delta K$  [32].

We created spatial interpolations of the  $Q$  matrix using the option in TESS to generate posterior predictive maps of admixture proportions. We created a template map of the extent of probable chimpanzee ranges in Nigeria and Cameroon in ArcMap Version 10 (ESRI Corp., Redlands, CA) in the ASCII-raster format and input into TESS in order to generate posterior predictive maps of admixture proportions. Using these predictive maps, we generated spatial interpolations of the  $Q$  matrices by implementing a matrix based vector algorithm in the program TESS Ad-Mixer [33]. These spatial interpolations use color mixing to predict admixture proportions across space in order to better visualize the spatial partitioning of genetic differentiation.

We calculated allele size range, observed and expected heterozygosity, and an M-Ratio [84] for each locus for three populations as identified by cluster analysis. We tested these values against 10,000 random permutations of the data using the Arlequin version 3.5 software package [37].

We calculated three measures of population genetic differentiation using the Arlequin version 3.5 software package [37]:  $D^2$ ,  $R_{ST}$  and  $\delta\mu^2$ . The  $D^2$  [34] genetic distance is based on a model in which genetic drift is the only force influencing allele frequency differences across populations and is sensitive to recent differentiation events.  $R_{ST}$  [35] and  $\delta\mu^2$  [36] are similar to  $D^2$ , but both assume a stepwise mutation model (SMM). Consequently,  $R_{ST}$  and  $\delta\mu^2$  are more likely to capture whether differences in the mutation processes are important in driving population differentiation and enable assessment of sensitivity to mutation model assumptions [83]. These latter models differ in that  $R_{ST}$  is based on the fraction of the total variance in allele size between populations and is analogous to  $F_{ST}$  [35], whereas  $\delta\mu^2$  is based on differences in the means of microsatellite allele sizes [36]. Recent work has shown convincing evidence that the loci typed for this study appear to follow the SMM in both chimpanzees and bonobos [64,85].  $D^2$  calculations were completed on untransformed allele size calls. Since  $R_{ST}$  and  $\delta\mu^2$  assume the SMM, we transformed allele sizes to repeat size units prior to analysis in Arlequin version 3.5 [37]. We transformed allele sizes such that the smallest allele for each locus was scored as  $n$  and each subsequent allele was scored as  $n + 1$ . In infrequent cases where repeat unit sizes did not follow the  $n + 1$  model, and instead repeat units skipped  $x$  repeat(s), we scored the next allele in the data as  $(n + x + 1)$ , as described in a previous study [39]. We determined each pairwise genetic distance calculation by 100,000 replications in Arlequin, and the significance of these pairwise population genetic distances were evaluated by a significance test at  $p < 0.05$ .

We calculated distributions of alleles within and between populations using ADZE [86]. ADZE is a generalized rarefaction approach that counts alleles private to populations as well as combinations of populations. We used this program to infer (i) the mean number of distinct

alleles per locus, per population; (ii) the mean number of private alleles per locus, per population; and (iii) the mean number of uniquely shared private alleles per locus, between populations.

#### Population history

We used the program IMA [38] to estimate: (i) the population mutation parameter,  $\theta$ , for both descendant populations ( $\theta_1$  and  $\theta_2$ ) and an ancestral population ( $\theta_A$ ); (ii) rates of migration from population 1 to population 2 ( $m_1$ ) and from population 2 to population 1 ( $m_2$ ); (iii) and time of population divergence ( $t$ ) for all possible pairs of the three assumed populations: *P. t. ellioti* in western Cameroon & Nigeria, *P. t. ellioti* in central Cameroon, and *P. t. troglodytes* in southern Cameroon. We used IMA instead of the more recently released IMA2 [40] for several reasons. The dataset is composed mostly of microsatellites with poorly characterized mutation rates [39,64,87]. Also, due to the complexity of the IMA2 model, the high number of microsatellite alleles, and few number of microsatellite loci, we used the version of the IM model with the fewest assumptions. Additionally, IMA2 requires the user to input a well resolved, rooted phylogeny of the populations to be tested, and based on the results of the cluster analysis, it was counterproductive to swap through an already well resolved, and simple phylogeny.

