AMERICAN JOURNAL OF PHYSICAL ANTHROPOLOGY 157:217–225 (2015)

Inference of Sex-Specific Expansion Patterns in Human Populations From Y-Chromosome Polymorphism

Carla Aimé,* Evelyne Heyer, and Frédéric Austerlitz

Laboratoire d'Eco-Anthropologie et Ethnobiologie, UMR 7206 (Muséum National d'Histoire Naturelle—Centre National de la Recherche Scientifique—Université Paris 7 Diderot), Museum National d'Histoire Naturelle, F-75231 Paris, France

KEY WORDS population genetics; demography; microsatellites; beast; sex-specific processes

ABSTRACT Studying the current distribution of genetic diversity in humans has important implications for our understanding of the history of our species. We analyzed a set of linked STR and SNP loci from the paternally inherited Y chromosome to infer the past demography of 55 African and Eurasian populations, using both the parametric and nonparametric coalescent-based methods implemented in the BEAST application. We inferred expansion events in most sedentary farmer populations, while we found constant effective population sizes for both nomadic hunter-gatherers and seminomadic herders. Our results differed, on several aspects, from previous results on mtDNA and autosomal markers. First, we found more recent expansion patterns in Eurasia than in Africa. This discrepancy, substantially stronger than the ones found with the other kind of markers, may result from a lower effective population size for men, which might have made male-transmitted markers more sensitive to the out-of-Africa bottleneck. Second, we found expansion signals only for sedentary farmers but not for nomadic herders in Central Asia, while these signals were found for both kind of populations in this area when using mtDNA or autosomal markers. Expansion signals in this area may result from spatial expansion processes and may have been erased for the Y chromosome among the herders because of restricted male gene flow. Am J Phys Anthropol 157:217–225, 2015.

As demographic processes are known to leave noticeable footprints on the current distribution of genetic diversity (e.g. Cavalli-Sforza et al., 1994), reconstructing the demographic history of human populations from contemporary genetic data is a challenge for population geneticists. Together with recent developments in sequencing technologies, modern statistical and bioinformatics tools can substantially improve our knowledge of past demographical events (Beaumont, 2004). In this context, numerical coalescent-based methods (Kingman, 1982) have been developed, allowing the inference of demographic parameters from molecular data (Excoffier and Heckel, 2006; Kuhner, 2008).

Population genetic studies have provided a substantial contribution to the understanding of the demographic history of *Homo sapiens* (Cavalli-Sforza and Feldman, 2003; Pakendorf and Stoneking, 2005), showing in particular the African origin of modern humans (Cann et al., 1987; Quintana-Murci et al., 1999). Other studies have revealed that demographic expansions started in many African and Eurasian populations during the Paleolithic period (e.g. Chaix et al., 2008; Batini et al., 2011). In particular, in a previous work on HVS-I and autosomal sequences, we inferred strong Paleolithic expansion events for sedentary farmers in Africa and Eurasia, weak expansions for nomadic herders in Eurasia and no expansions for nomadic hunter-gatherers in Africa (Aimé et al., 2013). These differences appeared to predate the emergence of agriculture and the sedentarization processes, suggesting that strong expansion events in some Paleolithic populations may have ultimately favored the emergence of farming in these populations. Nevertheless, another study on autosomal microsatellites revealed also expansion events in sedentary but not in nomadic populations, but the dating of these events was more consistent with the Neolithic period (Aimé et al., in press). Altogether, these results suggest that two successive expansion stages may have occurred in farmer populations, the first one starting before the emergence of farming, and the second one triggered by the Neolithic transition. It also highlights the fact that different markers can be more sensitive to ancient or to recent events.

Our previous studies have focused on maternally-transmitted and autosomal markers (Aimé et al., 2013, in press). In this context, it is interesting also to investigate whether Y-chromosome based inferences display

Additional Supporting Information may be found in the online version of this article.

Grant sponsor: ATM-MNHN grant "Les relations Sociétés-Natures dans le long terme," by the ANR grant "Altérité culturelle;"; Grant number: ANR-10-ESVS-0010; Grant sponsor: ANR grant DemoChips; Grant number: 12-BSV7-0012; Grant sponsor: Centre National de la Recherche Scientifique (to C.A.).

*Correspondence to: Carla Aimé, Laboratoire d'Eco-Anthropologie et Ethnobiologie, UMR 7206 (Muséum National d'Histoire Naturelle—Centre National de la Recherche Scientifique—Université Paris 7 Diderot), Museum National d'Histoire Naturelle, 57 rue Cuvier F-75231, Paris, France.
E-mail: mme.carla.aime@gmail.com

Received 10 July 2014; accepted 13 January 2015

DOI: 10.1002/ajpa.22707
Published online 7 February 2015 in Wiley Online Library (wileyonlinelibrary.com).

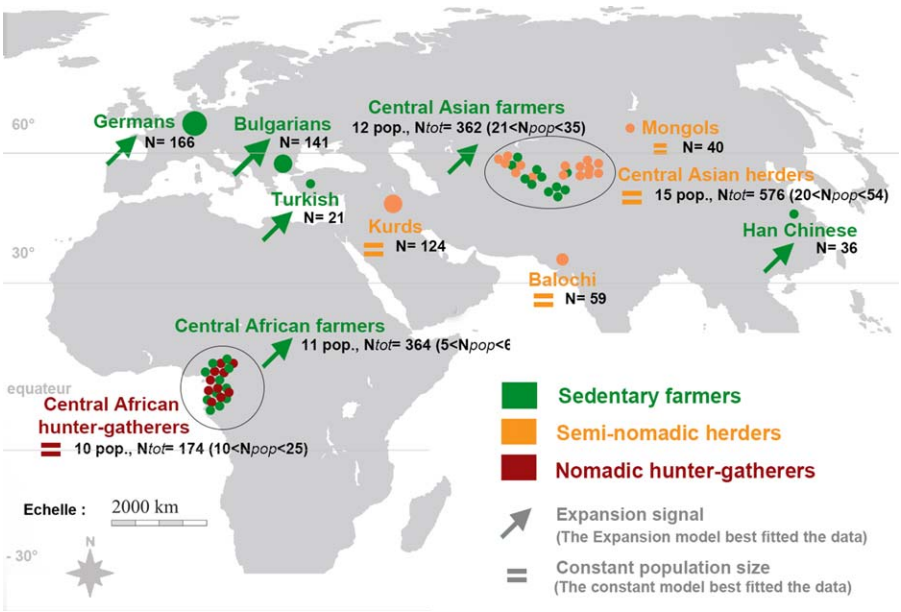

Fig. 1. Location of the analyzed populations, and sample sizes. Each population is indicated by a full circle, which size is proportional to the sample size of the corresponding population. These circles are color coded according to whether the populations are "Sedentary farmers," "Semi-nomadic herders," or "Nomadic hunter-gatherers.". [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

