Bayesian Inference of the Demographic History of Chimpanzees

Daniel Wegmann*[,1,2] and Laurent Excoffier[1,2]

[1]Department of Biology, Computational and Molecular Population Genetics Laboratory, Institute of Ecology and Evolution, University of Bern, Bern, Switzerland

[2]Swiss Institute of Bioinformatics, Lausanne, Switzerland

***Corresponding author:** E-mail: daniel.wegmann@ucla.edu.

Associate editor: Rasmus Nielsen

Abstract

Due to an almost complete absence of fossil record, the evolutionary history of chimpanzees has only been studied recently on the basis of genetic data. Although the general topology of the chimpanzee phylogeny is well established, uncertainties remain concerning the size of current and past populations, the occurrence of bottlenecks or population expansions, or about divergence times and migrations rates between subspecies. Here, we present a novel attempt at globally inferring the detailed evolution of the *Pan* genus based on approximate Bayesian computation, an approach preferentially applied to complex models where the likelihood cannot be computed analytically. Based on two microsatellite and DNA sequence data sets and adjusting simulated data for local levels of inbreeding and patterns of missing data, we find support for several new features of chimpanzee evolution as compared with previous studies based on smaller data sets and simpler evolutionary models. We find that the central chimpanzees are certainly the oldest population of all *P. troglodytes* subspecies and that the other two *P. t.* subspecies diverged from the central chimpanzees by founder events. We also find an older divergence time (1.6 million years [My]) between common chimpanzee and Bonobos than previous studies (0.9–1.3 My), but this divergence appears to have been very progressive with the maintenance of relatively high levels of gene flow between the ancestral chimpanzee population and the Bonobos. Finally, we could also confirm the existence of strong unidirectional gene flow from the western into the central chimpanzee. These results show that interesting and innovative features of chimpanzee history emerge when considering their whole evolutionary history in a single analysis, rather than relying on simpler models involving several comparisons of pairs of populations.

Key words: *Pan troglodytes*, *Pan paniscus*, approximate Bayesian computation, demographic inference.

Introduction

Based on morphological and geographical criteria, four distinct populations of chimpanzees, our closest living relative, have been described (Groves 2001): Bonobos (*Pan paniscus*), central chimpanzees (*P. troglodytes troglodytes*), western chimpanzees (*P. t. verus*), and eastern chimpanzees (*P. t. schweinfurthii*). Although the Bonobos are recognized as an independent species (Groves 2001), the taxonomic status of the common chimpanzee (*P. troglodytes*) is disputed (Morin et al. 1994; Gonder et al. 1997, 2006; Fischer et al. 2006). Due to an apparent lack of fossil records (McBrearty and Jablonski 2005), the understanding of the demographic history of contemporary chimpanzee populations relies almost exclusively on the analysis of genetic data.

Recently, studies have tried to estimate chimpanzee history and demography using different types of loci (Gagneux et al. 1999; Won and Hey 2005; Fischer et al. 2006; Becquet and Przeworski 2007; Caswell et al. 2008). They commonly suggest the same tree topology with divergence times around 1 million years ago (Ma) between the Bonobos and the common chimpanzees, around 0.5 Ma for the western chimpanzees, and around 0.2 Ma for the split between the central and the eastern chimpanzees. Some studies found evidence for ongoing bidirectional gene flow between the central and the eastern chimpanzees (Becquet and Przeworski 2007) and of unidirectional migration from the western into the central chimpanzees (Won and Hey 2005). Only suggestive evidence of recent gene flow between the Bonobos and common chimpanzees has been reported so far (Becquet and Przeworski 2007). Classical population genetic analyses revealed a recent population expansion in central chimpanzees, Bonobos (Fischer et al. 2006), and in eastern chimpanzees (Gagneux et al. 1999), as well as a potential substructure within the western chimpanzees (Becquet et al. 2007; Leuenberger and Wegmann 2010).

To date, a detailed picture of the chimpanzee history is known for individual events only, despite the availability of large genomic data sets, which is mainly due to methodological difficulties. Many insights came from the study of individual populations using standard population genetic approaches or pairwise comparisons of genetic diversity (Morin et al. 1994; Stone et al. 2002; Fischer et al. 2006). These approaches provide a detailed picture on the relationship between different populations, but parameter inference is difficult due to the complex interaction of

different demographic parameters such as population size changes, divergence, and gene flow. Bayesian parameter inference has been applied to the study of isolated pairs of populations using two different approaches (Won and Hey 2005; Becquet and Przeworski 2007), but an extension to more than two populations is difficult, mainly due to the challenge to obtain likelihood formulations in models with that many parameters (Cornuet et al. 2008; Nielsen and Beaumont 2009). First approaches to infer parameters in models of more than two populations failed to incorporate migration, which may have an influence on the estimation of some parameters, especially on population sizes (Caswell et al. 2008). Very recently, Hey (2009) estimated gene flow between all pairs of recent and ancestral population but assumed that the size of all these populations remained constant over time.

Here, we present a novel methodology to infer the demographic history of chimpanzees based on Approximate Bayesian Computation (ABC, Tavaré et al. 1997; Beaumont et al. 2002; Leuenberger and Wegmann 2010). Such an approach is preferentially applied to complex models for which the likelihood cannot be computed analytically (Fagundes et al. 2007; Patin et al. 2009). In brief, ABC consists of two steps (Beaumont et al. 2002): a large set of simulations is first performed and then used to estimate posterior distributions. Simulations are performed by 1) drawing parameter from prior distributions, 2) generating genetic data under these parameter values, and 3) calculation of summary statistics based on the resulting data. The posterior distributions are finally estimated on the basis of a set of simulations closest to the observed summary statistics using a postsampling adjustment (Beaumont et al. 2002; Leuenberger and Wegmann 2010). The advantage of the ABC methodology is that it allows the integration of detailed aspects of the demographic history of chimpanzees, such as a complex migration patterns, populations expansions, and the divergence of more than two populations. It also conveniently allows the simultaneous analysis of different data sets (microsatellites and DNA resequencing data) with their specific characteristics (sampling scheme, ploidy level, and different amount of missing data).

Materials and Methods

Samples and Loci

We based our analysis on a subset of 310 microsatellites examined in a large set of common chimpanzees and Bonobos (Becquet et al. 2007). In a first step, we excluded genotypes from captive born individuals known to be hybrids or reported to be hybrids based on the genetic analysis (Becquet et al. 2007). This resulted in the genotypes of 16 central chimpanzees (*P. t. troglodytes*), 50 western chimpanzees (*P. t. verus*), 6 eastern chimpanzees (*P. t. schweinfurthii*), and 6 Bonobos (*P. paniscus*). We discarded nonautosomal loci and discarded seven additional loci showing departure from a pure stepwise mutation model (SMM) (less than 50% of the sampled alleles with a common repeat length). Individual alleles at other loci disagreeing with the reported repeat length (Becquet et al. 2007) were considered as missing data. This increased the missing data level from 3.1% to 6.2%. We finally removed another 20 loci due to 100% missing data in at least one of the sampled populations, which let us with a total set of 265 microsatellite loci.

