Supporting Information

Gonder et al. 10.1073/pnas.1015422108

SI Methods

Dataset Preparation. DNA from chimpanzees housed at the Limbe Wildlife Centre (LWC) in Limbe, Cameroon was isolated from whole blood samples; and DNA extract yields were quantified as previously reported (1). DNA samples were transported from Cameroon to the United States in full compliance with Convention on International Trade in Endangered Species and Centers for Disease Control export and import regulations. This research was carried out with Institutional Animal Care and Use Committee approval from the University at Albany, State University of New York.

We produced microsatellite genotype profiles of 45 chimpanzees drawn from a subset of 310 microsatellite loci (2). This subset of autosomal loci (n = 27) we analyzed here has been shown previously (2) to have considerable power for: (*i*) distinguishing bonobos from chimpanzees, (*ii*) classifying individual chimpanzees into geographically disjunct populations (western, central, and eastern) that correspond to three of the recognized chimpanzee subspecies, and (*iii*) reliably detecting hybrid individuals. Details about the loci included in this study are given in Table S1.

PCR reactions were performed using the Qiagen Multiplex PCR kit (Qiagen, Valencia, CA) in Eppendorf Mastercyclers (Eppendorf, Westbury, NY), and were carried out using 1 ng of DNA for each reaction following the manufacturer's protocol. Microsatellite genotyping was carried out in four multiplex PCR reactions using the ABI G5 dye set (Applied Biosystems, Foster City, CA). Each multiplex PCR product was analyzed on an ABI 3130 capillary array genetic analyzer (Applied Biosystems). Fragment sizes were determined against Genescan 600 Liz size standard (Applied Biosystems). Allele sizes were determined using GENEMAPPER ID version 2.7 software (Applied Biosystems). Heterozygous and homozygous loci were scored a minimum of two and a maximum of four times by independent PCRs for each individual.

Dataset Integration. These allele size data were integrated with allele size data from individuals previously genotyped for the same loci reported by Becquet et al. (2). We corrected for allele size differences due to apparatus and protocol discrepancies (3) by retyping a subset of individuals (n = 10) reported previously (2). Table S1 lists marker and allele size adjustments made for each locus included in this study. Previous studies in humans have shown that such integration yields datasets suitable for making inferences about population history (4). Of the original 130 individuals genotyped by us and those reported by Becquet et al. (2), we removed the following individuals from subsequent analyses: 27 captive-born chimpanzees, two chimpanzees listed as wild-born but with untraceable/unreliable International Species Information System (ISIS) records, and one LWC individual that was missing alleles for >15% of the loci (despite repeated attempts to produce allele sizes suitable for scoring). The combined dataset contains genotype profiles for six bonobos and 94 wild-born chimpanzees with estimated origins from the following locations: Cameroon (n = 45) (1), western Africa (n = 31), central Africa (n = 12), and eastern Africa (n = 6) (2). Allele sizes newly generated for 45 LWC chimpanzees with estimated origins in Cameroon (1) are listed in Table S2.

Data Analysis. ***Cluster analysis.*** Population structure and individual ancestry were examined using a Bayesian clustering approach implemented in the STRUCTURE Version 2.3 software package (5). This program estimates the shared population history of individuals based solely on their genotypes under a model of Hardy–Weinberg equilibrium and linkage equilibrium in the ancestral populations, thereby making no a priori assumptions regarding population classifications. STRUCTURE 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. STRUCTURE runs were performed: (*i*) with a model that allows individuals to have ancestry in multiple populations ("admixture mode"); (*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 $K = 1$ to $K = 10$ were performed for all datasets. STRUCTURE outputs were processed with CLUMPP (6); and a G-statistic >99% was used to assign groups of runs to a common clustering pattern. CLUMPP output for each K value was plotted with DISTRUCT (7). We used a combination of methods to infer a maximum number of chimpanzee populations (K_{MAX}) including, (*i*) the K value at which the posterior probability distribution (PPD) values reached an apex before decreasing (5), (*ii*) high stability of clustering patterns between runs, (*iii*) the K_{MAX} value at which $K_{\mathrm{MAX}} + 1$ no longer split the cluster distinguished by K_{MAX} (4), (*iv*) correspondence between maximum PPD values from STRUCTURE runs and significant eigenvectors recovered by PCA, and (*v*) calculating an adhoc statistic, ΔK (8), as estimated by the STRUCTURE HARVESTER software package version 0.56.4 (9).

Principal components analysis (PCA). The EIGENSOFT software package (10) was used to perform PCA on individual genotypes to identify significantly different populations. We developed a script in MATLAB (The MathWorks, Natick, MA) that converted the 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). This file was processed in *SmartPCA*, which produced eigenvectors and eigenvalues. The statistical significance of each eigenvector was tested by 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 (10). All analyses using EIGENSOFT were performed blinded to a priori population labels.

Allele frequency differentiation. Three measures of population genetic differentiation were calculated using the ARLEQUIN 3.5 software package (11): D^2, R_{ST}, and $(\delta\mu)^2$. The D^2 (12) 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} (13) and $(\delta\mu)^2$ (14) 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 are also more sensitive for detecting ancient population separations (4). 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} (13), whereas $(\delta\mu)^2$ is based on differences in the means of microsatellite allele sizes (14). Recent work has shown convincing evidence that the loci typed for this study appear to follow the SMM in both chimpanzees and bonobos (2, 15).

D^2 calculations were completed on untransformed allele size calls. Because R_{ST} and $(\delta\mu)^2$ assume the SMM, allele sizes were transformed to repeat size units before analysis in ARLEQUIN (11). Allele sizes were transformed 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), the next allele in the data were scored as $(n + x + 1)$. Individuals with $\geq$25% membership ($n = 7$) in more than one ancestral cluster from the STRUCTURE analysis (Fig. 2*A*) were treated as potential hybrids and excluded from population pairwise genetic distance calculations. Each pairwise genetic distance calculation was determined by 100,000 replications in ARLEQUIN. The significance of these pairwise population genetic distance calculations were evaluated by a significance test at $P < 0.05$.

