

Human Genetic Data Reveal Contrasting Demographic Patterns between Sedentary and Nomadic Populations That Predate the Emergence of Farming

Carla Aimé,^{*1} Guillaume Laval,² Etienne Patin,² Paul Verdu,¹ Laure Ségurel,³ Raphaëlle Chaix,¹ Tatyana Hegay,⁴ Lluís Quintana-Murci,² Evelyne Heyer,¹ and Frédéric Austerlitz¹

¹Laboratoire Eco-Anthropologie et Ethnobiologie, UMR 7206, Muséum National d'Histoire Naturelle, Centre National de la Recherche Scientifique, Université Paris 7 Diderot, Paris, France

²Unit of Human Evolutionary Genetics, Institut Pasteur, CNRS URA3012, Paris, France

³Department of Human Genetics, University of Chicago

⁴Academy of Sciences, Institute of Immunology, Tashkent, Uzbekistan

*Corresponding author: E-mail: aime@mnhn.fr.

Associate editor: John Novembre

Abstract

Demographic changes are known to leave footprints on genetic polymorphism. Together with the increased availability of large polymorphism data sets, coalescent-based methods allow inferring the past demography of populations from their present-day patterns of genetic diversity. Here, we analyzed both nuclear (20 noncoding regions) and mitochondrial (HVS-I) resequencing data to infer the demographic history of 66 African and Eurasian human populations presenting contrasting lifestyles (nomadic hunter-gatherers, nomadic herders, and sedentary farmers). This allowed us to investigate the relationship between lifestyle and demography and to address the long-standing debate about the chronology of demographic expansions and the Neolithic transition. In Africa, we inferred expansion events for farmers, but constant population sizes or contraction events for hunter-gatherers. In Eurasia, we inferred higher expansion rates for farmers than herders with HVS-I data, except in Central Asia and Korea. Although isolation and admixture processes could have impacted our demographic inferences, these processes alone seem unlikely to explain the contrasted demographic histories inferred in populations with different lifestyles. The small expansion rates or constant population sizes inferred for herders and hunter-gatherers may thus result from constraints linked to nomadism. However, autosomal data revealed contraction events for two sedentary populations in Eurasia, which may be caused by founder effects. Finally, the inferred expansions likely predated the emergence of agriculture and herding. This suggests that human populations could have started to expand in Paleolithic times, and that strong Paleolithic expansions in some populations may have ultimately favored their shift toward agriculture during the Neolithic.

Key words: population genetics, inferences, coalescent, neolithic transition, expansions.

Introduction

Studying the current distribution of genetic diversity in human populations has important implications for our understanding of the evolution and history of our species. Indeed, within- and among-population genetic diversity has been shaped both by demographic forces, such as gene flow and genetic drift, and by selective processes (e.g., Balaesque et al. 2007). Cultural factors like social organization and technological innovation have also had a considerable indirect impact on patterns of genetic diversity, as they can influence both the demographic and adaptive history (e.g., Ambrose 2001; Oota et al. 2001; Kumar et al. 2006; Heyer et al. 2012).

The Neolithic revolution is thought to be one of the most important cultural and technological transitions in human history. During this period, different human populations domesticated plants and animals in several parts of the world, including Central Africa, the Middle Eastern Fertile Crescent, Eastern Asia, and Central America (Bocquet-Appel and

Bar-Yosef 2008). The emergence of farming occurred concomitantly with the sedentarization of most nomadic hunter-gatherer populations. Other populations remained nomadic, but some of them also developed new means of subsistence like nomadic herding. According to some archeologists and paleoanthropologists, the major human expansions would have started as a result from the Neolithic transition: sedentarized populations could have experienced strong demographic expansions (e.g., Bocquet-Appel 2011), whereas nomadic populations may have remained constant because of inherent constraints of their lifestyle (e.g., a longer inter-birth interval; Short 1982). However, a number of population genetic studies have reported evidence for more ancient expansion processes in many African and Eurasian populations, starting during the Paleolithic period (e.g., Chaix et al. 2008; Atkinson et al. 2009; Laval et al. 2010; Batini et al. 2011). These findings seem consistent with the “demographic theory” proposed by Sauer (1952), according to which human populations could have started to increase

before the Neolithic, and these Paleolithic expansions in some populations may have ultimately favored their shift toward farming.

The recent developments in sequencing technologies and bioinformatics tools have allowed the exploration of large multilocus polymorphism data sets. In combination with archeological and paleoanthropological records, it can substantially improve our ability to infer past demographical events (Beaumont 2004). Stemming from Kingman's (1982) coalescent theory, numerical coalescent-based methods have thus been developed, allowing the inference of demographic parameters from molecular data. Most of these methods assume a specific demographic model. Moreover, nonparametric approaches, such as Extended Bayesian Skyline Plots (EBSs, Heled and Drummond 2010), allow inference of the demographic history of populations without assuming a specific model, by using the time intervals between serial coalescent events (see Excoffier and Heckel 2006 and Ho and Shapiro 2011 for reviews).

Here, we used these methods to investigate 1) the relationship between lifestyle (i.e., sedentary farming, nomadic herding, or nomadic hunting-gathering) and demographic patterns in a large set of African and Eurasian populations, and 2) the chronology of demographic expansions and the emergence of farming, by comparing inferred expansion onset times with the dating of the most ancient archeological traces of farming and herding (potteries, irrigation structures, and animals bones) reported in Bocquet-Appel and Bar-Yosef (2008) for each region. In addition, by computing F_{ST} values and immigration rates, we investigated the extent to which the inferred demographic patterns could be explained by spatial expansion processes. Indeed, modeling studies (Ray et al. 2003; Excoffier 2004) have shown that such processes can produce signals on within-population diversity patterns similar to those obtained with pure demographic expansions. In particular, these studies argue that ancient spatial expansion signals could be attenuated or suppressed in isolated populations. Different expansion signals among populations as inferred from genetic data may thus in part reflect variation in immigration rates and extent of population isolation.

