

# Distinguishing between population bottleneck and population subdivision by a Bayesian model choice procedure

BENJAMIN M. PETER,\*† DANIEL WEGMANN\*<sup>1</sup> and LAURENT EXCOFFIER\*†

\*Computational and Molecular Population Genetics (CMPG), Institute of Ecology and Evolution, University of Bern, Baltzerstrasse 6, CH-3012 Bern, Switzerland, †Swiss Institute of Bioinformatics, 1015 Lausanne, Switzerland

## Abstract

Although most natural populations are genetically subdivided, they are often analysed as if they were panmictic units. In particular, signals of past demographic size changes are often inferred from genetic data by assuming that the analysed sample is drawn from a population without any internal subdivision. However, it has been shown that a bottleneck signal can result from the presence of some recent immigrants in a population. It thus appears important to contrast these two alternative scenarios in a model choice procedure to prevent wrong conclusions to be made. We use here an Approximate Bayesian Computation (ABC) approach to infer whether observed patterns of genetic diversity in a given sample are more compatible with it being drawn from a panmictic population having gone through some size change, or from one or several demes belonging to a recent finite island model. Simulations show that we can correctly identify samples drawn from a subdivided population in up to 95% of the cases for a wide range of parameters. We apply our model choice procedure to the case of the chimpanzee (*Pan troglodytes*) and find conclusive evidence that Western and Eastern chimpanzee samples are drawn from a spatially subdivided population.

*Keywords:* approximate Bayesian computations, bottleneck, demographic history, island model, model choice

Received 18 December 2009; revision revised 4 June 2010; accepted 22 June 2010

## Introduction

In conservation biology, information on current and past population size is crucial for determining appropriate action. For instance, if there is evidence that a population has recently declined, some appropriate restoration measures may be required. On the other hand, if a species had a low but stable population size for a long time, more simple conservation measures may be taken. Therefore, methods for assessing the demographic history of a species have a long tradition in applied ecol-

ogy and conservation biology. Typically, these methods are based on a survey of census data coupled with implicit or explicit demographic models (Besbeas *et al.* 2002). Unfortunately, data sets spanning over more than a couple of years are rare, and most of these methods do not allow for inference beyond the time of sampling. Especially for species with long generation times, such as large mammals or birds, the typical study duration of a few years might not encompass a species complete lifespan. A possible solution is to use paleontological or archaeological data, or even historical trade records (Jackson *et al.* 2001). However, these methods often only provide incomplete or unsuitable data. Moreover, because this data typically have not been recorded for ecological purposes, it might be difficult to obtain confidence intervals for inferred parameters.

Correspondence: Laurent Excoffier, Fax: +41 31 631 48 88; E-mail: laurent.excoffier@iee.unibe.ch

<sup>1</sup>Present address: Department of Ecology and Evolutionary Biology, University of California Los Angeles, Los Angeles, CA 90095, USA.

In the past decade, alternative methods based on the analysis of genetic data with coalescent methods have emerged (Kingman 1982; Wakeley 2008), allowing the estimation of effective population size (Kuhner *et al.* 1995), population divergence time (Hey & Nielsen 2004) or population size changes (Cornuet & Luikart 1996; Beaumont 1999; Drummond *et al.* 2005; Heled & Drummond 2008).

There are a number of different procedures for estimating or detecting population size changes. First, some approaches use a specific feature of the data expected under a given demographic model. For instance, rare alleles are more likely to be lost in a bottleneck than more frequent alleles (Cornuet & Luikart 1996; Garza & Williamson 2001), and this fact can be exploited to recognize past bottlenecks. Second, other approaches try to infer the past demography of a population from the rate of coalescent events along a single genealogy, when compared to that of the standard coalescence process in a stationary population (Pybus *et al.* 2000; Drummond *et al.* 2005). A third possibility is the use of a full likelihood approach to estimate the parameters of a model of population size change from multilocus data, as for instance implemented in the program *msvar* (Beaumont 1999; Storz & Beaumont 2002) specifically designed to deal with microsatellite data. While this latter approach should lead to accurate and reliable conclusion with sufficient data, it still requires the exploration of a complex multidimensional parameter space, usually performed using a Markov-Chain Monte Carlo (MCMC) approach which can become computationally very intensive when the number of loci is large. The Approximate Bayesian Computation (ABC) framework (Tavare *et al.* 1997; Beaumont *et al.* 2002) was developed to estimate parameters of more complex models in manageable computer time. ABC has then been successfully applied to infer demographic history for various organisms, such as humans (Fagundes *et al.* 2007), *Arabidopsis thaliana* (Francois *et al.* 2008), or *Drosophila melanogaster* (Thornton & Andolfatto 2006), and several implementations of the ABC algorithm have been published (Beaumont *et al.* 2002; Excoffier *et al.* 2005a; Cornuet *et al.* 2008; Jobin & Mountain 2008; Wegmann *et al.* 2009). In the past few years, various variations of the algorithm have been developed (Beaumont *et al.* 2002; Sisson *et al.* 2007; Blum & François 2010), but most ABC implementations make inferences based on a set of summary statistics (e.g. the number of alleles, heterozygosity) designed to capture most of the information contained in the original data. In the simplest ABC procedure, a large number of simulations are performed from the model under consideration, using random parameter values drawn from some prior distributions. The summary statistics based on these simu-

lations ( $S$ ) are then compared to the summary statistics of observed data ( $S^*$ ). Using some Euclidean distance measure  $\delta = \|S - S^*\|$ , simulations are retained if  $\delta < \epsilon$ , an arbitrarily small distance used as a rejection criterion. The posterior parameter distribution is then calculated from the distributions of the parameters of the retained simulations using a locally weighted linear regression step (Beaumont *et al.* 2002).

Like all model-based inference procedures, full likelihood approaches assume a fixed demographic model. If model assumptions are violated or if data have been generated under a different model, this may of course lead to erroneous inferences and conclusions. In this respect, Nielsen & Beaumont (2009) showed that when data are simulated under a finite island model with a constant population size, then the program *msvar* (Beaumont 1999) is likely to wrongly infer the occurrence of a strong population bottleneck, even though the subdivided population size did not change. This erroneous imputation can be understood by considering the coalescent process in an island model. In a subdivided population, the coalescence process may be divided into two distinct phases (Wakeley 1999). During the first phase, called the 'scattering phase', lineages either coalesce or migrate quickly to other demes than those being sampled, until there is a single lineage left per deme. In the second phase, called the 'collecting phase', the remaining lineages follow a standard coalescent process, like in a single population, but on a different time scale. Thus, the rate of coalescent events during the scattering phase is larger than in the collecting phase, which lasts much longer than the scattering phase. The resulting genealogy and the associated pattern of genetic diversity will thus look very much like that expected after a recent bottleneck, with an excess of recent coalescent events (Nielsen & Wakeley 2001; Ptak & Przeworski 2002). Note that structured and declining population dynamics may result in virtually identical genealogies. For instance, it has been shown that a single expanding population model can lead to any possible allele frequency spectrum (Myers *et al.* 2008). What differs between demographic models, however, is the likelihood with which different genealogies are produced, and we can therefore distinguish between different models using a probabilistic approach.

