



Recovering the geographic origin of early modern humans by realistic and spatially explicit simulations

Nicolas Ray, Mathias Currat, Pierre Berthier, et al.

Genome Res. 2005 15: 1161-1167

Access the most recent version at doi:[10.1101/gr.3708505](https://doi.org/10.1101/gr.3708505)

References This article cites 38 articles, 13 of which can be accessed free at:
<http://genome.cshlp.org/content/15/8/1161.full.html#ref-list-1>

License

Email Alerting Service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

To subscribe to *Genome Research* go to:
<https://genome.cshlp.org/subscriptions>

Cold Spring Harbor Laboratory Press

Methods

Recovering the geographic origin of early modern humans by realistic and spatially explicit simulations

Nicolas Ray,¹ Mathias Currat, Pierre Berthier, and Laurent Excoffier

Computational and Molecular Population Genetics Lab, Zoological Institute, University of Bern, 3012 Bern, Switzerland

Most genetic and archeological evidence argue in favor of a recent and unique origin of modern humans in sub-Saharan Africa, but no attempt has ever been made at quantifying the likelihood of this model, relative to alternative hypotheses of human evolution. In this paper, we investigate the possibility of using multilocus genetic data to correctly infer the geographic origin of humans, and to distinguish between a unique origin (UO) and a multiregional evolution (ME) model. We introduce here an approach based on realistic simulations of the genetic diversity expected after an expansion process of modern humans into the Old World from different possible areas and their comparison to observed data. We find that the geographic origin of the expansion can be correctly recovered provided that a large number of independent markers are used, and that precise information on past demography and potential places of origins is available. In that case, it is also possible to unambiguously distinguish between a unique origin and a multiregional model of human evolution. Application to a real human data set of 377 STR markers tested in 22 populations points toward a unique but surprising North African origin of modern humans. We show that this result could be due to ascertainment bias in favor of markers selected to be polymorphic in Europeans. A new estimation modeling this bias explicitly reveals that East Africa is the most likely place of origin for modern humans.

[Supplemental material is available online at www.genome.org.]

The problems of human origins are far from being settled, and no complete agreement has been reached on the process by which anatomically modern humans (*Homo sapiens sapiens*) appeared and occupied their present range. Current archeological, anthropological, and genetic evidence globally favor a complete replacement of previous representatives of the genus *Homo* by early modern humans having originated in Africa (the recent African origin [RAO] model) (see, e.g., Harpending and Rogers 2000; Kaessmann and Paabo 2002; Stringer 2002; Cavalli-Sforza and Feldman 2003; Currat and Excoffier 2004). This is mainly based on the observation of earliest forms of anatomically modern humans in Africa (White et al. 2003) and a very recent common ancestry of mitochondrial (Vigilant et al. 1991; Ingman et al. 2000) and Y-chromosome (Donnelly et al. 1996; Pritchard et al. 1999) genes for which the ancestral form is found in Africa. This view is also supported by the overall larger extent of genetic diversity (Excoffier 2002; Tishkoff and Williams 2002) and lower levels of linkage disequilibrium (e.g., Tishkoff et al. 1996; Gabriel et al. 2002) in Africa, as well as the persistence in Africa of most ancestral forms of nuclear genes (Takahata et al. 2001; Satta and Takahata 2004). This RAO model is challenged by the persistence of morphological characters from *Homo erectus* to modern humans on different continents (Wolpoff 1989; Wolpoff et al. 2000; but see Brauer et al. 2004), by the old ancestry of several nuclear genes (Hawks et al. 2000), or the inference of pre-*H. sapiens* range expansions from current patterns of molecular diversity (e.g., Templeton 2002).

While the RAO model or one of its extensions seems the most likely (Excoffier 2002), it has not been formally tested

against competing models. Alternative models could either be other locations for a recent and unique origin (UO) of modern humans, an incomplete replacement of *H. erectus* individuals by modern humans, or the multiregional evolution model (ME model) (see, e.g., Wolpoff 1989; Wolpoff et al. 2000). The latter model postulates that there was a gradual and simultaneous transition from *H. erectus* to modern forms on different continents, and that this synchronized process was possible because of continuous migrations between continents. While the exact parameters of this model have never been really defined, a very few attempts have been made to evaluate the relative likelihood of the UO and the ME models (Takahata et al. 2001; Satta and Takahata 2004). The observed prevalence of ancestral genes in Africa was found incompatible with the ME model, unless the African population size was much larger than that of the other continents (Takahata et al. 2001). More recently, it was also shown that a more explicit spatial model would better fit the data (Satta and Takahata 2004).

In the present study, we introduce a new approach to test between alternative models of human evolution. It consists of comparing the observed pattern of genetic diversity between populations to that expected under various models of human evolution obtained from spatially explicit simulations taking into account geographic and/or environmental constraints (Currat et al. 2004). We thus simulate a range expansion process under the UO model from arbitrary geographic origins or an ME model assuming various continental sizes, environmental heterogeneity, and different rates of exchange between continents, and see whether the patterns of genetic diversity allow one to distinguish between UO and ME models. Simulated data are first used to validate our approach, and we then apply our testing procedure to a data set of 377 short tandem repeat (STR) loci (Rosenberg et al. 2002).