We ran multiple iterations of IMA in parallel, for each paired population, using a multi-processor computer cluster located at the SUNY College of Nanoscale Science and Engineering. First, we ran the IMA analysis using “M-Mode” (MCMC Mode) with a full complement of model parameters, and a broad range of priors for all parameters ( $\theta_1$ ,  $\theta_2$ ,  $\theta_A$ ,  $m_1$ ,  $m_2$ , and  $t$ ). Each run was performed with heated chains using the two-step scheme [88]. For each run, we ran burn-in replicates for a total of five days. We then reduced the ranges of the model parameters over repeated runs in order to sample more densely their respective posterior distributions for each of the three population comparisons. This resulted in different lengths of analysis for each comparison: the western Cameroon/central Cameroon comparison generated 1,004,512 genealogies using 58 chains; the western Cameroon/southern Cameroon comparison generated 684,192 genealogies using 40 chains; and the central Cameroon/southern Cameroon comparison generated 947,723 genealogies using 38 chains. After the runs converged toward the same values, we loaded the saved genealogies from multiple runs into “L-Mode” (Load Trees Mode) in order to calculate values for model parameters for each pair of populations pooled across these multiple independent runs.

We converted the migration parameters,  $m_1$  and  $m_2$ , into the number of migrants exchanged per generation ( $Nm$ ) using the equation  $Nm = (\theta^*m)/2$ . We converted the demographic parameter ( $t$ ) into years by scaling by mutation rate ( $t/\mu$ ). We scaled the  $t$  parameter for mtDNA

assuming a mutation rate ( $\mu$ ) of  $1.64 \times 10^{-7}$  [89]. We scaled the  $t$  parameter for microsatellite loci using several microsatellite mutation rates, given the uncertainty in microsatellite mutation rates in the literature [36,39,64,87]. We used a slow mutation rate of  $3.53 \times 10^{-5}$  [64], an intermediate mutation rate of  $7.75 \times 10^{-5}$  (calculated from the geometric mean of rates from Wegmann and Excoffier [87]), and a fast mutation rate of  $1.6 \times 10^{-4}$  [36,39]. We scaled final values for demographic parameters assuming a 20-year generation time for chimpanzees [90]. We scaled effective population sizes ( $N_E$ ) using  $\theta$  and a per generation  $\mu$  using the equation  $N_E = \theta/(4*\mu*20)$ .

## Additional files

### Additional file 1: Sample locations and number of individuals included in study.

<sup>a</sup>Individuals determined by relatedness indices calculated using COANCESTRY [23]. <sup>b</sup>Samples from these locations were not genotyped for microsatellite loci.

**Additional file 2: Summary Statistics for mtDNA HVRI.** <sup>a</sup>Significant values of  $F_u$ 's FS shown as bold. <sup>b</sup> $p = 0.04$ . <sup>c</sup> $p < 0.001$ .

### Additional file 3: Summary statistics for microsatellite loci.

\*Observed heterozygosity calculated in Arlequin [37]. †Expected heterozygosity calculated in Arlequin [37]. ‡M-Ratio from Garza & Williamson [84].

### Additional file 4: Identification of neutral versus outlier loci.

Heterozygosity was plotted against  $F_{ST}$  for all 21 microsatellite loci using LOSITAN [67]. Ranges of values for balancing selection (yellow), positive selection (red) and neutrality (gray) were identified. All 21 microsatellite loci fell into the acceptable range of neutrality.

### Additional file 5: Spatial Analysis of Molecular Variance (SAMOVA) for mtDNA HVRI.

### Additional file 6: Mismatch distribution for two populations.

Mismatch distribution using the mtDNA HVRI locus for *P. t. ellioti* (A) and *P. t. troglodytes* (B). Harpending's Raggedness Index for *P. t. ellioti* (A) was 0.003 ( $p = 0.977$ ), and for *P. t. troglodytes* (B) it was 0.007 ( $p = 0.168$ ).