We complemented the data set with 26 unlinked intergenic regions ranging from 653 to 1626 bp (totaling 22,400 bp) and sequenced in 10 central chimpanzees born in Gabon, 10 western chimpanzees from Sierra Leone, and 10 eastern chimpanzees from the Sweetwater Reserve in Kenya (Fischer et al. 2006). The same loci were additionally typed in nine unrelated Bonobos sampled from European zoos (Fischer et al. 2006), as well as in three human populations (Voight et al. 2005), which served as an outgroup to infer locus-specific mutation rates.

Although our inference approach is capable of dealing with more data sets, we did not include the data set of Yu et al. (2003) for two reasons: 1) due to the small sample sizes of several populations, we did not expect to gain much precision in our estimations, but it would have considerably increased our computational load; 2) more importantly, since we based our model of chimpanzee evolution (and the priors of its parameters) to a great extent on previous results obtained from the analysis of this data set by Won and Hey (2005), we preferred not to use the same data twice in our Bayesian framework.

Assumed Demographic Model

Following previous studies (Becquet and Przeworski 2007; Becquet et al. 2007; Caswell et al. 2008), we assumed an evolutionary model between the chimpanzee populations (depicted in fig. 1), where the divergence between the ancestral population of all common chimpanzees and the Bonobo is assumed to predate all other events. In our Bayesian ABC analysis, we assumed that this parameter $T_{DIV}PAN$ has a uniform prior distribution between 35,000 and 100,000 generations, noted as $T_{DIV}PAN \sim U[35{,}000, 10^5]$. The ancestral common chimpanzee population further diverged into western chimpanzees ($T_{DIV}WEC \sim U[15{,}000, 60{,}000]$) and the ancestral population of the eastern and central chimpanzees, itself having diverged recently ($T_{DIV}EC \sim U[4000, 25{,}000]$). Note that in our simulations (described below), we only accepted parameter vectors agreeing with the assumed and well-supported chimpanzee topology (Morin et al. 1994; Kaessmann et al. 1999; Stone et al. 2002; Fischer et al. 2004). We assumed uniform priors on the $\log_{10}$ scale of all population sizes between 5,000 and 250,000 diploid individuals. We furthermore allowed the four current populations to have changed their size exponentially since their divergence form an ancestral population by assuming the size at the time of divergence to be a fraction x of the current size with a uniform prior on $\log_{10}(x)$ between -1 and 0.5.

In order to keep the number of parameters reasonably low, we assumed symmetric gene flow between eastern and central chimpanzees, between central chimpanzees and

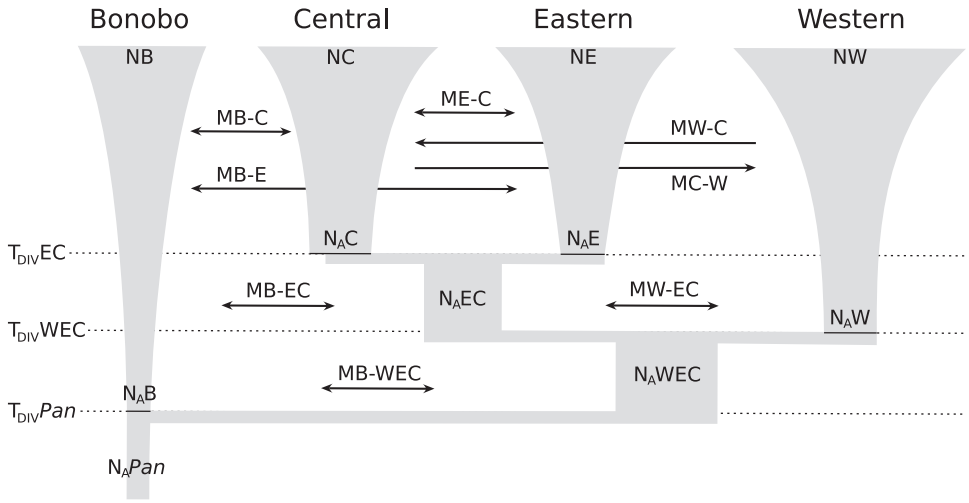

FIG. 1. Assumed evolutionary model. The evolutionary model assumed to describe the relationship between the Bonobos (*Pan paniscus*, B) and three populations of the common chimpanzee (*Pan troglodytes*): central (*P. troglodytes troglodytes*, C), eastern (*P. t. schweinfurthii*, E), and western chimpanzees (*P. t. verus*, W). Abbreviations: population sizes (N), ancestral population sizes (N_A), divergence times in generations (T_{DIV}), and effective number of immigrant genes ($M = Nm$).

Bonobos, between eastern chimpanzees and Bonobos, as well as between all pairs of ancestral populations ($M = Nm \sim 10^{U[-0.3,\ 1]}$). However, we used prior knowledge obtained from the analysis of a different data set (Won and Hey 2005) and thus allowed asymmetric migration between the western and the central chimpanzees. Note also that there is little hope to find strong signature of asymmetric gene flow among ancestral populations because many lineages will have coalesced by that time, thus limiting the information available about old demographic events.

Mutation rates of the DNA segments were inferred from alignments with human samples (Voight et al. 2005) assuming a divergence time of 7 million years between chimpanzees and humans (Caswell et al. 2008) and an average generation time of 20 years in both species (Fischer et al. 2004). Locus-specific mutation rates varied between 0.4×10^{-8} and 3.0×10^{-8} with an average of 1.6×10^{-8} per base pair per generation. The average mutation rate per generation of the microsatellites was drawn in $\bar{\mu} \sim U[2 \times 10^{-5}, 3 \times 10^{-4}]$, but we used locus-specific mutation rates distributed as a Gamma(α, α/μ), with the shape parameter α drawn in U[8,15]. We simulated the microsatellites using a pure SMM, because these loci were formerly shown to fit an SMM (Becquet et al. 2007). Note that we have only taken recombination between sequenced DNA regions into account but not recombination within the DNA sequence regions. However, the summary statistics used here (see below) are not thought to be affected by the low levels of recombination expected within our DNA loci (but see Thornton 2005 for an effect on Tajima's D with larger recombination rates). All prior distributions are summarized in table 2.

Substructure within Chimpanzee Populations

We observed elevated F_{IS} values in the different chimpanzee populations (table 1). Positive F_{IS} values have been attributed to inbreeding previously, but an alternative explanation is population subdivision (Wahlund 1928), where some individuals would have been sampled in different subdivisions and where the observed excess of homozygotes (F_{IS}) would be proportional to the extent of differentiation between subdivisions (F_{ST}). In the case of the western chimpanzees, it has indeed be found that an island model might easily explain the observed F_{IS} pattern, without need to invoke inbreeding (Leuenberger and Wegmann 2010). In order to account observed departures from random mating in chimpanzee populations, we reproduced the observed level of inbreeding by making individuals homozygote at any locus with probability F_{IS} (Nordborg and Donnelly 1997), calculated for microsatellites and DNA sequences independently. When estimating past demography, it is important to take population subdivision into account because it can give a wrong signal of population bottleneck (Nielsen and Beaumont 2009).

ABC

We performed a total of 2×10^6 simulations under a parallelized ABC-Markov chain Monte Carlo (MCMC) approach, as described in Wegmann et al. (2009) and implemented in the software package ABCtoolbox (Wegmann 2010). The ABC-MCMC approach is based on a likelihood-free MCMC

Table 1. Observed Population- and Marker-Specific F_{IS} Values.