distinct patterns. Studies comparing mitochondrial, Y-chromosome, and autosomal markers have shown indeed that men and women differ both in their effective population sizes and effective migration rates (Heyer et al., 2012). Indeed, effective population size is generally lower for men than for women (Wilder et al., 2004), which may result from cultural factors (for instance, see Heyer et al., 2012, for a review and discussion about polygyny), even if some exceptions were observed, for instance in Pygmy populations (Verdu et al., 2013).

In this context, it is thus interesting to investigate whether the expansion patterns inferred from genetic data differ between male-transmitted markers and their female or biparental counterparts. In particular, we found generally more recent expansions events in Eurasia than in Africa. This observation may stem from the out-of-Africa process (Aimé et al., 2013, in press), which may have attenuated or suppressed signals of older expansions in contemporary Eurasia. Men having generally lower effective population sizes, it is thus clearly interesting to investigate whether male-transmitted genes were more affected by this process or not. Moreover, differences between inferences from paternally and maternally inherited markers may occur due to sex-specific migration rates. In particular, migration levels impact demographic inferences, as simulations studies have shown that spatial expansion signals are attenuated or suppressed in isolated populations (Ray and Excoffier, 2003; Excoffier, 2004). Indeed, in a spatially-expanding population, the signal of population growth can be erased if the gene flow among populations is too low (less than 20 migrants). The extent of sex-specific differences in migration patterns vary among populations. For instance, migration is much more female-biased in exogamous patrilocal populations than in endogamous cognatic populations (Chaix et al., 2007; Ségurel et al., 2008) These differences could lead to contrasted inferences on the demographic histories of these populations.

In the present study, we performed coalescent-based inferences from Y chromosome STR polymorphism data from 21 African populations, 27 Central Asian populations, and 7 populations from several other regions in Eurasia (Europe, Middle East, Pamir, and East-Asia). As STRs can be affected by homoplasy, when SNP data were available (i.e. for 22 Central Asian populations), we also included them in the analyses, along with the STR data, in order to limit the effects of this homoplasy on our inferences. We combined a parametric approach (i.e. model testing) to infer demographic parameters assuming several demographic models, and a nonparametric (i.e. model-free) approach (Bayesian Skyline Plots, BSP, Drummond et al., 2005) to visualize the evolution of the effective population sizes (N_e) through time. These analyses aimed at addressing the following questions: Do the inferences on Y-chromosome microsatellites reveal contrasted demographic patterns between farmer, herder and hunter-gatherer populations? Do these patterns differ from those observed with mtDNA or nuclear DNA markers? In particular, for the populations that showed an expansion pattern, does the timing inferred with the Y chromosome differ from the timing obtained with other kind of markers?

MATERIAL AND METHODS

Population sampling and marker sets

We studied a large set of African and Eurasian populations (see Fig. 1 for the geographic locations and sample sizes for all analyzed populations). For Africa, we analyzed published data (Verdu et al., 2013) for 364 male individuals from 11 sedentary farmer populations and 174 individuals from 10 nomadic hunter-gatherer populations (Supporting Information Table S1). We used

five tetranucleotide microsatellites (DYS389b, DYS391, DYS390, DYS393, DYS19) on the nonrecombining Y chromosome (NRY). Note that DYS389b is obtained by subtracting DYS389I from DYS389II.

For Eurasia, we combined published data from several studies to analyze four sedentary farmer populations and three nomadic herder populations from Europe, Middle-East, Pamir and East-Asia (Supporting Information Table S1). For Middle East and Pamir, we studied the Turk (farmers, $N = 124$) and the Kurd *(herders,* $N = 21$) samples from Quintana-Murci et al. (2004), and the Baloch (herders, $N = 59$) sample from Qamar et al. (2002). For East Asia, we studied the Han Chinese (farmers, $N = 36$) and the Mongol (herders, $N = 40$) populations (Kayser et al., 2001). In addition, we analyzed two European farmer populations: the Germans ($N = 166$, Henke et al., 2001) and the Bulgarians ($N = 141$, Zaharova et al., 2001). For these populations, we used six microsatellite loci (DYS19, DYS389b, DYS390, DYS391, DYS392, and DYS393). Here again, DYS389b have been obtained by subtracting DYS389I from DYS389II.

For our detailed study of Central Asia, we used a data set of 576 individuals from 15 seminomadic herder populations and 362 individuals from 12 sedentary farmer populations (Supporting Information Table S1). These data consisted mainly of previously published data (Chaix et al., 2004; Ségurel et al., 2008, 2013; Heyer et al., 2009). In addition, the data for one farmer population (LUZ) and one herder population (TKY) were genotyped for this study under the same conditions as in Chaix et al. (2004). For all Central Asian populations, we used nine NRY microsatellite loci (DYS19, DYS388, DYS389b, DYS390, DYS391, DYS392, DYS393, DYS426, and DYS439). Here again, DYS389b have been obtained by subtracting DYS389I from DYS389II. We also used a set of 18 Y-SNP polymorphisms for 12 Central Asian populations (Supporting Information Table S1; Balaresque et al., unpublished data) and a set of 38 Y-SNP polymorphisms for eight other Central Asian populations (Supporting Information Table S1; Heyer et al., unpublished data). Finally, two populations (UZA and UZT) were genotyped for this study for the same 38 SNPs as in Heyer et al. (unpublished data), following the same protocol.

As detailed in Supporting Information Table S1, most of these populations were also analyzed in our previous studies on HVS-1 sequences (Aimé et al., 2013) and autosomal microsatellite markers (Aimé et al., in press). This allowed us to compare, for each geographic area, the general trends observed using the different kinds of genetic markers.