### Additional file 7: Mismatch distribution for three populations.

Mismatch distribution using the mtDNA HVRI locus for *P. t. ellioti* (Rainforest, A), *P. t. ellioti* (Ecotone, B) and *P. t. troglodytes* (C). Harpending's Raggedness Index for *P. t. ellioti* (Rainforest, A) was 0.002 ( $p = 0.989$ ), for *P. t. ellioti* (Ecotone, B) it was 0.008 ( $p = 1$ ) and for *P. t. troglodytes* (C) it was 0.007 ( $p = 0.168$ ). Model frequency for *P. t. ellioti* (Ecotone, B) was not plotted, because the analysis only generated a value for one sample point.

**Additional file 8: Estimating  $K_{MAX}$ .**  $K_{MAX}$  was inferred from 50 independent TESS [30] runs for each value of  $K$  from 1 to 8. (A) Estimated values for the Deviance Information Criteria (DIC) [31]. (B) Estimated  $\Delta K$  [32] values.

### Additional file 9: Genetic differentiation ( $D^2$ , $R_{ST}$ and $\delta\mu^2$ ) and migration ( $2Nm$ ) between populations from across the study area from 21 autosomal microsatellite loci.

<sup>a</sup>Measures of differentiation were computed between populations as inferred by the cluster analyses. <sup>b</sup>Values above the diagonal are based on Reynold's Coancestry Coefficient,  $D^2$ , which does not assume that mutations follow the SMM [34]. <sup>c</sup>Values in underlined italics are migrants exchanged between populations each generation ( $2Nm$ ). <sup>d</sup>Values below the diagonal are based on an  $R_{ST}$  [35] model of evolution that assumes that mutations follow the SMM. <sup>e</sup>Values in parentheses are based on  $\delta\mu^2$  model of evolution that also assumes that mutations follow the SMM [36]. <sup>f</sup>All population pairwise values were significantly different from null expectations from 10,000 permutations of the data in Arlequin [37].

### Additional file 10: Inference of microsatellite allelic diversity.

(A) Mean number of distinct alleles found in sampled populations.

(B) Mean number of private alleles found in sampled populations. (C) Mean number of uniquely shared alleles between sampled populations.

### Additional file 11: Summary of results from IMA analysis assuming slow, medium, and fast mutations rates for microsatellite loci.

\*Demographic parameters were scaled assuming a mtDNA mutation rate of  $1.64 \times 10^{-7}$  [89]. There is a great deal of uncertainty in microsatellite mutation rates. Thus, we scaled demographic parameter estimates using slow, intermediate and fast microsatellite mutation rates. Green cells show a demographic estimates calculated using a slow rate of  $3.53 \times 10^{-5}$  from Becquet et al. [64]. Blue cells show demographic estimates using an intermediate rate of  $7.75 \times 10^{-5}$  calculated from the geometric mean of rates from Wegmann and Excoffier [87]. Red cells show demographic estimates scaled with the fastest mutation rate of  $1.6 \times 10^{-4}$  [36,39]. Demographic parameters were scaled assuming a generation time of 20 years for chimpanzees [90].

### Additional file 12: Tests of nested models from IMA for three populations of chimpanzees.

<sup>a</sup>Description of model tested. <sup>b</sup>LLR test statistics are calculated as twice the difference in log-likelihood between the indicated model and the full model ( $\Theta_1 \Theta_2 \Theta_A m_1 m_2$ ), with degrees of freedom equal to the difference in the number of parameters free to be estimated from the data. <sup>c</sup> $P$  values are obtained by comparison to the Chi-square distribution. <sup>d</sup>Number of free parameters, or degrees of freedom for each model. <sup>e</sup>When the null model is true and has a parameter fixed at the boundary of the parameter space, the expected distribution is a mixture [38]. <sup>f</sup>Bold numbers indicate models that could not be rejected ( $p > 0.05$ ) by the 2LLR log-likelihood ratio test.