We used 20 a priori neutral autosomal regions and the hypervariable control region (HVS-1) of the mitochondrial DNA (mtDNA) sequenced in 404 individuals from 16 populations and 2,429 individuals from 61 populations, respectively (supplementary table S1, Supplementary Material online). Given their distinct properties and modes of transmission, we compared the inferences obtained with these two types of markers, in order to gain complementary insights into the past demography of the studied populations. By studying many populations from different geographic areas worldwide, we were able to determine which patterns were observed across all populations and which were specific to a given geographical region. First, we focused on Central Africa, where nomadic hunter-gatherer populations, commonly called Pygmies, coexist with sedentary farmer populations. These two groups are genetically differentiated and seem to have diverged about 60,000 years ago (Patin et al. 2009; Verdu et al. 2009), thus long before the Neolithic sedentarization of

farmer populations in this area (5,000–4,000 years before present [YBP]; Bocquet-Appel and Bar-Yosef 2008). Second, we analyzed a sample of populations from several distant geographical regions of Eurasia where sedentary farmers coexist with nomadic herders. This was of particular interest, as to our knowledge the differences in demographic processes between herders and farmers have not been studied yet. Third, we performed a more detailed study in Central Asia, another area of interest as it is thought to have been a major corridor during the successive Eurasian migration waves (Nei and Roychoudhury 1993).

Results

Neutrality Tests

Focusing first on Africa, all farmer populations showed at least one significantly negative value for one of the four neutrality tests (Tajima's [1989] D , Fu and Li's [1993] D and F and Fu's [1997] F_s , table 1), which can be interpreted as a signal of expansion. Conversely, hunter-gatherer populations showed no such expansion signals. Aka and Mbuti hunter-gatherers presented at least one significantly positive test, indicating a possible contraction event. Similarly, for HVS-I sequences, we found significantly negative Fu's F_s values for all farmer populations except the Ewondo, but no expansion signal for hunter-gatherers (supplementary table S2, Supplementary Material online). Kola hunter-gatherers showed a significantly positive Tajima's D , indicating a possible contraction event.

Similar analyses on autosomal sequences in Europe and East Asia revealed no significant expansion signals, neither in sedentary nor in nomadic populations (table 1). We even observed contraction signals in two sedentary populations, one East-Asian and one European. Indeed, we found significantly positive values for two neutrality tests for the Japanese and three neutrality tests for the Danes. Conversely, for HVS-I sequences from Eurasia (supplementary table S2, Supplementary Material online), we obtained significant signals of expansion for at least one test (Fu's F_s) for all populations (including Japanese and Danes). All sedentary populations except Koreans also showed significant signals of expansion for the three other tests, whereas the Koreans and all nomadic populations showed a significant expansion signal only for Fu's F_s .

Focusing on Central Asia, no neutrality test was significant for the autosomes in neither Tajik sedentary farmers (TAB) nor Kyrgyz nomadic herders (KIB) (table 1). Conversely, for HVS-I sequences (supplementary table S2, Supplementary Material online), all farmers and herders presented a significant expansion signal for at least one test, except one farmer population (TDS).

Coalescent-Based Inferences of Demographic History Africa: Pre-Neolithic Demographic Expansions in Sedentary Farmer Populations

Considering first the autosomal data, models consistent with an increase in population size best fitted the data for all African farmer populations (supplementary table S3, Supplementary Material online). The "expansion model"

Table 1. Summary Statistics and Neutrality Tests Computed from the Whole Autosomal Sequences.

Population	Area	Lifestyle	S ^a	K ^b	Tajima's D ^c	Fu & Li's D ^c	Fu & Li's F ^c	Fu's F ^c
Akele	Africa	Sedentary farmers	6.95	6.45	−0.35	−0.55	−0.57*	−1.12*
Chagga	Africa	Sedentary farmers	8.65	7.95	−0.48	−0.70*	−0.74*	−1.27*
Mozambicans	Africa	Sedentary farmers	8.80	9.55	−0.62*	−1.15**	−1.15**	−3.33*
Ngumba	Africa	Sedentary farmers	7.05	6.20	−0.20	−0.41*	−0.41	−0.68
Yoruba	Africa	Sedentary farmers	7.50	7.15	−0.14	−0.03	−0.03	−0.73
Aka	Africa	Nomadic HG ^d	6.95	6.60	0.12	0.34*	0.32	−0.30
G. Baka	Africa	Nomadic HG	6.30	6.00	0.008	0.17	0.14	−0.33
S. Baka	Africa	Nomadic HG	6.10	5.6	0.17	0.05	0.10	−0.03
Kola	Africa	Nomadic HG	6.55	6.25	−0.14	−0.03	−0.08	−0.75
Mbuti	Africa	Nomadic HG	6.60	6.10	0.25	0.35*	0.37*	0.16
Danes	Eurasia	Sedentary farmers	5.50	4.85	0.30*	0.16	0.24*	0.73*
Han	Eurasia	Sedentary farmers	5.20	4.70	−0.03	−0.01	−0.02	0.21
Japanese	Eurasia	Sedentary farmers	4.20	3.85	0.45*	0.22	0.34*	1.06*
Chuvash	Eurasia	Nomadic herders	5.70	5.05	0.09	0.11	0.12	0.34
Tajiks (TAB)	C. Asia	Sedentary farmers	9.00	9.00	0.19	0.03	0.10	0.24
Kyrgyz (KIB)	C. Asia	Nomadic herders	10.4	10.40	0.11	0.08	0.11	0.23

NOTE.—Values significantly higher than expected for a constant population size model are italicized, whereas significantly lower values are underlined.

^aNumber of polymorphisms.

^bNumber of haplotypes.

^cWe report the means over the 20 regions.

^dHG = Hunter-gatherers significance levels: * $P < 0.05$, ** $P < 0.01$ after FDR correction for multiple testing (Benjamini and Hochberg 1995).

best fitted the data for the two East-African farmer populations (namely Chagga and Mozambicans), whereas the “exponential model” best fitted the data for all West-African farmer populations (Akele, Ngumba, and Yoruba), with positive growth rates in all cases (supplementary table S4, Supplementary Material online). Conversely, no signals of expansion were found for hunter-gatherer populations, as the “constant model” always best fitted the data (supplementary tables S3 and S4, Supplementary Material online). Consistently, EBSPs showed signals of expansions for farmer populations (fig. 1A). 95% highest probability density (HPD) intervals for the estimated number of demographic changes did not include 0, indicating at least one significant change in population size (supplementary table S5, Supplementary Material online). Conversely, we found no evidence of population size changes for hunter-gatherers (fig. 1B and supplementary table S5, Supplementary Material online). We further dated the onset of farmer expansions from at least 62,275 YBP (assuming $\mu = 2.5 \times 10^{-8}$ /generation/site) or 124,550 YBP (assuming $\mu = 1.2 \times 10^{-8}$ /generation/site) for Mozambicans to 7,975 or 15,950 YBP for Yoruba. Visual examination of the 95% HPD intervals showed that the expansion event inferred for the Mozambican population was significantly older than those inferred for the other populations (supplementary table S6, Supplementary Material online).