Recent work has shown that microsatellite loci can be used to build robust phylogenies (16). We constructed phylogenetic trees using three measures of population genetic differentiation D^2 (12), R_{ST} (13), and $(\delta\mu)^2$ (14) described above. Trees based on D^2 were included here, because D^2 gives more reliable phylogenetic results compared with R_{ST} and $(\delta\mu)^2$ (16, 17), in cases where microsatellite alleles do not follow a stepwise mutation process, where other evolutionary forces such as genetic drift and/or gene flow have stronger influence on shaping diversity than mutation, and when the number of loci is relatively small, as is the case in this study (16).

For the D^2 analysis, we resampled the dataset 10,000 times to generate multiple distance matrices. We constructed unrooted neighbor joining phylograms for these matrices using the PHYLIP software package, version 3.5 (18) with the *Neighbor* program. *Consense* was used to obtain a consensus tree that was then used by *Contml* to generate branch lengths from allele frequency data using a maximum likelihood algorithm. For the R_{ST} and $(\delta\mu)^2$ analyses, single trees with branch lengths were produced using *Neighbor* from population pairwise differentiation values calculated with ARELQUIN. *Consense* calculated bootstrap values. Trees were plotted using the GENEIOUS software package (Version 4.8; ref. 19), and branches with at least 70% support were labeled.

Population divergence times. Calculating population divergence times is challenging using microsatellites, especially when the time to most recent common ancestor (T_{MRCA}) might be quite ancient (16). However, the microsatellite loci included in this study have been shown to be accurate molecular clocks for *Pan* compared with autosomal resequencing data (15). We calculated population divergence times based on $(\delta\mu)^2$ (14) assuming a mutation rate (μ) of 1.6×10^{-4}, the median μ for these loci reported by Wegemann and Excoffier (20). We further assumed a 20-y generation time (g), which is consistent with studies from the wild (21) and has been used in recent studies (22, 23). Population splitting times were calculated using the method described by Goldstein et al. (14): $(\delta\mu)^2 \times g/2\mu$.

Dataset validation. We evaluated the reliability of the analyses for the dataset reported here using three methods. First, our analyses were based on a dataset containing only 9% of the loci reported by Becquet et al. (2), and unlike that study, this dataset included only wild-born chimpanzees. We examined how well this reduced dataset captured the genetic structure reported by Becquet et al. (2), including bonobos and chimpanzees from Upper Guinea (western), central Africa, and eastern Africa, but excluding those from Cameroon. Second, our analyses are also based on unequal sample sizes for chimpanzees from different regions of Africa. In particular, the samples size of chimpanzees reported to originate from eastern Africa ($n = 6$) was much smaller than for the other three regions: Upper Guinea ($n = 31$), Cameroon ($n = 45$), and central Africa ($n = 12$). Unequal population sample sizes can greatly bias estimates of genetic differentiation, as well as the numbers of distinct and private alleles found per locus in different populations (24, 25). We used two rarefaction procedures to explore the possibility that our results were the result of unequal sample sizes instead of real population structure. First, we constrained sample sizes to be equal in all populations identified in the full dataset by creating randomized data subsets ($n = 10$) that included six each of chimpanzees from Upper Guinea, the Gulf of Guinea region, southern Cameroon, central Africa, eastern Africa, and bonobos. We carried out cluster analyses including, generating STRUCTURE (5) and PCA plots (10), along with reevaluating K_{MAX} for each randomized dataset. Finally, we applied a rarefaction procedure developed by Kalinowski (25) for counting alleles private to combinations of populations corrected for unequal sample sizes between populations as implemented in the ADZE software package (26).

SI Results and Discussion

The PCA (Fig. S3) for the 27-locus genotype profiles including only wild-born Upper Guinea (western), central, and eastern chimpanzees recovered three significant principal components (PCs). These axes recapitulate the major population clusters reported previously by Becquet et al. (2). PC 1 separated bonobos from chimpanzees, extracting 36.7% of the observed variation. PC 2 separated Upper Guinea chimpanzees from chimpanzees occupying equatorial Africa, extracting for 44.1% of the genetic variation. PC 3 separated the central and eastern populations, accounting for 19.2% of the variation. We did not detect the fourth axis of variation reported by Becquet et al. (2), possibly due to the lack of captive-born individuals reported as "hybrids" by these authors. Alternatively, the reduced number of loci may lack the power to resolve subtle population differences at higher values of K. Table S3 compares R_{ST} and F_{ST} values between the full dataset including 310 loci as reported by Becquet et al. (2) compared with these including only bonobos along with the Upper Guinea (western), central, and eastern populations for the 27-locus dataset. Allele frequency differentiation values for the 27-locus dataset (this study) versus the 310-locus dataset reported by Becquet et al. (2) were highly correlated ($R_{ST}\ r^2 = 0.96$, $P < 0.5$; $F_{ST}\ r^2 = 0.96$, $P < 0.5$). Based on these findings, we concluded that the suite of microsatellite loci included in this study adequately captured the population structure reported by Becquet et al. (2) for Upper Guinea (western), central, and eastern chimpanzees. Consequently, the 27-locus dataset should yield a reliable picture regarding how chimpanzees from Cameroon contribute to the population structure of this species.

The population structure inferred for the ten randomized datasets of equal population size was highly consistent across runs. Each dataset returned identical K_{MAX} values as well as the same number of significant PCs by PCA. Fig. S4 *A–C* shows results for the inferred population structure for one of these randomized datasets composed of equal sized populations. The cluster analysis in STRUCTURE (Fig. S4*B*) revealed that K_{MAX} was 5 using both the PPD and ΔK criteria (Fig. S4*C*), instead of $K_{MAX} = 4$ or $K_{MAX} = 6$ for the full dataset including all 100 individuals. Also in contrast to the full dataset, chimpanzees originating in eastern Africa were distinguished from the others at lower values of K ($K = 4$) by STRUCTURE analysis than for the full dataset, and those from southern Cameroon clustered with chimpanzees from other parts of central Africa at $K_{MAX} = 5$. The PCA (Fig. S4*A*) captured four significant PCs that distinguished five significantly different populations of chimpanzees, whereas the PCA for the full dataset recovered six significantly different populations. The difference between the datasets containing equal population sizes compared with the full dataset was that the equal population size datasets did not recover a statistically significant PC that distinguished chimpanzees originating from southern Cameroon from chimpanzees originating from other areas of central Africa, as was found by PCA of the full dataset. We concluded based on these results that oversampling of chimpanzees from Cameroon did not greatly bias the results we obtained from the full dataset. However, unequal population sample sizes appear to have influenced discerning subtle population structure across equatorial Africa. In particular, the full dataset may be an underestimate of the allele frequency differentiation that separates central and east African chimpanzees, or alternatively, may be an overestimate of the subtle distinction that separates chimpanzees originating from southern Cameroon from those originating elsewhere in central Africa.