**Additional file 13: IMA probability distributions.** Posterior probability distributions for three parameters: (i) migration ( $2Nm$ ); (ii) effective population size ( $4N_e\mu$ ); and time of population divergence ( $T_{MRCA}$ ). Parameters are not scaled according to a mutation rate or chimpanzee generation time. Parameters were tested for three population comparisons: (A) *P. t. ellioti* (Rainforest) – *P. t. troglodytes*, (B) *P. t. ellioti* (Ecotone) – *P. t. troglodytes*, and (C) *P. t. ellioti* (Rainforest) – *P. t. ellioti* (Ecotone).

**Additional file 14: Microsatellite loci information.** <sup>a</sup>Polymorphic Information Content estimated in CERVUS version 3.0 [91].

## Competing interests

The authors declare that they have no competing interests.

## Authors' contributions

MWM, SL and MKG designed the research. MWM, SL, LG and MKG performed research. MWM, SL, PRSC and MKG analyzed data. SL, AAP, EEA, AN, LN, NMA, BJM, RF, MP, BHH and MKG contributed samples. MWM, SL, PRSC and MKG wrote the paper. All authors read and approved the final manuscript.

## Acknowledgements

We thank the governments of Cameroon and Nigeria for permission to conduct this research. We thank the Zoological Society of San Diego, the Ebo Forest Research Project, the Wildlife Conservation Society and the World Wildlife Fund for their support in Cameroon. We thank Joe Kujawski, Anne Shelton, and the State University of New York – College of Nanoscale Science and Engineering (CNSE) for their assistance with running IMA on the CNSE servers. We thank Caro-Beth Stewart, George Robinson, Jeremy Kirchman and Adam Freedman for helpful comments on early drafts of this manuscript. National Science Foundation awards 0755823 and 1243524 (to MKG), National Institutes of Health award R37 AI 50529 (to BHH and MP), funding from National Geographic (to AN), the Offield Family Foundation (to BJM), the USFWS Great Apes Conservation Fund (to BJM), the Arcus Foundation (to BJM), and the University at Albany – State University of New York (to MKG) supported this work.

## Author details

<sup>1</sup>Department of Biology, Drexel University, Philadelphia, Pennsylvania 19104, USA. <sup>2</sup>Department of Biological Sciences, University at Albany - State University of New York, Albany, New York 12222, USA. <sup>3</sup>Institut de Recherche pour le Développement (IRD) and Université Montpellier 1 (UM1), 34394 Montpellier, France. <sup>4</sup>Forensic Biology Section, North Carolina State Crime

Laboratory, Raleigh, North Carolina 27603, USA. <sup>5</sup>Wildlife Conservation Society, Bronx, New York 10460, USA. <sup>6</sup>Institute for Conservation Research, Zoological Society of San Diego, Escondido, CA 92027, USA. <sup>7</sup>Ebo Forest Research Project, BP 3055, Messa, Yaoundé, Cameroon. <sup>8</sup>Wildlife Conservation Society - Tanzania, Ruaha-Katavi Landscape, Iringa, Tanzania. <sup>9</sup>Environment and Rural Development Foundation, BP 189, Buea, Cameroon. <sup>10</sup>Department of Biological Sciences, University of New Orleans, New Orleans, Louisiana 70148, USA. <sup>11</sup>School of Natural Sciences, University of Stirling, Stirling FK9 4LA, UK. <sup>12</sup>Wildlife Conservation Society - Cameroon, BP 3055, Messa, Yaoundé, Cameroon. <sup>13</sup>Departments of Medicine and Microbiology, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA.