Sample	DNA	Microsatellites
Bonobo	−0.054*	0.023
Eastern chimpanzee	0.049*	0.093*
Central chimpanzee	0.111*	0.057*
Western chimpanzee	0.096*	0.026*

NOTE.—F_{IS} values were calculated with Arlequin 3.0. Only significant positive F_{IS} values (P value < 0.05, *) were reproduced in the simulation step of the ABC procedure.

doi:10.1093/molbev/msq028

Table 2. Characteristics of Prior and Posterior Distributions.

	Prior[a]			Estimation accuracy					Posterior			
						Relative RMISE[b]						
Parameters[c]	Minimum	Maximum	R^2 [d]	*P* value[e]	$RMSE_{mode}$[f]	STR	DNA	Half	Mode	HPDI 50[g]	HPDI 90[g]	HPDI 95[g]
log(NB)	3.70	5.40	0.72	0.233	0.20	1.16	0.98	1.05	3.90	[3.80, 4.00]	[3.70, 4.13]	[3.70, 4.20]
log(NC)	3.70	5.40	0.50	0.081	0.27	1.07	1.15	1.03	5.13	[5.03, 5.25]	[4.88, 5.40]	[4.80, 5.40]
log(NE)	3.70	5.40	0.65	0.003	0.23	1.16	1.21	1.06	4.13	[4.02, 4.26]	[3.83, 4.45]	[3.78, 4.50]
log(NW)	3.70	5.40	0.76	0.004	0.20	1.26	1.25	1.05	3.99	[3.88, 4.11]	[3.72, 4.25]	[3.70, 4.28]
$\log(N_A B/NB)$	−1.00	0.50	0.03	0.240	0.69	1.01	1.00	1.00	−0.71	[−0.95, −0.37]	[−1.00, 0.18]	[−1.00, 0.30]
$\log(N_A C/NC)$	−1.00	0.50	0.04	0.010	0.6	1.01	1.00	1.00	−0.25	[−0.56, 0.01]	[−0.92, 0.27]	[−0.98, 0.33]
$\log(N_A E/NE)$	−1.00	0.50	0.08	0.098	0.57	1.01	0.99	1.00	−0.89	[−1.00, −0.69]	[−1.00, −0.22]	[−1.00, −0.06]
$\log(N_A W/NW)$	−1.00	0.50	0.02	0.063	0.68	1.01	0.99	1.00	−0.86	[−0.99, −0.71]	[−1.00, −0.25]	[−1.00, −0.08]
$\log(N_A EC)$	3.70	5.40	0.06	0.077	0.42	1.01	0.99	1.00	4.48	[4.17, 4.73]	[3.78, 5.11]	[3.73, 5.21]
$\log(N_A WEC)$	3.70	5.40	0.06	0.146	0.40	1.01	0.98	1.00	3.95	[3.73, 4.20]	[3.70, 4.86]	[3.70, 5.05]
$\log(N_A Pan)$	3.70	5.40	0.83	**<0.001**	0.22	1.41	1.22	1.21	4.95	[4.78, 5.16]	[4.56, 5.39]	[4.45, 5.40]
$T_{DIV}EC$	2,000	25,000	0.01	0.105	6,305	1.01	1.00	1.00	21,879	[17,834, 24,537]	[9,975, 24,999]	[8,010, 24,999]
$T_{DIV}WEC$	15,000	50,000	0.01	0.582	9,476	1.01	0.99	1.00	27,311	[22,914, 36,808]	[17,111, 45,426]	[15,879, 46,833]
$T_{DIV}Pan$	35,000	100,000	0.01	0.133	20,539	1.03	0.99	1.00	80,075	[66,683, 90,200]	[50,678, 99,999]	[44,799, 99,999]
log(ME-C)	−0.30	1.20	0.66	0.046	0.23	0.90	1.20	1.11	0.18	[0.06, 0.30]	[−0.11, 0.48]	[−0.16, 0.54]
log(MW-C)	−0.30	1.20	0.26	0.012	0.30	0.93	1.00	1.02	1.07	[0.92, 1.20]	[0.51, 1.20]	[0.41, 1.20]
log(MC-W)	−0.30	1.20	0.52	0.004	0.26	1.06	1.10	1.03	−0.30	[−0.30, −0.23]	[−0.30, −0.12]	[−0.30, −0.07]
log(MB-C)	−0.30	1.20	0.59	0.139	0.22	0.90	1.24	1.10	−0.30	[−0.30, −0.28]	[−0.30, −0.23]	[−0.30, −0.22]
log(MB-E)	−0.30	1.20	0.45	0.631	0.26	0.92	1.13	1.11	−0.11	[−0.21, −0.01]	[−0.30, 0.14]	[−0.30, 0.21]
log(MB-EC)	−0.30	1.20	0.01	0.164	0.33	1.00	0.99	1.00	0.03	[−0.20, 0.29]	[−0.30, 0.78]	[−0.30, 0.94]
log(MW-EC)	−0.30	1.20	0.02	0.063	0.35	1.01	1.00	1.00	0.48	[0.22, 0.78]	[−0.08, 1.15]	[−0.15, 1.19]
log(MB-WEC)	−0.30	1.20	**<0.01**	0.051	0.33	1.00	1.00	1.00	0.89	[0.54, 1.08]	[−0.19, 1.20]	[−0.21, 1.20]
$\mu \times 10^5$	2.00	30.00	0.91	**<0.001**	1.56	3.26	6.01	1.59	8.89	[10.11, 7.54]	[8.58, 9.21]	[9.89, 7.89]
α	8.00	15.00	0.01	0.180	2.44	0.98	1.00	1.00	9.79	[8.46, 11.20]	[8.00, 13.91]	[8.00, 14.33]

[a] All priors are uniformly distributed.

[b] Geometric mean over all pseudo-observed data sets of the RMISE of a reduced data set (only microsatellites [STR], DNA sequences, or half of both types or markers), relative to the RMISE obtained with the full data set. A value larger than one indicates a lower performance for the reduced data sets.

[c] See figure 1 for more information on the parameters. Log refers here to the logarithm with base 10.

[d] Coefficient of determination (R^2) obtained when regressing a parameter against the 10 used PLS components, which is indicative for the power to estimate individual parameters (Neuenschwander et al. 2008).

[e] Coverage *P* value computed using a Kolmogorov–Smirnoff test. Bold values are significant deviations from uniformity after Bonferroni correction. See text for details.

[f] Average RMSE.

[g] HPDI is chosen as the continuous interval of parameter values with highest posterior density.

sampler introduced recently (Marjoram et al. 2003). We used the two freely available programs SIMCOAL 2.0 (Laval and Excoffier 2004) and Arlequin 3 (Excoffier et al. 2005) to simulate genetic data and calculate summary statistics, respectively. In order to match both, the observed level of missing data and the observed Wahlund effect, we modified the output of SIMCOAL 2.0 prior to the calculation of summary statistics.