In this study, we propose to use an ABC model choice procedure to distinguish between populations that are structured from populations that are panmictic, but which recently changed in size. We show here by simulation that our model choice procedure has high power to assign simulated data sets to the correct evolutionary model, even with a moderate number of loci. We finally illustrate our method on a

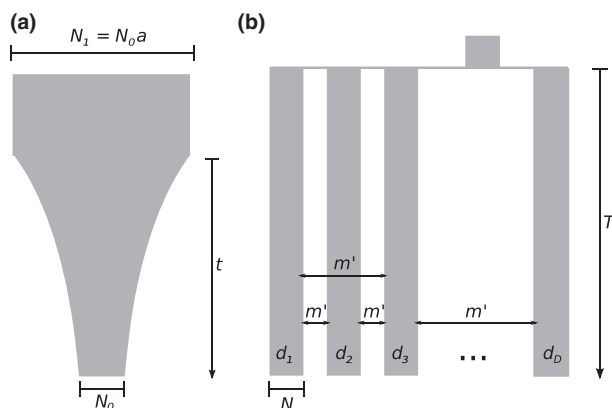
large microsatellite data set of the chimpanzee (*Pan troglodytes*).

## Material and methods

We aim here at evaluating whether observed patterns of genetic diversity can be better explained by a model of recent population size change or by a model of subdivided populations. To do this, we propose to apply previously described ABC model selection procedures (Pritchard *et al.* 1999; Beaumont 2008) to assign test data sets to one of our two alternative models.

### Models

**Population size change (PSC) model.** The PSC model we use is similar to the model with exponential population size change described in Beaumont (1999), which is also implemented in the msvar program (Beaumont 1999) (see Fig. 1a). This model allows for exponential growth/decline at a constant rate starting from the current population with size  $N_0$ , and, going backwards in time, to vary exponentially over  $t$  generations to a population of size  $N_1 = aN_0$ . Thus, if  $a > 1$ , the population has declined in size, if  $a < 1$  the population has expanded, and if  $a = 1$ , the population is obviously remaining constant in size. We follow Beaumont (1999) by estimating the scaled parameters  $\theta = N_0\mu$  and  $\tau = t/N_0$  of the PSC model, but it was judged more convenient to set up priors on the natural parameters  $N_0$  and  $t$ . We therefore used the following prior distributions for the model parameters:  $N_0 \sim LU(100, 50\,000)$ ,  $a \sim LU(10^{-3}, 10^3)$  and  $t \sim LU(1, 10^3)$ ,



**Fig. 1** Schematic representation of the two alternative models. (a) Model of population size change with current population size  $N_0$ , ancient population size  $N_1$  and  $t$  is the time since population size change. We estimate the scaled parameters  $\theta = 2N_0\mu$  and  $\tau = t/N_0$  from this model, where  $\mu$  is the mutation rate. (b) Recent finite island model with  $D$  demes (denoted  $d_1 \dots d_D$ ) of identical haploid size  $N$ , the migration rate  $m' = m/(D-1)$  is identical between all demes, and  $T$  denotes the age of the island model.

where  $LU(a,b)$  is a log-uniform distribution with density  $f(x) = \{[\log(b) - \log(a)]x\}^{-1}$  if  $x \in [a; b]$  and 0 otherwise. In all simulations, we used a sample size of  $n = 25$  diploid individuals.

**Recent Finite Island (RFI) model.** To realistically model a subdivided population, we have simulated data from a finite island model with  $D$  demes of size  $N$  genes (Wright 1931) where current demes are assumed to have originated from a single deme of size  $N_A = N$  some time  $T$  ago (see Fig. 1b), which seems biologically more realistic than assuming an infinite age for the island model. The origin of this subdivision could for instance be considered as the time of a speciation event, the origin of some invasive species, or a migration out of an ice age refugium. In our RFI model, we simulated 100 demes of haploid size  $N$ , with immigration rate  $m$ , so that the total number of genes arriving in any deme is  $Nm$  per generation. While coalescent simulations are performed for a haploid population, we simulated a diploid sample of size  $n$ . As the sampling scheme may have a major impact on the resulting genetic diversity (Städler *et al.* 2009), we defined an additional sampling parameter  $S$ , specifying that  $nS$  individuals are sampled from a single focal deme and that the remaining  $n(1-S)$  individuals are sampled from demes chosen at random from the whole subdivided population. Thus, if  $S = 0$ , individuals are sampled randomly over all demes, and if  $S = 1$ , all individuals are sampled from the same deme. We have additionally chosen to constrain the product  $nS$  to integer values, and  $S$  was thus restricted to  $n + 1$  possible values, between 0 and 1. Prior distributions for this model are  $Nm \sim LU(0.1, 100)$ ,  $T \sim LU(10^2, 10^5)$  and  $nS \sim dUI(0, n)$ , where  $dUI(a,b)$  is a discrete uniform distribution on the integers between  $a$  and  $b$ . Like under the PSC model, we used a sample size of  $n = 25$  diploid individuals for all simulations.

### Genetic data

Throughout this study, we have simulated and analysed microsatellite data made up of 5, 10, 20, 50, 100, or 200 unlinked loci. Microsatellite data were generated under a generalized stepwise mutation model as implemented in SIMCOAL2 (Laval & Excoffier 2004). The mutation rate was allowed to vary between loci, with locus-specific mutation rates being drawn from a Gamma distribution  $\text{Gamma}(\alpha, \alpha/\bar{\mu})$  (Voight *et al.* 2005; Fagundes *et al.* 2007), where  $\bar{\mu}$  is the mean mutation rate and  $\alpha$  is a shape parameter. Unless mentioned otherwise  $\bar{\mu}$  was set to  $5 \times 10^{-4}$  per generation, which is a typical value for mammalian microsatellites (Ellegren 2004), and  $\alpha$  was treated as a nuisance parameter with values taken from  $U(5, 20)$ .

## ABC

We used the ABC procedure with postsampling local linear regression (2002) to estimate the parameters of each model. We used the program ABCsampler from the ABCTOOLBOX package (Wegmann *et al.* 2009) to specify the priors and to automatically launch simulations, which were done using the program SIMCOAL2 (Laval & Excoffier 2004). Six summary statistics were calculated using a console version of ARLEQUIN 3.12, called arlsumstat (Excoffier & Lischer 2010):  $F_{IS}$ , Garza & Williamson's (2001)  $M$ , the number of alleles  $K$  and its standard deviation over loci, as well as the heterozygosity  $H$  and its standard deviations over loci. These summary statistics were chosen for the following reasons: For a given sample size, the number of alleles  $K$  is informative not only on population size (Ewens 1972; Wakeley 1998) but also on inbreeding levels (Fu 1997).  $M$  was specifically developed as a measure for detecting signals of bottlenecks, and we used its modified definition  $K/(R + 1)$  (Excoffier *et al.* 2005b), where  $R$  is the allelic range (i.e. the difference in number of repeat between the longest and the smallest allele) of the microsatellite. During a bottleneck,  $R$  decreases less than  $K$ , leading to  $M$  values lower than 1 (Garza & Williamson 2001). Heterozygosity is also assumed to decrease more slowly than  $K$  after a bottleneck (Cornuet & Luikart 1996), and it is reduced in a subdivided population (Wahlund 1928). Finally,  $F_{IS}$  typically takes positive values if the sample is drawn from a subdivided population (Wahlund 1928).

Rejection sampling, postsampling locally weighted linear regression and calculation of posterior distributions were performed using the program ABCEST2 (Excoffier *et al.* 2005b). For each model, we performed  $10^6$  simulations with values randomly drawn from the prior. For parameter estimation, we retained the 5000 (0.5%) simulations with associated Euclidean distance between observed and simulated summary statistics closest to the observed data set. Parameters were transformed using a log(tan) transformation (Hamilton *et al.* 2005), to keep posteriors within the boundaries of the priors.