We found similar results for the HVS-I sequences from Central Africa (supplementary tables S3 and S7, Supplementary Material online). Indeed, the exponential model with positive growth rates best fitted the data for all farmer populations, indicating expansion events. Conversely, the exponential model with negative modal values for growth rate (i.e., contraction event) provided the best fit for all hunter-gatherer populations. However, as the 95% HPD

intervals for growth rates included 0, we could not conclude any significant contraction events for these populations. Similarly, EBSPs indicated a significant expansion event for all farmer populations (fig. 2A and supplementary table S5, Supplementary Material online), whereas we found no evidence of population size changes for hunter-gatherers (fig. 2B and supplementary table S5, Supplementary Material online). We dated farmer populations expansions from 31,350 or 62,700 YBP (assuming $\mu = 10^{-5}$ or 5×10^{-6} /generation/site, respectively) to 45,319 or 90,638 YBP (supplementary table S6, Supplementary Material online).

Finally, both with autosomes and HVS-I, all hunter-gatherer populations had lower current effective population size (N_0) values than farmer populations (supplementary tables S4 and S7, Supplementary Material online). Furthermore, the inferred expansion onsets for all farmer populations largely predated the emergence of farming in Central Africa (5,000–4,000 YBP; Bocquet-Appel and Bar-Yosef 2008) (figs. 3 and 4).

Eurasia: Contrasting Demographic Patterns for Farmer Populations with Autosomes and Stronger Pre-Neolithic Expansions for Farmers Than Herders with HVS-I

The coalescent-based analyses of autosomes in East-Asian and European populations showed contrasting demographic patterns across sedentary populations (supplementary tables S3 and S4, Supplementary Material online). Using the parametric BEAST analysis, the expansion model best fitted the data for Han Chinese, indicating an expansion event. Conversely, we inferred that Japanese and Danes either underwent a contraction event or remained at constant size. Indeed, the exponential model with negative growth rates best fitted the data for these two populations, but the

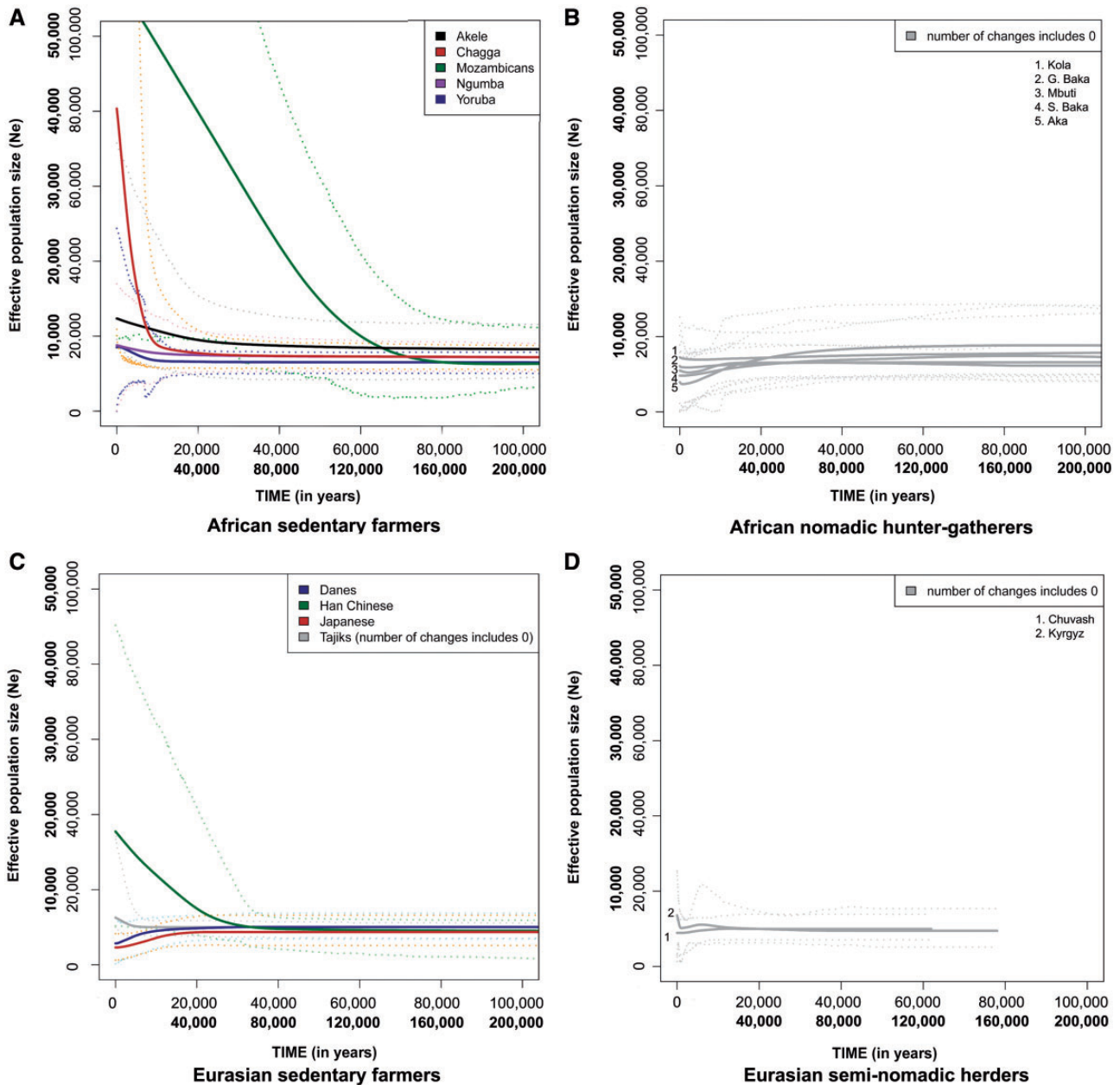


Fig. 1. EBSs inferred from autosomal sequences in African sedentary farmers (A), African nomadic hunter-gatherers (B), Eurasian sedentary farmers (C), and Eurasian nomadic herders (D). The values indicated in bold on the axes are obtained assuming a mutation rate of $\mu = 1.2 \times 10^{-8}$ /generation/site (measured from trios parents–children by Conrad et al. 2011), and the other values correspond to $\mu = 2.5 \times 10^{-8}$ /generation/site (derived from the sequence divergence human–chimpanzee by Pluzhnikov et al. 2002). Although time was expressed in generations for the analyses, we represented time in years here, assuming a generation time of 25 years. Time 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. Populations for which the estimated number of demographic changes include 0 (i.e., no significant signal of expansion or decline) are represented in light gray.