The ABC inference was based on a total of 96 summary statistics calculated within and between the population samples (see full list in supplementary table S1, Supplementary Material online). On DNA sequence data, we computed for each population the number of segregating sites S, the number of private segregating sites S_{pr}, Tajima's D (Tajima 1989), Fu's F_S (Fu 1997), and the number of pairwise differences π on the concatenated sequences. Between all pairs of populations, we also computed pairwise F_{ST}, computed as θ_W (Weir and Cockerham 1984; Excoffier et al. 1992), as well as the average number of pairwise differences π_{xy}. We also computed the total number of segregating sites and global F_{ST} over all populations. On microsatellite data, we computed in each population the mean and standard deviation (SD) over loci of the number of alleles K, the range of the allele size R, the expected heterozygosity H, the Garza–Williamson statistic (Garza and Williamson 2001) modified as $GW = K/(R + 1)$ (Excoffier et al. 2005) and another modification of GW computed as $GW^* = K/(R_{Tot} + 1)$, where R_{Tot} is the range in allele size computed over all sampled populations (Wegmann et al. 2009). The same statistics and their SD over populations were also computed over the pooled populations, except GW^* because $GW = GW^*$ in that case. We additionally computed the differentiation index Φ_{ST} (Michalakis and Excoffier 1996) and the genetic distance $(\delta\mu)^2$ (Goldstein et al. 1995) between all pairs of populations and as well the global Φ_{ST}. Following Wegmann et al. (2009), we extracted partial least squares (PLSs) components from these summary statistics based on the 10,000 simulations performed in the calibration step of the ABC-MCMC algorithm. We used the R package "PLS" (Mevik and Wehrens 2007) to find the appropriate number of PLS component to use (11 in our cases). The same set of 96 summary statistics were also calculated on the observed data set, PLS transformed and used to calculate their Euclidean distance to each simulation. The reduction of the summary statistics space using PLS has been shown to be beneficial when dealing with large sets of potentially correlated summary statistics, mainly because the calculation of Euclidean distances is not very meaningful in too many dimensions (Wegmann et al. 2009). We then retained a total of 10,000 simulations closest to the observed data and applied the regression adjustment ABC–GLM proposed by Leuenberger and Wegmann (2010), as implemented in the software package ABCtoolbox (Wegmann 2010). This approach assumes that the likelihood function can be locally approximated by a general linear model (GLM) around the observed values, which makes its estimation possible.

Validation of the Estimation Procedure

Following previous recommendations (Wegmann et al. 2009), we checked if we were able to generate unbiased posterior distributions with the chosen estimation procedure, based on 1,000 pseudo-observed data sets generated with known parameter values drawn according to the prior distributions. The way we generated the simulations used to estimate posterior distributions differs between our validation step and the application to the chimpanzee data set: although an ABC-MCMC approach leads to the same set of simulations as rejection sampling (Wegmann et al. 2009) with less effort, the simulations generated may not be reused for another observed data set. We therefore used the ABC-MCMC approach only to generate the larger set of simulations used for the chimpanzee data set and generated a total of 10^6 simulations using a conventional rejection sampling approach (Tavaré et al. 1997; Beaumont et al. 2002) with parameter values drawn from the prior distributions. We again computed the same set of 96 summary statistics and transformed the statistics with the same PLS components as for the observed data set. Using the software package ABCtoolbox (Wegmann 2010), we computed the coverage property of the posterior distributions obtained with an ABC-GLM regression adjustment based on the 5,000 simulations closest to the given pseudo-observed data set. The coverage is the proportion of times a true parameter value is present in a given credible interval. For instance, 80% and 95% credible intervals should include the true parameter with probabilities 0.8 and 0.95, respectively. In other words, the posterior quantiles of the true parameter values should be uniformly distributed in [0, 1] (Cook et al. 2006; Wegmann 2010). The uniformity of the posterior quantiles of the true parameter was assessed with a classical Kolmogorov–Smirnov test for each parameter independently and significance was attributed after Bonferroni correction.

We also computed the coefficient of variation (R^2) obtained when regressing the chosen PLS components against a model parameter, which is indicative of the power to estimate individual parameters (Neuenschwander et al. 2008). We further computed the root mean squared error of the mode ($\text{RMSE}_{\text{mode}}$) for each parameter, based on the 1,000 pseudo-observed data sets. The $\text{RMSE}_{\text{mode}}$ provides a crude estimate of the accuracy of the mode, as a point estimate. Note that we also report the whole posterior distributions and root mean integrated squared error (RMISE, see below), which gives the most detailed information about the uncertainty of our point estimates.

In order to evaluate the benefit of combining two data sets with different marker types, we calculated the RMISE of the estimation (Wegmann 2010) on the 1,000 pseudo-observed data sets with microsatellite or DNA sequence data alone, as well as with the combined data sets. Because the interpretation of individual RMISE values is difficult, we report the RMISE computed on the reduced data set relative to that computed over the complete data set, and this as a geometric mean over all pseudo-observed data set.

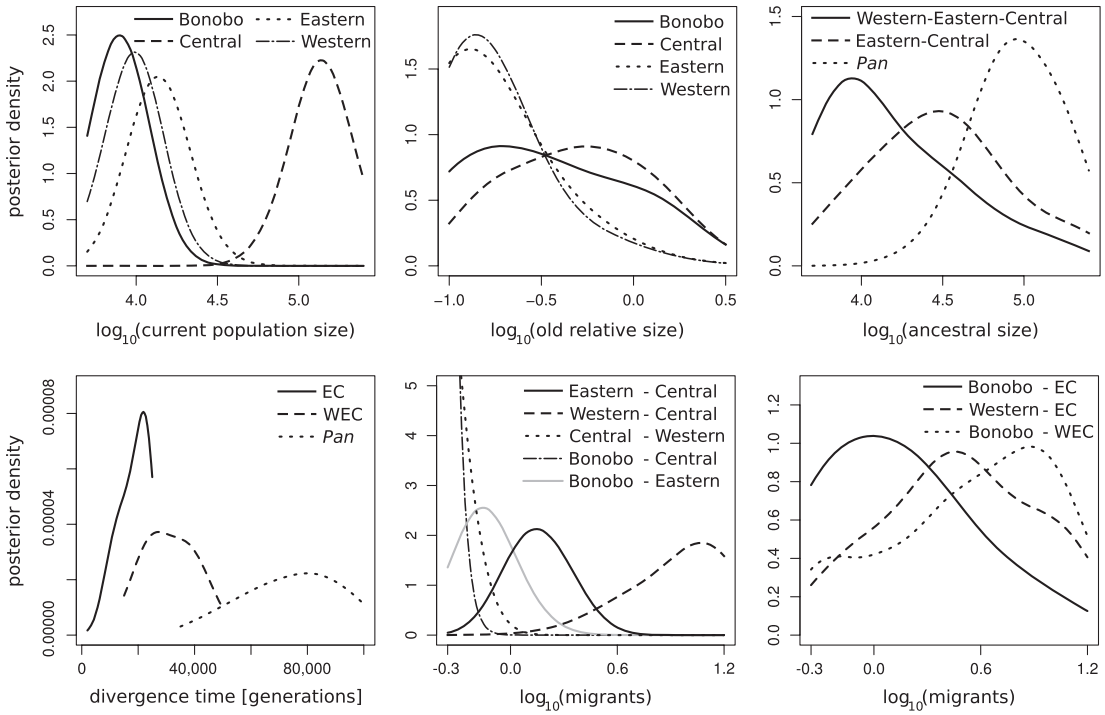

Fig. 2. Posterior distributions of selected parameters. Posterior estimates are based on the ABC-GLM approach applied to the 10,000 simulations closest to the observed data, out of a total of 2×10^6 simulations performed under an ABC-MCMC algorithm. All posterior distributions are shown over the whole range of their respective prior distribution. Note that the "old relative size" in the upper-middle pane refers to the ratio of the initial size after divergence and the current size of the three chimpanzee populations: a value smaller than 0 of $\log_{10}$(old relative size) implies that a population expansion occurred.

