

ARTICLE

Different kinds of genetic markers permit inference of Paleolithic and Neolithic expansions in humans

Carla Aimé^{*,1,2} and Frédéric Austerlitz¹

Recent population genetic studies have provided valuable insights on the demographic history of our species. However, some issues such as the dating of the first demographic expansions in human populations remain puzzling. Indeed, although a few genetic studies argued that the first human expansions were concomitant with the Neolithic transition, many others found signals of expansion events starting during the Palaeolithic. Here we performed a simulation study to show that these contradictory findings may result from the differences in the genetic markers used, especially if two successive expansion events occurred. For a large majority of replicates for each scenario tested, microsatellite data allow only detecting the recent expansion event in that case, whereas sequence data allow only detecting the ancient expansion. Combined with previous real data analyses, our results bring support to the ideas that (i) a first human expansions started during the Palaeolithic period, (ii) a second expansion event occurred later, concomitantly with the Neolithic transition.

European Journal of Human Genetics (2017) 25, 360–365; doi:10.1038/ejhg.2016.191; published online 21 December 2016

INTRODUCTION

Demographic changes are known to leave footprints on allelic frequencies. Together with the increased availability of large genetic data sets, numerical methods stemming from Kingman's coalescent theory¹ allow inferring the past demography of populations from their present-day patterns of genetic diversity.^{2,3} Several types of genetic markers can be used, each category of markers having its own specificities. For instance, as they are uni-parentally transmitted, mitochondrial sequences are informative only about maternal lineages, whereas the Y-chromosome provides information about paternal lineages. In Aimé *et al.*,⁴ it has been shown that these two types of markers sometimes provide different inferences on the demographic history of human populations, especially as females generally exhibit higher effective population sizes and higher migration rates than males.^{5,6}

Conversely, autosomal markers provide synthetic information about both maternal and paternal lineages. When these loci are carefully selected in order to avoid pairwise linkage disequilibria, they also offer the possibility to consider each of them as an independent replicate. Among these autosomal markers, sequences or microsatellites data also exhibit contrasted properties. In particular, microsatellites have higher mutation rates than sequences.^{7,8} Moreover, microsatellites are subject to strong homoplasy, which can lead to false signals when using models where this phenomenon is not efficiently controlled for. One simulation-based study⁹ showed that when microsatellites are included into an Approximate Bayesian Computation analysis,¹⁰ they provide substantially better estimation for recent admixture events than sequence data. However, they did not investigate the question of inferring other kind of demographic events such as expansions. Moreover, an empirical study on a non-human species (the black sea porpoise) showed that integrating microsatellites in an ABC analysis allowed inferring very recent expansion events.¹¹ In the latter

study, the recent events occurred a few generations ago, so the question remains whether contrasting analyses based on either microsatellites and on sequences can help disentangle the question of the Neolithic *vs* Paleolithic expansion.

Furthermore, although DNA sequences used in previous studies were limited to some parts of the genome, next-generation data including whole-genome sequence data, which are now increasingly available, will certainly offer much greater power to make precise demographic inferences. However, up to now, their use is still limited by informatics issues. For instance, most coalescent-based applications allowing inferences of the demographic history of populations from contemporary genetic data are still limited in terms of amount of data that can be processed (eg, five individuals for MCMC¹²).

The timing of demographic expansions in humans is a long-standing question historically addressed by archeologists and paleoanthropologist. In particular, during the Neolithic period, the emergence of farming and animal domestication occurred in several parts of the world (Central Africa, Middle-East, Eastern Asia and Central America¹³), concomitantly with the sedentarization of most nomadic hunter-gatherer populations. This transition was one of the most important cultural and technological revolutions in our history, which affected many aspects of lifestyle (diet, technologies and social organization). According to most archaeologists and paleoanthropologists (eg, Bocquet-Appel¹⁴), the first major expansions in most Eurasian populations would have occurred as a result from this transition. Bocquet-Appel¹⁴ notably showed an increase in the number of enclosures and in the proportion of subadults in Eurasian burial sites during the Neolithic and this was interpreted as a proof of a natality increase. Conversely, in some other areas, archeological data have shown traces of demographic growth preceding this transition. For instance, in Africa, radiocarbon dating suggests that a demographic expansion started about 60 000 to 80 000 YBP.¹⁵

¹UMR 7206, EcoAnthropologie et Ethnobiologie, CNRS/MNHN/Université Paris Diderot, Paris, France; ²UMR 5554, Institut des Sciences de l'Évolution, CNRS—Université de Montpellier, Montpellier, France