95% HPD intervals also included $g = 0$. The constant model best fitted the data for the Chuvash, a traditionally nomadic population. EBSs showed a significant expansion event for the Han population: the value of 0 was not included in the 95% HPD interval of the number of demographic changes (supplementary table S5, Supplementary Material online). These expansion events started at least 36,025 or 72,050 YBP (fig. 1C and supplementary table S6, Supplementary Material online), clearly predating the emergence of farming in East Asia, about 9,000 YBP (Bocquet-Appel and Bar-Yosef

2008) (fig. 3). Japanese showed a significant contraction event (i.e., the value of 0 was not included in the 95% HPD interval of the number of demographic changes; supplementary table S5, Supplementary Material online) starting at least 21,350 or 42,700 YBP. Danes also showed a significant contraction event, starting at least 26,440 or 52,880 YBP (fig. 1C and supplementary tables S5 and S6, Supplementary Material online). EBS analyses showed no significant demographic changes for the Chuvash (fig. 1D and supplementary table S5, Supplementary Material online).

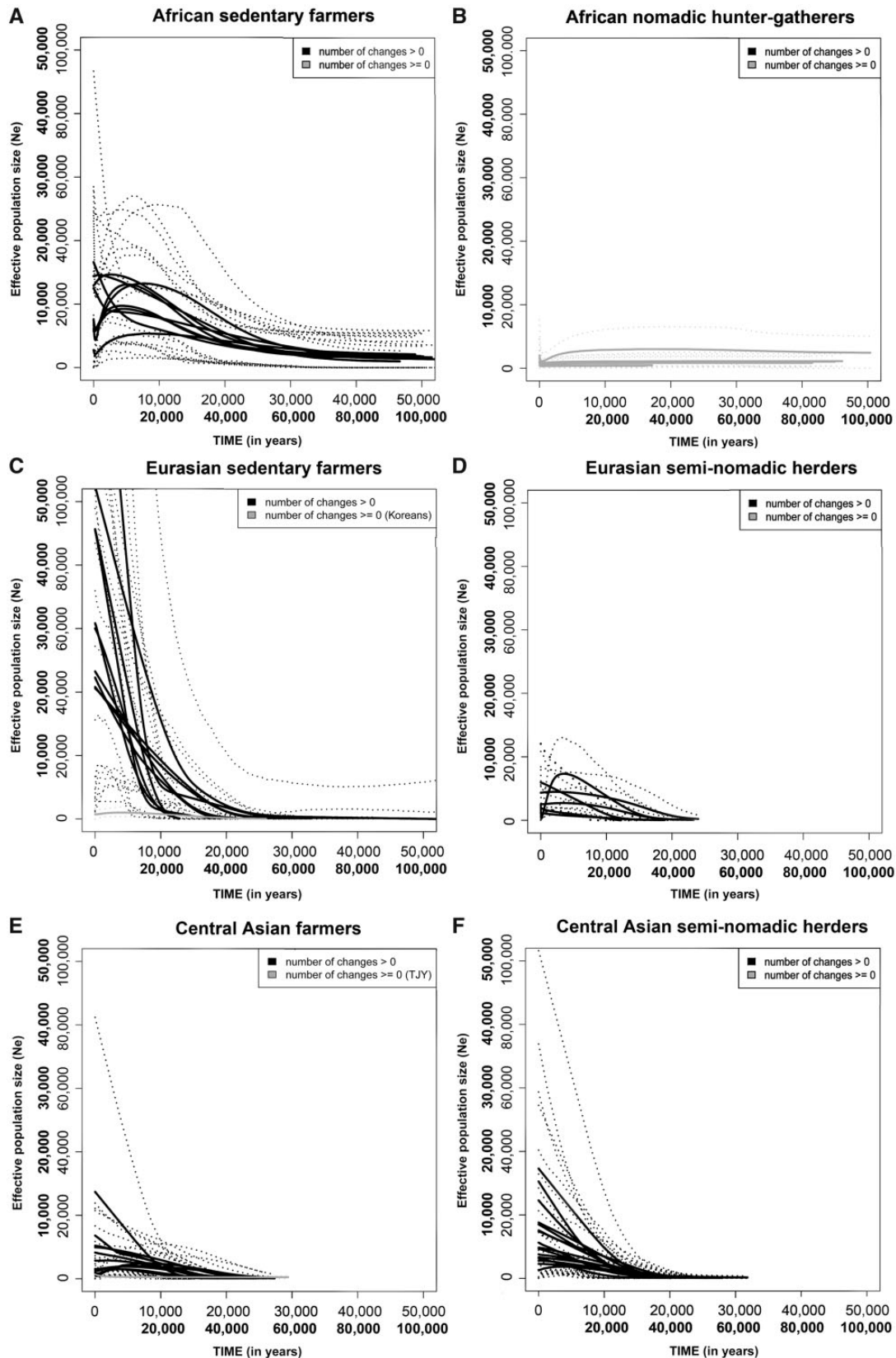


FIG. 2. EBSPs inferred from HVS-I sequences in African sedentary farmers (A), African nomadic hunter-gatherers (B), Eurasian sedentary farmers (C), Eurasian nomadic herders (D), Central Asian sedentary farmers (E), and Central Asian nomadic herders (F). The values indicated in bold are obtained assuming a mutation rate of $\mu = 5 \times 10^{-6}$ /generation/site (transitional changes rate, Forster et al. 1996), and the others correspond to $\mu = 10^{-5}$ /generation/site (pedigree-based, Howell et al. 1996; Heyer et al. 2001). Time is represented time 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. Populations for which the estimated number of demographic changes include 0 (i.e., no significant signal of expansion or decline) are represented in light gray and the others in black.