¹Corresponding author.

E-mail nicolas.ray@zoo.unibe.ch; fax 41 31 631 48 88.

Article and publication are at <http://www.genome.org/cgi/doi/10.1101/gr.3708505>.

Methods

Population samples

The simulations performed in this paper were designed to match a previous study of 377 STR markers analyzed in 52 human worldwide populations (Rosenberg et al. 2002). From this data set we considered only populations with a sample size of more than 20 individuals in order to avoid potential sample errors for smaller sample size. The 22 population samples meeting this criterion will be referred to hereafter as the “Rosenberg22 data set,” and their locations and sizes are reported in Figure 1.

Demographic and environmental simulations

The land surface of a Hammer-Aitoff projection of the Old World was divided into a grid of 9226 subpopulations (or demes), each occupying an area of $100 \times 100 \text{ km}^2$, as shown in Figure 1. Using the software SPLATCHE (Currat et al. 2004), we performed simulations of a range expansion from 25 different geographic origins evenly distributed every 2000 km on the surface of the Old World, as shown in Figure 1. These simulations are based on a two-step process. Environmental information is first used to simulate, forward in time, a demographic and spatial expansion from an initial population. We record, for each generation, the number of individual genes present in a deme, as well as the number of immigrant genes coming from each of the four nearest-neighboring demes on the grid. This demographic and migration history is stored in a database, which is then used to generate, backward in time, the genealogy and the diversity of genes sampled at given locations. The details of the demographic

implementation can be found in Currat et al. (2004), but we describe below the specificities of the present simulations.

Unless specified otherwise, the origin of each expansion consisted in a single deme assumed to have an initial size of 50 diploid individuals (100 nuclear genes). The onset of the expansion was set to 4000 generations, which corresponds to 120,000 yr ago assuming a generation time of 30 yr for modern humans (Tremblay and Vezina 2000). Note that this latter estimate is larger than the commonly used values of 20–25 yr, but generation times larger than 28 yr have now been reported in a series of modern societies (Gage 1998; Tremblay and Vezina 2000; Helgason et al. 2003), and generation times of 27 yr have also been estimated for prehistoric societies (Gage 1998), which is only slightly more than the estimates available for chimpanzees (23 yr) (Gage 1998). Each generation, the occupied demes are subject to a growth phase followed by an emigration phase. The growth phase is logistic with a constant growth rate of $r = 0.3$ (Cavalli-Sforza et al. 1994; Steele et al. 1998), and a carrying capacity (K) that depends on the environment in which the deme is located (see below). The emigration phase consisted in sending a total of $0.05N_t$ emigrants distributed among the four nearest-neighboring demes, where N_t is the size of the deme at time t . The exact number of emigrants sent to each neighboring deme i (E_i) was controlled through friction values (F_i) assigned to each neighboring deme. Friction here expresses the relative difficulty of moving through a deme, and is kept within a range of 0.1 (lowest friction, easiest migration) to 1 (highest friction, most difficult migration). E_i is computed from a multinomial distribution, with directional probabilities P_i proportional to the relative frictions F_i of four neighboring demes obtained as

$$P_i = \left(F_i \cdot \sum_{j=1}^4 \frac{1}{F_j} \right)^{-1}.$$

This formula implies that the number of emigrants sent to any deme is inversely proportional to its relative friction. A sea was considered here as a complete barrier to migration.

Environmental heterogeneity

We considered environmental heterogeneity by relating the type of vegetation associated to each deme to carrying capacity and friction values (see Supplemental Table 1). A digital map of present potential vegetation (around 4000 yr ago) (Ray 2003) was used to that end. The relative carrying capacity and friction maps are shown in Figure 1, C and D.

Genetic simulations

Each genetic simulation consisted in generating the molecular diversity at a given number of STR loci for each of the 22 population samples shown in Figure 1. A spatially explicit coalescent approach was used to simulate the gene genealogies of the 22 samples and their molecular diversity at STR loci with a