*Correspondence: Dr C Aimé, UMR 5554, Institut des Sciences de l'Évolution, CNRS—Université de Montpellier, Place E. Bataillon—C.C.065, 34095 Montpellier cedex 05, France. Tel: +33 674906360; Fax: +33 140793231; E-mail: carla.aimé@univ-montp2.fr

Received 29 August 2016; revised 12 October 2016; accepted 22 November 2016; published online 21 December 2016

Recently, population genetics and coalescent-based methods have also been used to address major issues about the demographic history of human populations in multiple areas. For instance, although when and how early modern humans reached America for the first time was subject to debates (eg, Waters and Stafford¹⁶ and Goebel *et al.*¹⁷), demographic inferences using mitochondrial DNA (mtDNA) sequences indicated an initial differentiation from Asian populations ended with a moderate bottleneck in Beringia during the last glacial maximum (LGM), around 23 000 to 19 000 years ago.¹⁸ Then, toward the end of the LGM, a strong spatial and demographic population expansion started 18 000 and finished 15 000 years ago (thus well before the Neolithic transition).

Interestingly, some studies using mtDNA and/or autosomal DNA sequences inferred expansion events predating the Neolithic transition even in Eurasian populations. For instance, estimated expansion times range between 63 000 and 17 000 YBP using HVS-I data.¹⁹ In Aimé *et al.*,²⁰ using nuclear and HVS-I data, we inferred also expansion events that predated the Neolithic transition in farmer and herder populations from Africa and Eurasia. Conversely, we did not find any signal of past expansion in contemporary hunter-gatherer populations from these areas. We thus suggested that previous Paleolithic demographic expansions may have promoted the emergence of farming during the Neolithic period. For Africa, several authors also inferred Palaeolithic expansion events, with onsets ranging from 80 000 to 25 000 YBP.^{21–24} However, in a study on microsatellite data,²⁵ we did not find signals of Palaeolithic expansions, either in Eurasia or Africa: the observed signals of expansion events were indeed consistent with the Neolithic transition. Finally, in Aimé *et al.*,⁴ we also found signals of expansion events that were consistent with the Neolithic transition for Eurasia using Y-chromosome microsatellite data.

In Aimé *et al.*,^{4,25} we suggested that these contrasted findings may result from the specificities of each type of genetic markers and indicate two successive expansion events in the studied populations. Indeed, as explained above, the higher mutation rate of microsatellites as compared to DNA sequences may increase their sensitivity to recent events. In turn, if two successive expansions occurred in the studied population, signals of the more ancient event might be masked by more recent signals and thus be undetectable. Conversely, more ancient expansions signals may be detected using more slowly evolving markers such as sequence data. Using Simcoal version 2.1.2,²⁶ we simulated here both DNA sequences (mitochondrial or autosomal) and microsatellite (autosomal or from the Y chromosome) data sets under several scenarios involving either one or two successive expansion events, starting at different points in time, consistent with

either a Paleolithic or a Neolithic expansion. These data sets were similar to those used in Aimé *et al.*^{4,20,25} in terms of numbers of loci.

Then, we used the program Beast²⁷ to obtain a *posteriori* estimations of these dates and their highest probability density (HPD) interval (95% HPD), in order to compare the estimated values with the true values used to simulate the data. This allowed us to investigate (i) to which extent the expansion could be inferred for each kind of markers in the single-expansion scenarios and (ii) whether the older or the younger expansion events were detected for each kind of markers in the scenarios with two successive expansions.

MATERIALS AND METHODS

Data simulation

We generated a large amount of simulated population genetics samples using the coalescent-based program Simcoal version 2.1.2.²⁶ These simulations were performed under 10 different scenarios (Table 1 and Figure 1). For each scenario, a simulated population underwent either a single expansion event ('single expansion' scenarios) or two expansion events ('successive expansions' scenarios), with two possible expansion rates g (10^{-3} and 10^{-2} per generation) and three possible starting times t (200, 800 or 2000 generations ago) for the expansions. These starting times corresponded to 5000, 20 000 or 50 000 YBP, assuming a generation time of 25 years, as usually assumed in human population genetics studies.^{19,22,28} For all scenarios, samples of 100 individuals were simulated.