### *Inference of population size change from samples drawn from a subdivided population*

In order to check that our ABC procedure recovers the same effect as that described by Nielsen & Beaumont (2009) using a full-likelihood approach, namely the inference of a recent bottleneck if the sample is drawn from a subdivided population, we estimated the parameters of a population size change model via ABC from two series of 20 000 data sets generated under an RFI

model. In the first series, sampled genes were chosen randomly from the whole subdivided population ( $S = 0$ ), and in the second series, all genes were sampled from a single deme ( $S = 1$ ). In both series,  $T$  and  $Nm$  were drawn randomly from their respective priors. For each data set, we estimated the marginal posterior distribution of the population size change parameter  $a$  of a PSC model assuming no subdivision. To assess the significance of  $a$ , we followed a procedure described in Lee (2004, but see also Lindley 1965; Zellner 1971), and calculated  $P = \Pr(a < 1)$  for all data sets by numerical integration over the estimated marginal density of  $a$ . We assumed that values of  $P$  smaller than 0.025 to be considered as a significant signal of recent bottleneck, and values of  $P > 0.975$  as a significant signal of population expansion under a PSC model.

### *Estimation precision*

As ABC is an approximate inference method, it is important to check for possible biases in the estimates. One way to do this is to simulate data sets with known parameters and then check if we are able to correctly estimate them. We did this for a total of 21 sets of parameters, where we simulated 100 data sets with 10, 50 and 200 loci each.

Unbiased posterior distributions have a well-balanced coverage property, such that the true value of a parameter should be found in  $q$  per cent of the times in a  $q$ % credible interval. We checked this by creating 10 000 artificial data sets of 200 unlinked loci for each model, randomly drawing each time parameters from our prior distribution. We then estimated the parameter posterior distributions with ABC and estimated the proportion of the true parameters contained in 50%, 90% and 95% credible intervals. To detect regions of the parameter space where we have potentially inaccurate posteriors, we further examined local coverage properties for different parameter values. We divided the 10 000 simulated data sets into 20 discrete bins according to the value of the estimated posterior mode, for each parameter independently. The coverage property was then assessed for each bin by computing the proportion of simulated data sets where the true parameter value falls in 50, 90 or 95% HPD credible interval.

### *Model comparison*

We aim here at calculating the relative probability of two models  $M_1$  and  $M_2$  given an observed set of summary statistics  $S^*$ , which should lead to the calculation of a Bayes Factor  $BF = \Pr(data|M_1)/\Pr(data|M_2)$ . Because we do not compute the model likelihoods in our ABC approach, BF cannot be explicitly computed,

and we have therefore used two alternative ABC approaches to estimate the relative probabilities of the two models  $M_1$  and  $M_2$ .

The first procedure is based on an extension of logistic regression leading to the estimation of the relative posterior probabilities of each model as proposed by Beaumont (2008), which has previously been applied in an evolutionary context (Fargundes *et al.* 2007). The second procedure follows Pritchard *et al.* (1999). Simulations performed under the two models are sorted according to their Euclidean distance  $\delta$  to the observed summary statistics, and the relative posterior probabilities of the two models are estimated as the proportion of simulations coming from each model among a fixed number  $f$  of simulations with smallest distance  $\delta$ . We followed Estoup *et al.* (2004) in fixing  $f$  to 200, but different values of  $f$  had very little impact on the results (results not shown).

In our Bayesian setting, the decision on which model is better supported is based on posterior model probability or on the Bayes Factor  $BF = \Pr(M_1)/\Pr(M_2)$ , which differs from a decision process under a frequentist approach based on a  $P$ -value. The  $P$ -value in a frequentist settings gives a direct estimate of the probability to wrongly reject a correct hypothesis (type-I error), and the Bayes Factor can be used similarly to predict the probability to wrongly choose a given scenario as  $P = 1/(BF + 1)$  (Lee 2004). Indeed, a decision in favour of a model based on  $BF = 2$  should be wrong in 1/3 of the cases. In order to assess if the BFs resulting from our model choice procedures are unbiased, we simulated 10 000 data sets under both the RFI and PSC model and computed their associated BFs. We then

allocated our data sets to discrete bins of BFs, and we checked if the proportions of data sets generated under RFI and PSC models were equal to  $P$  and  $1-P$ , respectively, as predicted by the average computed BFs. This procedure has the advantage not to require us to make any decision on the origin of each data set.

In order to decide if an observed data set is more likely to have been generated under model  $M_1$  or under model  $M_2$ , we used an arbitrary threshold value  $k \geq 0.5$  as follows: if the relative posterior probability of any model was larger than  $k$ , then the data set was assigned to that model. If no model posterior probability exceeded  $k$ , then we assumed the outcome to be inconclusive and did not assign the data set to any model. Thus, the number of false positives decreases with increasing value of  $k$ , but at the expense of an increasing number of unassigned data sets.

### Application to chimpanzee

We illustrate our model choice procedure by analysing samples of chimpanzee (*Pan troglodytes*) from various origins and analysed for a large number of microsatellites (Becquet *et al.* 2007). For our analysis, we used only individuals that were congruently allocated by zoo records and genetics (Becquet *et al.* 2007) to either the western (*P. t. verus*) or eastern (*P. t. schweinfurthii*) chimpanzee populations. This procedure resulted in data set of sizes  $n_W = 50$  individuals and  $n_E = 6$  individuals for the western and eastern population, respectively. Each of these two samples was then analysed separately.

**Table 1** Prior distributions

Model	Parameter	Distribution	Mean	Mode	Quantile		
					5%	50%	95%
PSC	$N_0$	Log-uniform [ $100, 5 \times 10^4$ ]	8029	100	136	2236	36646
PSC	$a$	Log-uniform [ $10^{-3}, 1000$ ]	62	0.0001	0.0019	1.0	501.2
PSC	$t$	Log-uniform [ $1, 10^3$ ]	144	1	1.41	31.6	708
PSC	$\tau$	$\tau = tN_0^{-1}$	0.23	0.013	0.00016	0.014	1.258
PSC/RFI	$n$	Constant	25	25	25	25	25
PSC/RFI	$\mu$	Gamma ( $\alpha, \alpha/0.0005$ )	$5 \times 10^{-4}$	—	$2.7 \times 10^{-4}$	$4.9 \times 10^{-4}$	$7.7 \times 10^{-4}$
PSC/RFI	$\alpha$	Uniform [5,20]	12.5	—	5.75	12.5	12.75
PSC	$\theta$	$\theta = 2\mu N_0$	8	0.1	0.14	2.24	36.6
RFI	$D$	Constant	100	100	100	100	100
RFI	$N$	Constant	200	200	200	200	200
RFI	$T$	Log-uniform [ $100, 10^5$ ]	14462	100	141	3162	70798
RFI	$Nm$	Log-uniform [0.1,100]	14.46	0.1	0.141	3.16	70.8
RFI	$S$	Uniform [0,1]	0.5	—	0.05	0.5	0.95

RFI, Recent Finite Island model; PSC, Population Size Change model.  $N_0$ , current population size;  $N_1$ , population size before change;  $a$ ,  $N_1/N_0$ ;  $t$ , time since population size change;  $n$ , sample size (number of diploid individuals);  $\mu$ , mutation rate;  $\alpha$ , shape parameter of Gamma distribution;  $D$ , number of demes;  $N$ , number of gene copies per deme;  $T$ , age of the island model;  $Nm$ , Number of immigrant genes per deme per generation;  $S$ , Sampling parameter (proportion of individuals sampled in a given deme, see text).