Received: 9 September 2014 Accepted: 15 December 2014

Published online: 21 January 2015

## References

1. Caldecott JO, Miles L. United Nations Environment Programme., World Conservation Monitoring Centre. World Atlas of Great Apes and their Conservation. Berkeley: University of California Press, in association with UNEP-WCMC, Cambridge, UK; 2005.
2. Lambert JE. Seasonality, fallback strategies, and natural selection: A chimpanzee and cercopithecoid model for interpreting the evolution of hominin diet. In: Ungar PS, editor. Evolution of the human diet: The known, the unknown, and the unknowable. Oxford; New York: Oxford University Press; 2007. p. 324–43.
3. Stumpf R. Chimpanzees and Bonobos, Diveristy within and between species. In: Campbell CJ, Fuentes A, MacKinnon KC, Panger M, Bearder SK, editors. Primates in perspective. New York: Oxford University Press; 2007. p. 321–44.
4. Gagneux P, Gonder MK, Goldberg TL, Morin PA. Gene flow in wild chimpanzee populations: what genetic data tell us about chimpanzee movement over space and time. *Philos Trans Royal Soc B: Biol Sci.* 2001; 356(1410):889–97.
5. Whiten A, Goodall J, McGrew WC, Nishida T, Reynolds V, Sugiyama Y, et al. Cultures in chimpanzees. *Nature.* 1999;399(6737):682–5.
6. Bjork A, Liu W, Wertheim JO, Hahn BH, Worobey M. Evolutionary history of chimpanzees inferred from complete mitochondrial genomes. *Mol Biol Evol.* 2011;28(1):615–23.
7. Bowden R, MacFie TS, Myers S, Hellenenthal G, Nerrienet E, Bontrop RE, et al. Genomic tools for evolution and conservation in the chimpanzee: *Pan troglodytes ellioti* is a genetically distinct population. *PLoS Genet.* 2012;8(3): e1002504.
8. Prado-Martinez J, Sudmant PH, Kidd JM, Li H, Kelley JL, Lorente-Galdos B, et al. Great ape genetic diversity and population history. *Nature.* 2013; 499(7459):471–5.
9. Gonder MK, Oates JF, Disotell TR, Forstner MR, Morales JC, Melnick DJ. A new west African chimpanzee subspecies? *Nature.* 1997;388(6640):337.
10. Gonder MK, Disotell TR, Oates JF. New genetic evidence on the evolution of chimpanzee populations, and implications for taxonomy. *Int J Primatol.* 2006;27(4):1103–27.
11. Oates JF, Groves CP, Jenkins PD. The type locality of *Pan troglodytes vellerosus* (Gray, 1862), and implications for the nomenclature of West African chimpanzees. *Primates.* 2009;50(1):78–80.
12. Anthony NM, Johnson-Bawe M, Jeffery K, Clifford SL, Abernethy KA, Tutin CE, et al. The role of Pleistocene refugia and rivers in shaping gorilla genetic diversity in central Africa. *Proc Natl Acad Sci.* 2007; 104(51):20432–6.
13. Ghebrial L, Lankester F, Kiyang JA, Akh AE, de Vries S, Foto R, et al. Tracing the origins of rescued chimpanzees reveals widespread chimpanzee hunting in Cameroon. *BMC Ecology* 2010, 10(1):2.
14. Groves CP. Primate Taxonomy. Washington, D.C: Smithsonian Institution; 2001.
15. Grubb P, Butynski TM, Oates JF, Bearder SK, Disotell TR, Groves CP, et al. Assessment of the Diversity of African Primates. *Int J Primatol.* 2003; 24(6):1301–57.
16. Kingdon J. The Kingdon Field Guide to African Mammals. Princeton: Princeton University Press; 2003.
17. Clifford SL, Anthony NM, Bawe-Johnson M, Abernethy KA, Tutin CE, White LJ, et al. Mitochondrial DNA phylogeography of western lowland gorillas (*Gorilla gorilla gorilla*). *Mol Ecol.* 2004;13(6):1551–65.
18. Oates JF, Bergl R, Linder J, Oates JF. New genetic evidence on the evolution of chimpanzee populations, and implications for taxonomy. *Int J Primatol.* 2006;27(4):1103–27.
19. Gonder MK, Disotell TR, Oates JF. New genetic evidence on the evolution of chimpanzee populations, and implications for taxonomy. *Int J Primatol.* 2006;27(4):1103–27.
20. White F. The Vegetation of Africa. Paris: UNESCO; 1983.
21. Maisels F. Mbam Djerem National Park, Cameroon: at the forest's edge. *Canopee.* 2005;27:2–6.