Four marker types were simulated for each scenario: autosomal sequences, mitochondrial sequences, autosomal microsatellites and Y-chromosome microsatellites. For the autosomal sequences, we simulated 20 unlinked diploid sequences of 1300 base pairs (bp) to be consistent with existing short neutral sequence marker sets, such as the one developed by Patin *et al.*²⁸ used in Aimé *et al.*²⁰ We assumed a mutation rate of 2.5×10^{-8} per generation and per site.²⁹ For the mitochondrial sequences, we simulated a haploid sequence of 400 bp, corresponding to the HVS-I region, assuming a mutation rate of 10^{-5} per generation per site.^{30,31} For the autosomal microsatellites, we simulated 20 unlinked diploid loci per individual, assuming a mutation rate of 10^{-4} per locus (ie, the lower bound of the uniform distribution that is generally used in the literature³²). Finally, for the Y-chromosome microsatellites, we simulated ten linked haploid loci per individual, assuming a mutation rate of 2.1×10^{-3} for each locus.³³ For all markers, we assumed a current effective population size of $N_e = 50\,000$ (or $2 N_e = 100\,000$ for autosomal markers). This value is consistent with estimated values for the current effective population size of post-Neolithic populations (eg, African farmer populations³²). One hundred replicates were performed per scenario and per marker type.

Data analysis

The simulated data were analysed using the parametric approach implemented in BEAST v1.8, following the same procedure as in Aimé *et al.*^{4,20,25} Four demographic models are implemented in BEAST: constant effective population size (N_0 ; constant model), population expansion with an increasing growth rate

Table 1 Description of each scenario

Scenario	Starting time t (in generations)	Growth rate g (per generation)
A1. Slow ancient expansion	2000	10^{-3}
A2. Rapid ancient expansion	2000	10^{-2}
B1. Slow intermediate expansion	800	10^{-3}
B2. Rapid intermediate expansion	800	10^{-2}
C1. Slow recent expansion	200	10^{-3}
C2. Rapid recent expansion	200	10^{-2}
D1. Slow ancient+recent rapid expansion	2000+200	$10^{-4}+10^{-2}$
D2. Rapid ancient+recent rapid expansion	2000+200	$10^{-3}+10^{-2}$
E1. Slow intermediate+recent rapid expansion	800+200	$10^{-4}+10^{-2}$
E2. Rapid intermediate+recent rapid expansion	800+200	$10^{-3}+10^{-2}$