Fig. S5 *A* and *B* shows allele richness by region and private alleles found in each population corrected for unequal population sample size. Allele richness did not vary considerably between regions, whereas private alleles occurred more frequently among chimpanzees originating in central and eastern Africa compared with Upper Guinea, the Gulf of Guinea region, or southern Cameroon. Fig. S6 shows shared private alleles between various population pairs extrapolated to equal population size. Intriguingly, more shared private alleles were found among the central-eastern population pair and among the Gulf of Guinea/southern Cameroon population pair than any other population combinations. The relatively high number of shared private alleles between central and eastern chimpanzees taken together with their low allele frequency differentiation values (Table 1) suggests that these populations probably share much of their recent genetic history and might be characterized by ongoing gene flow as suggested by some previous studies (27). In contrast, the relatively high number of shared private alleles between the Gulf of Guinea/southern Cameroon chimpanzee population pair considered jointly with their higher allele frequency differentiation (Table 1) suggests the possibility that these two lineages may be characterized by a pattern of recent introgression between lineages, as expected based limited evidence reported in previous studies (1, 28, 29).

1. Ghobrial L, et al. (2010) Tracing the origins of rescued chimpanzees reveals widespread chimpanzee hunting in Cameroon. *BMC Ecol* 10:2.
2. Becquet C, Patterson N, Stone AC, Przeworski M, Reich D (2007) Genetic structure of chimpanzee populations. *PLoS Genet* 3:e66.
3. Pasqualotto AC, Denning DW, Anderson MJ (2007) A cautionary tale: Lack of consistency in allele sizes between two laboratories for a published multilocus microsatellite typing system. *J Clin Microbiol* 45:522–528.
4. Tishkoff SA, et al. (2009) The genetic structure and history of Africans and African Americans. *Science* 324:1035–1044.
5. Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155:945–959.
6. Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* 23:1801–1806.
7. Rosenberg NA (2004) DISTRUCT: A program for the graphical display of population structure. *Mol Ecol Notes* 4:137–138.
8. Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol* 14:2611–2620.
9. Earl DA (2009) Structure Harvester (Department of Ecology and Evolution Biology, University of California, Los Angeles), Version 0.3.
10. Patterson N, Price AL, Reich D (2006) Population structure and eigenanalysis. *PLoS Genet* 2:e190.
11. Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour* 10:564–567.
12. Reynolds J, Weir BS, Cockerham CC (1983) Estimation of the coancestry coefficient: basis for a short-term genetic distance. *Genetics* 105:767–779.
13. Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. *Genetics* 139:457–462.
14. Goldstein DB, Ruiz Linares A, Cavalli-Sforza LL, Feldman MW (1995) Genetic absolute dating based on microsatellites and the origin of modern humans. *Proc Natl Acad Sci USA* 92:6723–6727.
15. Sun JX, Mullikin JC, Patterson N, Reich DE (2009) Microsatellites are molecular clocks that support accurate inferences about history. *Mol Biol Evol* 26:1017–1027.
16. Takezaki N, Nei M (2008) Empirical tests of the reliability of phylogenetic trees constructed with microsatellite DNA. *Genetics* 178:385–392.
17. Balloux F, Lugon-Moulin N (2002) The estimation of population differentiation with microsatellite markers. *Mol Ecol* 11:155–165.
18. Felsenstein J (1989) PHYLIP–Phylogeny Inference Package (Version 3.2). *Cladistics* 5: 164–166.
19. Drummond A, et al. (2009) Geneious, (Biomatters Limited, Auckland, New Zealand), Version 4.8.
20. Wegmann D, Excoffier L (2010) Bayesian inference of the demographic history of chimpanzees. *Mol Biol Evol* 27:1425–1435.
21. Gage TB (1998) The comparative demography of primates: with some comments on the evolution of life histories. *Annu Rev Anthropol* 27:197–221.
22. Caswell JL, et al. (2008) Analysis of chimpanzee history based on genome sequence alignments. *PLoS Genet* 4:e1000057.
23. Hey J (2010) The divergence of chimpanzee species and subspecies as revealed in multipopulation isolation-with-migration analyses. *Mol Biol Evol* 27:921–933.
24. Leberg PL (2002) Estimating allelic richness: effects of sample size and bottlenecks. *Mol Ecol* 11:2445–2449.
25. Kalinowski ST (2004) Counting alleles with rarefaction: Private alleles and hierarchical sampling designs. *Conserv Genet* 5:539–543.
26. Szpiech ZA, Jakobsson M, Rosenberg NA (2008) ADZE: a rarefaction approach for counting alleles private to combinations of populations. *Bioinformatics* 24:2498–2504.
27. Fischer A, Pollack J, Thalmann O, Nickel B, Pääbo S (2006) Demographic history and genetic differentiation in apes. *Curr Biol* 16:1133–1138.
28. Gagneux P, Gonder MK, Goldberg TL, Morin PA (2001) Gene flow in wild chimpanzee populations: what genetic data tell us about chimpanzee movement over space and time. *Philos Trans R Soc Lond B Biol Sci* 356:889–897.
29. Gonder MK, Disotell TR, Oates JF (2006) New genetic evidence on the evolution of chimpanzee populations, and implications for taxonomy. *Int J Primatol* 27:1103–1127.