22. Smith TB, Wayne RK, Girman DJ, Bruford MW. A role for ecotones in generating rainforest biodiversity. *Science.* 1997;276(5320):1855–7.
23. Wang J. COANCESTRY: A program for simulating, estimating and analysing relatedness and inbreeding coefficients. *Mol Ecol Resour.* 2011;11(1):141–5.
24. Wang J. Triadic IBD coefficients and applications to estimating pairwise relatedness. *Genet Res.* 2007;89(3):135–53.
25. Maley J. The African rain forest - Main characteristics of changes in vegetation and climate from the Upper Cretaceous to the Quaternary. *Proc R Soc Edinb Biol Sci.* 1996;104:31–73.
26. Nichol JE. Geomorphological evidence and Pleistocene refugia in Africa. *The Geogr J.* 1999;165(1):79–89.
27. Goldberg TL, Ruvolo M. The geographic apportionment of mitochondrial genetic diversity in east African chimpanzees, *Pan troglodytes schweinfurthii*. *Mol Biol Evol.* 1997;14(9):976–84.
28. Dupanloup I, Schneider S, Excoffier L. A simulated annealing approach to define the genetic structure of populations. *Mol Ecol.* 2002; 11(12):2571–81.
29. Rogers AR, Harpending H. Population growth makes waves in the distribution of pairwise genetic differences. *Mol Biol Evol.* 1992;9(3):552–69.
30. Chen C, Durand E, Forbes F, François O. Bayesian clustering algorithms ascertaining spatial population structure: A new computer program and a comparison study. *Molecular Ecology Notes.* 2007;7(5):747–56.
31. Spiegelhalter DJ, Best NG, Carlin BP, Van Der Linde A. Bayesian measures of model complexity and fit. *J Royal Stat Soc: Series B (Statistical Methodology).* 2002;64(4):583–639.
32. Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Mol Ecol.* 2005;14(8):2611–20.
33. Mitchell MW, Rowe B, Sesink Clea PR, Gonder MK. TESS Ad-Mixer: A novel program for visualizing TESS Q matrices. *Conserv Genet Resour.* 2013;5(4):1075–8.
34. Reynolds J, Weir BS, Cockerham CC. Estimation of the coancestry coefficient: basis for a short-term genetic distance. *Genetics.* 1983;105(3):767–79.
35. Slatkin M. A measure of population subdivision based on microsatellite allele frequencies. *Genetics.* 1995;139(1):457–62.
36. Goldstein DB, Ruiz Linares A, Cavalli-Sforza LL, Feldman MW. Genetic absolute dating based on microsatellites and the origin of modern humans. *Proc Natl Acad Sci.* 1995;92(15):6723–7.
37. Excoffier L, Lischer HEL. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour.* 2010;10(3):564–7.
38. Hey J, Nielsen R. Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. *Proc Natl Acad Sci.* 2007;104(8):2785–90.
39. Gonder MK, Locatelli S, Ghebrial L, Mitchell MW, Kujawski JT, Lankester FJ, et al. Evidence from Cameroon reveals differences in the genetic structure and histories of chimpanzee populations. *Proc Natl Acad Sci.* 2011; 108(12):4766–71.
40. Hey J. The divergence of chimpanzee species and subspecies as revealed in multipopulation Isolation-with-Migration analyses. *Mol Biol Evol.* 2010;27(4):921–33.
41. Millicent E, Thoday JM. Effects of disruptive selection. *Heredity.* 1961;16:199–217.
42. Maynard Smith J. Sympatric speciation. *Am Nat.* 1966;100(916):637–50.
43. Endler JA. Geographic Variation, Speciation, and Clines. Princeton, New Jersey: Princeton University Press; 1977.
44. Oates JF. The distribution of Cercopithecus monkeys in West African forests. In: Gautier-Hion A, Boulière F, Kingdon J, editors. A Primate Radiation: Evolutionary Biology of the African Guenons. Cambridge: Cambridge University Press; 1988. p. 79–103.
45. Harcourt AH, Wood MA. Rivers as barriers to primate distributions in Africa. *Int J Primatol.* 2012;33(1):168–83.
46. Haffer J. Speciation in Amazon forest birds. *Science.* 1969;165:131–7.
47. Endler JA. Pleistocene forest refuges: Fact or fancy? In: Prance GT, editor. Biological Diversification in the Tropics. New York: Columbia University Press; 1982. p. 641–57.