From a total of 310 available microsatellite loci, we discarded all loci that were monomorphic for the studied populations or those present on non-autosomal chromosomes. Additionally, we removed all loci with very low levels of heterozygosity ( $H < 0.2$ ), as they might characterize loci that are no longer STRs. Alleles that were imperfect repeats were considered as missing data, and loci with more than 5% missing data were further discarded, resulting in a set of 236 loci for the western population, and of 233 loci for the eastern population.

For each data set, we run ABC under the RFI and PSC models, with priors customized to more closely fit the natural population as follows: because the effective population sizes have been estimated to be around 7000–10 000 individuals for the western population (Won & Hey 2005; Becquet & Przeworski 2007; Becquet *et al.* 2007; Caswell *et al.* 2008; Wegmann & Excoffier 2010), and to about 15 000 individuals with fairly wide confidence intervals for the eastern population (Becquet & Przeworski 2007; Wegmann & Excoffier 2010), we used a prior of  $N_0 \sim LU(10^2, 10^6)$  under the PSC model, and we set the prior of local deme size to  $N \sim U(100, 1000)$  under the RFI model, with the number of demes kept constant at 100. In order to reflect our uncertainty in chimpanzee mean mutation rate  $\bar{\mu}$  we used a prior on  $\bar{\mu} \sim U(10^{-4}, 5 \times 10^{-4})$  and we used a wider shape parameter  $\alpha \sim U(2, 20)$ . Additionally, we allowed for uncertainty in the size of the ancestral pop-

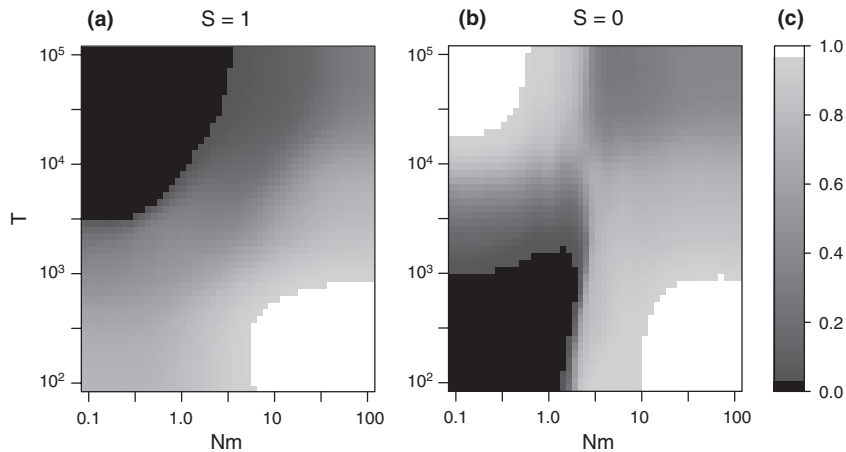
ulation in the RFI model by taking a prior  $N_A \sim LU(1000, 50\ 000)$ . We also set the sample size of simulated data equal to those of observed data sets,  $n_W$  and  $n_E$ . We accounted for missing data in the chimpanzee data set by removing the same number of alleles missing for each locus and each simulation before calculating the summary statistics (Wegmann & Excoffier 2010).

## Results

### *Effect of population subdivision on estimation of population size change*

We tested the effect of assuming an absence of population subdivision when inferring population size change from genetic data by analysing data sets generated under an RFI model assuming a PSC model. As shown in Fig. 2, depending on the type of sampling in the subdivided population, the age of the island model, and the amount of gene flow between demes, very different demographic histories may be inferred, ranging from evidence for large demographic expansions to evidence for strong population decline.

If all the samples are taken from a single deme ( $S = 1$ , Fig. 2a),  $T$  is large ( $T > 5000$  generations) and gene flow is limited ( $Nm < 5$ ), which closely corresponds to the situation studied by Nielsen & Beaumont (2009), we also find evidence for a strong population



**Fig. 2** Relative size change estimated by assuming a panmictic population, when data sets are actually generated under the RFI model. We plot in Fig. 2a,b the value of  $P$ , the probability of a population increase, defined as  $P = \Pr(a < 1)$ , where  $a$  is the ratio between the ancient and the current population size. A large value of  $P$  suggests that the population has been increasing, while a small value of  $P$  would indicate that the population has been declining. Parameters of the simulated data sets are  $S$ : sampling parameter,  $Nm$ : number of immigrants genes per generation,  $T$ : age of the island model (see text). (a)  $S = 1$ : the individuals are sampled from a single deme; (b)  $S = 0$ : the individuals are sampled randomly over the whole subdivided population; (c) Colour code for probability  $P$ : Pure black and pure white areas represent values of  $P$  outside of  $[0.025; 0.975]$ . Each plot is based on the analysis of 20 000 data sets of 200 unlinked microsatellite loci simulated under the RFI model, with parameters randomly drawn from prior distributions shown in Table 1.

decline. This can be understood by considering the coalescent process where we expect a large number of recent coalescent events in the scattering phase, and a few much older coalescent events occurring in the collecting phase or in the ancestral population before the creation of the islands. Under the same conditions, but with a different sampling scheme, the inference is completely reversed, as we get significant evidence for population expansion. This is because of the absence of the scattering phase, very few coalescent events in the island phase, and most coalescences occurring in the ancestral population before the onset of the island phase (see Excoffier 2004 for an analytical treatment of a similar case under the infinite island model).

Inference appears not to depend on sampling conditions when  $Nm$  is above a certain threshold ( $Nm > 10$ ). In this scenario, migration is frequent enough that the effect of sampling becomes negligible. When  $T$  is small ( $T < 5000$  generations), most coalescent events will be in the ancestral population, leading to a signal of a population expansion corresponding to the creation of the island model. When  $T$  is larger ( $T > 5000$  generations), most coalescent events will occur during the island phase and less in the ancestral population and the population will thus be close to an equilibrium island model without size change (Wakeley 1999).

#### *Precision of parameter estimates and coverage properties*

The accuracy of parameter estimation under the island model is reported in Fig. S1 (Supporting Information). Not surprisingly, we find estimations to be more precise with larger number of studied loci. The age of the island model ( $T$ ) and the number of gene exchanged between demes ( $Nm$ ) are very well estimated over the whole parameter range. The parameter describing the sampling scheme  $S$  is also relatively well recovered for low to moderate amount of gene flow ( $Nm = 0.3$  and  $3$ , respectively). However, this parameter is more difficult to estimate when gene flow is high between demes ( $Nm = 30$ ), because in that case the level of genetic diversity is independent of the exact place of sampling as genes move rapidly between demes and the whole subdivided population resembles more a panmictic population.