(Wegmann 2010). In order to check if the observed data is in agreement with the assumed model, we computed the distribution of the marginal densities (the likelihood of the model based on the postsampling adjustment implemented in ABCtoolbox) of the retained 10,000 for posterior estimation (see above). The fraction of simulation with smaller marginal densities than the observed data serves as a *P* value (Wegmann 2010).

Results

Validation of the ABC Approach

We used a total of 1,000 pseudo-observed data sets to check the coverage property of the marginal posterior distributions estimated with our approach. Although most marginal posterior distributions pass the test, estimates of both, the ancestral population size of all chimpanzees N_A*Pan* and of the average microsatellite mutation rate μ are slightly overestimated on average, as can be seen from the histograms of the posterior quantiles (supplementary fig. S1, Supplementary Material online). We further assessed whether observed data is in agreement with the assumed model. We found the observed values to fall within the simulate data (*P* value 0.078), which suggests that the assumed model is capable of reproducing the observed summary statistics (the 11 PLS components in our case).

Inference on Chimpanzee Past Demography

Posterior distributions of the parameters of the chimpanzee evolutionary model shown in figure 1 were generated based on 2×10^6 simulations and obtained under the ABC-MCMC approach and shown in figure 2. Some of their additional properties are summarized in table 2. The point estimates (modes) of the estimated parameters on a natural scale are further given in table 3 and plotted onto a sketch of the evolutionary model (supplementary fig. S2, Supplementary Material online). We find the central chimpanzees to have the largest current effective population size of roughly 130,000 diploid individuals. The other three populations are of smaller effective size: Bonobo size ~8,000, eastern chimpanzee ~13,000, and western chimpanzee size ~10,000 diploid individuals. All four populations show a signal of a recent exponential population expansion. The strongest signal (~8-fold) was found for the eastern chimpanzees, followed by the western chimpanzees (~7-fold), and the Bonobo (~5-fold). The central chimpanzees seem to have only doubled their size since the split from the eastern chimpanzees. Unlike the central chimpanzees, we find that the ancestral eastern and western population sizes are much smaller than the size of the population they originated from before divergence. It suggests that the eastern and western populations went through a bottleneck just after their divergence and before expanding to their current range. Since we find no evidence of a bottleneck in the demographic history of the central chimpanzee, it suggests that the central chimpanzee population is the likely ancestral population of all chimpanzees and that the other common chimpanzee populations diverged from the central chimpanzees by founder events.

Table 3. Mode and 90% HPDI of Key Parameters of Chimpanzee History.

Parameter	Current study[a]		Becquet and Przeworski (2007)[b,c]		Won and Hey (2005)[c,d]		Caswell et al. (2008)[e]		Hey (2009)[f]	
NB	7,900	[5,000, 13,500]	10,400	[7,900, 15,200]	8,300	[6,100, 10,900]	—	—	9,900	[7,500, 12,800]
NB/N_AB	5.13	[0.66, 10.00]	—	—	—	—	—	—	—	—
NC	134,900	[75,900, 251,200]	23,100	[8,600, 59,700]	22,000	[14,900, 35,600]	118,000	[91,000, 159,000]	31,400	[18,800, 51,200]
NC/N_AC	1.78	[0.54, 8.31]	—	—	—	—	—	—	—	—
NE	13,500	[6,800, 28,200]	17,700	[5,000, 71,800]	—	—	—	—	9,600	[5400, 15300]
NE/N_AE	7.76	[1.66, 10.00]	—	—	—	—	—	—	—	—
NW	9,800	[5,200, 17,800]	10,100	[7,700, 21,100]	7,000	[4,600, 9,800]	9,100	[8,100, 10,000]	8600	[6300, 11,700]
NW/N_AW	7.24	[1.78, 9.77]	—	—	—	—	—	—	—	—
N_AEC	30,200	[6,000, 128,800]	46,000	[33,500, 75,100]	—	—	—	—	36,900	[21,700, 63,000]
N_AWEC	8,900	[5,000, 72,400]	15,000	[6,100, 22,400]	4,600	[200, 9,900]	16,000	[12,400, 19,600]	8,300	[4,100, 14,600]
N_APan	89,100	[36,300, 245,500]	32,900	[22,200, 48,700]	13,800	[0, 25,900]	20,900	[16,400, 25,500]	19,600	[8,800, 32,700]
$T_{DIV}EC$ (Ma)	0.44	[0.20, 0.50]	0.22	[0.14, 1,40]	—	—	—	—	0.11	[0.05, 0.18]
$T_{DIV}WEC$ (Ma)	0.55	[0.34, 0.91]	0.44	[0.32, 1.10]	0.49	[0.30, 0.73]	0.51	[0.43, 0.59]	0.54	[0.41, 0.76]
$T_{DIV}Pan$ (Ma)	1.60	[1.01, 2.00]	0.86	[0.62, 1.35]	1.02	[0.69, 1.55]	1.29	[1.14, 1.45]	1.09	[0.79, 1.80]
ME-C symmetric	1.51	[0.78, 3.02]	0.80	[0.08, 1.39]	—	—	—	—	n.s.	—
MW-C	11.75	[3.24, 15.85]	—	—	0.38		—	—	n.s.	—
MC-W	0.50	[0.50, 0.76]	—	—	0.00		—	—	n.s.	—
MW-C symmetric	—	—	0.32	[0.10, 0.52]	—	—	—	—	n.s.	—
MB-C symmetric	0.50	[0.50, 0.59]	0.01	[0.01, 0.04]	0.00	—	—	—	n.s.	—
MB-E symmetric	0.78	[0.50, 1.38]	0.06	[0.00, 0.10]	0.00	—			n.s.	—
MB-EC symmetric	1.07	[0.50, 8.71]	—	—	—	—	—	—	n.s.	—
MW-EC symmetric	3.02	[0.83, 14.13]	—	—	—	—	—	—	n.s.	—
MW-EC	—	—	—	—	—	—	—	—	0.44	—
MB-WEC symmetric	7.76	[0.65, 15.85]	—	—	—	—	—	—	n.s.	—

[a] We estimated $M = Nm$, but report here $M = 2Nm$ to be consistent with other studies.

[b] This study reported the 95% HPDI.

[c] Wherever several estimates were available (because parameters were estimated in several pairwise analyses), we report the mean of the point estimates and the lowest and largest estimate of the HPDI, respectively.

[d] Won and Hey (2005) used a divergence time of 6 Ma (rather than 7 Ma) and 15 years per generation (rather than 20). To make the estimates comparable, we adjusted the population size estimates and all the ages by 7/6 and further multiplied the population sizes and the migration parameters by 15/20, following Caswell et al. (2008).