48. Mayr E, O'Hara RJ. The biogeographic evidence supporting the Pleistocene forest refuge hypothesis. *Evolution*. 1986;40(1):55–67.
49. Haffer J. Hypotheses to explain the origin of species in Amazonia. *Braz J Biol*. 2008;68(4):917–47.
50. Simard F, Ayala D, Kamdem G, Pombi M, Etouna J, Ose K, et al. Ecological niche partitioning between *Anopheles gambiae* molecular forms in Cameroon: the ecological side of speciation. *BMC Ecol*. 2009;9(1):17.
51. Freedman AH, Thomassen HA, Buermann W, Smith TB. Genomic signals of diversification along ecological gradients in a tropical lizard. *Mol Ecol*. 2010;19(17):3773–88.
52. Smith TB, Thomassen HA, Freedman AH, Sehgal RNM, Buermann W, Saatchi S, et al. Patterns of divergence in the olive sunbird *Cyanomitra olivacea* (Aves: Nectariniidae) across the African rainforest-savanna ecotone. *Biol J Linn Soc*. 2011;103(4):821–35.
53. Sesink Clee PR, Abwe EE, Ambahe RD, Anthony NM, Fotso R, Locatelli S, et al. Chimpanzee population genetic structure in Cameroon and Nigeria is associated with habitat variation that may be lost under climate change. *BMC Evolutionary Biology*. 2015;15:2.
54. Wiens JJ. Speciation and ecology revisited: Phylogenetic niche conservatism and the origin of species. *Evolution*. 2004;58(1):193–7.
55. Blair ME, Sterling EJ, Dusch M, Raxworthy CJ, Pearson RG. Ecological divergence and speciation between lemur (*Eulemur*) sister species in Madagascar. *J Evol Biol*. 2013;26(8):1790–801.
56. Mitchell MW, Locatelli S, Sesink Clee PR, Thomassen HA, Gonder MK. Environmental variation and rivers govern the structure of chimpanzee genetic diversity in a biodiversity hotspot. *BMC Evolutionary Biology*. 2015;15:1.
57. Wang IJ, Bradburd GS. Isolation by environment. *Mol Ecol*. 2014;23(23):5649–5662.
58. Goldberg TL. *Genetics and Biogeography of East African Chimpanzees, Pan troglodytes schweinfurthii*. Cambridge, Massachusetts: Harvard University; 1996.
59. Keele BF, Van Heuverswyn F, Li Y, Bailes E, Takehisa J, Santiago ML, et al. Chimpanzee Reservoirs of Pandemic and Nonpandemic HIV-1. *Science*. 2006;313(5786):523–6.
60. Walsh PS, Metzger DA, Higuchi R. Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. *Biotechniques*. 1991;10(4):506–13.
61. Quantifiler® Human DNA Quantification Kit and Quantifiler® Y Human Male DNA Quantification Kit Users Manual. Foster City, CA: Applied Biosystems®; 2006.
62. Neel C, Etienne L, Li Y, Takehisa J, Rudicell RS, Bass IN, et al. Molecular epidemiology of simian immunodeficiency virus infection in wild-living gorillas. *J Virol*. 2010;84(3):1464–76.
63. Ghebranious N, Vaske D, Yu A, Zhao C, Marth GT, Weber JL. STRP Screening Sets for the human genome at 5 cM density. *BMC Genomics*. 2003;4(1):6.
64. Becquet C, Patterson N, Stone AC, Przeworski M, Reich D. Genetic structure of chimpanzee populations. *PLoS Genet*. 2007;3(4):e66.
65. Pasqualotto AC, Denning DW, Anderson MJ. A cautionary tale: Lack of consistency in allele sizes between two laboratories for a published multilocus microsatellite typing system. *J Clin Microbiol*. 2007;45(2):522–8.
66. Bonin A, Bellemain E, Bronken Eidesen P, Pompanon F, Brochmann C, Taberlet P. How to track and assess genotyping errors in population genetics studies. *Mol Ecol*. 2004;13(11):3261–73.