Except for the population size ratio  $a$ , the precision of estimates under the population change model depends heavily on the timing of the decline  $\tau = t/N_0$  (Fig. S2, Supporting Information). As we fixed the onset of population decline  $t$  to 100 generations, the first set of simulations with  $N_0 = 100$  corresponds to  $\tau = 1$ , which results in good estimates for both  $N_0$  and  $t$ . This is because most coalescent events will occur during the population size change, and the varying rate of

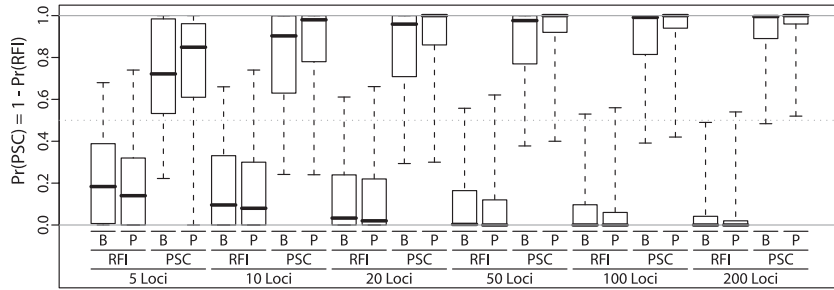
coalescent events will provide information on the population size change and the timing of this change. On the other hand, if  $\tau$  is too small (implying that the population decline occurred very recently), there will be very few coalescent events during the population size change and thus little information to infer this process, and the rate of coalescence will only provide information on the ancestral size  $N_1$  and give little signal of population size change. As mentioned earlier,  $a$  seems to be well estimated over the whole parameter space, with credible intervals declining with both increasing  $a$  and increasing number of loci.

The coverage of the posterior distributions seems overall relatively correctly estimated for the PSC model, whereas the posteriors obtained under the RFI model are slightly too narrow as the coverage of the 90% and 95% credible intervals is underestimated by about 5% for the three parameters of this model (see Table S1). A closer look at Fig. S3 (Supporting Information), where we report the coverage of the posteriors for different parameter values, shows that this is mainly because of small values of  $T$ , whereas the posteriors of larger  $T$  are correctly estimated.  $Nm$  posteriors are also too narrow for values of  $Nm < 1$ , but much better estimated for larger values (Fig. S3d, Supporting Information). Furthermore, we observe that posteriors are too narrow for some extreme values of several parameters, such as very small values of  $\tau$  and large values of  $\theta$  and  $a$  under the PSC model, as well as for small  $Nm$  values and large  $S$  values under the RFI model.

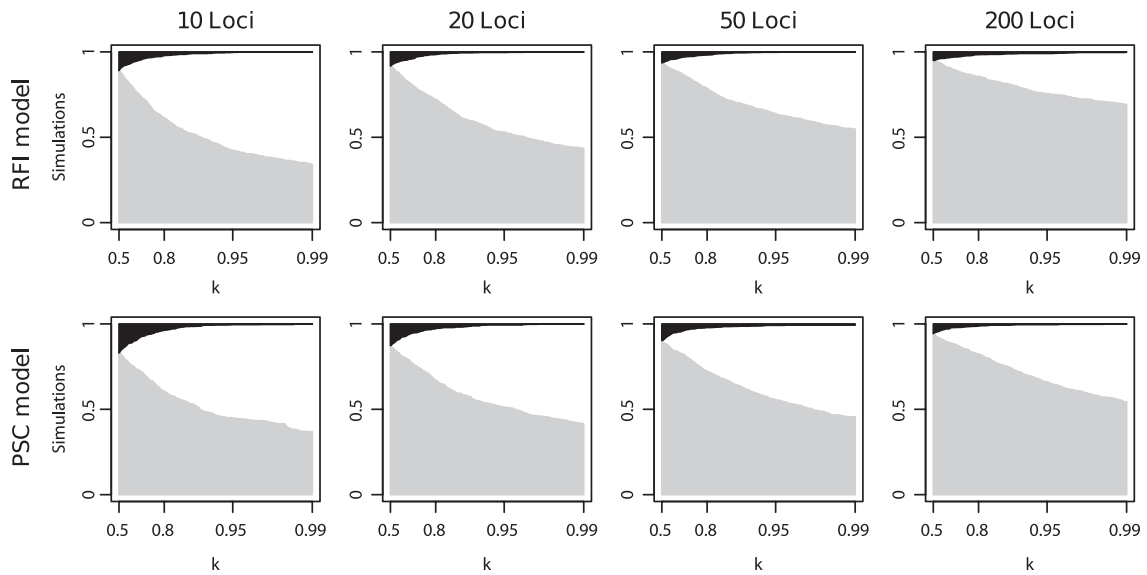
#### *Model choice accuracy*

In Fig. 3, we report the posterior probability of the PSC model estimated for various data sets generated under PSC and RFI models and in Fig. 4 the accuracy of the model choice procedure for different  $k$  values and different number of loci, as inferred with the logistic regression method of Beaumont (2008) (results based on Pritchard *et al.* (1999) method give very similar results and are given in the supplementary Fig. S4, Supporting Information).

As expected, the accuracy of the model choice procedure increases with the number of loci and the models are very well discriminated with more than 20 loci (Fig. 3). In Fig. 4, we see that for the least stringent threshold of  $k = 0.5$ , 89.5% and 95.1% of the pseudo observed data sets are correctly assigned when using 10 and 200 loci, respectively. A  $k$  value of 0.5 implies that all the remaining simulations are assigned to the 'wrong' model. Fortunately, the number of false positives quickly decreases with  $k$ . For  $k = 0.8$ , <5% of all simulations are assigned incorrectly for all data sets, and <2% for data sets are incorrectly assigned with 50



**Fig. 3** Accuracy of the model choice procedure. The boxplots show the lower and upper quartiles, the median and the limits of a 95% interval.  $\Pr(PSC)$ : Relative probability that data were generated under a PSC model;  $\Pr(RFI)$ : Relative probability that data were generated under an RFI model. Each boxplot is based on 1000 replicates, with model parameters drawn randomly from the priors shown in Table 1. B: Procedure of Beaumont (2008), P: Procedure of Pritchard *et al.* (1999) (see text).



**Fig. 4** Accuracy of model assignment using the model choice procedure of Beaumont (2008). For a given  $k$  value ( $x$ -axis), the grey area gives the proportion of data sets that were correctly assigned to the model they were simulated from. The black area gives the proportion of data sets that were assigned to the wrong model, and the white areas correspond to simulations that could not be assigned unambiguously for the respective  $k$  value. Note that the  $x$ -axis is on a logarithmic scale.

or more loci. With  $k = 0.95$ , <1% of all data sets are assigned incorrectly.

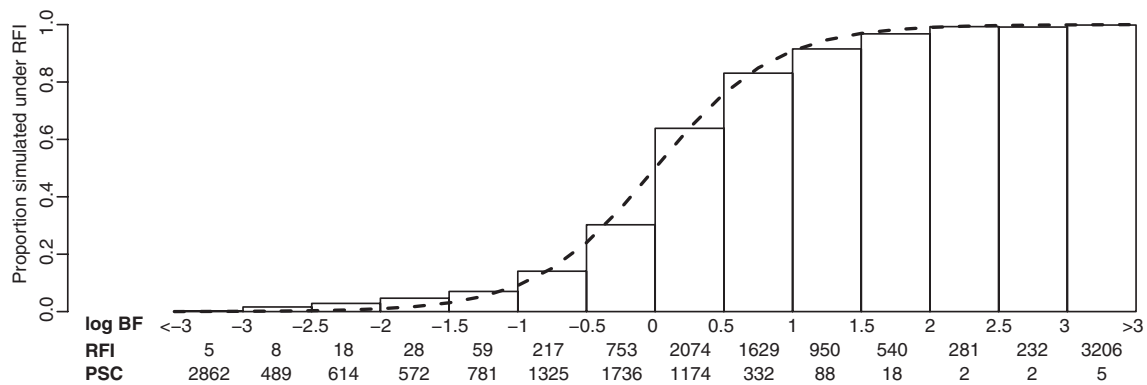
To see to which extent the Bayes Factors computed from our ABC approach are informative about the accuracy of the model choice, we performed our model choice procedure on 10 000 data sets randomly generated under the RFI model and on 10 000 others generated under the PSC model. We computed the Bayes Factor associated with each of these data sets, and report in Fig. 5 the proportion of data sets generated under a PSC model showing a given Bayes Factor in favour of the RFI model vs. its expectation based on the Bayes Factor. We can see in Fig. 5 that the simulated proportions match very well those expected from the Bayes Factors. We note however a slight excess of data sets generated under the RFI model in the bins below

$1/30$  ( $\log_{10}(BF) < -1.5$ ), implying that we overestimate large Bayes Factors favouring the PSC model. However, this effect would only bear on a relatively small number of simulations (0.6% of all simulations) that would not be correctly assigned if a decision was made on the basis of these Bayes factors. Nevertheless, this bias should be considered when decisions are made. On the other hand, we do not observe a similar bias for data sets leading to large Bayes factors in favour of the RFI model.