[e] Caswell et al. (2008) reported the 90% credible interval of the mode, estimated by a bootstrapping procedure.

[f] Hey (2009) used a divergence time of 6 Ma (rather than 7 Ma). To make the estimates comparable, we adjusted the population size estimates and all the ages by 7/6, following Caswell et al. (2008). Note that Hey (2009) estimated asymmetric migration rates between all pairs of recent and ancestral populations but reported only significant point estimates: n.s. means that the point estimates were estimated but not significantly different from zero.

Despite assuming a given topology, the timing of the different divergence events was free to vary in our evolutionary model. Assuming a generation time of 20 years in all chimpanzee populations (Fischer et al. 2004), we find a divergence time between Bonobos and common chimpanzee populations to be around 1.60 Ma, which is slightly older than previous estimates (table 3), but we note that the 90% highest posterior density interval (HPDI) is quite large for this estimate (1.33–1.80 Ma). The western chimpanzees seem to have diverged from the other common chimpanzees about 0.55 Ma, and the split between the eastern and central chimpanzees would have occurred already 0.44 Ma.

Compared with previous approaches, an interesting feature of our model is the possibility to infer various migration parameters at the same time. As was reported previously, most chimpanzee populations exchange migrants at very low rates. The Bonobos, for instance, seem to be almost completely isolated from the central and the eastern chimpanzees (<0.5 and ~ 0.8 migrants per generation, respectively) nowadays. We find, however, evidence for limited gene flow between the Bonobos and the ancestral populations of the central and the eastern chimpanzees and even higher levels of gene flow between the Bonobos and the ancestral population of all common chimpanzees (~ 0.8 and ~ 7.7 migrants per generation, respectively). This is in agreement with a progressive divergence between the Bonobos and the common chimpanzees. Among the current chimpanzee populations, only the eastern and central chimpanzees seem to exchange migrants at notable rates (~ 1.5 migrants per generation). We also find asymmetric migration between the central and the western chimpanzees with an immigration of about 11.7 individuals per generation from the western into the central chimpanzees, but a complete absence of gene flow in the opposite direction. We further report some evidence for an exchange of migrants prior to the eastern–central split (~ 3.0 migrants per generation).

Discussion

Comparison with Previous Studies

We report our estimates of the parameters of chimpanzee evolution along those obtained in previous studies (Won and Hey 2005; Becquet and Przeworski 2007; Caswell et al. 2008; Hey 2009) in table 3. Although most previous attempts to infer the demographic history of the chimpanzee populations either failed to incorporate migration (Caswell et al. 2008) or were done on a series of simpler models involving pairs of populations (Won and Hey 2005; Becquet and Przeworski 2007), most of our population size estimates are in perfect agreement with previous estimates. We also confirm the large effective size of the central chimpanzees reported by Caswell et al. (2008), which is at odds with previous findings (Won and Hey 2005; Becquet and Przeworski 2007; Hey 2009). We estimate the size of the ancestral population of all chimpanzees N_A*Pan* to be larger than that previously reported, but this may be due to an overestimation of this parameter (supplementary fig. S1, Supplementary Material online).

Interestingly, we do not find evidence for a recent population contraction in the western chimpanzee (Caswell et al. 2008), despite the fact that our model explicitly allows such a contraction to happen. We rather report here a more complex history of the western chimpanzees: a recent population expansion, which occurred after a founder event following the population split from the other chimpanzee populations. However, the apparent discrepancy between our results and those of Caswell et al. (2008) may be due to two factors: First, there is evidence for substructure in the western chimpanzees (Leuenberger and Wegmann 2010), which is explicitly taken into account in our approach, and a recent population expansion signal might be missed when population subdivision is not properly taken into account (Nielsen and Beaumont 2009; Stadler et al. 2009). Second, and more important, the model envisioned by Caswell et al. (2008) is much simpler than the scenario examined here, and their inference of a population contraction was based on a comparison of the western chimpanzee with the size of the ancestral population of all common chimpanzees, both assumed to have remained constant through time. It seems thus likely that the contraction signal found by Caswell et al. (2008) is due to the founder effect we have identified before the split of the two species, and our results are therefore not incompatible with those of Caswell et al. (2008) about a smaller average size of the western chimpanzees than the ancestral population of all chimpanzees.

Our analysis further suggests that the divergence between the common chimpanzees and the Bonobo has occurred clearly more than 1 Ma, which is close to the upper limit of previous estimates. Although our estimate of the divergence time of the western chimpanzees is in good agreement with previous estimates, we found the divergence of the central and eastern chimpanzees to be slightly older. We note, however, that some previous estimates had a very large credible interval (Becquet and Przeworski 2007). Our divergence time estimates are generally in good agreement with those reported by Caswell et al. (2008), despite the assumption of no gene flow in their inference. This is not surprising because migrations between the chimpanzee populations generally happen at very low levels, pairwise divergence time should not be affected by the presence of additional populations (Caswell et al. 2008), and because the assumption of an absence of migration was not found to affect the estimate of the divergence time between the Bonobo and the common chimpanzee (Caswell et al. 2008).

Our analysis generally confirms previously reported migration patterns, including strong and unidirectional gene flow from western to central chimpanzees. Since we did not allow asymmetric migration to occur between the western and the ancestors of the eastern and central chimpanzees, we cannot infer if this asymmetry is recent, but we find some evidence for an exchange of migrants prior to the eastern–central split. If the asymmetry is recent, it is

therefore due to a lack of migration from the central into the western chimpanzees. Some evidence for an old asymmetry has been reported recently when considering the common chimpanzees only (Hey 2009). However, the same study could not confirm this pattern when the ancestors of the eastern and central chimpanzees were allowed to exchange migrants with the Bonobo, as suggested by our study.

Although we could confirm previous evidence of gene flow between the Bonobos and the eastern chimpanzees (Becquet and Przeworski 2007; Hey 2009), we find evidence for a noninstantaneous divergence between the Bonobos and the common chimpanzees. Our estimates of gene flow before the divergence of central and eastern chimpanzees are indeed found larger than recent rates. The highest level of gene flow, however, is found between the Bonobos and the ancestral population of all common chimpanzees, suggesting a prolonged period of secondary contacts between the two species. Our estimates of gene flow are generally much larger than those previously reported, in agreement with the findings that migration events are, in this context, much easier to detect based on microsatellites than on DNA sequences, which were used in all previous analyses reported in table 3. In keeping with these finding, the only study attempting to estimate gene flow between ancestral populations of chimpanzees so far concluded that substantially more data were required to make clear statements on past gene flow (Hey 2009).

Although we find that our estimates of the average mutation rate of the microsatellites $\bar{\mu}$ are slightly biased toward too large values, we still find a relatively small mutation rate of $\bar{\mu} < 10^{-4}$, which is well below the generally assumed average mutation rate of $\bar{\mu}$ around 5×10^{-4} for mammals (Ellegren 1995; Ellegren et al. 1997), and which is slightly lower than a recent estimate in humans (Wegmann et al. 2009). However, these results are in good agreement with a recent unbiased estimate of microsatellite mutability based on human–chimpanzee sequence alignments (Webster et al. 2002). Indeed this study found a larger mutability of microsatellites in humans than in chimpanzees and reported mutation rates 2×10^{-5} for tri- and tetranucleotide microsatellites in both species.