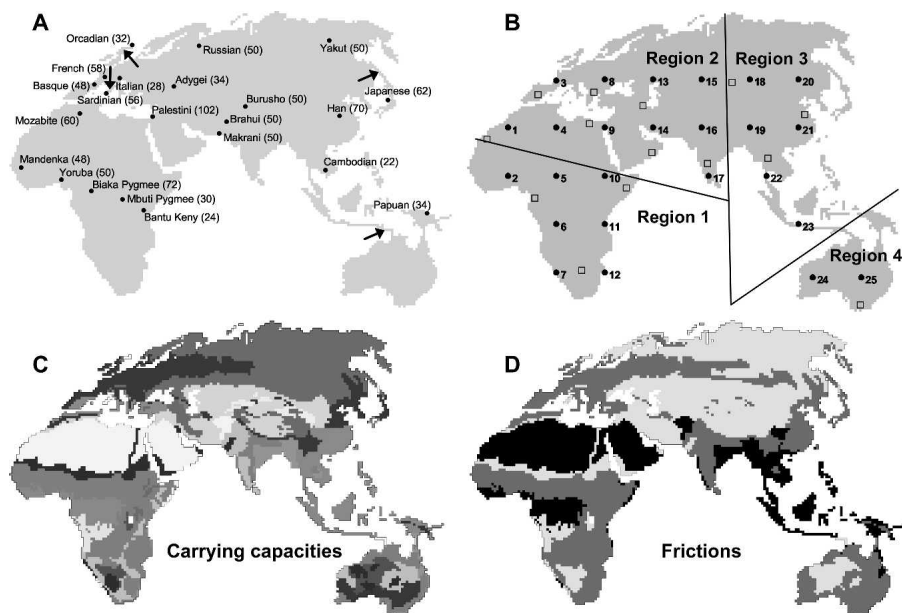


Figure 1. (A) Location, name, and size (within parentheses) of the 22 populations drawn from Rosenberg et al. (2002) for which genetic diversity was simulated. Arrows indicate the location of the four land bridges introduced to allow the complete colonization of the Old World and Oceania, since four samples (Orcadian, Sardinian, Japanese, and Papuan) are localized in islands. We imposed a low carrying capacity ($K = 10$) for demes located on these bridges to simulate a likely Founder Effect associated with the settlement of these islands. (B) Location of the 25 simulated origins (black dots). Open squares indicate the location of the 14 pseudo-observed origins. The four genetic regions defined in Rosenberg et al. (2002) are defined by the straight black lines. (C) Representation of relative carrying capacity values. (D) Representation of relative friction values. In C and D, darker colors indicate relatively higher carrying capacity or friction values.

modified version of the SPLATCHE program (Currat et al. 2004). For each of the 25 geographic origins and each evolutionary scenario described below, we performed 10,000 coalescent simulations of the genetic diversity of the 22 samples at a single STR locus under a strict stepwise mutation model. Multilocus data at L independent STR loci ($L = 1, 20, \text{ or } 377$, as reported below) were obtained by sampling (with replacement) L of the 10,000 loci. This bootstrap procedure was repeated 10,000 times to obtain 10,000 data sets of L loci. For each data set, the index of population differentiation R_{ST} (Slatkin 1995) was then computed between all pairs of populations, assuming a strict stepwise mutation model (Michalakis and Excoffier 1996), and using the software ARLEQUIN ver. 2.0 (Schneider et al. 2000).

Tested evolutionary scenarios

The UO and ME scenarios that we have simulated are shown and described in Figure 2. Note that the total time depth of both evolutionary scenarios was kept identical and set to 30,000 generations or about 900,000 yr. Under the ME model, we distinguished nine possible combinations of population sizes and migration rates between continental groups, as listed in Table 1, to represent different scenarios of the ME model. In the first three scenarios (listed as nos. 26, 27, and 28 in Table 1), continental groups have a constant size of $2N = 7000$, but exchange migrants at various rates. The next three scenarios (listed as nos. 29, 30, and 31 in Table 1), the African population was assumed to have a much larger size ($2N = 20,000$) than the two other populations ($2N = 1000$). Migration rates out of continental groups (m) were adjusted to keep the absolute number of emigrants (Nm) identical for each continent. These three scenarios are similar to those reported in Takahata et al. (2001). Finally, scenarios 32, 33, and

34 listed in Table 1 consider identical emigration probabilities from all continents, but owing to a large size of the African population, Africa is sending 10 times more migrants to the other continents than it receives from these regions.

Identification of a given geographical origin

If we assume that a given genetic data set is the product of a particular evolutionary scenario, one would ideally like to estimate the likelihood of all possible scenarios that can possibly generate the data, and choose the one maximizing this likelihood. Although this approach is not really feasible because of the large number of potential scenarios and the difficulty in computing their likelihood, an approximation can be envisioned as follows: (1) Select a restricted number of scenarios with unique or multiple origin. (2) For each of them, replace the likelihood by a measure of goodness of fit of the genetic data to the model. (3) Finally, select the evolutionary scenario that maximizes the goodness of fit as a point estimate.

In this study, we use simulations to determine the goodness of fit between an observed data set and a model as follows: The Pearson correlation coefficient is first computed between a matrix of pairwise indices R_{ST} calculated from observed data and a matrix of pairwise R_{ST} s calculated from data simulated under a given model. By repeating this procedure for many simulations per model, one obtains an empirical distribution of the correlation coefficient under a given model. Based on previous experience resulting from extensive simulations (Ray 2003), the 90% quantile value of the distribution was taken as our goodness-of-fit index, and will be referred to as the R90 statistic in the following.

Results

Finding the geographic origin of an expansion under the single origin model

Simulations were used to evaluate the possibility to correctly recover the geographic origin of an expansion from multiloci STR data, assuming that past demographic information was available. The 10,000 simulations performed for each scenario were divided into two sets of 5000 simulations. The first 5000 simulations performed under each scenario were used as pseudo-observed data and were used for comparison to the second set of 5000 simulations generated under all scenarios. The geographic origin of the expansion was assigned to the scenario showing the largest R90 statistic. The frequency of the pseudo-observed simulations (over a total of 125,000) that were correctly assigned to the true origin of the range expansion was equal to 0.149 (standard deviation over the 25 origins, $SD_{25} = 0.136$) for one locus. Although this frequency is very low, it is still three to four times higher than a completely random assignment over 25 putative origins. It suggests that some information on the geographic origin of an expansion can be extracted from a single locus, despite the high stochasticity of the coalescent process. When using 20 loci, this frequency increases substantially to 0.657 ($SD_{25} = 0.230$), and reaches 0.989 ($SD_{25} = 0.035$) with 377 loci. In Figure 3, we report the assignment success rate by geographic location for the three numbers of loci. It appears that there is a large heterogeneity in correct assignment across the different geographic origins when a few loci are available. For instance, with a single locus, the origins numbered 3, 7, 17, 20, and 23 have a much larger probability to be correctly assigned than the other origins, which may