### Application

For both the eastern and western chimpanzee population, we simulated 1 million data sets under the PSC and RFI models. We then applied both model choice





**Fig. 5** Assessing the ability of Bayes Factor to correctly predict Type I error. We performed the model choice procedure of Beaumont (2008) on 10 000 simulated data sets generated under the RFI model and 10 000 data sets generated under the PSC model. The Bayes Factor BF was estimated for each data set. We pooled simulations with similar BFs into bins and computed the proportion of data sets generated under the RFI model in each bin (histogram), which should be equal to  $1/(\overline{BF} + 1)$  (dashed line), where  $\overline{BF}$  is the average BF in the bin.

procedures. The RFI model was strongly favoured in both populations. For the western population, the probability of the RFI model was 0.98 using Pritchard *et al.* (1999) method, and  $>0.99$  using the method of Beaumont (2008). For the eastern population, the method of Beaumont (2008) again strongly favoured the RFI ( $>99\%$  probability), and support from Pritchard *et al.*'s (1999) method was slightly lower at 94% probability for the RFI model. This higher support for this model of population subdivision is also confirmed by a visual inspection of pair-wise scatter plots of the summary statistics (Figs S5 and S6, Supporting Information). For the RFI model, the retained simulations nicely cluster around the observed values, whereas the retained simulations are much more widely spread for the PSC. Furthermore, as our results are strongly supporting the RFI model, they should not be affected by the possible bias reported in Fig. 5, Supporting Information, as this only occurs for the PSC model.

Our analysis supports previous results of Becquet *et al.* (2007) who found evidence of internal population structure within the western chimpanzee subpopulation based on a principal components analysis, and of Leuenberger & Wegmann (2009), who obtained a similar result using a different Bayesian approach. Our findings further support the hypothesis of Wegmann & Excoffier (2010) that the signal of chimpanzee population decline reported by Caswell *et al.* (2008) may have indeed been caused by internal population structure.

It has to be noted that individuals belonging to the eastern chimpanzee population have been captured several hundred kilometres apart from each other (Becquet *et al.* 2007), and that the true origin of many individuals of the western population is unknown. Therefore, the assumption of population subdivision and heterogeneous sampling from several demes is very likely.

We clearly see in Figs S5 and S6 (Supporting Information) that the high observed  $F_{IS}$  values are extremely difficult to obtain under a simple model of population size change. Such high values of  $F_{IS}$  are typically attributed to inbreeding, and we cannot rule out the possibility of inbreeding, but given the sampling scheme, the presence of hidden population structure seems like a more plausible explanation for these results (Leuenberger & Wegmann 2009).

## Discussion

### *Effect of population subdivision on population size change inference*

We globally find that inferences on population size change can be drastically affected when population subdivision is not properly accounted for. While the interpretation of inferences under a wrong model makes little sense *per se*, it is interesting to see that data drawn from an island model can show extremely different patterns, concordant with declining, constant or increasing population sizes. Whereas we confirm that the analysis of a sample of genes drawn from a single deme belonging to an island population gives a signal of a recent bottleneck (Nielsen & Beaumont 2009), very different inferences can be obtained if the sample includes individuals from several demes, if the population has been recently subdivided, or if the amount of gene flow between demes is large (see Fig. 2). While it remains to be tested, we believe that many real samples are drawn from subdivided populations, and it might therefore be appropriate to be cautious when interpreting the output of programs inferring past population size changes assuming population panmixia (e.g. msvar Beaumont 1999; or beast Drummond *et al.* 2005; see also Leblois *et al.* 2006).

### Model choice accuracy

Not surprisingly, the available number of loci is one of the main factors determining the precision of the model choice procedures, and even better results would certainly be obtained with more than the 200 loci we simulated (see Figs 3 and 4). Interestingly, even for large numbers of loci there are still quite a number of simulations that cannot be assigned to the correct model. For the RFI model, we show in Fig. S7 (Supporting Information) that this is primarily because of simulated cases when both  $T$  and  $Nm$  are large. This result makes sense because a subdivided population exchanging many migrants tends to behave like a large panmictic population. For data generated under the PSC model, we find that models are more difficult to distinguish when  $a < 1$  (expansion), and both  $N_0$  and  $t$  are large (Fig. S8, Supporting Information). This assignment problem is likely due to the fact that star-shaped genealogies obtained under such old and strong expansions look indeed very similar to those obtained under our RFI model. Additionally, we find that data sets where populations do not drastically change in size ( $a \approx 1$ ), are also prone to wrong assignment, as these data sets may resemble those generated under an old island model with large levels of gene flow.

Overall, our ABC model choice provides satisfying results (Figs 3 and 4), even when used with as few as 10 loci. In that case, when using a threshold value of  $k = 0.8$  for model probability assignment, the error rate is below 5% for the logistic regression method of Beaumont (2008), and about 60% of the simulations can be unambiguously assigned to the correct model. With 50 microsatellite loci, which is still a reasonable figure for conservation genetics studies, the error rate goes down to 1.8% for  $k = 0.8$ , and 78.6% of the simulations are unambiguously assigned to the correct model.

Interestingly, the specific structure of the island model has little impact on the outcome of the test: When we simulate data sets under finite island models with different number of demes and different deme size than those used to test the reference RFI model, the results of the model choice procedure are barely affected (Fig. S9, Supporting Information). It suggests that the composite parameter  $Nm$  is more important in defining the population structure than are the actual number of demes and deme size, provided that the number of demes is sufficiently large.

Quite remarkably, we find that the performance of our model choice procedure is not very sensitive to the total number of simulations performed under each model. As shown in Fig. S10 (Supporting Information), the distribution of model posterior probabilities estimated from just 10 000 simulations are overall very similar to those

estimated with 1 million simulations, and the simulation of 100 000 data sets under each model should therefore be enough to obtain accurate model relative probabilities. It suggests that even though precise parameter estimation still requires a large number of simulations (the exact number depending on the complexity of the underlying model), models can be correctly told apart even if their parameters are not accurately estimated. This makes sense because model choice based on ABC is actually nothing more than estimating a single parameter with only two possible states: our two models.