We finally find that our credible intervals compare well with to those reported by previous Bayesian analysis, despite the much more complex evolutionary model assumed here (Won and Hey 2005; Becquet and Przeworski 2007). Note that the credible intervals reported by Caswell et al. (2008) are not really comparable because they report the credible interval of the best-fitted point estimate.

Recently, a fourth chimpanzee population or subspecies of western Cameroon and southern Nigeria has been recognized on the basis of mitochondrial DNA (mtDNA) data (Gonder et al. 1997) and named *P. t. ellioti* (Oates et al. 2009). So far only limited mtDNA data are available for this population, and none was included in this study. While geographically close to the central chimpanzees, this population was reported to be related to the western chimpanzees. If the central chimpanzees exchanged migrants with an unsampled population genetically close to the western chimpanzees, it would also appear as if there was unidirectional gene flow from western to the central population. However, a recent morphometric analysis grouped putative *P. t. ellioti* populations closer to the central chimpanzees than to the western chimpanzees (Pilbrow 2006), and the analysis of microsatellite data in three samples showing *P. t. ellioti* mtDNA haplotypes revealed that these individuals clustered within the central chimpanzees (Becquet et al. 2007), suggesting potential discrepancies between mtDNA and nuclear markers about population affinities of *P. t. ellioti*. We feel thus confident that the inferred asymmetrical gene flow from central to western chimpanzees is not due to the presence of this unsampled population.

Methodological Considerations

The ABC approach enables us to take specific features of the sampling scheme of the data sets into account: using the observed F_{IS} values, we replicated the Wahlund effect occurring when diploid individuals are sampled over the whole range of a structured population. In order to check if the simultaneous use of the two types of markers improves our estimates, we repeated the estimation procedure of our pseudo-observed data sets based on the DNA sequences or microsatellites alone, as well as on a data set using half of the microsatellites and half of the DNA sequences. We then compared the accuracy of these estimates by computing the RMISE obtained with the reduced data sets, relative to the RMISE obtained using the full data (see Relative RMISE columns in table 2). We generally find only a slightly reduced accuracy when using our reduced data sets, but we note that these data sets already incorporate a large number of loci. It suggests that it may not be necessary to sample more than about 100 loci to make these demographic inferences. The largest increase in accuracy using the full data was for those parameters with large R^2 values, which seem thus to be good predictors of the estimability of individual parameters. Although we find population sizes to be better estimated with the DNA sequences than with the microsatellites alone, migration rates show the opposite pattern. Indeed, microsatellites alone lead to more accurate estimations than using the full data, at the cost of less accurate estimations of the population sizes. However, several parameters clearly benefit from the use of both types of markers simultaneously, especially the population sizes because we obtain better estimates with half of both markers than with one marker type alone. It is interesting to note that, as expected, the average mutation rate of the microsatellites $\bar{\mu}$ is much better estimated when using DNA sequences with known mutation rates concurrently.

We report further characteristics of the posterior distributions obtained from the observed reduced data sets in supplementary table S2 (Supplementary Material online). As expected from the accuracy results above, we generally

find very similar modes estimated on the two observed reduced data sets but slightly wider credible intervals as compared with the full data set. We find some evidence for different point estimates for three parameters only (log(NAE/NE); T_{DIV}EC; log(MW-EC)). However, we note that the HPDIs of these parameters computed on the two separate data sets are broadly overlapping, and thus the two data types do not point toward drastically different estimates.

Conclusions

The flexible ABC approach has enabled us to infer parameters of a complex evolutionary model for chimpanzees based on two available data sets with different marker types (Fischer et al. 2006; Becquet et al. 2007). By simultaneously inferring population sizes, divergence times and complex migration patterns, some interesting differences with previous findings, mostly based on population pairs, became apparent. For instance, we do not find evidence for a population contraction in the western chimpanzee (Caswell et al. 2008), despite the fact that our model explicitly allows such a contraction to have happened. We rather report here a recent population expansion of the western chimpanzees, which occurred after this population split from the other chimpanzee populations. We further find evidence for a noninstantaneous divergence between the Bonobos and the common chimpanzees, suggesting a more complex speciation than previously anticipated. Note that nonallopatric speciation may not be uncommon among primates: Patterson et al. (2006), for instance, proposed that humans and chimpanzees diverged first but then hybridized later before separating permanently. Since the central chimpanzees are the only chimpanzee population having expanded recently without showing any signal of a bottleneck prior to this expansion, we can speculate that the central chimpanzees are the ancestral population of all common chimpanzees. Our results are thus compatible with the view that the western and the eastern chimpanzees have diverged from the central chimpanzees and expanded into a new range. This inferred scenario is in perfect agreement with the assumed recent expansion of the eastern chimpanzees north of the Congo River (Goldberg and Ruvolo 1997; Gagneux et al. 1999; Eriksson et al. 2004).

Supplementary Material

Supplementary tables S1–S2 and supplementary figures S1–S2 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

We thank Olaf Thalmann for providing us with DNA sequence data and estimates of mutation rates for these loci based on comparison with humans, as well as for very helpful comments on an earlier draft of the manuscript. We also thank two anonymous reviewers for their constructive comments on the manuscript. This work was supported by Swiss National Foundation grants no 3100A0-112072 and 3100A0-126074 to L.E.

References

Beaumont MA, Zhang W, Balding DJ. 2002. Approximate Bayesian computation in population genetics. *Genetics* 162:2025–2035.

Becquet C, Patterson N, Stone AC, Przeworski M, Reich D. 2007. Genetic structure of chimpanzee populations. *PLoS Genet.* 3:10.

Becquet C, Przeworski M. 2007. A new approach to estimate parameters of speciation models with application to apes. *Genome Res.* 17:1505–1519.

Caswell JL, Mallick S, Richter DJ, Neubauer J, Gnerre CSS, Gnerre S, Reich D. 2008. Analysis of chimpanzee history based on genome sequence alignments. *PLoS Genet.* 4:14.

Cook SR, Gelman A, Rubin DB. 2006. Validation of software for Bayesian models using posterior quantiles. *J Comput Graph Stat.* 15:675–692.

Cornuet JM, Santos F, Beaumont MA, Robert CP, Marin JM, Balding DJ, Guillemaud T, Estoup A. 2008. Inferring population history with DIY ABC: a user-friendly approach to approximate Bayesian computation. *Bioinformatics* 24:2713–2719.

Ellegren H. 1995. Mutation rates at porcine microsatellite loci. *Mamm Genome.* 6:376–377.

Ellegren H, Moore S, Robinson N, Byrne K, Ward W, Sheldon BC. 1997. Microsatellite evolution—a reciprocal study of repeat lengths at homologous loci in cattle and sheep. *Mol Biol Evol.* 14:854–860.

Eriksson J, Hohmann G, Boesch C, Vigilant L. 2004. Rivers influence the population genetic structure of bonobos (*Pan paniscus*). *Mol Ecol.* 13:3425–3435.

Excoffier L, Estoup A, Cornuet JM. 2005. Bayesian analysis of an admixture model with mutations and arbitrarily linked markers. *Genetics* 169:1727–1738.