67. Antao T, Lopes A, Lopes RJ, Beja-Pereira A, Luikart G. LOSITAN: A workbench to detect molecular adaptation based on a Fst-outlier method. *BMC Bioinformatics*. 2008;9(1):323.
68. Beaumont MA, Nichols RA. Evaluating loci for use in the genetic analysis of population structure. *Proc Royal Soc B: Biol Sci*. 1996;263(1377):1679–26.
69. Guo SW, Thompson EA. Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics*. 1992;48(2):361–72.
70. Lynch M. Estimation of relatedness by DNA fingerprinting. *Mol Biol Evol*. 1988;5(5):584.
71. Queller DC, Goodnight KF. Estimating relatedness using molecular markers. *Evolution*. 1989;43(2):258–75.
72. Li CC, Weeks DE, Chakravarti A. Similarity of DNA fingerprints due to chance and relatedness. *Hum Hered*. 1993;43(1):45–52.
73. Ritland K. Estimators for pairwise relatedness and inbreeding coefficients. *Genet Res*. 1996;67(2):175–85.
74. Lynch M, Ritland K. Estimation of pairwise relatedness with molecular markers. *Genetics*. 1999;152(4):1753–66.
75. Wang J. An estimator for pairwise relatedness using molecular markers. *Genetics*. 2002;160(3):1203–15.
76. Milligan BG. Maximum-likelihood estimation of relatedness. *Genetics*. 2003;163(3):1153–67.
77. Bandelt HJ, Forster P, Rohl A. Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol*. 1999;16:37–48.
78. Posada D, Crandall KA. Intraspecific gene genealogies: trees grafting into networks. *Trends Ecol Evol*. 2001;16(1):37–45.
79. Patterson N, Price AL, Reich D. Population structure and eigenanalysis. *PLoS Genet*. 2006;2(12):e190.
80. Durand E, Jay F, Gaggiotti OE, François O. Spatial inference of admixture proportions and secondary contact zones. *Mol Biol Evol*. 2009;26(9):1963–73.
81. Jakobsson M, Rosenberg NA. CLUMPP: A cluster matching and permutation program for dealing with label and switching and multimodality in analysis of population structure. *Bioinformatics*. 2007;23(14):6.
82. Rosenberg NA. DISTRUCT: A program for the graphical display of population structure. *Mol Ecology Notes*. 2004;4(1):2.
83. Tishkoff SA, Reed FA, Friedlaender FR, Ehret C, Ranciaro A, Froment A, et al. The genetic structure and history of Africans and African Americans. *Science*. 2009;324(5930):10.
84. Garza JC, Williamson EG. Detection of reduction in population size using data from microsatellite loci. *Mol Ecol*. 2001;10(2):305–18.
85. Sun JX, Mullikin JC, Patterson N, Reich DE. Microsatellites are molecular clocks that support accurate inferences about history. *Mol Biol Evol*. 2009;26(5):1017–27.
86. Szpiech ZA, Jakobsson M, Rosenberg NA. ADZE: A rarefaction approach for counting alleles private to combinations of populations. *Bioinformatics*. 2008;24(21):2498–504.
87. Wegmann D, Excoffier L. Bayesian inference of the demographic history of chimpanzees. *Mol Biol Evol*. 2010;27(6):1425–35.
88. Hey J. IMA documentation: Using the IMA program. Camden, New Jersey: Rutgers University; 2009.
89. Soares P, Ermini L, Thomson N, Mormina M, Rito T, Röhl A, et al. Correcting for purifying selection: An improved Human mitochondrial molecular clock. *The Am J Human Gen*. 2009;84(6):740–59.
90. Gage TB. The comparative demography of primates: With some comments on the evolution of life histories. *Annual Rev Anthropol*. 1998;27:197–221.
91. Kalinowski ST, Taper ML, Marshall TC. Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Mol Ecol*. 2007;16(5):1099–106.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