An important summary statistic which certainly allows one to distinguish between the two competing models is the inbreeding coefficient  $F_{IS}$ . Under the RFI model,  $F_{IS}$  may take values up to 0.8 (see Figs S5 and S6, Supporting Information), while  $F_{IS}$  hardly ever reaches values higher than 0.05 without population subdivision (Leuenberger & Wegmann 2009). This Wahlund effect occurs here because we sampled at least two genes per deme to mimic the sampling of diploid individuals, and this pair of genes is obviously more related than a pair taken at random when gene flow is limited between demes, while non-zero  $F_{IS}$  values only occur because of the sampling variance in panmictic populations (Wegmann & Excoffier 2010). However, the other statistics still provide information for model choice, as shown in Fig. S11 (Supporting Information) where we report model posterior probabilities computed without the use of  $F_{IS}$ . PSC model posterior probabilities for data generated under the PSC model are hardly affected by not taking  $F_{IS}$  into account, while RFI model posterior probabilities of data generated under the RFI model are lowered, but still sufficiently high to allow correct model choice (Fig. S11, Supporting Information). Indeed, for  $k = 0.5$ , the difference caused by  $F_{IS}$  is small, with only about 1% less simulations correctly assigned when  $F_{IS}$  is removed, independent of the number of loci used. For a  $k$  value of 0.8, however, the pattern is strikingly different, while the addition of  $F_{IS}$  has little effect for data sets simulated under the PSC model, up to 35% of data sets previously assigned correctly to the RFI can no longer be assigned. Thus, the inclusion of  $F_{IS}$  considerably increases the power to correctly identify the RFI model.

### Suitability of the ABC approach

The ABC approach has been used here to distinguish between two alternative models and to estimate parameters under these models. A study of the properties of the inferred parameter estimates (Figs S1 and S2, Supporting Information) and their posteriors (Table S1 and Fig. S3, Supporting Information) shows that the parameters of these two models are reasonably estimated and that the coverage of the posteriors is quite accurate over

a wide range of the estimated parameter values. Inferred posterior distributions appear too narrow for some extreme values of the parameters when there is not much information in the data about these parameters (e.g. when a population size change is too recent to affect patterns of diversity, Fig. S3a, Supporting Information), and in that case may depend too much on the prior we used. We note that full-likelihood approaches (Beaumont 1999) may reveal more appropriate than our ABC approach to estimate parameter and their associated posterior distributions, but as mentioned earlier, the full evaluation of the properties of these methods and their integration into model choice procedure remain to be carried out. While we are indeed able to correctly discriminate between panmixia and an island model with our ABC approach (Figs 3 and 4), other and potentially more powerful approaches to compare models and estimate parameters might be developed. Analytical approaches to calculate parameter likelihoods under an island model with an instantaneous population size change (Wakeley 1999; Wakeley *et al.* 2001) could be extended to incorporate additional mutation models and be compared to results from a PSC model (Beaumont 1999) with AIC- or BIC-based tests (Schwarz 1978; Akaike 1981). Another way to compare models might be the use of a Bayesian model choice procedure based on reversible jump MCMC (Green 1995; Huelsenbeck *et al.* 2004; Foll & Gaggiotti 2008), which is more powerful than ABC when it can be implemented (for instance, see Appendix in Fagundes *et al.* 2007). However, we have shown here that the power of the present ABC approach was already quite high with sufficient number of markers (Fig. 3), and we did not find any evidence for a significant bias in our model choice results (Fig. 5). Moreover, ABC approaches are easy to customize to one's own need as demonstrated with the chimpanzee data set where we could add further complexity in the mutation model and implement a correction for missing data. An interesting advantage of ABC approaches is that they can be easily validated by using the same set of simulations to estimate parameters and estimate model posterior probabilities over the whole parameter space, which is unfortunately impossible for most full-likelihood methods taking days of computations for a single inference. ABC approaches seem therefore well suited to extend model choice procedures to more realistic situations and to customize them for species having particularly complex demography or peculiar life-history traits.

### Acknowledgements and software availability

We are grateful to two anonymous reviewers, François Roussel, and Mark Beaumont for their useful comments and sug-

gestions. A series of bash scripts and the executable files necessary to perform the simulations described in this paper can be found on <http://cmpg.unibe.ch/software/PSC-vs-RFI/>. The scripts should run directly under Linux/Mac OS, but require a bash emulator such as cygwin (<http://www.cygwin.com>) to run under Microsoft Windows. Note that the time required for the model choice procedure is not short but perfectly doable on a single computer; testing a data set with 20 loci and 100 000 simulations per model requires approximately 30 h of computation time on a Linux PC with a 3.2 GHz dual-core CPU. This work has been partly supported by a Swiss NSF Grant No. 3100A0-126074 to LE.

### References

- Akaike H (1981) Likelihood of a Model and Information Criteria. *Journal of Econometrics*, **16**, 3–14.
- Beaumont MA (1999) Detecting population expansion and decline using microsatellites. *Genetics*, **153**, 2013–2029.
- Beaumont M (2008) Joint determination of topology, divergence time and immigration in population trees. In: *Simulations, Genetics and Human Prehistory* (eds Matsumura S, Forster P, Renfrew C), pp. 135–154. McDonald Institute for Archaeological Research, Cambridge.
- Beaumont MA, Zhang WY, Balding DJ (2002) Approximate Bayesian computation in population genetics. *Genetics*, **162**, 2025–2035.
- Becquet C, Przeworski M (2007) A new approach to estimate parameters of speciation models with application to apes. *Genome Research*, **17**, 1505–1519.
- Becquet C, Patterson N, Stone AC, Przeworski M, Reich D (2007) Genetic structure of chimpanzee populations. *PLoS Genetics*, **3**, e66.
- Besbeas P, Freeman SN, Morgan BJT, Catchpole EA (2002) Integrating mark-recapture-recovery and census data to estimate animal abundance and demographic parameters. *Biometrics*, **58**, 540–547.
- Blum MGB, François O (2010) Non-linear regression models for Approximate Bayesian Computation. *Statistics and Computing*, **20**, 63–73.
- Caswell JL, Mallick S, Richter DJ *et al.* (2008) Analysis of chimpanzee history based on genome sequence alignments. *PLoS Genetics*, **4**, e10000057.
- Cornuet JM, Luikart G (1996) Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics*, **144**, 2001–2014.
- Cornuet JM, Santos F, Beaumont MA *et al.* (2008) Inferring population history with DIY ABC: a user-friendly approach to approximate Bayesian computation. *Bioinformatics*, **24**, 2713–2719.
- Drummond AJ, Rambaut A, Shapiro B, Pybus OG (2005) Bayesian coalescent inference of past population dynamics from molecular sequences. *Molecular Biology and Evolution*, **22**, 1185–1192.
- Ellegren H (2004) Microsatellites: Simple sequences with complex evolution. *Nature Reviews Genetics*, **5**, 435–445.
- Estoup A, Beaumont M, Sennedot F, Moritz C, Cornuet JM (2004) Genetic analysis of complex demographic scenarios: spatially expanding populations of the cane toad, *Bufo marinus*. *Evolution*, **58**, 2021–2036.