Excoffier L, Laval G, Schneider S. 2005. Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evol Bioinform Online.* 1:47–50.

Excoffier L, Smouse PE, Quattro JM. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondria1 DNA restriction data. *Genetics* 131:479–491.

Fagundes NJR, Ray N, Beaumont M, Neuenschwander S, Salzano FM, Bonatto SL, Excoffier L. 2007. Statistical evaluation of alternative models of human evolution. *Proc Natl Acad Sci U S A.* 104:17614–17619.

Fischer A, Pollack J, Thalmann O, Nickel B, Paabo S. 2006. Demographic history and genetic differentiation in apes. *Curr Biol.* 16:1133–1138.

Fischer A, Wiebe V, Paabo S, Przeworski M. 2004. Evidence for a complex demographic history of chimpanzees. *Mol Biol Evol.* 21:799–808.

Fu Y-X. 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* 147:915–925.

Gagneux P, Wills C, Gerloff U, Tautz D, Morin PA, Boesch C, Fruth B, Hohmann G, Ryder OA, Woodruff DS. 1999. Mitochondrial sequences show diverse evolutionary histories of African hominoids. *Proc Natl Acad Sci U S A.* 96:5077–5082.

Garza JC, Williamson EG. 2001. Detection of reduction in population size using data from microsatellite loci. *Mol Ecol.* 10:305–318.

Goldberg TL, Ruvolo M. 1997. The geographic apportionment of mitochondrial genetic diversity in east African chimpanzees, *Pan troglodytes schweinfurthii*. *Mol Biol Evol.* 14:976–984.

Goldstein DB, Ruiz-Linares A, Cavalli-Sforza LL, Feldman MW. 1995. Microsatellite loci, genetic distances, and human evolution. *Proc Natl Acad Sci U S A.* 92:6723–6727.

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.

Gonder MK, Oates JF, Disotell TR, Forstner MRJ, Morales JC, Melnick DJ. 1997. A new west African chimpanzee subspecies? *Nature* 388:337–337.

Groves C. 2001. Primate taxonomy. Washington (DC): Smithsonian Institution Press.

Hey J. 2009. The divergence of chimpanzee species and subspecies as revealed in multi-population isolation-with-migration analyses. *Mol Biol Evol.* Advance Access published December 2, 2009, doi:10.1093/molbev/msp298.

Kaessmann H, Wiebe V, Pääbo S. 1999. Extensive nuclear DNA sequence diversity among chimpanzees. *Science* 286:1159–1162.

Laval G, Excoffier L. 2004. SIMCOAL 2.0: a program to simulate genomic diversity over large recombining regions in a subdivided population with a complex history. *Bioinformatics* 20:2485–2487.

Leuenberger C, Wegmann D. 2010. Bayesian computation and model selection in population genetics. *Genetics* 184:243–252.

Marjoram P, Molitor J, Plagnol V, Tavare S. 2003. Markov chain Monte Carlo without likelihoods. *Proc Natl Acad Sci U S A.* 100: 15324–15328.

McBrearty S, Jablonski NG. 2005. First fossil chimpanzee. *Nature* 437:105–108.

Mevik BH, Wehrens R. 2007. The pls package: principal component and partial least squares regression in R. *J Stat Softw.* 18.

Michalakis Y, Excoffier L. 1996. A generic estimation of population subdivision using distances between alleles with special reference for microsatellite loci. *Genetics* 144:1061–1064.

Morin PA, Moore JJ, Chakraborty R, Jin L, Goodall J, Woodruff DS. 1994. Kin selection, social-structure, gene flow, and the evolution of chimpanzees. *Science* 265:1193–1201.

Neuenschwander S, Largiadèr C, Ray N, Currat M, Vonlanthen P, Excoffier L. 2008. Colonization history of the Swiss Rhine basin by the bullhead (*Cottus gobio*): inference under a Bayesian spatially explicit framework. *Mol Ecol.* 17:757–772.

Nielsen R, Beaumont MA. 2009. Statistical inferences in phylogeography. *Mol Ecol.* 18:1034–1047.

Nordborg M, Donnelly P. 1997. The coalescent process with selfing. *Genetics* 146:1185–1195.

Oates J, Groves C, Jenkins P. 2009. The type locality of Pan troglodytes vellerosus (Gray, 1862), and implications for the nomenclature of West African chimpanzees. *Primates* 50:78–80.

Patin E, Laval G, Barreiro LB, et al. (15 co-authors). 2009. Inferring the demographic history of African farmers and pygmy hunter-gatherers using a multilocus resequencing data set. *PLoS Genet.* 5:e1000448.

Patterson N, Richter DJ, Gnerre S, Lander ES, Reich D. 2006. Genetic evidence for complex speciation of humans and chimpanzees. *Nature* 441:1103–1108.

Pilbrow V. 2006. Population systematics of chimpanzees using molar morphometrics. *J Hum Evol.* 51:646–662.

Stadler T, Haubold B, Merino C, Stephan W, Pfaffelhuber P. 2009. The impact of sampling schemes on the site frequency spectrum in nonequilibrium subdivided populations. *Genetics* 182: 205–216.

Stone AC, Griffiths RC, Zegura SL, Hammer MF. 2002. High levels of Y-chromosome nucleotide diversity in the genus Pan. *Proc Natl Acad Sci U S A.* 99:43–48.

Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123:585–595.

Tavaré S, Balding DJ, Griffiths RC, Donnelly P. 1997. Inferring coalescence times from DNA sequence data. *Genetics* 145: 505–518.

Thornton K. 2005. Recombination and the properties of Tajima's D in the context of approximate-likelihood calculation. *Genetics* 171:2143–2148.

Voight BF, Adams AM, Frisse LA, Qian YD, Hudson RR, Di Rienzo A. 2005. Interrogating multiple aspects of variation in a full resequencing data set to infer human population size changes. *Proc Natl Acad Sci U S A.* 102:18508–18513.

Wahlund S. 1928. Zusammensetzung von Populationen und Korrelationerscheinungen vom Standpunkt der Vererbungslehre aus betrachtet. *Hereditas* 11:65–106.

Webster MT, Smith NGC, Ellegren H. 2002. Microsatellite evolution inferred from human–chimpanzee genomic sequence alignments. *Proc Natl Acad Sci U S A.* 99:8748–8753.

Wegmann D, Leuenberger C, Excoffier L. 2009. Efficient approximate Bayesian computation coupled with Markov Chain Monte Carlo without likelihood. *Genetics* 182:1207–1218.

Wegmann D, Leuenberger C, Neuenschwander S, Excoffier L, Excoffier L. Forthcoming 2010. ABCtoolbox: A versatile toolkit for approximate Bayesian computations. *BMC Bioinformatics.*

Weir BS, Cockerham CC. 1984. Estimating F-statistics for the analysis of population structure. *Evolution* 38:1358–1370.

Won YJ, Hey J. 2005. Divergence population genetics of chimpanzees. *Mol Biol Evol.* 22:297–307.

Yu N, Jensen-Seaman MI, Chemnick L, Kidd JR, Deinard AS, Ryder O, Kidd KK, Li W-H. 2003. Low nucleotide diversity in chimpanzees and bonobos. *Genetics* 164:1511–1518.