Data analyses

Parametric approach. We computed the Bayesian posterior distribution of demographic parameters using the MCMC algorithm implemented in BEAST v1.7.5 (Drummond and Rambault, 2007), assuming for STR data a single step mutational model (SSM) which takes homoplasy into account, implemented in BEAST by Wu and Drummond (2011). When available (22 populations, Supporting Information Table S1), we also included simultaneously the data for the SNP polymorphisms, assuming a biallelic model. Indeed, even if all the performed analyses were designed to take homoplasy into account, using a limited number of Y-STR loci introduces the risk of hidden variation. As SNP data were available for the same individuals as for the STR data for these Central Asian populations, we decided to analyze both SNP and STR data simultaneously for these populations, in order to better differentiate between identical by state (IBS) and identical by descent (IBD) haplotypes. As all SNP and STR loci were on the Y chromosome, we assumed that they were all fully linked. We tested the four demographic models implemented in BEAST: constant effective population size ("Constant model") with a single parameter (population size N_0), expansion with a constant growth rate ("Exponential model") with two parameters (current population size N_0 and growth rate g), expansion with a decreasing growth rate ("Logistic model") with two parameters (current population size N_0 and growth rate g), and the "expansion model" with three parameters (N_0, N_1, g), 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.

The BEAST application estimates composite parameters for each model, namely $N_0\mu$ and g/μ, where μ is the mutation rate. In addition, for the "Expansion model," the ratio between the current (N_0) and ancestral (N_1) effective population sizes is also estimated. We inferred N_0 and g from these composite parameters, assuming for STR data a rate of 2.1×10^{-3} /generation/locus, found both in a pedigree-based study (Heyer et al., 1997) and in a study combining population and father-son pair data (Burgarella and Navascués, 2011). For the SNP data used for Central Asian populations, we used a wide uninformative uniform prior between 10^{-8} and 10^{-4}/generation/locus. Indeed, these loci were chosen as they are highly polymorphic. For the populations for which the "Expansion model" best-fitted the data, we inferred the dates of expansion onsets (t) using: $t = (1/g) \times \ln (N_1/N_0)$, applied to each step of the MCMC algorithm. This formula makes the approximate assumption that the population started to grow from its initial size N_1 to reach its final size N_0 after t generations.

We performed three runs of 6×10^7 steps per population and per demographic model for the African populations, and three runs of 1.2×10^8 steps for the other populations, allowing thus for least three runs of at least 10^7 steps per locus for each population. We recorded one tree every 1,000 steps, resulting in a total of 10^5 trees per locus and per run. We then removed the first 10% of each run (burn-in period) and combined the runs to obtain reasonably high effective sample sizes (ESS of 100 or above). The convergence of these runs was assessed by visual inspection of traces using Tracer v1.5 (Rambaut and Drummond, 2014) to check for concordance between runs, and also by the computation of the Gelman and Rubin's convergence diagnostic (Gelman and Rubin, 1992) using R v2.14.1 (R Development Core Team, 2011) with the function *gelman.diag* available in the package *coda* (Plummer et al., 2006). In order to facilitate a large exploration of the parameter space, we chose uniform priors between 1 and 10^5 for N_0 and N_1 and between -0.1 and 0.1 per generation for g. We assumed a generation time of 25 years, permitting the comparison with previous studies (Aimé et al., 2013, in press).

For each population and demographic model, we obtained the mode and the 95% Highest Probability

TABLE 1. Modes and 95% HPD of expansion onset times (t) inferred from the parametric method, and t inferred from the nonparametric method (BSP), compared with datations of the emergence of farming

Area	Analyzed populations	Life-style	Emergence of farming/herding	*t* 95% lower (yrs BP)	*t* mode (yrs BP)	*t* 95% upper (yrs BP)	*t* from BSPs (yrs BP)
Africa	Central-African farmers[a]	Sedentary farmers	5,000 YBP	28,321	86,235	417,878	73,391
East-Asia	Han Chinese	Sedentary farmers	9,000 YBP	1,446	18,093	42,822	8,600
Middle-East	Turkish	Sedentary farmers	11,000 YBP	7,404	11,598	17,796	7,975
Europe	Bulgarians	Sedentary farmers	9,000 YBP	5,006	8,442	15,248	6,800
Europe	Germans	Sedentary farmers	9,000 YBP	5,571	8,321	12,403	6,400
Central-Asia	Indo-Iranian farmers[a]	Sedentary farmers	9,000 YBP	4,432	8,277	29,682	9,560

[a] For these area, we report the mean values over populations, considering only populations for which we inferred a signal of expansion. Detailled results are presented for each population in Supporting Information Table S2 for the parametric method and Supporting Information Table S4 for the nonparametric method. We assumed a mutation rate of $\mu = 2.1 \times 10^{-3}$/generation/site (Heyer et al., 1997; Burgarella and Navascués, 2011) and a generation time of 25 years.

Density (HPD) interval of each parameter, inferred from the posterior distributions obtained using the R package Locfit (Loader, 1999). Then, we computed the Deviance Information Criteria (DIC) of each model, as the model with lower DIC was considered as the best-fitting model (Spiegelhalter et al., 2002). DIC was computed as in Aimé et al. (in press). A difference of five points in DIC was considered as significant (Spiegelhalter et al., 2002).

Non-parametric approach. Bayesian Skyline Plots (BSPs, Drummond et al., 2005), also implemented in BEAST, estimate demographic changes occurring continuously through time within a population, using the time intervals between successive coalescent events. As above, we combined three runs of 6×10^7 steps for African populations and three runs of 1.2×10^8 steps for Eurasian populations. We assumed also the same mutation rates as above, and a generation time of 25 years. Outputs were also analyzed with Tracer v1.5 (Rambaut and Drummond, 2014) to visually check for convergence and ESS, and Gelman and Rubin's (1992) convergence diagnostic was computed as above. Finally, we used the population growth curves generated from Tracer to assess the time at which populations began to expand. Each Skyline plot consisted of smoothed data points at ≈10–20 generation intervals. We considered that the population increased (or decreased) when both the median and HPD values for N_e increased (or decreased) between more than two successive data points. Although this method did not allow providing a confidence interval for inferred expansion timings, this conservative approach ensured that we considered only relevant expansion signals.