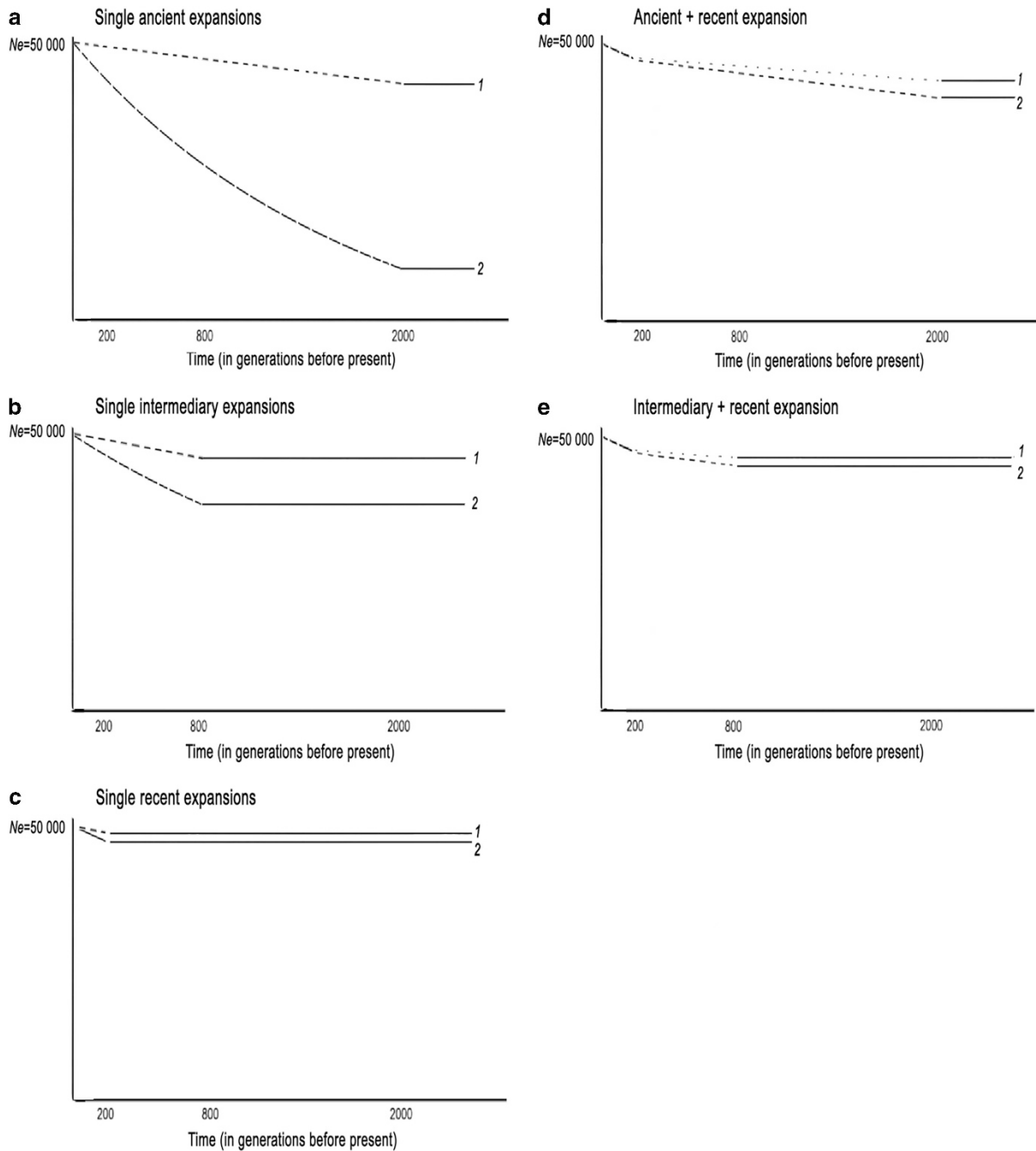


Figure 1 Description of each scenario. (a) Single ancient expansions (1: slow, 2: rapid), (b) single intermediary expansions (1: slow, 2: rapid), (c) single recent expansions (1: slow, 2: rapid), (d) ancient (slow) + recent (1: slow, 2: rapid) expansions and (e) intermediary (slow) + recent (1: slow, 2: rapid) expansions.

(g) (exponential model), population expansion with a decreasing growth rate (logistic model) and the expansion model, in which N_0 is the present day population size, N_1 the population size that the model asymptotes to going into the distant past and g the exponential growth rate that determines how fast the transition is from near the N_1 population size to N_0 population size. As in Aimé *et al.*,²⁰ we selected the best-fitting demographic model by estimating marginal likelihoods using two methods: path sampling and stepping-stone sampling.³⁴ The model with the larger marginal likelihood was the expansion model for all cases, consistently with the parameters used for the simulations. It was thus considered as the best-fitting model. To infer the current effective population size N_0 and the growth rate g from the composite parameters estimated with BEAST ($N_0 \mu$ and g/μ , where μ is the mutation rate), we used the μ values used

to simulate the data (see before). We then inferred the dates of expansion onsets (t) using the following formula: $t = (1/g) \times \ln(N_1/N_0)$, applied to each step of the MCMC algorithm.²⁵

RESULTS

Single expansion scenarios

We found that the three parameters (N_e , g and t) of these scenarios were always correctly estimated. Indeed, the true values of the three parameters were almost always included in the 95% HPD interval estimated on the simulated data sets (Table 2). For each parameter set and each marker type, this was indeed the case for at least 96 of the 100

Table 2 Number of replicates out of 100 for which the three parameters set in the simulations (N_e , g and t) were included in the 95% HPD interval of the estimations in the single expansion scenarios

Marker	A1	A2	B1	B2	C1	C2
Autosomal sequences	99	100	98	97	99	96
HVS-I sequences	97	100	98	99	98	99
Autosomal microsatellites	98	97	100	97	100	99
Y-chromosome microsatellites	97	99	97	100	99	100

Abbreviation: HPD, highest probability density.

replicates. The estimates of N_e and t were unbiased in most cases, as their mean estimated values over the 100 replicates were close to the true values (see Supplementary Tables 1), whereas the estimated growth rate showed some upper or lower biases depending upon the case. Moreover, consistently with the simulated conditions, we found significant signals of expansion in all cases (ie, for each scenario and each replicate), as the HPD intervals for growth rates never included the value of 0.

Successive expansion' scenarios

When performing the BEAST analysis on the simulated mtDNA or autosomal sequences, under the expansion model, only the more ancient expansion event was detected in a vast majority of cases (Figure 2 and Supplementary Table S2) for these scenarios. Indeed, in at least 87 replicates out of 100, the 95% HPD intervals of the estimates for the three parameters included only the true values of the parameters of the ancient expansion, but not those of the recent expansion event. In the majority of the other cases (between 2 and 13 replicates depending on scenarios and markers), the true values of the parameters of the ancient and the recent expansion were all included in the 95% HPD intervals. In the few remaining cases (0 to 6 replicates), neither the parameters of the ancient event nor those of the recent event were included in the 95% HPD interval. Finally, it never occurred that only the true values of the recent event were included in the 95% HPD intervals.