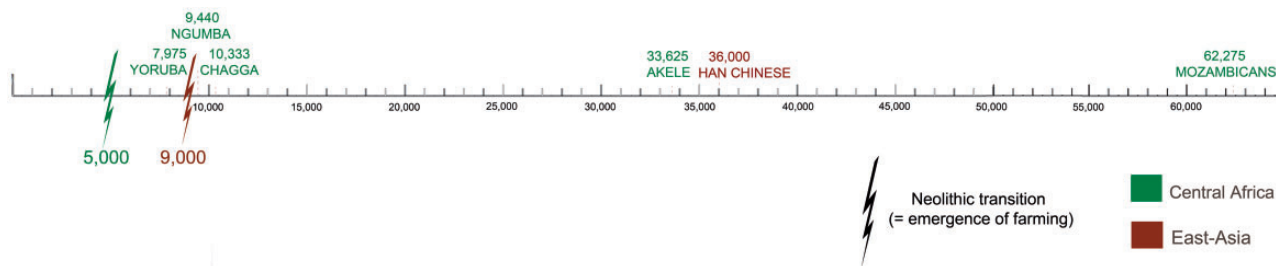


Fig. 3. Comparison of estimated times for expansion onsets using autosomes and dating of the first archeological traces of farming in Africa and China. Time is represented backward (in YBP). Only populations for which the EBS analysis showed a significant expansion event are represented. We reported the time values estimated with the highest mutation rate that we used for the autosomes ($\mu = 2.5 \times 10^{-8}$ /generation/site). Thus, these time values can be considered as a lower bound for the expansion onsets. The dates for the emergence of farming come from the review by Bocquet-Appel and Bar-Yosef (2008). They are based on archeological remains.

For the HVS-I sequences from Eurasia (supplementary tables S7 and S8, Supplementary Material online), the parametric BEAST analyses showed that models consistent with an increase in population size (expansion model or exponential model with positive growth rates) best fitted the data for all sedentary populations except Koreans. Conversely, the constant model best fitted the data for all nomadic populations as well as Koreans. EBSs showed, however, significant expansion events for both farmers and herders, but not for Koreans (fig. 2C and D and supplementary table S5, Supplementary Material online). Nevertheless, there was a tendency toward stronger expansion rates and higher N_e values in sedentary than in nomadic populations (fig. 2C and D), although the 95% HPD intervals for N_e were quite large for sedentary populations. The estimated expansion onset times inferred from the EBSs (supplementary table S6, Supplementary Material online) followed an east-to-west gradient: they appeared more ancient in Eastern populations, in both sedentary and nomadic populations (supplementary fig. S1, Supplementary Material online). They also clearly predated the Neolithic transition in all geographic areas (fig. 4).

The Central Asian Exception: Similar Demographic Patterns in Farmers and Herders

For autosomes, the constant model best fitted the data for both sedentary farmers (TAB) and traditionally nomadic herders (KIB) (supplementary tables S3 and S4, Supplementary Material online). EBSs showed also no significant demographic changes for these populations (figs. 1C and D and supplementary table S5, Supplementary Material online).

For HVS-I (supplementary tables S3 and S7, Supplementary Material online), the exponential model best fitted the data for six of the 12 sedentary farmer populations (including TAB), whereas the constant model was preferred for the other farmers. Unlike the rest of Eurasia, a model indicating expansion (the exponential model with positive growth rates) was also selected for all nomadic herders. Moreover, EBSs showed significant expansion signals for both herder and farmer populations except TJY (Yagnobs from Dushanbe), since at least 13,860 YBP (or 27,720 YBP) for farmers and 16,546 YBP (or 33,092 YBP) for herders, on average (fig. 2E and F and supplementary tables S5 and S6, Supplementary

Material online). Again, these inferred expansion onsets predated the emergence of farming in the area, about 8,000 YBP (Bocquet-Appel and Bar-Yosef 2008) (fig. 4). Inferred expansions for Central Asian sedentary farmers seemed overall weaker (i.e., lower growth rate and lower N_e) than those observed for other sedentary populations in Eurasia, although we observed important variations in growth rates and N_e among populations and large 95% HPD intervals for some of them (fig. 2C and E).

Degrees of Isolation and Migration Patterns

African farmer populations appeared less isolated and received more migrants than hunter-gatherer populations. Indeed, the population-specific F_{ST} values (supplementary table S9, Supplementary Material online) were, on average, significantly lower for farmers than for hunter-gatherers (mean_[farmers] = 0.058; mean_[HG] = 0.192; Wilcoxon two-sided test P value = 0.0002). Moreover, the estimated number of immigrants was significantly higher for sedentary farmers than for nomadic hunter-gatherers (mean_[farmers] = 31.4; mean_[HG] = 2.21; P value = 0.0001) (supplementary table S10, Supplementary Material online). For hunter-gatherers, the F_{ST} values were negatively correlated with the negative growth rates that we inferred from the parametric method ($\rho = -0.893$; P value = 0.012) (fig. 5B), meaning that less isolated populations showed weaker contraction events (i.e., less negative growth rates). Conversely, there was no significant correlation between F_{ST} values and inferred growth rates for sedentary farmers ($\rho = 0.433$; P value = 0.249) (fig. 5A). However, we found a significant positive correlation between the number of immigrants and the inferred growth rates (supplementary fig. S2, Supplementary Material online) among sedentary farmer populations ($\rho = 0.867$; P value = 0.004) but not among nomadic hunter-gatherers ($\rho = 0.536$; P value = 0.235).

For Eurasia, we found no significant difference in F_{ST} values between farmers and herders (mean_[farmers] = 0.039; mean_[herders] = 0.043; P value = 0.77; supplementary table S9, Supplementary Material online), except in Central Asia, for which we found significantly lower F_{ST} values for nomadic herders than for sedentary farmers (mean_[farmers] = 0.018; mean_[herders] = 0.008; P value = 0.017). We report a significant