RESULTS

For Africa, the parametric method showed clear expansion events for all sedentary farmer populations but one. Indeed, the "Expansion model" best fitted the data (i.e. had the lowest DIC value) for all these sedentary populations (Supporting Information Table S2), with differences in DIC higher than five points for all these populations, except for the Ewondo. For this population, we could not distinguish between an "Expansion model" or a "Constant" model (Supporting Information Table S2), which may be due to a low sample size (10 individuals). Conversely, the "Constant model" or the "Exponential model" with negative modal values for growth rates best fitted the data for all nomadic hunter-gatherer populations, which showed thus either a constant population size or slight contraction events (Supporting Information Table S2). Nevertheless, as the HPD interval for the growth rate *g* included 0 for all these hunter-gatherer populations (Supporting Information Table S3), we considered that we did not infer any significant demographic changes for these populations.

The inferred growth rates for the sedentary farmer populations ranged between 3.99×10^{-4} [4.96×10^{-5} to 8.76×10^{-4}] and 4.11×10^{-3} [5.19×10^{-4} to 6.97×10^{-3}] per year (Supporting Information Table S3). As the "expansion model" best fitted the data, we could estimate the expansion onset times under this parametric model for all African farmer populations. It ranged between 38,327 [10,025–144,496] years before present (YBP) and 119,531 [38,522–487,667] YBP (Table 1 and Supporting Information Table S3). Using the nonparametric method, the BSP graphs (Fig. 2a) showed large HPD intervals including a potential constant population size for all populations. However, when considering the median values of N_e through time, five sedentary farmer populations (Gabonese Fang, Kota, Nzebi, Teke, and Tsogho) stood out from the others by showing an increase in N_e starting between 64,492 YBP and 79,867 YBP depending on the population (Supporting Information Table S4), followed by a stabilization of N_e or a slight contraction at about 25,000 YBP. Thus, with both parametric and nonparametric methods, all estimated dates clearly predated the emergence of farming and the Neolithic transition in Central Africa (Table 1). Conversely, for the nomadic populations, in agreement with the parametric method, we showed no expansion signal for these populations with the nonparametric method (Fig. 2b).

For Eurasia, the "Expansion model" best fitted the data and the differences in DIC between this model and the others were higher than five points (Supporting Information Table S2) for all sedentary farmer populations, indicating a clear expansion signal. Conversely, the "Constant model" best fitted the data for the three nomadic herder populations (Supporting Information Table S2). The inferred growth rates for sedentary farmer populations ranged between 6.28×10^{-4} [3.79×10^{-6} to 1.83×10^{-3}] and 1.65×10^{-3} [1.06×10^{-3} to 2.39×10^{-3}] per year (Supporting Information Table S3). The BSP graph showed stronger expansions (i.e. higher N_e and growth rates) for sedentary farmers than

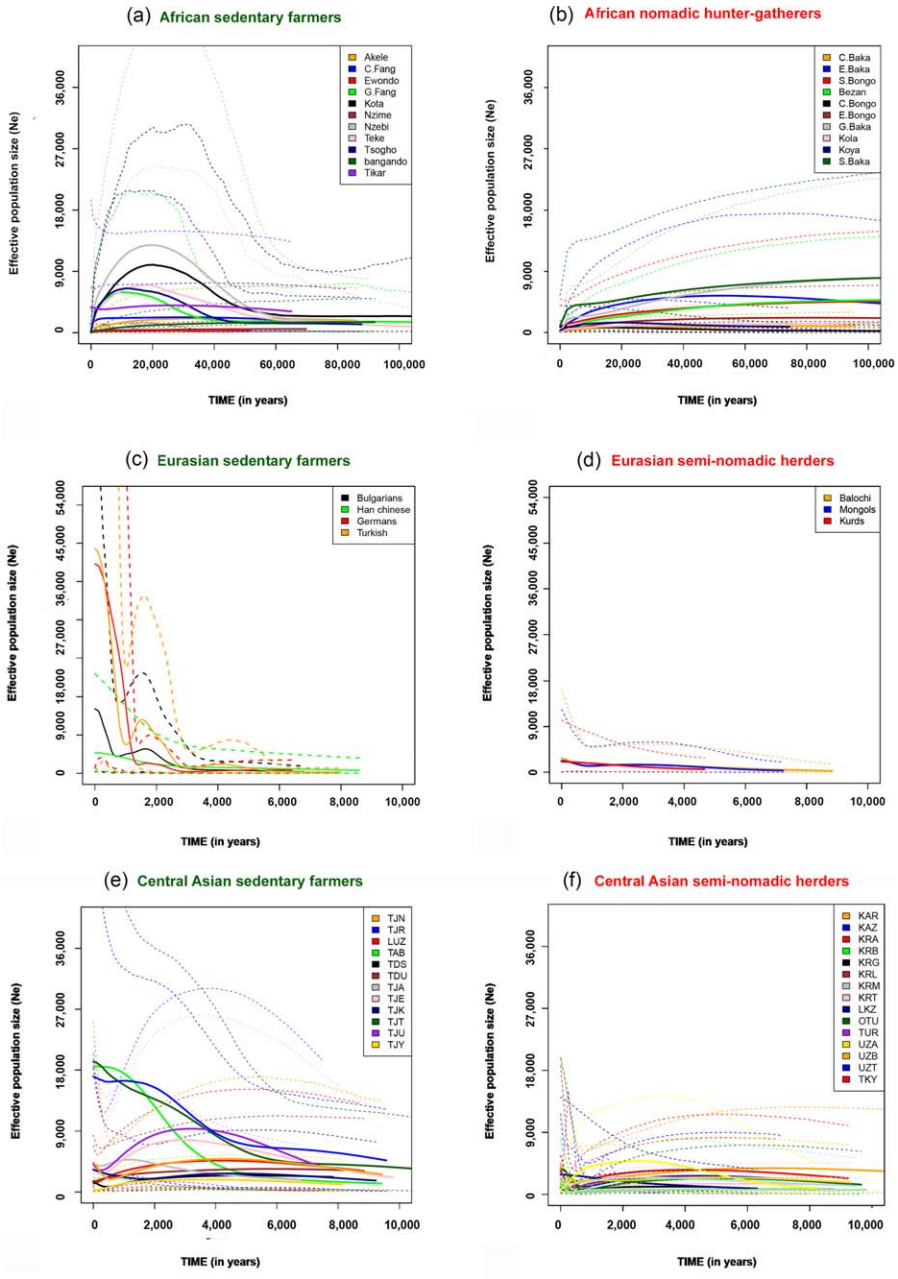

Fig. 2. Bayesian Skyline Plots inferred for African sedentary farmers (**a**), African nomadic hunter-gatherers (**b**), Eurasian sedentary farmers (**c**), Eurasian seminomadic herders (**d**), Central Asian sedentary farmers (**e**), and Central Asian seminomadic herders (**f**). Time is represented in years, assuming a generation time of 25 years. It is represented backward on the *X* axis: from present to the left to the most distant past on the right. 95% lower and upper HPD are represented by dashed lines. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