The pattern was strikingly opposite for the Y-chromosome and autosomal microsatellite markers, where only the recent expansion event was detected in the vast majority of cases. Indeed, for these markers, in at least 81 out of 100 replicates, only the true parameters of the most recent expansion event were included in the 95% HPD intervals. In most other cases, the HPD intervals of the estimations for the three parameters included those corresponding to both ancient and recent expansions (between 5 and 16 replicates). In a few cases (two to four replicates), neither the parameters of the ancient expansion event nor those of the recent one were included in the 95% HPD interval. Finally, it never occurred for these markers that the ancient expansion was included in the 95% HPD interval but not the recent one. Finally, when considering the mean estimated values and the mean 95% HPD interval over the 100 replicates, the mean modal estimate of the expansion onset time (t) was closer to the onset time of the ancient event and its 95% HPD interval included this onset time, but not the onset time of the more recent event, for the sequence data (autosomal or mitochondrial). Conversely, for the autosomal and Y-chromosome microsatellite data, the mean modal estimate of t was closer to the onset time of the recent event and the mean 95% HPD interval included this onset time, but not that of the more ancient one.

The mean estimated growth rate (g) was intermediate between the growth rates of the ancient and recent expansions in all cases, while its HPD interval included the values corresponding to the most recent but not those corresponding to the more ancient expansion.

DISCUSSION

First, we showed that using a moderate number of neutral markers, the MCMC method implemented in BEAST allowed us to efficiently detect signals of expansion events and provided reliable estimates of effective population sizes, growth rates and starting times in at least 96% of cases under quite simple scenarios involving a single expansion event. As these scenarios were very simple (single expansion, absence of admixture, etc.), this correct estimations of the parameters in these cases could be expected to some extent. However, we had to check that it was indeed the case with the sample sizes and number of markers assumed here, before interpreting our main results, which were obtained under conditions involving two successive expansions.

When we simulated these two successive expansion events, we found strikingly contrasting results for the different types of genetic markers. In particular, under several demographic scenarios involving two successive expansions, we detected only the most recent expansion event in at least 87 replicates out of 100 replicates when using Y-chromosome or autosomal microsatellite data. Conversely, only the oldest expansion event was detected in in at least 81 cases over 100 replicates using mtDNA or autosomal sequence data. It is noteworthy that, for the range of parameters tested in this study, these effects depended neither on growth rates nor on time intervals between the two expansions.

Considering together the results of this simulation study with those from our previous studies on real data in human populations^{4,20,25} provides rather interesting insights on the demographic history of our species. Indeed, using mitochondrial and autosomal DNA sequences, we detected expansion events predating the Neolithic transition in multiple African and Eurasian populations.²⁰ These results were consistent with previous genetic studies.^{19,21–24} Conversely, in Aimé *et al.*,²⁵ we detected signals of expansion events concomitant with the Neolithic transition in the same African and Eurasian populations using microsatellite data. In Aimé *et al.*,⁴ we also found expansion events concomitant with the Neolithic transition in Eurasia using Y-chromosome microsatellites. We suggested that these apparently contrasted results might be explained by two successive expansion events, one during the Palaeolithic and one during the Neolithic. The results from the present simulation study demonstrate clearly that it is a plausible scenario.

It is worth noting that the finding of a Paleolithic expansion event in Africa is consistent with some paleoanthropological data. Indeed radiocarbon dating suggested a demographic expansion in Africa 60 000–80 000 YBP.¹⁵ This Paleolithic demographic expansion could be linked to a rapid environmental change towards a dryer climate and/or to the emergence of new hunting technologies,^{15,35} which may have increased food availability. However, for Eurasia, the idea that demographic growth in most populations started during the Neolithic period with the emergence of farming and sedentarization is largely accepted among archaeologists and paleoanthropologists¹⁴ Nevertheless, paleoanthropological remains might be too scarce for detecting the Palaeolithic expansion. It is often hard, indeed, to evaluate population densities based on archeological remains only, especially in more ancient times as the archaeological records become more fragmented.