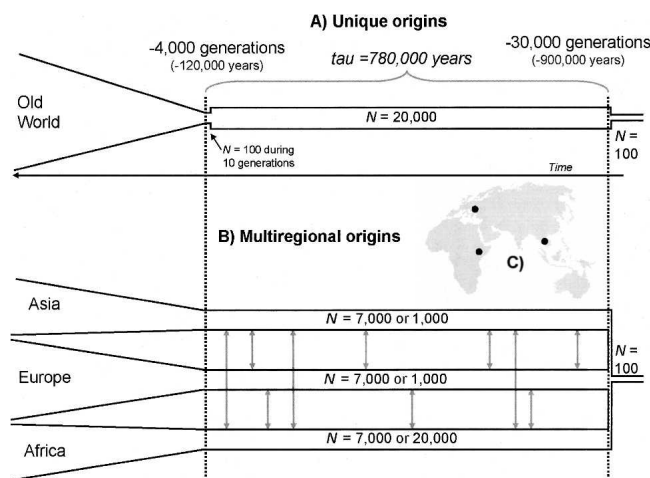


Figure 2. Demography and timeline of the different simulated evolutionary models. (A) Unique origin (UO) model. In this model, 30,000 generations ago, a small population ($N = 100$ genes) went through a demographic expansion after a first speciation event. Then, 4000 generations ago, a range expansion followed a bottleneck of 10 generations to mimic a second speciation event. The large population preceding the speciation and range expansion can be considered to be a large subdivided population. (B) Multiregional evolution (ME) model. As in A, a small population went through a speciation event and instantaneously colonized the three continents 30,000 generations ago. For 26,000 generations the continents harbored relatively large populations and exchanged occasional migrants (see Table 1 for continent population sizes and migration rates under different scenarios). Then, 4000 generations ago, three range expansions were initiated from the three different origins shown in C.

Table 1. Description of the nine ME scenarios (nos. 26–34)

Scenario no.		N			$m (\times 10^4)$	
		Africa	Europe	Asia	Africa	Europe/Asia
26	$Nm = 0.1$ (all regions)	7000	7000	7000	0.143	0.143
27	$Nm = 1$ (all regions)	7000	7000	7000	1.43	1.43
28	$Nm = 10$ (all regions)	7000	7000	7000	14.3	14.3
29	$Nm = 0.1$ (all regions)	20,000	1000	1000	0.05	1
30	$Nm = 1$ (all regions)	20,000	1000	1000	0.5	10
31	$Nm = 10$ (all regions)	20,000	1000	1000	5	100
32	$Nm = 0.1$ (in Africa)	20,000	1000	1000	0.05	0.05
33	$Nm = 1$ (in Africa)	20,000	1000	1000	0.5	0.5
34	$Nm = 10$ (in Africa)	20,000	1000	1000	5	5

N is the effective number of genes in a continent.

m is the probability of emigration per gene per generation.

Nm is the number of emigrant genes per generation per continent.

be related to their position on continental edges. With 20 loci, these results are more contrasted. The origins with lowest assignment success are mainly inland origins, with the exceptions of Australian origins, which are often mistakenly assigned to South-east Asia.

Distinguishing between the unique origin and the multiregional evolution models

The same procedure aiming at finding the geographic origin of a range expansion can be used to distinguish between data sets generated under a UO or under an ME model. In this context, a data set is correctly assigned if the scenario chosen on the basis of the R90 statistic belongs to the same evolutionary model as that used to generate it, regardless of the location of the origin or the type of ME scenario. In Table 2, we show the assignment scores of data sets simulated under the 25 UO scenarios and those simulated under the nine ME scenarios. It clearly appears that the evolutionary models are extremely well discriminated. With a single locus, correct assignment is 79.0%, but there is a sharp difference between the correct assignment of data sets generated under a single or multiple origins, with a much lower recovery rate (38.8%) of the ME model compared to 93.5% correct assignment to the UO model. With 20 loci, correct evolutionary model assignment is around 99% and reaches 100% with 377 loci.

Dealing with unknown origins

We have assumed so far that the potential geographic origins were known without error, but it is unlikely to be the case in practice. We investigate here the consequence of assuming an incorrect geographical origin of the expansion on the probability of recovering the source of a spatial expansion. The locations of 14 alternative and assumed “true” positions for the origins of an expansion are reported in Figure 1B as empty squares. A series of 10,000 simulations was performed from each of these origins. The 25 potential origins defined in Figure 1B were used as simulated origins, and the resulting genetic data sets were compared to those generated from “true” origins. Since the true and simu-

lated origins differ, as should be the case in reality, we measured the probability to recover the correct geographic region of origin. We therefore partitioned the Old World into four regions according to the results of Rosenberg et al. (2002), and as reported in Figure 1B. Note that this spatial partition in our study does not reflect continental discontinuities in human genetic variation, but is rather a practical mean of defining areas of putative origins. The frequencies of correct assignment per region between simulated and “true” origins were equal to 0.771 with 20 loci, and 0.852 with 377 loci. These frequencies are much lower than their

equivalents computed when the origins are known (0.882 for 20 loci and 0.999 for 377 loci).