for nomadic herders (Fig. 2c,d). The expansion onsets time estimated with both parametric and nonparametric methods for sedentary farmers were quite more recent than the ones observed in Africa, and compatible with the Neolithic transition (Table 1). Indeed, using the parametric method, we inferred expansions starting between 8,321 [5,571–12,403] YBP and 18,093 [1,446–42,822] YBP (Supporting Information Table S3). Using the nonparametric method, we inferred expansions starting between 6,400 YBP and 8,600 YBP (Supporting Information Table S4).

For Central Asia, we also inferred expansion events for the sedentary farmer populations, but not for the nomadic herder populations using the parametric method. Indeed, the "Expansion model" best fitted the data, with a difference in DIC with the other models higher than five points for all sedentary populations except TJY (Table 1 and Supporting Information Table S2). Conversely, the "Constant model" or the "Exponential model" with negative modal values for growth rates best fitted the data for all nomadic herder populations and the TJY (Table 1 and Supporting

Information Table S2). The inferred growth rates for the expanding populations ranged between 3.96×10^{-4} [4.64×10^{-6} to 1.65×10^{-2}] and 2.96×10^{-2} [6.58×10^{-3} to 4.56×10^{-2}] per year (Supporting Information Table S3). As for the BSPs, although the HPD intervals were rather wide, at least seven farmer populations (TAB, TJE, TJR, TJT, TJU, LUZ, and TJN), but no nomadic populations, showed a tendency for expansion (Fig. 2e,f). Again, using the parametric method, the estimates for expansion onset dates for sedentary farmers were compatible with the Neolithic transition or the upper Paleolithic period (Table 1). Indeed, we inferred expansion onsets times varying between 7,259 [4,265–28,179] YBP and 9,782 [3,812–29,176] YBP (Supporting Information Table S3). We inferred similar values using the nonparametric method, between 8,762 YBP and 11,098 YBP (Supporting Information Table S4).

DISCUSSION

In this study, using Y-chromosome microsatellite markers, we inferred for African populations Paleolithic expansion events in most sedentary farmer populations and no significant changes in effective population sizes for nomadic hunter-gatherers. These results are mostly consistent with our previous results on mitochondrial and autosomal sequences (Aimé et al., 2013; see Supporting Information Table S1 for more details about our previous results obtained for each population). They contrast, nevertheless, with the more recent (Neolithic) expansion events inferred for sedentary farmers using autosomal microsatellites (Aimé et al., in press).

The situation was quite different for Eurasia, where we inferred here much more recent expansion events for the Y chromosome than in Africa, with dating compatible with the Neolithic transition. These expansions were inferred for sedentary farmers but not for nomadic herders. Except for Central Asia, these inferences were, in this case, consistent with those obtained from autosomal microsatellites in Aimé et al. (in press) (see Supporting Information Table S1 for more details), but not with the more ancient (Paleolithic) expansion events inferred using mitochondrial and autosomal sequences in Aimé et al. (2013). For Central Asia, we had found expansion events for both sedentary farmers and nomadic herders (Aimé et al., 2013, in press), while we found here these events only for sedentary farmers using Y-chromosome data.

As Y-chromosome microsatellites, mitochondrial DNA, autosomal sequences, and autosomal microsatellites are four very different sets of genetic loci, each with different transmission and mutational issues, formal testing of the differences between our inferences would be certainly inappropriate. However, especially because of these differences, comparing the general trends observed can give us insights about how contrasted modes of transmission can lead to different or similar inferences. In particular, as the mitochondrial DNA is maternally inherited and the Y chromosome is paternally inherited, contrasted patterns obtained from these markers can be interpreted as signals of sex-specific processes (Heyer et al., 2012). We will discuss below the differences observed with the Y chromosome as compared to the other systems.

Impact of sex-specific processes on demographic inferences

Investigating the paternally-inherited Y chromosome and contrasting the observed trends with those found with other systems (Aimé et al., 2013, in press), in particular with the maternally-inherited mitochondrial DNA (mtDNA), allowed us to observe several male-specific patterns. We found, in particular, for the Y chromosome, much more recent expansion times for the Eurasian populations than for the African populations. This difference was quite strong, since, while the expansion onset times appeared to be generally consistent with rather ancient Paleolithic times for the African populations, they pointed rather toward Neolithic expansions times or late Paleolithic expansion times for the Eurasian populations. This discrepancy between Africa and Eurasia seemed substantially stronger than for the other genetic markers (mtDNA and autosomal sequences, Aimé et al., 2013; autosomal microsatellites, Aimé et al., in press). This might be connected with male-specific processes. In particular, male effective population size is known to be generally lower than its female counterpart in human populations (Wilder et al., 2004; Lippold et al., 2014). Thus, the Out-of-Africa bottleneck might have affected more the male-transmitted genes than the female- or biparentally-transmitted ones. Traces of expansion events that occurred before the Out-of-Africa migration might therefore have been attenuated in Eurasia for the Y chromosome, because of this bottleneck effect. This process could explain that we could only detect the most recent expansions in Eurasia using Y-chromosome data, which may translate into an apparently more recent expansion time for the male line.