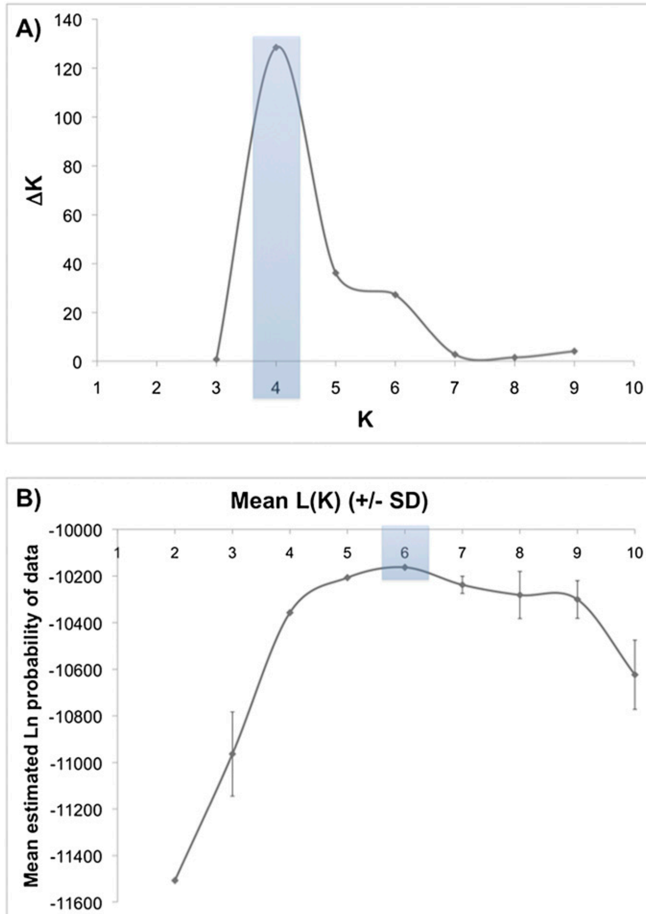


Fig. S1. Estimates of K_{MAX} from 50 independent STRUCTURE (1) runs for each value of *K* from 1 to 10. (*A*) Estimated ΔK (2) values calculated with STRUCTURE HARVESTER (3) for the *Pan* dataset. (*B*) Estimated Ln probability of data [Ln P(D)].

1. Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155:945–959.
2. Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol* 14:2611–2620.
3. Earl DA (2009) Structure Harvester (Department of Ecology and Evolution Biology, University of California, Los Angeles), Version 0.3.

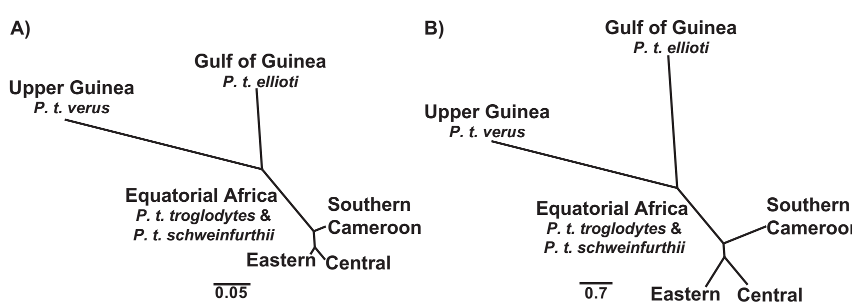


Fig. S2. Neighbor-Joining phylograms based on the R_{ST} genetic distances (1) (*A*), and $(\delta\mu)^2$ genetic distances (2) (*B*).

1. Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. *Genetics* 139:457–462.
2. Goldstein DB, Ruiz Linares A, Cavalli-Sforza LL, Feldman MW (1995) Genetic absolute dating based on microsatellites and the origin of modern humans. *Proc Natl Acad Sci USA* 92: 6723–6727.

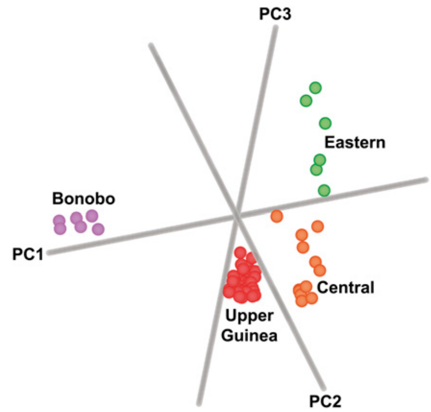


Fig. S3. PCA plot for western, central and eastern chimpanzees from Becquet et al. (1) for the 27-locus microsatellite dataset included in this study.

1. Becquet C, Patterson N, Stone AC, Przeworski M, Reich D (2007) Genetic structure of chimpanzee populations. *PLoS Genet* 3:e66.