Application to human nuclear STRs

In Figure 4A, we show the R90 goodness-of-fit statistics between the Rosenberg22 data set and data generated under the various evolutionary scenarios. The values of the statistics are reported in Supplemental Table 2. While scenarios assuming a UO overall better fit the data than multiregional scenarios, we find that multiregional scenario 26 (three continental groups of equal size and low Pleistocene migration rates between continents), and multiregional scenario 29 (much larger African effective size and low Pleistocene migrations) lead to better fits than Western European or Central Asian unique origins. However, origins 1 (northwestern Africa) and 9 (Near East) give the best fit, followed by North African origins. Note that the best values of the R90 statistics are here much lower than the mean values of these statistics computed from simulated scenarios (0.30 vs. 0.98).

Effect of ascertainment bias

The relatively unexpected results obtained from the analysis of the Rosenberg22 data sets pointing toward a North African origin of modern humans have led us to envision the possibility of an ascertainment bias in the choice of the 377 STRs. Even though it has been claimed that STRs were free of ascertainment bias because of their high level of polymorphism (Rogers and Jorde 1996; Harpending and Rogers 2000), the fact that these STRs have been primarily ascertained on a European CEPH panel (J. Weber, pers. comm.) makes this hypothesis worth checking. To

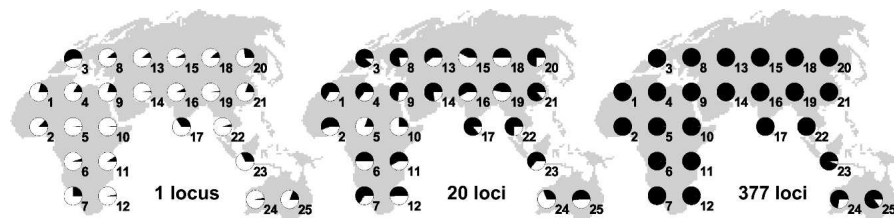


Figure 3. Relative frequencies of correct assignment for 25 simulated origins and for different numbers of available STR loci. The black area represents the proportion of simulations for which the origin of demographic expansion was correctly recovered from the R90 statistic. The frequencies of correct assignment for the complete set of simulations are given in Table 2.

Table 2. Frequencies of correct assignment to the single origin or multiregional evolution models

	Frequencies of correct assignment		
	Single origin	Multiregional	Total
1 locus	0.935 (0.438)	0.388 (0.420)	0.790 (0.433)
20 loci	0.995 (0.884)	0.992 (0.877)	0.994 (0.882)
377 loci	1.000 (0.991)	1.000 (0.991)	1.000 (0.991)

Model assignment was declared as correct if a data set generated under any scenario assuming a single origin was assigned to any geographic origin under the UO model, or if a data set generated under a ME scenario was assigned to any ME scenario. In all, 125,000 simulated data sets were generated under the UO model (over all possible geographic origins), and 45,000 data sets were generated under all considered ME scenarios. Values within parentheses are the mean values of the R90 statistic computed when the inferred origin is correctly assigned.

test whether our results would be affected by an ascertainment bias, we mimicked such a bias for a set of pseudo-observed simulations. We performed 20,000 simulations from an East African origin (no. 10 in Fig. 1). For each simulation, we computed the expected heterozygosity of the five European samples (French, Basque, Italian, Sardinian, and Orcadian), and we selected the 1000 simulations with the highest heterozygosity for this subset of samples to constitute a biased set of STR loci. We also randomly selected 1000 simulations to constitute an unbiased set of STR loci. Both types of loci were sampled with replacement to produce biased or unbiased data sets of 377 loci. Assignment scores for region no. 1 (Subsaharan Africa) for simulations from origin no. 1 were computed for the biased and the unbiased data sets. We find that the assignment to region no. 1 is correct in 100% of 1000 simulated data sets when using unbiased loci, but this figure drops to 81% with ascertainment bias, showing that the use of a biased set of loci can dramatically lower the probability to infer the correct region of origin of an expansion.

To check if it was possible to remove this bias from our analyses, we compared biased data sets to pseudo-observed equally biased simulated data sets, and obtained in that case an improved assignment score for region no. 1 of 86%. We reanalyzed the Rosenberg22 data set using an artificially biased set of simulations and report the new results in Figure 4B and in Supplemental Table 2. This time, an East African origin for modern humans is clearly favored, and the R90 statistic is improved compared to the previous analysis (Fig. 4A), showing that a selection of STR loci with high diversity in European populations improves the fit between observed and simulated data.