This explanation is consistent with Tang et al. (2002), who found the Y-chromosome most recent common ancestor (MRCA) to be of the order of half that for mtDNA for Eurasia, and who suggested that this discrepancy could result from a lower effective population size for men than for women. It is also consistent with Lippold et al. (2014), who demonstrated that the ratio of female effective population size to male effective population size has been greater than one throughout the history of modern humans. Lippold et al. (2014) also found more recent expansion signals in Eurasia than in Africa. Nevertheless, the expansion signals that they inferred for Eurasia was more ancient than in our study (40,000–80,000 years BP), and they did not find a major discrepancy between mitochondrial-based and Y-chromosome based inferred dates of expansions onsets. However, they mentioned that these results should be interpreted with caution, both because of the small sample sizes leading to wide confidence intervals, and also as, unlike in our study, they merged samples from several populations with different life-styles and demographic histories, which can produce spurious signals of population growth (Gunnarsdottir et al., 2011). Furthermore, our estimates for expansion onset times in Eurasia are consistent with those found by Pritchard et al. (1999) using a set of eight Y-chromosome microsatellites with a rejection algorithm approach. However, Pritchard et al. detected also recent expansion events (starting between 5,000 and 37,000 years) for Africa. Some constraints in their model, especially the choice of a prior distribution for expansion onset times assuming a mean value of 20,000 years might explain the fact that they did not detect more ancient expansions. Together, our results and those from Pritchard et al. (1999) are also consistent with our previous conclusions reported in Aimé et al. (2013, in press), in which we suggested two successive expansion waves in both African and Eurasian farmer populations, the first starting during Paleolithic

times and the second resulting from the Neolithic transition.

Moreover, our results for Central Asia differed from our previous inferences from maternally and biparentally transmitted markers (Aimé et al., 2013, in press), for which we found signals of demographic expansions for both sedentary farmers and seminomadic herders in this area. Using the male-transmitted Y chromosome, we found here expansion signals for Central Asian sedentary farmers but not for Central Asian nomadic herders. This discrepancy between the Y chromosomes and the other kind of markers appears to be quite specific to Central Asia. Indeed, our results for Africa and the rest of Eurasia are consistent with our previous studies by showing expansions in sedentary but not in nomadic populations.

This discrepancy observed in Central Asia might be connected with differential migration effects. Indeed simulation studies (Ray and Excoffier, 2003; Excoffier, 2004) have shown that expansion signals can be observed in populations that have undergone a spatial expansion process, in which populations colonize successively new habitats, but that these signals will only be observed in populations where the number of immigrants is high enough. In our previous study on mtDNA data (Aimé et al., 2013), we found that the expansion signal in this area could result, to some extent, from a spatial expansion process. In fact, farmers and herders may both have experienced such spatial expansion events, but the remaining signal of these expansions might not be observed nowadays for the Y chromosome in herder populations, because of the very low male migration rate for these populations in this area. Indeed these populations are highly patrilocal, which leads to a very limited level of male migration (Chaix et al., 2007), which thus precludes the observation of an expansion signal on the Y chromosome (Ray and Excoffier, 2003). Together with this phenomenon, the specific patrilineal social organization in herders populations further reduces the male effective population size (Ségurel et al., 2008), which may also participate to lower expansion signals for the Y chromosome as compared to mtDNA.

Possible confounding factors

First, the BEAST analyses that we performed here make the implicit assumption that the studied populations are isolated, which is questionable for human populations. Especially, the processes leading to the Bantu expansion (i.e. the spatial expansion of Bantu-speaking farmer populations in Africa during the Neolithic period, see e.g. Nurse and Philippson, 2003) led to large population migrations. However, to date, this phenomenon is difficult to explicitly take into account, as we still lack crucial knowledge concerning the geographical origins of some of the major Y-lineages, which we nowadays find widely dispersed in sub-Saharan African populations.

Second, selection may have occurred on the whole NRY region. However, we analyzed here a large set of populations sampled in very distant geographical regions with contrasted environments (i.e Central Africa, East Africa, Europe, Middle-East, Central Asia, Pamir, Siberia, and East-Asia). The main conclusions of this study rely on consistent patterns between these areas, and it seems unlikely that processes such as selection could have biased the estimates in the same way for all studied populations within each group. Moreover, as explained above, our result of different expansion patterns between sedentary and nomadic populations have been also demonstrated using neutral autosomal markers in previous studies (Aimé et al., 2013, in press). Finally, note that data availability did not allow us to use the same loci for all geographic areas, precluding formal comparison between each of them. Our conclusions are, therefore, based on comparisons of general trends which are, nevertheless, highly consistent among populations within each group.

As for the dating of expansion events, we have used here a generation time of 25 years, permitting the comparison with previous studies (Aimé et al., 2013, in press). Using a generation time of 29 years (Tremblay and Vezina, 2000) instead of 25 lead to slightly more ancient estimates for all populations, not changing our main conclusions. The time estimates are also quite dependent upon the mutation rate. We used here a value of 2.1×10^{-3}/generation/locus, which was obtained both in a pedigree study (Heyer et al., 1997) and in a study combining population and father-son pair data (Burgarella and Navascués, 2011). Zhivotovsky et al. (2004) estimated a rate of 6.9×10^{-4}/generation/locus, with an indirect method that might be more sensitive to selection and homoplasy (Burgarella and Navascués, 2011). In this context, it is interesting to note that we performed the BEAST analyses assuming a SSM mutation model that takes homoplasy into account. Using the rate of Zhivotovsky et al. (2004) or the recalibrated rate of Shi et al (2010) on the BEAST outputs would lead us to very ancient expansion onset times in Africa (up to 300,000 years), which appear rather in contradiction with our own estimates and the estimates by other authors (e.g. Batini et al., 2011) on mtDNA and autosomes.

Our conclusion on that aspect differ thus from a recent paper of Wei et al. (2013b), which concluded that the inferences obtained using this mutation rate with other programs (NETWORK and BATWING) better correlate with those obtained with re-sequencing data (high-coverage complete sequences of 36 diverse human Y chromosomes from Africa, Europe, South Asia, East Asia, and the Americas). This might be linked with the assumptions made by the different programs. Nevertheless, the inferences obtained by Wei et al. (2013a) on these complete sequence data are quite consistent with our results. First, they found an ancient expansion event (57,000–74,000 YBP), which we have also detected using mitochondrial DNA and autosomal sequences for both Africa and Eurasia (Aimé et al., 2013). Second, they found a more recent expansion event (4,300–13,000 YBP), which we have also detected using autosomal microsatellite data for both Africa and Eurasia (Aimé et al., in press). Thus, using complete Y chromosome sequences on a small set of individuals allowed them to detect both expansion events. On the opposite, using our methodology on a worldwide set of Y-chrosomome STR data, we did not detect the most ancient expansion event in Eurasia, while we detected the more recent one. As explained above, traces of expansion events which occurred before the Out-of-Africa migration might have been attenuated (and therefore might be more difficult to detect) in Eurasia for the Y chromosome because of a bottleneck effect. Studying a limited number of markers on a large set of populations allowed us, nevertheless, to demonstrate contrasted patterns between different regions of the World and to study in details the Central Asian area.