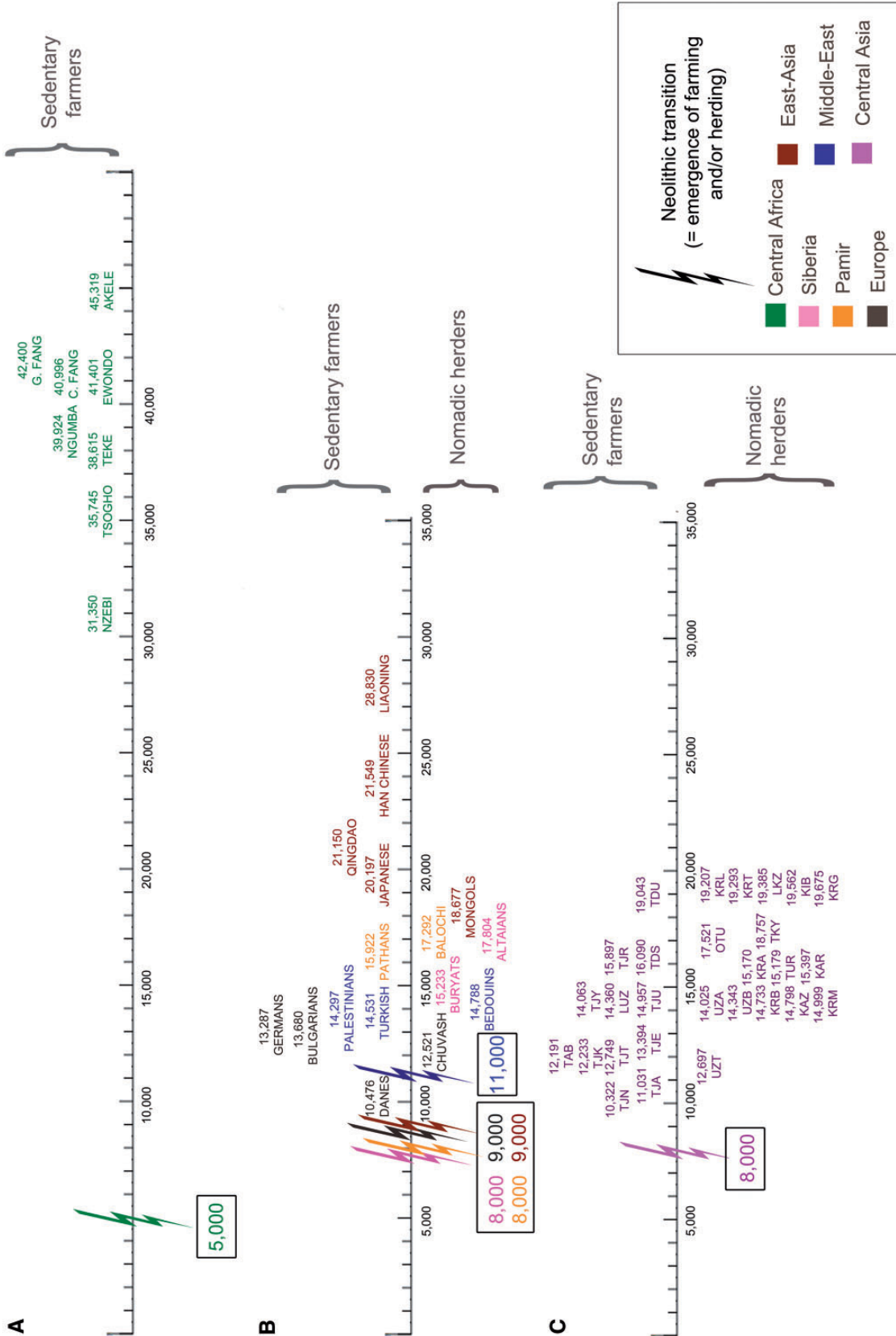


Fig. 4. Comparison of estimated times for expansion onsets using HVS-I and dating of the first archeological traces of farming or herding in Central Africa (A), Eurasia (B), and Central Asia (C). Time is represented backward (in YBP). Only populations for which the EBSP analysis showed a significant expansion event are represented. We reported the time values estimated with the highest mutation rate that we used for the HVS-I sequences ($\mu = 10^{-5}$ /generation/site). Thus, these time values can be considered as a lower bound for the expansion onsets. The dates for the emergence of farming come from the review by Bocquet-Appel and Bar-Yosef (2008). They are based on archeological remains.

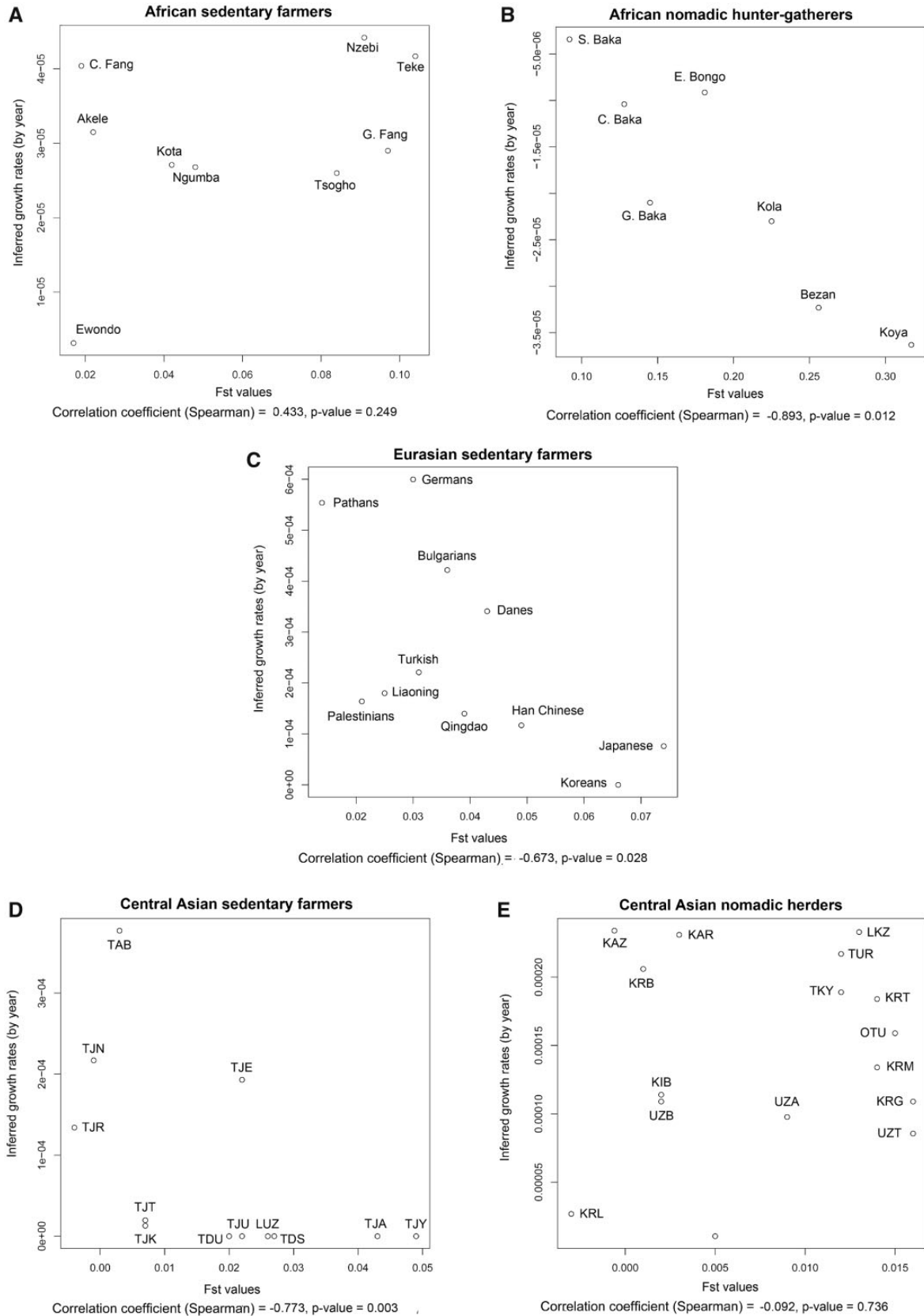


FIG. 5. Correlations between population-specific F_{ST} values and inferred growth rates in African farmer (A) and hunter-gatherer (B) populations, Eurasian farmer populations (C), and Central Asian farmer (D) and herder populations (E). Population-specific F_{ST} values were computed with ARLEQUIN v3.11 (Excoffier et al. 2005). The growth rates were inferred under the best-fitting model from the parametric method using BEAST (Drummond and Rambaut 2007). When the best-fitting model was the constant model, we assumed a growth rate of 0. Note that we did not represent Eurasian herder populations as the constant model best-fitted the data for all of them. Plots and correlation tests were performed using R v2.14.1 (R Development Core Team 2011).