- Ewens WJ (1972) Sampling Theory of Selectively Neutral Alleles. *Theoretical Population Biology*, **3**, 87–112.
- Excoffier L (2004) Patterns of DNA sequence diversity and genetic structure after a range expansion: lessons from the infinite-island model. *Molecular Ecology*, **13**, 853–864.
- Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, **10**, 564–567.
- Excoffier L, Estoup A, Cornuet JM (2005a) Bayesian analysis of an admixture model with mutations and arbitrarily linked markers. *Genetics*, **169**, 1727–1738.
- Excoffier L, Laval G, Schneider S (2005b) Arlequin ver. 3.0: an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online*, **1**, 47–50.
- Fagundes NJR, Ray N, Beaumont M *et al.* (2007) Statistical evaluation of alternative models of human evolution. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 17614–17619.
- Foll M, Gaggiotti O (2008) A Genome-Scan Method to Identify Selected Loci Appropriate for Both Dominant and Codominant Markers: a Bayesian Perspective. *Genetics*, **180**, 977–993.
- Francois O, Blum MGB, Jakobsson M, Rosenberg NA (2008) Demographic history of European populations of *Arabidopsis thaliana*. *PLoS Genetics*, **4**, e10000075.
- Fu YX (1997) Coalescent theory for a partially selfing population. *Genetics*, **146**, 1489–1499.
- Garza JC, Williamson EG (2001) Detection of reduction in population size using data from microsatellite loci. *Molecular Ecology*, **10**, 305–318.
- Green PJ (1995) Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, **82**, 711–732.
- Hamilton G, Currat M, Ray N *et al.* (2005) Bayesian estimation of recent migration rates after a spatial expansion. *Genetics*, **170**, 409–417.
- Heled J, Drummond AJ (2008) Bayesian inference of population size history from multiple loci. *BMC Evolutionary Biology*, **8**, 289.
- Hey J, Nielsen R (2004) Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. *Genetics*, **167**, 747–760.
- Huelsenbeck JP, Larget B, Alfaro ME (2004) Bayesian phylogenetic model selection using reversible jump Markov chain Monte Carlo. *Molecular Biology and Evolution*, **21**, 1123–1133.
- Jackson JBC, Kirby MX, Berger WH *et al.* (2001) Historical overfishing and the recent collapse of coastal ecosystems. *Science*, **293**, 629–638.
- Jobin MJ, Mountain JL (2008) REJECTOR: software for population history inference from genetic data via a rejection algorithm. *Bioinformatics*, **24**, 2936–2937.
- Kingman JFC (1982) The Coalescent. *Stochastic Processes and Their Applications*, **13**, 235–248.
- Kuhner MK, Yamato J, Felsenstein J (1995) Estimating Effective Population-Size and Mutation-Rate from Sequence Data Using Metropolis-Hastings Sampling. *Genetics*, **140**, 1421–1430.
- 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.
- Leblois R, Estoup A, Streiff R (2006) Genetics of recent habitat contraction and reduction in population size: does isolation by distance matter? *Molecular Ecology*, **15**, 3601–3615.
- Lee PM (2004) *Bayesian Statistics—An Introduction*. Hodder Arnold, London.
- Leuenberger C, Wegmann D (2009) Bayesian Computation and Model Selectin without Likelihoods. *Genetics*, **184**, 243–252.
- Lindley DV (1965) *Introduction to Probability and Statistics from a Bayesian Viewpoint*. University Press, Cambridge.
- Myers S, Fefferman C, Patterson N (2008) Can one learn history from the allelic spectrum? *Theoretical Population Biology*, **73**, 342–348.
- Nielsen R, Beaumont MA (2009) Statistical inferences in phylogeography. *Molecular Ecology*, **18**, 1034–1047.
- Nielsen R, Wakeley J (2001) Distinguishing migration from isolation: a Markov chain Monte Carlo approach. *Genetics*, **158**, 885–896.
- Pritchard JK, Seielstad MT, Perez-Lezaun A, Feldman MW (1999) Population growth of human Y chromosomes: a study of Y chromosome microsatellites. *Molecular Biology and Evolution*, **16**, 1791–1798.
- Ptak SE, Przeworski M (2002) Evidence for population growth in humans is confounded by fine-scale population structure. *Trends in Genetics*, **18**, 559–563.
- Pybus OG, Rambaut A, Harvey PH (2000) An integrated framework for the inference of viral population history from reconstructed genealogies. *Genetics*, **155**, 1429–1437.
- Schwarz G (1978) Estimating Dimension of a Model. *Annals of Statistics*, **6**, 461–464.
- Sisson SA, Fan Y, Tanaka MM (2007) Sequential Monte Carlo without likelihoods. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 1760–1765.
- Städler 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.
- Storz JF, Beaumont MA (2002) Testing for genetic evidence of population expansion and contraction: an empirical analysis of microsatellite DNA variation using a hierarchical Bayesian model. *Evolution*, **56**, 154–166.
- Tavare S, Balding DJ, Griffiths RC, Donnelly P (1997) Inferring coalescence times from DNA sequence data. *Genetics*, **145**, 505–518.
- Thornton K, Andolfatto P (2006) Approximate Bayesian inference reveals evidence for a recent, severe bottleneck in a Netherlands population of *Drosophila melanogaster*. *Genetics*, **172**, 1607–1619.
- Voight BF, Adams AM, Frisse LA *et al.* (2005) Interrogating multiple aspects of variation in a full resequencing data set to infer human population size changes. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 18508–18513.
- Wahlund S (1928) Composition of populations and correlation appearances viewed in relation to the studies of inheritance. *Hereditas*, **11**, 65–106.
- Wakeley J (1998) Segregating sites in Wright's island model. *Theoretical Population Biology*, **53**, 166–174.
- Wakeley J (1999) Nonequilibrium migration in human history. *Genetics*, **153**, 1863–1871.

- Wakeley J (2008) *Coalescent Theory. An Introduction*. Ben Roberts, Greenwood Village, Colorado.
- Wakeley J, Nielsen R, Ardlie K, Liu-Cordero SN, Lander ES (2001) The discovery of single nucleotide polymorphisms and inferences of human demographic history. *American Journal of Human Genetics*, **67**, 210.
- Wegmann D, Excoffier L (2010) Bayesian Inference of the Demographic History of Chimpanzees. *Molecular Biology and Evolution*, **27**, 1425–1435.
- Wegmann D, Leuenberger C, Neuenschwander S, Excoffier L (2009) ABCtoolbox: a versatile toolkit for Approximate Bayesian Computation. *BMC Bioinformatics*, **11**, 116.
- Won YJ, Hey J (2005) Divergence population genetics of chimpanzees. *Molecular Biology and Evolution*, **22**, 297–307.
- Wright S (1931) Evolution in Mendelian populations. *Genetics*, **16**, 0097–0159.
- Zellner A (1971) *An Introduction to Bayesian Inference in Econometrics*. Wiley, New York.

### Supporting information

Additional supporting information may be found in the online version of this article.

**Table S1** Overall coverage properties of credible intervals for model parameters

**Fig. S1** Accuracy of parameter estimations with ABC under the population size change (PSC) model.

**Fig. S2** Accuracy of parameter estimations with ABC under a recent finite island (RFI) model.

**Fig. S3** Coverage properties of HPD credible intervals over the parameter ranges.

**Fig. S4** Accuracy of model assignment using the model choice procedure of Pritchard *et al.* (1999).

**Fig. S5** Distribution of the summary statistics for the eastern chimpanzee population.

**Fig. S6** Distribution of the summary statistics for the western chimpanzee population.

**Fig. S7** Quality of model choice depending on parameter values of the RFI model.

**Fig. S8** Quality of model choice for different parameter values of the PSC model.

**Fig. S9** Model choice is largely insensitive about the number of demes and the size of the RFI model.

**Fig. S10** Outcome of the model choice procedure of Beaumont (2008), based on ten thousands, hundred thousand or 1 million simulations.

**Fig. S11** Effect of  $F_{IS}$  as summary statistic on model choice accuracy.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.