Finally, the nonparametric method (BSPs) gave sometimes less clear results than the parametric method, as the HPD intervals on BSP graphs were very wide, thus often not allowing to distinguish between a constant population size and an expansion event. However, the use of BSPs allowed us to infer the demographic history of populations with no prior assumption of a particular demographic model. This is especially interesting for analyzing populations for which previous knowledge about demographic history is scarce, which was often the case here (especially for Eurasian nomadic herders and African hunter-gatherers). Here, the fact that BSP graphs showed expansions in some sedentary populations but no nomadic populations, together with the agreement between expansion onset times inferred from the parametric and the nonparametric methods for these populations, confirmed that the a priori assumption of particular demographic models seem unlikely to have strongly impacted our conclusions with the parametric method.

CONCLUSION

Our Y-chromosome based inferences allowed us to infer expansion patterns in most farmer populations in Africa and Eurasia, and no expansion signals in hunter-gatherer and most herder populations. We were, in particular, able to identify specific male-related patterns. In particular, we inferred a strong discrepancy between the expansion times inferred for African and Eurasian populations, the former being found to be much more ancient than the latter. As this discrepancy is much stronger than for mtDNA and autosomal DNA, we suggest that male-specific patterns may be responsible for it. In particular, the bottleneck connected with the out-of-Africa process may have more strongly affected male-transmitted genes, as the effective population size is usually smaller for men than for women in humans.

Moreover, focusing on Central Asia, we found a strong impact of migration patterns and sex-specific processes on demographic inferences in this area. These processes resulted in a discrepancy between the inferences from paternally-inherited markers, on one side, and maternally or biparentally markers, on the other side. In fact, farmers and herders may both have experienced spatial expansion events in Central Asia, but the remaining signal of these expansions cannot be observed for the Y chromosome in herder populations because of the very low male migration rate for herders in this area (Ray et al., 2003). These results highlight the importance of analyzing several types of markers when performing demographic inferences from genetic data. In addition, as the low male effective population size and migration rates for herders results in part from patrilocality (Chaix et al., 2007; Ségurel et al., 2008), our study also highlights the importance of considering cultural factors when studying the repartition of genetic diversity in humans.

LITERATURE CITED

Aimé C, Laval G, Patin E, Verdu P, Ségurel L, Chaix R, Hegay T, Quintana-Murci L, Heyer E, Austerlitz F. 2013. Human genetic data reveal contrasting demographic patterns between sedentary and nomadic populations that predate the emergence of farming. Mol Biol Evol 30:2629-2644.

Aimé C, Verdu P, Ségurel L, Martinez-Cruz B, Heyer E, Austerlitz F. In press. Microsatellite data show recent demographic expansions in sedentary but not in nomadic human populations in Africa and Eurasia. Euro J Hum Genet 22: 1201-1207.

Batini C, Lopes J, Behar DM, Calafell F, Jorde LB, Van Der Veen L, Quintana-Murci L, Spedini G, Destro-Bisol G, Comas D. 2011. Insights into the demographic history of African Pygmies from complete mitochondrial genomes. Mol Biol Evol 28:1099-1110.

Beaumont MA. 2004. Recent developments in genetic data analysis: what can they tell us about human demographic history? Heredity 92:365-379.

Burgarella C, Navascués M. 2011 Mutation rate estimates for 110 Y-chromosome STRs combining population and father-son pair data. Eur J Hum Genet 19:70-75.

Cann RL, Stoneking M, Wilson AC. 1987. Mitochondrial DNA and human evolution. Nature 325:31-36.

Cavalli-Sforza LL, Feldman MW. 2003. The application of molecular genetic approaches to the study of human evolution. Nat Genet 33:266-275.

Cavalli-Sforza LL, Menozzi P, Piazza A. 1994. The history and geography of human genes. Princeton: Princeton University Press.

Chaix R, Austerlitz F, Hegay T, Quintana-Murci L, Heyer E. 2008. Genetic traces of east-to-west human expansion waves in eurasia. Am J Phys Anthropol 136:309-317.

Chaix R, Austerlitz F, Khegay T, Jacquesson S, Hammer MF, Heyer E, Quintana-Murci L. 2004. The genetic or mythical ancestry of descent groups: lessons from the Y chromosome. Am J Hum Genet 75:1113-1116.

Chaix R, Quintana-Murci L, Hegay T, Hammer MF, Mobasher Z, Austerlitz F, Heyer E. 2007. From social to genetic structures in Central Asia. Curr Biol 17:43-48.

Drummond AJ, Rambaut A. 2007. Beast: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol 7:214.

Drummond AJ, Rambaut A, Shapiro B, Pybus OG. 2005. Bayesian coalescent inference of past population dynamics from molecular sequences. Mol Biol Evol 22:1185-1192.

Excoffier L. 2004. Patterns of DNA sequence diversity and genetic structure after a range expansion: lessons from the infinite-island model. Mol Ecol 13:853-864.

Excoffier L, Heckel G. 2006. Computer programs for population genetics data analysis: a survival guide. Nature Rev Genet 7: 745-758.

Gelman A, Rubin DB. 1992. Inference from iterative simulation using multiple sequences (with discussion). Stat Sci 7:457-511.

Gunnarsdottir ED, Li M, Bauchet M, Finstermeier K, Stoneking M. 2011. High-throughput sequencing of complete human mtDNA genomes from the Philippines. Genome Res 21:1-11.

Henke J, Henke L, Chatthopadhyay P, Kayser M, Dulmer M, Cleef S, Pöche H, Felske-Zech H. 2001. Application of Y-chromosomal STR haplotypes to forensic genetics. Croat Med J 42:292-297.

Heyer E, Balaresque P, Jobling MA, Quintana-Murci L, Chaix R, Ségurel L, Aldashev A, Hegay T. 2009. Genetic diversity and the emergence of ethnic groups in Central Asia. BMC Genet 10:49.

Heyer E, Chaix R, Pavard S, Austerlitz F. 2012. Sex-specific demographic behaviours that shape human genomic variation. Mol Ecol 21:597-612.

Heyer E, Puymirat J, Dietjes P, Bakker E, De Knijff P. 1997. Estimating Y chromosome specific microsatellite mutation frequencies using deep rooting pedigrees. Hum Mol Genet 6:799-803.