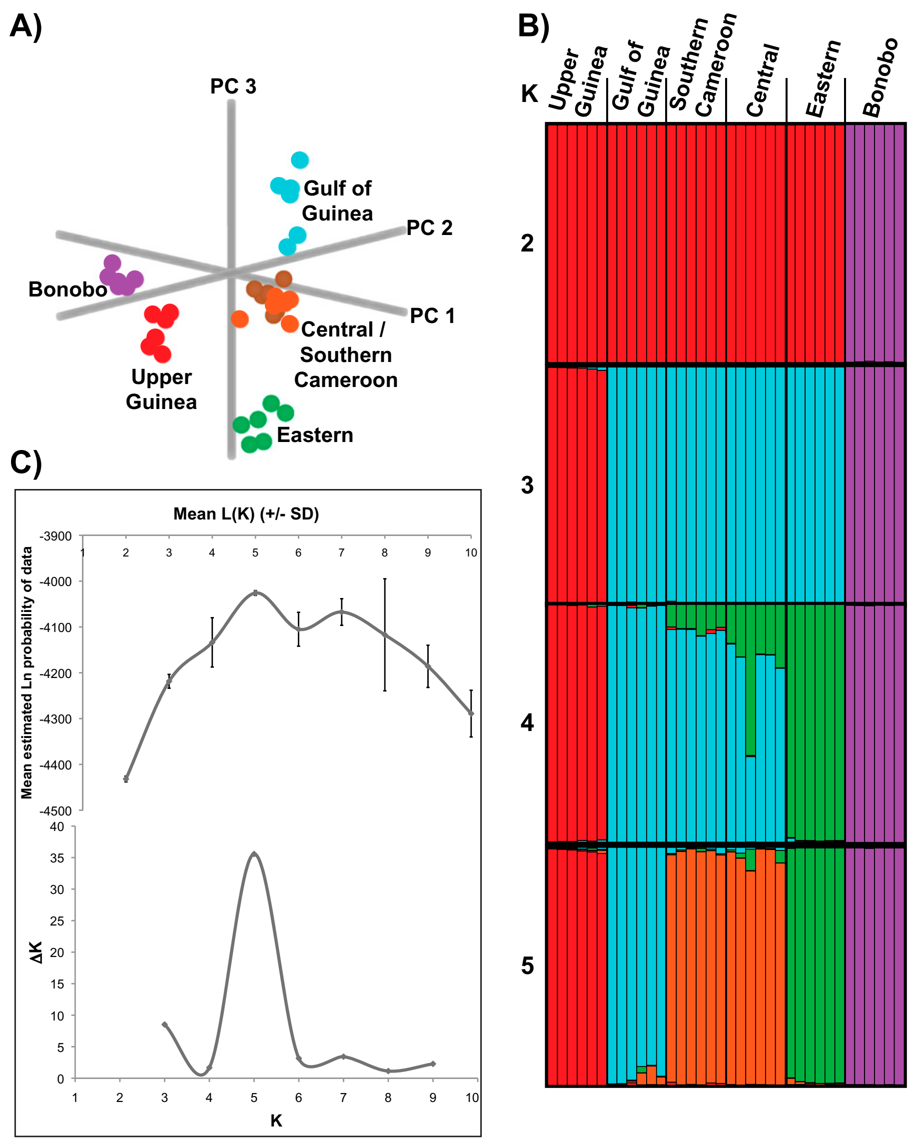
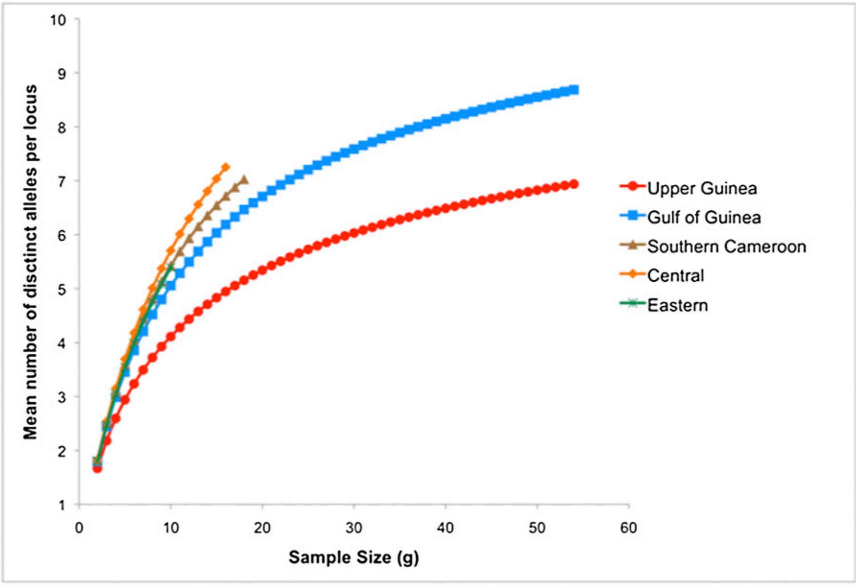


Fig. S4. Population structure inferred for the constrained sample size dataset. Six individuals from each of the geographically separated populations inferred from the full dataset were included in this analysis. (*A*) PCA Plot. (*B*) STRUCTURE plots for $K = 2–5$. (*C*) Estimated ΔK and Ln probability of data [Ln P(D)].

A)



B)

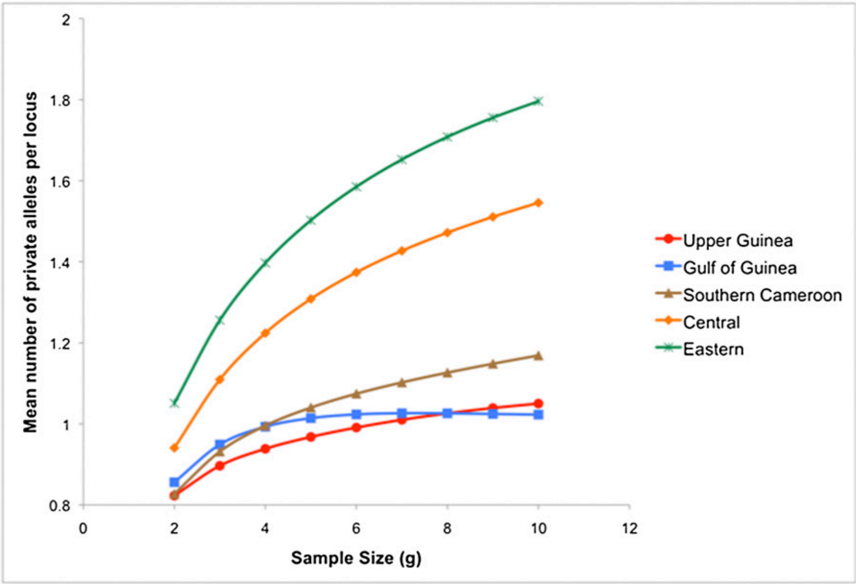


Fig. S5. Inference of allele richness and private alleles in chimpanzee populations as implemented in ADZE software package (1). (*A*) Mean number of distinct alleles found in different chimpanzee populations. (*B*) Mean number of private alleles found in different chimpanzee populations.

1. Szpiech ZA, Jakobsson M, Rosenberg NA (2008) ADZE: a rarefaction approach for counting alleles private to combinations of populations. *Bioinformatics* 24:2498–2504.