Our study also highlights the advantages of simultaneously analysing different types of genetic markers when inferring the past

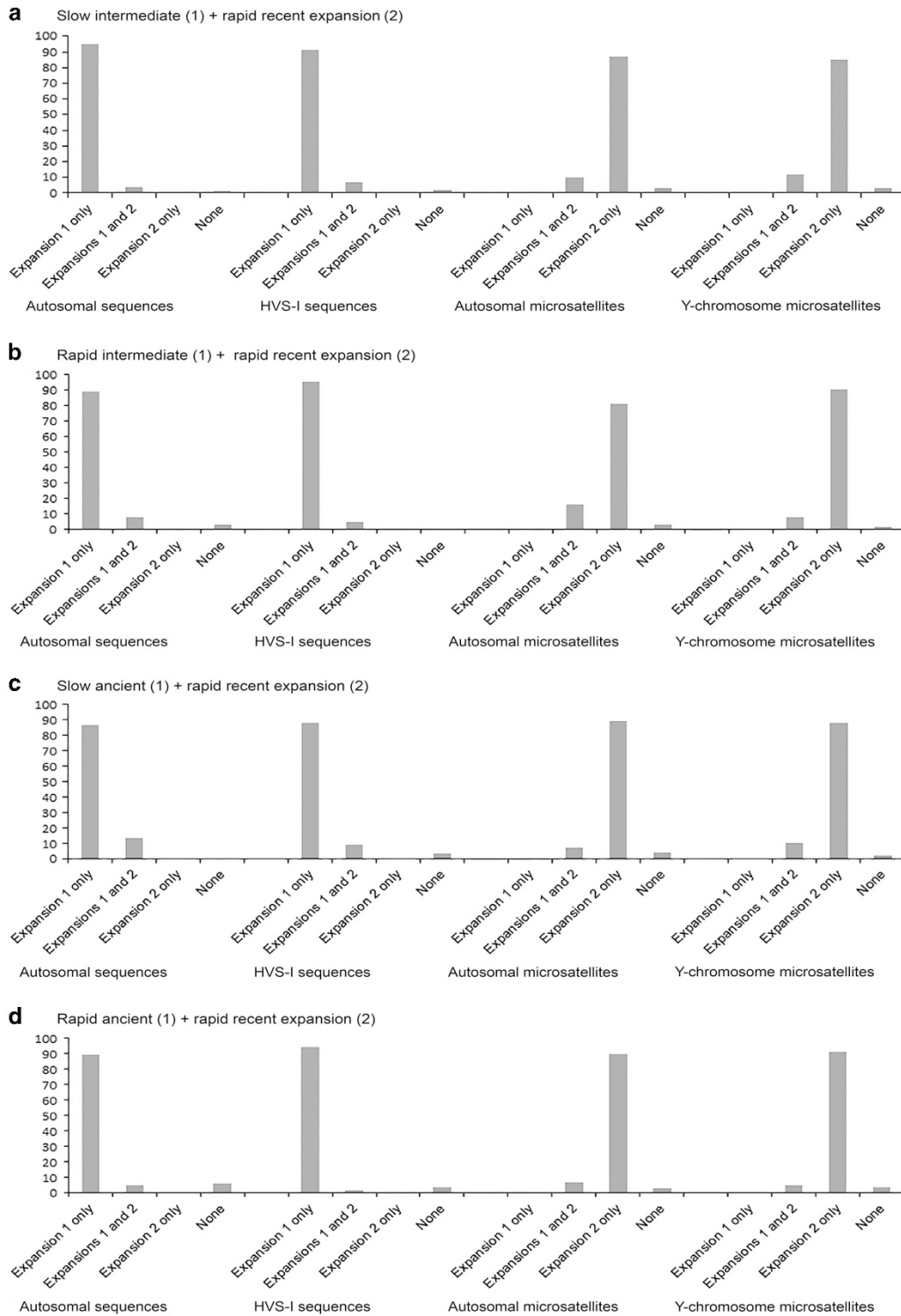


Figure 2 Distributions of the number of replicates over 100 for which the oldest, the most recent, both and none of the simulated expansion events were correctly estimated, for each marker and scenario. An expansion event was considered to be correctly estimated when the three parameters set in the simulations (N_e , g and t) were included in the HPD (95% highest probability density) interval of the estimations. (a) Slow intermediate + rapid recent expansion, (b) rapid intermediate + rapid recent expansion, (c) slow ancient + rapid recent expansion and (d) rapid ancient + rapid recent expansion.

demographic history of populations for any species. As for microsatellites, it is well known that mutation rates are variable among microsatellite markers.³⁶ As we chose to consider a mutation rate of 10^{-4} per generation per site, which is the lower bound of the generally

used uniform distribution,³² we certainly underestimated the difference in mutation rates between microsatellites and sequence data. This conservative assumption strengthens our conclusion that microsatellite markers and sequence data provide very different insights on the

demographic history of human populations. Moreover, as the aim of this study was to investigate the consequences of using one type of marker rather than another when inferring the starting dates of past human expansions, we chose here to use growth rates and starting times close to those inferred in empirical studies in humans.^{19–25} It would be interesting to investigate other kind of scenarios in future studies.