Discussion

We present here a first attempt at finding the geographic origin of modern humans from patterns of current worldwide nuclear genetic diversity, taking explicitly into account the physical constraints to dispersion existing for humans. Our present simulation results show that patterns of genetic diversity after a range expansion in a heterogeneous environment depend on the geographical origin of the expansion, which is not the case in a homogeneous and continuous environment (Ray et al. 2003). It implies that extant patterns of genetic diversity can be used to recover the place of origin of modern humans, if one assumes that they speciated at a single and precise location. Assuming that the pattern of environmental heterogeneity is known to a

certain extent, our results show that this location can only be well recovered provided that a large number of markers are available, that these markers do not suffer from ascertainment bias or that ascertainment bias can be taken into account, and that the simulated origin is close to the true origin. Our analysis taking ascertainment bias into account points toward an East African origin for modern humans, in agreement with another recent study showing that gene diversity was decreasing linearly with distance from East Africa, and thus suggesting a progressive range expansion from this source (Prugnolle et al. 2005). It is also interesting to note that in addition to Europe and Asia as unlikely sources for modern humans, southern Africa receives surprisingly very little support as the cradle of mankind.

The simulations and comparisons of molecular diversity stemming from UO and ME models suggest that the two evolutionary scenarios could be clearly distinguished from each other with as few as 20 loci (Table 2), even when the details of each model were not perfectly defined (geographic origin for the UO model, or migration rates and continental effective sizes under the ME model). The application of our methodology to real data clearly globally favors a unique origin for modern humans rather than multiregional evolution, as the proportion of explained variance in population differentiation is four times higher for the best UO scenario ($R90^2 = 0.1$, origin 10) than for the best ME scenario ($R90^2 = 0.023$, origin 26). Among the ME scenarios, it is worth noting that the best fit to the data is found for scenarios 26 and 29, which corresponds to cases in which the Pleistocene migration rates between continents are very limited ($Nm = 0.1$, or one migrant gene exchanged every 10 generations). This very limited amount of gene flow between continents makes it difficult to understand how humans could have evolved simultaneously toward modernity, as advocated by the supporters of the multiregional evolution hypothesis (Hawks et al. 2000; Wolpoff et al. 2000). It is also interesting to note that the best supported ME scenario (no. 26) assumes equal continental sizes, whereas previous studies (Takahata et al. 2001; Satta and Takahata 2004)

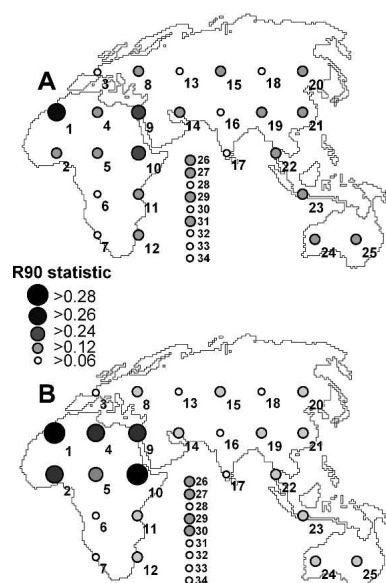


Figure 4. Values of the R90 statistic for 25 UO scenarios (nos. 1–25) and nine ME scenarios (nos. 26–34) computed from the Rosenberg22 data set under the unbiased (A) and biased (B) simulated data sets. Exact values of the R90 statistics are reported in Supplemental Table 2.

had found a better support for ME scenarios in which the African continent would have had a much larger population size than other continents. It supports the view that the pattern of genetic differentiation between human populations is more affected by the geography (migration corridors, contours) of the continents than by their population densities. Thus, our simulations suggest that the best alternative to a unique (likely Sub-Saharan African) origin would be a multiregional scenario assuming equal continental densities and very limited but equal gene flow between continents before the maximum Riss glaciation.

Compared to simulations (see Table 2), the fit of the different models to the observed pattern of genetic structure among populations as measured by the R90 statistic is not very large. This discrepancy suggests that despite the additional spatial realism of our simulation model, a simple model of population expansion from a given origin may still be too simplistic to fully explain the current pattern of human diversity (e.g., Harpending and Rogers 2000; Przeworski et al. 2000; Yu et al. 2001). Indeed, human genetic diversity has certainly been affected by factors that we could not incorporate into our simulation, such as historical events like population admixture, population contractions, and bottlenecks during the last glacial maximum (Harding and McVean 2004), a possible earlier appearance of modern humans (e.g., 160,000 yr ago) (White et al. 2003), subsequent range expansions like the spread of farmers from Neolithic cradles (e.g., Chikhi et al. 2002; Cordaux et al. 2004; Wen et al. 2004), but also long-distance migrations, which are believed to have an impact on current diversity (Templeton 2002), or interpopulation competition, which certainly delayed the colonization of Europe by Neanderthals (Currat and Excoffier 2004; Mellars 2004), or even differential selection on different continents (e.g., Harpending and Rogers 2000; Wall and Przeworski 2000; Hammer et al. 2004; Storz et al. 2004). The fact that the R90 statistic increases after a correction for ascertainment bias suggests that corrections for other potential sources of ascertainment or the use of more realistic mutation models could also contribute to improve the fit between observed and simulated data.