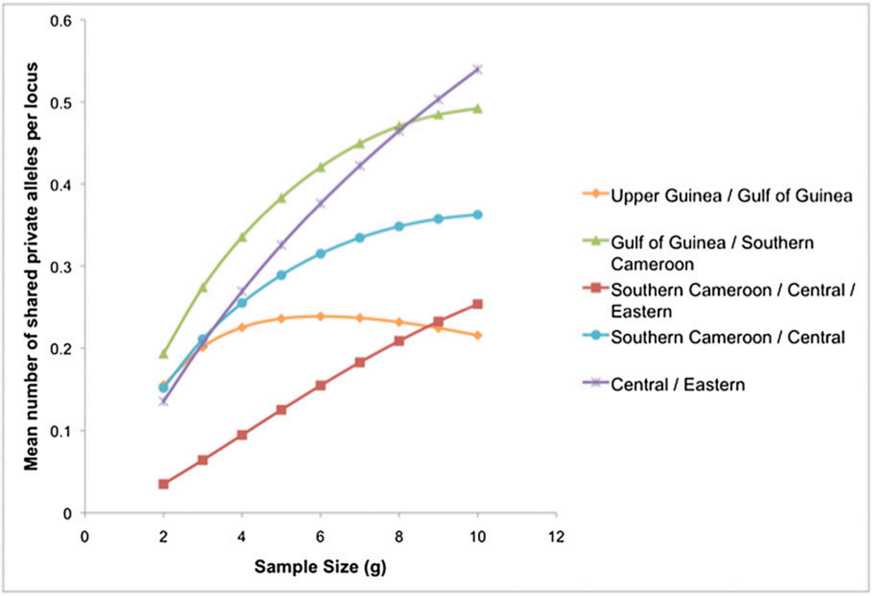


Fig. S6. Inference of uniquely shared alleles between various chimpanzee populations as implemented in ADZE software package (1).

1. Szpiech ZA, Jakobsson M, Rosenberg NA (2008) ADZE: a rarefaction approach for counting alleles private to combinations of populations. *Bioinformatics* 24:2498–2504.

Table S1. Information for markers used in this study

Name*	Chromosome	Locus	Repeat number	No of. alleles†	Allele size range		Calibration bp shift‡
GATA91H06M	12	D12S1301	4	22	87	146	+8
ATA27A06P	12	D12S1042	3	18	104	167	+12
GATA29A01	6	D6S1959	4	12	134	182	+5
GATA11A06	18	D18S542	4	35	156	210	+4
GATA104	7		4	32	169	227	–
GATA176C01	2	D2S2972	4	37	198	279	+5
GGAA4B09N	3	D3S2403	4	18	204	269	–
AGAT120	22	SNP343411	4	22	251	293	+1
ATTT030	6		4	12	104	142	+6
GGAA3A07M	1	D1S1612	4	27	123	189	+3
TCTA017M	9		4	27	146	209	+4
GATA25A04	17	D17S1299	4	28	172	226	+6
GATA8C04	17	D17S974	4	12	173	217	+1
GATA164B08P	3	D3S4545	4	41	193	258	+13
GATA28F03	4	D4S3248	4	14	223	271	–
GATA129D11N	21	D21S2052	4	11	109	153	+9
GATA43A04	1	D1S1653	4	36	107	229	+4
GATA116B01N	2	D2S2952	4	23	142	207	+3
UT7544	19	D19S559	4	20	136	177	−1
GATA129H04	1	D1S3721	4	39	159	260	+3
GATA61E03	6	D6S1051	4	15	207	268	+7
GATA71H05	16	D16S769	4	26	242	300	+6
GGAA21G11L	14	D14S617	4	18	111	201	+6
GATA14E09	8	D8S2324	4	16	180	220	+6
GATA50G06	15	D15S643	4	24	187	287	+3
GATA43C11	7	D7S1804	4	22	196	298	–
GATA7F05	3	D3S3039	4	17	246	312	–

*27 microsatellite loci located on the autosomes typed for this study from Becquet et al. (1).
†Allele number includes all raw allele calls from Becquet et al. (1) and this study.
‡The number of base pairs added to alleles to match the allele sizes reported in Becquet et al. (1).

1. Becquet C, Patterson N, Stone AC, Przeworski M, Reich D (2007) Genetic structure of chimpanzee populations. *PLoS Genet* 3:e66.