Kayser M, Brauer S, Weiss G, Schiefenhovel W, Underhill PA, Stoneking M. 2001. Independent histories of human Y chromosomes from Melanesia and Australia. Am J Hum Genet 68:173-190.

Kingman JFC. 1982. The coalescent. Stochast Proc App 13:235-248.

Kuhner MK. 2008. Coalescent genealogy samplers: windows into population history. Tr Ecol Evol 24:86-93.

Lippold S, Xu H, Ko A, Li M, Renaud G, Butthof A, Schroeder R, Stoneking M. 2014. Human paternal and maternal

demographic histories: insights from high-resolution Y chromosome and mtDNA sequences. Investig Genet 5:13.
Loader C. 1999. Local regression and likelihood. New York: Springer.
Nurse D, Philippson G. 2003 The Bantu languages. Routledge language family series 4. London, UK: Routledge.
Pakendorf B, Stoneking M. 2005. Mitochondrial DNA and human evolution. Annu Rev Genomics Hum Genet 6:165-183.
Plummer M, Best N, Cowles K, Vines K. 2006. CODA: Convergence diagnosis and output analysis for MCMC. R News 6:7-11.
Pritchard JK, Seielstad MT, Perez-Lezaun A, Feldman MW. 1999. Population growth of human Y chromosomes: a study of Y chromosome microsatellites. Mol Biol Evol 16:1791-1798.
Qamar R, Ayub Q, Mohyuddin A, Helgason A, Mazhar K, Mansoor A, Zerjal T, Tyler-Smith C, Mehdi SQ. 2002. Y-chromosomal DNA variation in Pakistan. Am J Hum Genet 70: 1107-1124.
Quintana-Murci L, Chaix R, Wells RS, Behar DM, Sayar H, Scozzari R, Rengo C, Al-Zahery N, Semino O, Santachiara-Benerecetti AS, Coppa A, Ayub Q, Mohyuddin A, Tyler-Smith C, Qasim Mehdi S, Torroni A, McElreavey K. 2004. Where west meets east: the complex mtDNA landscape of the Southwest and Central Asian corridor. Am J Hum Genet 74: 827-845.
Quintana-Murci L, Semino O, Bandelt HJ, Passarino G, Mcelreavey K, Santachiara-Benerecetti AS. 1999. Genetic evidence of an early exit of *Homo sapiens* from Africa through eastern Africa. Nat Genet 23:437-441.
R.Development.Core.Team. 2011. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.
Rambaut A, Suchard MA, Xie D, Drummond AJ. 2014. Tracer v1.6, Available from http://beast.bio.ed.ac.uk/Tracer.
Ray N, Currat M, Excoffier L. 2003. Intra-deme molecular diversity in spatially expanding populations. Mol Biol Evol 20:76-86.
Ségurel L, Austerlitz F, Toupance B, Gautier M, Kelley JL, Pasquet P, Lonjou C, Georges M, Voisin S, Cruaud C, Couloux A, Hegay T, Aldashev A, Vitalis R, Heyer E. 2013. Positive selection of protective variants for type 2 diabetes from the Neolithic onward: a case study in Central Asia. Eur J Hum Genet 21:1146–1151.
Ségurel L, Martinez-Cruz B, Quintana-Murci L, Balaresque P, Georges M, Hegay T, Aldashev A, Nasyrova F, Jobling MA, Heyer E, Vitalis R. 2008. Sex-specific genetic structure and social organization in central Asia: insights from a multilocus study. PLoS Genet 4:e1000200.
Shi W, Ayub Q, Vermeulen M, Shao RG, Zuniga S, van der Gaag K, de Knijff P, Kayser M, Xue Y, and Tyler-Smith C. 2010. A worldwide survey of human male demographic history based on Y-SNP and Y-STR data from the HGDP-CEPH populations. Mol Biol Evol 27:385-393.
Short R. 1982. The biological basis for the contraceptive effects of breast feeding. Int J Gynaecol Obstet 25:207-217.
Spiegelhalter DJ, Best NG, Carlin BR, Van Der Linde A. 2002. Bayesian measures of model complexity and fit. J R Stat Soc Ser B Stat Meth 64:583-616.
Tang H, Siegmund DO, Peidong S, Oefner PJ, Feldman MW. 2002. Estimation of coalescence times from nucleotide sequence data using a tree-based partition. Genetics 161:447–459.
Tremblay M, 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.
Verdu P, Becker NS, Froment A, Georges M, Grugni V, Quintana-Murci L, Hombert JM, Van der Veen L, Le Bomin S, Bahuchet S, Heyer E, Austerlitz F. 2013. Sociocultural behavior, sex-biased admixture, and effective population sizes in Central African Pygmies and non-Pygmies. Mol Biol Evol 30:918-937.
Wei W, Ayub Q, Chen Y, McCarthy S, Hou Y, Carbone I, Xue Y, and Tyler-Smith C. 2013a. A calibrated human Y-chromosomal phylogeny based on resequencing. Genome Res 23:388-395.
Wei W, Ayub Q, Xue Y, Tyler-Smith C. 2013b. A comparison of Y-chromosomal lineage dating using either resequencing or Y-SNP plus Y-STR genotyping. Forensic Sci Int: Genet 7:568-572.
Wilder JA, Mobasher Z, Hammer MF. 2004. Genetic evidence for unequal effective population sizes of human females and males. Mol Biol Evol 21:2047-2057.
Wu CH, Drummond AJ. 2011. Joint inference of microsatellite mutation models, population history and genealogies using transdimensional Markov chain monte carlo. Genetics 188: 151-164.
Zaharova B, Andonova S, Gilissen A, Cassiman JJ, Decorte R, Kremensky I. 2001. Y-chromosomal STR haplotypes in three major population groups in Bulgaria. Forensic Sci Int 124: 182-186.
Zhivotovsky LA, Underhill PA, Cinnioglu C, Kayser M, Morar B, Kivisild T, Scozzari R, Cruciani F, Destro-Bisol G, Spedini G, Chambers GK, Herrera RJ, Yong KK, Gresham D, Tournev I, Feldman MW, Kalaydjieva L. 2004. The effective mutation rate at Y chromosome short tandem repeats, with application to human population-divergence time. Am J Hum Genet 74:50-56.