In this context, as we aimed here to perform a simulation study to better understand the results of our previous empirical studies on microsatellites and short DNA sequences, we simulated here such kind of markers in order to be consistent. It will be also interesting in future works to simulate much larger data sets such as whole genome data, which are now becoming increasingly common and might offer sufficient power to infer several successive expansions. In this context, it will be interesting to analyze the efficiency of methods like MSMC¹² or PopsizerABC,³⁷ which assume a model in which populations go through successive events of instantaneous changes in population size through time. It is noteworthy however that these methods cannot assume a parametric model with for example one or two expansions with an exponential growth rate and that such models will need to be developed to be able to detect successive expansions, in order to compare directly with the work performed here.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

This work was supported by the Agence Nationale de la Recherche grant Demochips (ANR-12-BSV7-0012). CA was financed by a PhD grant from the 'Centre National de la Recherche Scientifique'. We thank Paul Verdu, the editor and two anonymous reviewers for helpful comments and suggestions on the manuscript.

- 1 Kingman JFC: The coalescent. *Stochast Proc App* 1982; **13**: 235–248.
- 2 Excoffier L, Heckel G: Computer programs for population genetics data analysis: a survival guide. *Nat Rev Genet* 2006; **7**: 745–758.
- 3 Ho SYW, Shapiro B: Skyline-plot methods for estimating demographic history from nucleotide sequences. *Mol Ecol Res* 2011; **11**: 423–434.
- 4 Aimé C, Heyer E, Austerlitz F: Inference of sex-specific expansion patterns in human populations from Y-chromosome polymorphism. *Am J Phys Anthropol* 2015; **157**: 217–225.
- 5 Wilder JA, Mobasher Z, Hammer MF: Genetic evidence for unequal effective population sizes of human females and males. *Mol Biol Evol* 2004; **21**: 2047–2057.
- 6 Heyer E, Chaix R, Pavard S, Austerlitz F: Sex-specific demographic behaviours that shape human genomic variation. *Mol Ecol* 2012; **21**: 597–612.
- 7 Zhivotovskiy LA, Underhill PA, Cinnioglu C *et al*: The effective mutation rate at y chromosome short tandem repeats, with application to human population-divergence time. *Am J Hum Genet* 2004; **74**: 50–61.
- 8 Ségurel L, Wyman MJ, Przeworki M: Determinants of mutation rate variation in the human germline. *Annu Rev Genomics Hum Genet* 2014; **15**: 47–70.
- 9 Cornuet JM, Ravigne V, Estoup A: Inference on population history and model checking using DNA sequence and microsatellite data with the software diyabc (v1.0). *BMC Bioinformatics* 2010; **11**: 401.