Our simulation scheme could theoretically be extended to incorporate more elaborate demographic and migration patterns, and one could increase the resolution of the grid such as to come arbitrarily close to the true origin of potential range expansions. However, the complexity of the scenarios we could explore is here mainly limited by computing cost, as well as by the availability of plausible estimates for the extra parameters that would need to be incorporated. A possible next step would consist of integrating our simulations into an Approximate Bayesian Computation (ABC) (Beaumont et al. 2002; Marjoram et al. 2003) framework, which would enable us to actually estimate parameters that were presently considered as fixed (such as the timing of the expansion, or demes' carrying capacities, or frictions in different environments) at the same time as estimating the geographic origin of the expansion.

Acknowledgments

We are grateful to Mark Stoneking for discussing this subject with us, to James Weber for giving us information on the criteria used to select the STR markers at the Marshfield center, and to two anonymous reviewers for their comments that helped to improve the manuscript. This work was supported by a Swiss NSF grant No. 3100A0-100800 to L.E.

References

- Beaumont, M.A., Zhang, W., and Balding, D.J. 2002. Approximate Bayesian computation in population genetics. *Genetics* **162**: 2025–2035.
- Brauer, G., Collard, M., and Stringer, C. 2004. On the reliability of recent tests of the Out of Africa hypothesis for modern human origins. *Anat. Rec.* **279A**: 701–707.
- Cavalli-Sforza, L.L. and Feldman, M.W. 2003. The application of molecular genetic approaches to the study of human evolution. *Nat. Genet.* **33**: 266–275.
- Cavalli-Sforza, L.L., Menozzi, P., and Piazza, A. 1994. *The history and geography of human genes*. Princeton University Press, Princeton, NJ.
- Chikhi, L., Nichols, R.A., Barbujani, G., and Beaumont, M.A. 2002. Y genetic data support the Neolithic demic diffusion model. *Proc. Natl. Acad. Sci.* **99**: 11008–11013.
- Cordaux, R., Deepa, E., Vishwanathan, H., and Stoneking, M. 2004. Genetic evidence for the demic diffusion of agriculture to India. *Science* **304**: 1125.
- Currat, M. and Excoffier, L. 2004. Modern humans did not admix with Neanderthals during their range expansion into Europe. *PLoS Biol.* **2**: 2264–2274.
- Currat, M., Ray, N., and Excoffier, L. 2004. SPLATCHE: A program to simulate genetic diversity taking into account environmental heterogeneity. *Mol. Ecol. Notes* **4**: 139–142.
- Donnelly, P., Tavaré, S., Balding, D.J., and Griffiths, R.C. 1996. Estimating the age of the common ancestor of men from the ZFY intron. *Science* **272**: 1357–1359.
- Excoffier, L. 2002. Human demographic history: Refining the recent African origin model. *Curr. Opin. Genet. Dev.* **12**: 675–682.
- Gabriel, S.B., Schaffner, S.F., Nguyen, H., Moore, J.M., Roy, J., Blumenstiel, B., Higgins, J., DeFelice, M., Lochner, A., Faggart, M., et al. 2002. The structure of haplotype blocks in the human genome. *Science* **296**: 2225–2229.
- Gage, T.B. 1998. The comparative demography of primates: With some comments on the evolution of life histories. *Annu. Rev. Anthropol.* **27**: 197–221.
- Hammer, M.F., Garrigan, D., Wood, E., Wilder, J.A., Mobasher, Z., Bigham, A., Krenz, J.G., and Nachman, M.W. 2004. Heterogeneous patterns of variation among multiple human X-linked loci: The possible role of diversity-reducing selection in non-Africans. *Genetics* **167**: 1841–1853.
- Harding, R.M. and McVean, G. 2004. A structured ancestral population for the evolution of modern humans. *Curr. Opin. Genet. Dev.* **14**: 667–674.
- Harpending, H. and Rogers, A. 2000. Genetic perspectives on human origins and differentiation. *Ann. Rev. Gen. Hum. Gen.* **1**: 361–385.
- Hawks, J., Hunley, K., Lee, S.H., and Wolpoff, M. 2000. Population bottlenecks and Pleistocene human evolution. *Mol. Biol. Evol.* **17**: 2–22.
- Helgason, A., Hrafnkelsson, B., Gulcher, J.R., Ward, R., and Stefansson, K. 2003. A populationwide coalescent analysis of Icelandic matrilineal and patrilineal genealogies: Evidence for a faster evolutionary rate of mtDNA lineages than Y chromosomes. *Am. J. Hum. Genet.* **72**: 1370–1388.
- Ingman, M., Kaessmann, H., Paabo, S., and Gyllensten, U. 2000. Mitochondrial genome variation and the origin of modern humans. *Nature* **408**: 708–713.
- Kaessmann, H. and Paabo, S. 2002. The genetical history of humans and the great apes. *J. Intern. Med.* **251**: 1–18.
- Marjoram, P., Molitor, J., Plagnol, V., and Tavaré, S. 2003. Markov chain Monte Carlo without likelihoods. *Proc. Natl. Acad. Sci.* **100**: 15324–15328.
- Mellars, P. 2004. Neanderthals and the modern human colonization of Europe. *Nature* **432**: 461–465.
- Michalakis, Y. and Excoffier, L. 1996. A generic estimation of population subdivision using distances between alleles with special reference for microsatellite loci. *Genetics* **142**: 1061–1064.
- Pritchard, J., Seielstad, M., Perez-Lezaun, A., and Feldman, M. 1999. Population growth of human Y chromosomes: A study of Y chromosome microsatellites. *Mol. Biol. Evol.* **16**: 1791–1798.
- Prugnolle, F., Manica, A., and Balloux, F. 2005. Geography predicts neutral genetic diversity of human populations. *Curr. Biol.* **15**: R159–R160.
- Przeworski, M., Hudson, R.R., and Di Rienzo, A. 2000. Adjusting the focus on human variation. *Trends Genet.* **16**: 296–302.
- Ray, N. 2003. “Modélisation de la démographie des populations humaines préhistoriques à l’aide de données environnementales et génétiques.” Ph.D. thesis, Université de Genève, Geneva, Switzerland.