Table S2. Microsatellite genotypes of chimpanzees newly generated for this study

Sample ID	GGA A3A 07M	GATA 129H 04	GAT A43 A04	GAT A116 B01N	GATA 176 C01	GATA 164B 08P	GGA A4B 09N	GAT A7F 05	ATA 28F 03	AT TT0 30	GAT A29 A01	GAT A61 E03	GAT A43 C11	GA TA1 04	GAT A14 E09	TCT A01 7M	ATA 27A 06P	GATA 91H 06M	GGAA 21G 11L	GATA 50G 06	GAT A71 H05	GAT A8C 04	GAT A25 A04	GAT A11 A06	UT7 544	GATA 129D 11N	AGA T120
LWC001	145	194	135	146	202	221	215	250	255	116	162	231	226	180	192	193	131	102	135	215	278	189	200	176	144	117	275
	169	204	139	154	218	229	253	296	259	119	174	231	242	188	192	193	131	106	143	231	278	209	211	180	144	125	287
LWC002	153	194	135	146	222	217	215	262	255	112	150	227	242	180	192	193	131	110	139	207	250	209	180	172	144	117	267
	169	196	139	150	226	221	227	296	255	123	162	231	246	188	192	193	131	110	139	243	282	213	180	172	152	117	267
LWC003	153	212	135	150	218	217	253	262	263	112	162	227	226	188	188	185	122	102	139	207	282	213	192	180	152	113	267
	169	216	139	154	251	217	253	296	263	123	166	235	238	200	192	193	131	106	151	235	282	213	204	204	152	121	275
LWC004	165	194	147	154	255	217	215	280	251	119	146	215	230	188	184	169	128	106	135	215	278	189	211	176	144	125	287
	165	196	147	161	259	254	227	296	259	119	174	227	250	220	192	197	128	106	151	231	278	213	211	196	152	133	287
LWC005	157	190	135	154	251	246	219	262	255	116	162	227	214	188	184	177	131	114	139	215	242	209	196	180	144	125	275
	161	190	135	158	255	267	227	296	259	116	174	231	250	196	192	177	134	114	143	227	242	217	196	180	144	125	279
LWC006	161	200	147	146	218	250	219	270	259	116	162	231	214	177	184	177	131	94	139	239	278	189	204	176	140	117	267
	165	212	167	146	247	254	253	296	259	116	174	231	222	208	196	189	131	98	143	243	282	213	204	196	148	125	271
LWC007	161	200	143	146	214	254	227	262	251	116	146	223	222	188	192	189	131	98	139	211	278	189	204	184	140	115	267
	169	200	147	146	222	283	227	296	259	119	146	231	222	196	192	189	143	138	139	235	278	209	204	196	164	117	283
LWC008	165	196	135	150	218	254	−9	−9	263	116	154	219	230	188	196	181	131	110	143	−9	−9	217	200	196	144	117	267
	173	196	139	158	218	258	−9	−9	263	123	174	219	246	196	200	193	152	110	143	−9	−9	217	208	196	152	125	267
LWC009	157	212	135	154	218	233	215	270	251	116	146	231	208	188	184	177	122	106	139	215	246	189	196	188	144	117	275
	165	216	139	173	251	267	227	296	267	123	162	239	208	188	192	177	131	110	143	239	278	213	204	196	160	125	279
LWC010	145	224	143	150	247	217	215	254	247	116	166	227	230	208	192	189	122	98	139	223	282	189	192	184	144	117	271
	145	228	147	158	247	250	219	296	247	116	174	231	254	208	196	189	122	110	139	235	282	205	192	196	148	117	271
LWC011	161	208	139	150	210	250	215	−9	251	112	166	227	222	177	184	177	131	106	143	−9	278	209	180	172	152	117	267
	169	212	143	169	247	267	215	−9	259	123	174	231	226	216	196	193	131	110	151	−9	278	209	204	196	160	125	267
LWC012	157	208	135	150	222	250	219	262	251	116	170	231	226	180	184	161	131	102	139	215	278	213	192	172	152	117	267
	165	212	147	158	226	263	224	296	259	123	174	231	250	188	196	185	134	110	139	219	282	217	204	196	152	125	275
LWC013	145	240	135	165	247	205	219	250	243	108	134	227	218	188	192	185	125	102	139	203	262	209	192	−9	144	109	271
	153	248	145	173	247	233	224	296	251	116	146	239	226	208	196	185	128	106	163	207	278	213	200	−9	160	121	279
LWC014	165	194	139	150	202	221	215	270	251	112	162	215	226	188	184	189	131	102	143	215	246	189	196	172	144	113	267
	165	208	139	173	222	246	253	296	251	123	174	227	246	188	192	197	131	106	147	215	278	217	196	196	160	125	283
LWC015	157	216	135	146	222	233	215	258	255	116	146	217	222	188	184	189	125	102	159	211	−9	209	192	176	144	−9	275
	161	248	139	165	243	267	221	296	263	119	170	223	226	208	200	197	128	110	163	223	−9	213	200	180	160	−9	287
LWC016	153	212	139	154	214	254	219	296	263	116	146	227	208	173	188	189	122	91	139	235	282	205	196	188	144	113	275
	161	216	143	154	259	279	224	296	263	116	146	231	208	212	208	189	122	110	143	235	282	209	204	188	156	117	275
LWC017	145	194	139	150	214	221	215	254	259	112	162	215	230	180	184	193	125	102	151	215	278	189	180	196	152	117	271
	169	194	147	169	218	246	227	296	267	119	166	227	250	200	184	193	131	114	151	243	278	209	200	196	156	125	275
LWC018	169	200	135	146	210	217	219	284	247	116	162	227	246	188	188	181	125	110	139	227	278	189	200	172	140	117	275
	169	248	143	169	218	229	227	296	259	119	166	235	250	196	188	193	128	110	143	239	278	189	204	176	144	125	287
LWC019	157	216	139	146	210	225	253	266	247	116	146	231	214	−9	196	185	128	102	135	207	278	189	200	184	144	113	267
	173	224	143	150	210	246	253	296	247	116	146	235	226	−9	200	197	155	110	143	243	292	189	200	204	156	125	275
LWC020	149	216	145	158	218	271	219	276	243	116	146	223	208	188	192	181	125	94	139	203	266	209	200	184	144	109	271
	181	244	147	161	218	279	219	296	251	132	166	227	230	188	192	185	131	110	159	223	290	209	200	184	160	121	279
LWC021	149	228	135	146	210	246	215	262	259	112	174	227	208	188	196	185	122	106	139	207	282	205	196	192	152	113	271
	165	228	143	158	226	291	215	296	263	119	178	231	222	204	196	193	122	110	139	227	282	217	196	192	156	121	275
LWC022	145	240	135	173	255	233	215	280	251	116	134	223	208	188	200	177	128	87	139	215	274	213	192	184	152	117	271
	173	252	143	173	255	254	219	296	255	127	146	262	222	204	200	205	128	106	167	215	274	213	204	184	156	129	279
LWC023	137	196	139	165	214	229	215	258	259	116	146	231	218	188	188	161	125	106	143	219	278	209	196	192	152	109	263
	169	204	143	165	218	267	227	296	259	123	146	235	242	208	196	193	134	106	143	231	278	217	211	196	164	125	279
LWC024	157	194	135	154	214	233	215	254	247	123	174	227	222	−9	184	185	125	106	139	−9	242	213	200	172	144	117	267
	161	194	135	169	255	254	253	254	247	123	174	231	246	−9	196	193	131	110	143	−9	278	213	200	172	152	117	275

Table S2. Cont.