- 10 Beaumont MA, Zhang W, Balding DJ: Approximate Bayesian computation in population genetics. *Genetics* 2002; **162**: 2025–2035.
- 11 Fontaine MC, Snirc A, Frantzis A *et al*: History of expansion and anthropogenic collapse in a top marine predator of the black sea estimated from genetic data. *Proc Natl Acad Sci USA* 2012; **109**: E2569–E2576.
- 12 Schiffels S, Durbin R: Inferring human population size and separation history from multiple genome sequences. *Nat Genet* 2014; **46**: 919–925.
- 13 Bocquet-Appel JP, Bar-Yosef O: *The Neolithic Demographic Transition and its Consequences*. Dordrecht, Netherlands: Springer, 2008.
- 14 Bocquet-Appel JP: When the world's population took off: The springboard of the Neolithic demographic transition. *Science* 2011; **333**: 560–561.
- 15 Mellars P: A new radiocarbon revolution and the dispersal of modern humans in Eurasia. *Nature* 2006; **439**: 931–935.
- 16 Waters MR, Stafford TW: Redefining the age of Clovis: implications for the peopling of the Americas. *Science* 2007; **315**: 1122–1126.
- 17 Goebel T, Waters MR, O'Rourke DH: The late Pleistocene dispersal of modern humans in the Americas. *Science* 2008; **319**: 1497–1502.
- 18 Fagundes NJ, Kanitz R, Eckert R *et al*: Mitochondrial population genomics supports a single pre-Clovis origin with a coastal route for the peopling of the Americas. *Am J Hum Genet* 2008; **82**: 583–592.
- 19 Chaix R, Austerlitz F, Hegay T, Quintana-Murci L, Heyer E: Genetic traces of east-to-west human expansion waves in Eurasia. *Am J Phys Anthropol* 2008; **136**: 309–317.
- 20 Aimé C, Laval G, Patin E *et al*: Human genetic data show contrasting demographic patterns between sedentary and nomadic populations predating the emergence of farming. *Mol Biol Evol* 2013; **30**: 2629–2644.
- 21 Cox MP, Morales DA, Woerner AE, Sozanski J, Wall JD, Hammer MF: Autosomal resequencing data reveal Late Stone Age signals of population expansion in sub-Saharan African foraging and farming populations. *PLoS One* 2009; **4**: e6366.
- 22 Laval G, Patin E, Barreiro LB, Quintana-Murci L: Formulating a historical and demographic model of recent human evolution based on resequencing data from noncoding regions. *PLoS One* 2010; **5**: e10284.
- 23 Batini C, Lopes J, Behar DM *et al*: Insights into the demographic history of African pygmies from complete mitochondrial genomes. *Mol Biol Evol* 2011; **28**: 1099–1110.
- 24 Soares P, Alshamali F, Pereira JB *et al*: The expansion of mtDNA haplogroup L3 within and out of Africa. *Mol Biol Evol* 2011; **29**: 915–927.
- 25 Aimé C, Verdu P, Ségurel L, Martinez-Cruz B, Heyer E, Austerlitz F 2014 Microsatellite data show recent demographic expansions in sedentary but not in nomadic human populations in Africa and Eurasia. *Eur J Hum Genet* 2014; **22**: 1201–1207.
- 26 Laval G, Excoffier L: Simcoal 2.0: A program to simulate genomic diversity over large recombining regions in a subdivided population with a complex history. *Bioinformatics* 2004; **20**: 2485–2487.
- 27 Drummond AJ, Rambaut A: Beast: Bayesian evolutionary analysis by sampling trees. *Bmc Evol Biol* 2007; **7**: 214.
- 28 Patin E, Laval G, Barreiro LB *et al*: Inferring the demographic history of African farmers and pygmy hunter-gatherers using a multilocus resequencing data set. *PLoS Genet* 2009; **5**: e1000448.
- 29 Pluzhnikov A, Di Rienzo A, Hudson R: Inferences about human demography based on multilocus analyses of noncoding sequences. *Genetics* 2002; **161**: 1209–1218.
- 30 Howell N, Kubacka I, Mackey DA: How rapidly does the human mitochondrial genome evolve? *Am J Hum Genet* 1996; **59**: 501–509.
- 31 Heyer E, Zietkiewicz E, Rochowski A, Yotov V, Puymirat J, Labuda D: Phylogenetic and familial estimates of mitochondrial substitution rates: study of control region mutations in deep-rooting pedigrees. *Am J Hum Genet* 2001; **63**: 1113–1126.
- 32 Verdu P, Austerlitz F, Estoup A *et al*: Origins and genetic diversity of Pygmy hunter-gatherers from western central Africa. *Curr Biol* 2009; **19**: 312–318.
- 33 Heyer E, Puymirat J, Dietjes P, Bakker E, De Knijff P: Estimating y chromosome specific microsatellite mutation frequencies using deep rooting pedigrees. *Hum Mol Genet* 1997; **6**: 799–803.
- 34 Baele G, Lemey P, Bedford T, Rambaut A, Suchard MA, Alekseyenko AV: Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. *Mol Biol Evol* 2012; **29**: 2157–2167.
- 35 Shea JJ: The impact of projectile weaponry on Late Pleistocene hominin evolution. In: Hublin JJ, Richards MP (eds): *Evolution of Hominin Diets*. Dordrecht: Springer, 2009.
- 36 Weber JL, Wong C: Mutation of human short tandem repeats. *Hum Mol Genet* 1993; **2**: 1123–1128.
- 37 Boitard S, Rodriguez W, Jay F, Mona S, Austerlitz F: Inferring population size history from large samples of genome-wide molecular data—an approximate Bayesian computation approach. *PLoS Genet* 2016; **12**: e1005877.

Supplementary Information accompanies this paper on European Journal of Human Genetics website (<http://www.nature.com/ejhg>)