- Ray, N., Currat, M., and Excoffier, L. 2003. Intra-deme molecular diversity in spatially expanding populations. *Mol. Biol. Evol.* **20**: 76–86.
- Rogers, A.R. and Jorde, L.B. 1996. Ascertainment bias in estimates of average heterozygosity. *Am. J. Hum. Genet.* **58**: 1033–1041.
- Rosenberg, N.A., Pritchard, J.K., Weber, J.L., Cann, H.M., Kidd, K.K., Zhivotovsky, L.A., and Feldman, M.W. 2002. Genetic structure of human populations. *Science* **298**: 2381–2385.
- Satta, Y. and Takahata, N. 2004. The distribution of the ancestral haplotype in finite stepping-stone models with population expansion. *Mol. Ecol.* **13**: 877–886.
- Schneider, S., Roessli, D., and Excoffier, L. 2000. *Arlequin: A software for population genetics data analysis*. User manual ver. 2.000. Genetics and Biometry Lab, Department of Anthropology, University of Geneva, Geneva, Switzerland.
- Slatkin, M. 1995. A measure of population subdivision based on microsatellite allele frequencies. *Genetics* **139**: 457–462.
- Steele, J., Adams, J.M., and Sluckin, T. 1998. Modeling Paleindian dispersals. *World Arch.* **30**: 286–305.
- Storz, J.F., Payseur, B.A., and Nachman, M.W. 2004. Genome scans of DNA variability in humans reveal evidence for selective sweeps outside of Africa. *Mol. Biol. Evol.* **21**: 1800–1811.
- Stringer, C. 2002. Modern human origins: Progress and prospects. *Phil. Trans. Roy. Soc. London (B)* **357**: 563–579.
- Takahata, N., Lee, S.H., and Satta, Y. 2001. Testing multiregionality of modern human origins. *Mol. Biol. Evol.* **18**: 172–183.
- Templeton, A. 2002. Out of Africa again and again. *Nature* **416**: 45–51.
- Tishkoff, S.A. and Williams, S.M. 2002. Genetic analysis of African populations: Human evolution and complex disease. *Nat. Rev. Genet.* **3**: 611–621.
- Tishkoff, S.A., Dietzsch, E., Speed, W., Pakstis, A.J., Kidd, J.R., Cheung, K., Bonne-Tamir, B., Santachiara-Benerecetti, A.S., Moral, P., and Krings, M. 1996. Global patterns of linkage disequilibrium at the CD4 locus and modern human origins. *Science* **271**: 1380–1387.
- Tremblay, M. and Vezina, H. 2000. New estimates of intergenerational time intervals for the calculation of age and origins of mutations. *Am. J. Hum. Genet.* **66**: 651–658.
- Vigilant, L., Stoneking, M., Harpending, H., Hawkes, K., and Wilson, A.C. 1991. African populations and the evolution of mitochondrial DNA. *Science* **253**: 1503–1507.
- Wall, J.D. and Przeworski, M. 2000. When did the human population size start increasing? *Genetics* **155**: 1865–1874.
- Wen, B., Li, H., Lu, D., Song, X., Zhang, F., He, Y., Li, F., Gao, Y., Mao, X., Zhang, L., et al. 2004. Genetic evidence supports demic diffusion of Han culture. *Nature* **431**: 302–305.
- White, T.D., Asfaw, B., DeGusta, D., Gilbert, H., Richards, G.D., Suwa, G., and Howell, F.C. 2003. Pleistocene *Homo sapiens* from Middle Awash, Ethiopia. *Nature* **423**: 742–747.
- Wolpoff, M.H. 1989. Multiregional evolution: The fossil alternative to Eden. In *The human revolution: Biological perspectives in the origins of modern humans* (eds. P. Mellars and C. Stringer), pp. 62–108. Princeton University Press, Princeton, NJ.
- Wolpoff, M.H., Hawks, J., and Caspari, R. 2000. Multiregional, not multiple origins. *Am. J. Phys. Anthropol.* **112**: 129–136.
- Yu, N., Zhao, Z., Fu, Y.X., Sambuughin, N., Ramsay, M., Jenkins, T., Leskinen, E., Patthy, L., Jorde, L.B., Kuromori, T., et al. 2001. Global patterns of human DNA sequence variation in a 10-kb region on chromosome 1. *Mol. Biol. Evol.* **18**: 214–222.

Received January 17, 2005; accepted in revised form May 18, 2005.