Downloaded from https://academic.oup.com/mbe/article/30/12/2629/1013982 by guest on 19 April 2024

negative correlation between F_{ST} values and inferred growth rates for sedentary farmers in Eurasia ($\rho = -0.673$; P value = 0.028) (fig. 5C) and Central Asia ($\rho = -0.773$; P value = 0.003) (fig. 5D), thus meaning that less isolated populations showed higher inferred growth rates. There was no significant correlation for Central Asian herders ($\rho = -0.092$; P value = 0.736) (fig. 5E). Note that this analysis could not be performed for the other Eurasian herder populations, as the constant model best fitted the data with the parametric method.

The estimation of the proportion of immigrants did not converge for 11 Eurasian populations (Han Chinese, Liaoning, Qingdao, Palestinians, Pathans, Mongols, as well as three Central Asian farmer populations and two Central Asian herder populations). Regarding the other populations, we showed no significant difference in the proportion of immigrants between farmers and herders, both in Central Asia (mean_[farmers] = 82.868; mean_[herders] = 260.265; P value = 0.12) and in the rest of Eurasia (mean_[farmers] = 576,757; mean_[herders] = 201,311; P value = 0.51) (supplementary table S10, Supplementary Material online). We also found no significant correlation between this proportion and the inferred growth rates for Eurasian farmers ($\rho = -0.238$; P value = 0.48), Central Asians farmers ($\rho = 0.386$; P value = 0.30), and Central Asian herders ($\rho = -0.42$; P value = 0.139) (supplementary fig. S2, Supplementary Material online).

Discussion

In this study, using a large set of populations from distant geographic areas, we report contrasted demographic histories that correlate with lifestyle. Moreover, the inferred expansion signals in both African and Eurasian farmer and herder populations predated the Neolithic transition and the sedentarization of these populations.

Contrasted Demographic Histories in Sedentary and Nomadic Populations

For Africa, both mtDNA and autosomal data revealed expansion patterns in most sedentary farmer populations, as indicated by neutrality tests and the parametric and nonparametric BEAST methods. Conversely, we found constant effective population sizes (or possibly contraction events) for all hunter-gatherer populations. Among the farmers, results were least clear for the Yoruba and the Ewondo populations, as no neutrality test was significant for these populations, whereas they showed evidence of expansion events when analyzed with BEAST. This indicates that these populations may have undergone weaker expansion dynamics (i.e., lower growth rates and N_e) than the others. These remarkable results are of particular importance for the Yoruba, as it is a reference population in many databases (HapMap, 1000 genomes). This also demonstrates the higher sensitivity of MCMC methods such as BEAST to detect expansions, in comparison to neutrality tests.

The contrasted patterns inferred between sedentary and nomadic populations in Africa suggest strong differences between the demographic histories of these two groups of populations. The question is whether this pattern results mostly

from differences in local expansion dynamics or whether spatial expansion processes at a larger scale were also involved. As shown by Ray et al. (2003), negative values for the neutrality tests will be observed in a spatial expansion process if the rate of migrants (Nm) is high enough (at least 20), but not otherwise. As in previous studies (e.g., Verdu et al. 2013), we report a higher degree of isolation (higher population-specific F_{ST} values) in hunter-gatherer populations than in farmer populations. Using the spatial expansion model of Excoffier (2004) also leads to higher estimates of the number of immigrants into farmer populations. Thus, both farmers and hunter-gatherers may have been subject to a spatial expansion process, but the limited number of migrants among hunter-gatherers may have resulted in an absence of expansion signals for them. This would be consistent with the positive correlation that we observe between the growth rates estimated with BEAST and the inferred number of immigrants in the sedentary farmer populations. However, this spatial expansion process seems unlikely to completely explain the strong association that we observed between lifestyle and expansion patterns, as some farmer populations (Teke, Gabonese Fang) displayed F_{ST} values similar to those of hunter-gatherers but a clear signal of expansion with relatively high growth rates. This suggests that even rather isolated farmer populations show substantial level of expansions. Moreover, F_{ST} values and inferred growth rates in farmer populations were not significantly correlated. Therefore, our results suggest that the expansion patterns observed in sedentary populations result not only from a spatial expansion pattern. In addition, local dynamics connected with the higher capacity of food production by farmers also explain their much stronger expansion signatures, relative to their neighboring hunter-gatherer populations.

For Eurasia, when considering the mtDNA data, all three methods (neutrality tests, parametric BEAST analyses, and EBSPs) yielded expansion signals for all sedentary farmer populations except Koreans. Conversely, only EBSPs and Fu's F_s test showed expansion signals for nomadic herders, but not the parametric BEAST method nor other neutrality tests. This result points toward weaker expansion dynamics in herders than in farmers, as supported also by the tendency for lower growth rates and N_e in herder populations than in farmer populations on the EBSP graphs (fig. 2C and D). It thus seems that the flexibility and nonparametric nature of EBSP analyses allows one to detect weaker expansion events than the parametric method. Moreover, Fu's F_s is known to be more sensitive than the other neutrality tests to detect expansions (Ramos-Onsins and Rozas 2002). Again, these inferred expansions may result at least in part from spatial expansion processes. The population-specific F_{ST} values are indeed rather low in Eurasia. Moreover, we found a significant negative correlation between F_{ST} values and inferred growth rates for the sedentary farmers, indicating that less isolated populations showed stronger expansion signals. However, although we inferred much stronger expansion patterns for the farmers than for the herders, we did not observe any differences in Eurasia between the farmers and the herders in the population-specific F_{ST} values or in the estimated number of

immigrants, suggesting that spatial processes alone cannot explain the strong difference that we observed between the expansion patterns of these two groups of populations. This indicates that the intrinsic demographic growth patterns are different between these two kinds of populations, the farmers showing much higher growth rates than the herders.