Sample ID	GGA A3A 07M	GATA 129H 04	GAT A43 A04	GAT A116 B01N	GATA 176 C01	GATA 164B 08P	GGA A4B 09N	GAT A7F 05	ATA 28F 03	AT TT0 30	GAT A29 A01	GAT A61 E03	GAT A43 C11	GA TA1 04	GAT A14 E09	TCT A01 7M	ATA 27A 06P	GATA 91H 06M	GGAA 21G 11L	GATA 50G 06	GAT A71 H05	GAT A8C 04	GAT A25 A04	GAT A11 A06	UT7 544	GATA 129D 11N	AGA T120
LWC026	165	194	135	154	255	225	227	250	247	116	162	227	226	180	184	177	131	102	143	211	250	189	196	172	152	117	267
	169	200	139	161	259	271	253	296	267	122	166	239	250	188	188	193	134	106	159	215	278	213	200	196	160	125	283
LWC027	141	212	139	158	198	225	215	262	243	116	146	223	208	188	196	177	122	102	143	227	278	209	200	188	148	117	275
	157	216	171	184	198	267	227	296	255	116	150	227	250	204	208	189	128	102	159	235	282	213	200	192	148	129	283
LWC028	165	204	139	158	210	237	219	246	247	116	162	227	218	177	192	181	131	110	139	215	278	189	196	188	140	109	267
	165	212	143	169	218	237	219	296	251	116	166	227	246	196	196	193	131	110	143	235	278	209	211	196	144	125	283
LWC029	145	256	135	154	255	233	219	250	251	116	166	227	258	188	184	181	119	106	139	211	246	209	188	180	144	121	263
	157	256	145	154	255	233	219	296	251	119	170	227	258	204	204	201	137	114	147	211	290	213	208	192	152	129	287
LWC030	157	194	139	146	214	221	224	266	247	112	154	215	222	188	192	193	131	106	143	223	246	205	180	192	152	117	267
	165	196	143	150	218	254	227	296	259	123	166	227	226	196	192	193	131	106	147	235	278	217	200	196	160	117	275
LWC031	141	220	143	161	222	229	215	296	255	112	150	219	222	188	196	185	125	87	143	211	270	213	184	188	160	109	263
	157	244	175	173	247	229	224	296	255	116	162	225	258	220	208	197	137	118	179	231	274	213	184	196	160	121	287
LWC032	149	208	143	165	251	221	219	280	255	116	146	231	226	177	186	181	122	91	179	215	270	209	200	180	144	113	287
	165	216	175	196	275	221	224	296	263	116	154	243	258	188	199	189	128	106	163	247	274	209	200	180	156	121	287
LWC033	161	196	131	150	210	254	215	270	247	112	162	227	230	180	192	193	128	98	143	215	278	209	204	176	144	113	267
	173	208	139	150	218	275	215	296	251	116	174	235	246	188	200	193	131	102	143	243	278	213	208	200	144	117	271
LWC034	153	200	143	146	214	213	215	270	255	112	166	231	246	200	192	193	131	94	139	211	278	213	200	176	152	117	275
	165	216	155	146	218	250	227	296	267	116	174	231	246	204	192	193	131	110	143	227	282	213	208	192	156	125	287
LWC035	165	−9	135	165	259	213	206	−9	255	112	146	−9	258	−9	184	181	122	87	139	211	−9	193	184	196	160	117	287
	173	−9	147	165	259	213	219	−9	255	132	146	−9	258	−9	204	181	134	102	147	211	−9	209	184	196	160	125	287
LWC036	153	194	139	154	214	229	215	296	259	112	146	215	208	188	196	185	122	106	139	195	278	213	196	188	144	117	267
	169	216	139	169	226	229	215	296	263	116	174	231	208	208	200	193	131	110	139	199	292	213	196	196	144	125	283
LWC037	141	236	135	184	247	237	219	276	247	116	146	231	230	188	200	193	122	91	143	227	246	209	200	188	152	121	275
	149	244	147	200	255	246	224	296	247	116	170	251	242	188	208	193	128	110	163	231	266	209	211	188	152	121	279
LWC038	161	190	139	154	214	225	215	266	251	116	166	227	246	188	192	195	125	110	139	207	278	189	200	188	144	117	267
	165	190	155	165	251	237	215	296	251	116	174	243	250	192	192	195	128	110	147	211	282	189	200	188	148	125	283
LWC039	161	224	143	169	243	258	227	270	243	116	146	231	226	188	184	189	125	87	139	219	242	209	192	176	144	113	287
	185	236	143	173	251	279	227	296	259	132	174	239	226	196	188	201	128	91	143	243	246	213	208	184	148	121	290
LWC040	145	232	147	154	263	217	215	250	243	116	134	219	226	188	184	181	128	91	159	203	262	209	220	156	152	113	279
	181	246	179	158	266	217	219	296	263	116	146	239	258	204	196	197	137	106	163	203	286	209	220	188	156	121	283
LWC041	147	204	135	146	214	233	215	250	255	119	146	231	246	188	192	189	131	106	139	207	250	189	200	176	144	113	279
	165	208	135	146	218	237	227	296	259	123	174	239	250	188	192	193	134	106	147	207	278	213	200	204	156	121	283
LWC042	157	248	135	154	198	229	215	280	243	116	146	219	208	169	184	181	125	91	159	215	266	213	192	184	144	113	283
	173	248	165	158	259	241	219	296	255	132	154	227	258	177	200	205	128	110	163	215	266	213	208	184	160	121	287
LWC043	165	212	139	154	218	221	215	−9	251	116	146	223	218	184	192	181	131	106	135	211	246	189	196	188	140	117	271
	173	220	143	158	218	237	227	−9	251	123	146	227	246	208	192	189	131	110	143	223	282	209	196	196	152	125	283
LWC044	157	188	131	154	218	217	215	254	247	112	162	215	246	177	184	181	122	106	139	207	278	189	172	180	152	113	275
	157	194	135	158	251	229	227	296	251	123	166	231	246	188	196	189	125	114	143	215	278	213	172	188	160	117	279
LWC045	−9	204	131	154	214	221	215	284	259	116	166	227	222	188	184	189	131	106	143	211	278	189	200	192	140	113	267
	−9	212	143	169	243	229	215	296	259	116	170	235	258	192	192	193	134	110	143	227	284	213	208	204	152	121	279
LWC046	165	204	127	154	218	213	−9	258	247	116	146	227	214	200	192	185	122	102	139	207	288	213	196	176	156	113	267
	165	216	127	154	233	262	−9	296	247	132	178	239	262	204	192	193	122	110	139	223	288	217	196	176	156	125	275

Table S3. Comparison of genetic differentiation among populations

Location*	Western	Eastern	Central	Bonobo
Western		**0.44 (0.26)**	**0.42 (0.19)**	**0.82 (0.34)**
Eastern	0.31 (0.32)		**0.025 (0.07)**	**0.70 (0.26)**
Central	0.31 (0.32)	0.05 (0.09)		**0.70 (0.21)**
Bonobo	0.68 (0.68)	0.57 (0.54)	0.51 (0.49)	

*Microsatellite genotypes for western, central, and eastern chimpanzees were reported by Becquet et al. (1). Pairwise R_{ST} (versus F_{ST}) is shown. Numbers in bold above the diagonal were calculated from the subset of 27 autosomal microsatellite loci from Becquet et al. (1). Numbers below the diagonal appear in Becquet et al. (1).

1. Becquet C, Patterson N, Stone AC, Przeworski M, Reich D (2007) Genetic structure of chimpanzee populations. *PLoS Genet* 3:e66.