To our knowledge, although other studies have found different patterns between hunter-gatherers and farmers (e.g., Verdu et al. 2009), our study is the first to show differences between farmers and herders, the two major post-Neolithic human groups. A plausible explanation could be that nomadic herders and hunter-gatherers share several of the constraints of a nomadic way of life. For instance, birth intervals are generally longer (at least 4 years) in nomadic populations than in sedentary populations (e.g., Short 1982). According to Bocquet-Appel (2011), these longer birth intervals may be mainly determined by diet differences. Indeed, Valeggia and Ellison (2009) demonstrated that birth interval is mainly determined by the rapidity of postpartum energy recovery, which may be increased by high carbohydrate food (like cereals) consumption. Moreover, the nomadic herder way of life may offer less food security than sedentary farming, the latter facilitating efficient long-term food storage.

However, unlike in Africa, we did not find systematically consistent patterns between the autosomal and mtDNA data in Eurasia. The possible contraction events that our results suggest for two sedentary populations (Japanese and Danes) with autosomes appeared concomitant with historical events that could have led to bottleneck processes. For the Japanese population, this contraction signal could indeed result from a founder effect due to the Paleolithic colonization of Japan by a subset of the Northern Asiatic people (especially from Korea; Nei 1995). Similarly, a bottleneck process may also have occurred in the Danish population, linked with the last glacial maximum occurring between 26,500 and 19,500 YBP (Clark et al. 2009). Reasons why these processes impacted the autosomes but not the mtDNA data remain to be determined, for instance through simulation studies. In any case, our study clearly emphasizes the utility of combining mtDNA and autosomal sequences, as they allow access to different aspects of human history. A recent study on harbor porpoises has similarly shown that nuclear markers were sensitive to a recent contraction event, whereas mtDNA allowed inferring a more ancient expansion (Fontaine et al. 2012).

Interestingly, Central Asia displayed a distinct pattern from the rest of Eurasia. Indeed, we did not infer higher expansion rates for sedentary farmers than for nomadic herders in that area. It could result from harsh local environmental conditions due to the arid continental climate in this area. Indeed, using pollen records, Dirksen and van Geel (2004) showed that the paleoclimate in Central Asia was very arid from at least 12,000 to 3,000 YBP, which could have limited the amount of suitable areas for farming and impacted human demography. Spatial expansion processes may also have played a role in this difference, as population-specific F_{ST} values were higher for the farmers than for the herders. This may indicate that more migrants were involved in the spatial expansion process for the herders than for the farmers,

yielding a weaker expansion signal (i.e., lower inferred growth rate) for the latter (Ray et al. 2003). This is supported by the negative correlation between the F_{ST} values and the inferred growth rates in the farmer populations. The Korean population also stood out as an exception in Eurasia. Even though it is a population of sedentary farmers, it showed no significant expansion signal with both the parametric and nonparametric methods with HVS-I. This could be explained by a later sedentarization of this population. The Korean Neolithic is notably defined by the introduction of Jeulmun ware ceramics about 8,000 YBP, but the people of the Jeulmun period were still predominantly semi-nomadic fishers and hunter-gatherers until about 3,000 YBP, when Koreans started an intensive crop production implying a sedentary lifestyle (Nelson 1993).

Inferred Expansion Signals Predate the Emergence of Farming

EBSP analyses revealed that the inferred expansion events in farmers and herder populations were more ancient than the emergence of farming and herding. Therefore, the differences in demographic patterns between farmers and herders seem to predate their divergence in lifestyle, which raises the question of the chronology of demographic expansions and the Neolithic transition. These findings appear to be quite robust to the choice of the scaling parameters. We used here both the lower and the higher mutation rate estimates in humans for autosomes (Pluzhnikov et al. 2002; Conrad et al. 2011) and for the HVS-I sequence (Forster et al. 1996; Howell et al. 1996). Despite this uncertainty in mutation rates, which lead to a 2-fold uncertainty in our time estimates, the inferred expansion signals predated the emergence of agriculture in both cases for all populations. Similarly, using a generation time of 29 years (Tremblay and Vezina 2000) instead of 25 years lead to slightly more ancient estimates, thus do not change our conclusions (data not shown). However, note that for HVS-I, using the higher bound of the credibility interval for the highest estimated mutation rate (2.75×10^{-5} /generation/site; Heyer et al. 2001) instead of the mean value (i.e., 10^{-5} /generation/site) leads to expansion time estimates consistent with the Neolithic transition in Eurasian populations (supplementary table S11, Supplementary Material online). Nevertheless, these estimates still clearly predated the Neolithic for the African populations. However, 10^{-5} /generation/site is by far the highest estimation of mutation rate in the literature (Howell et al. 1996). To infer Neolithic expansions in most Eurasian populations, one needs to assume a mutation rate of at least 2×10^{-5} /generation/site, much higher than other estimates from the literature, and is thus probably unrealistic. Moreover, our method for determining the expansion onset time using EBSP graph is very conservative and also tends to favor the lower bound of expansion onset times. Finally, for autosomes, using similarly 4.74×10^{-8} /generation/site instead of 2.5×10^{-8} /generation/site (Pluzhnikov et al. 2002) lead to an inferred expansion onset time that is not compatible with the Neolithic transition for all Eurasian and African populations, except for one

African population, the Yoruba (supplementary table S12, Supplementary Material online). Consequently, it seems very likely that the expansions inferred in this study correspond to Paleolithic rather than Neolithic demographic events, in agreement also with most previous studies, as detailed later.

In Africa, the emergence of agriculture has been dated between 5,000 and 4,000 YBP in the Western part of Central Africa and subsequently rapidly expanded to the rest of sub-Saharan Africa (Phillipson 1993). However, using HVS-I, we showed expansion events in farmer populations since about 30,000 or 60,000 YBP, thus largely predating the emergence of agriculture in the area. Similarly, using autosomes, especially in Eastern African populations, we inferred expansion signals that clearly predated the Neolithic. Notably, we inferred an expansion signal for Mozambicans since at least 80,000 YBP. Several genetic studies have already highlighted that expansion events occurred in African farmers before the Neolithic transition (e.g., Atkinson et al. 2009; Laval et al. 2010; Batini et al. 2011). This finding is also consistent with paleoanthropological data (i.e., radiocarbon dating), suggesting an expansion event in Africa 60,000–80,000 YBP (Mellars 2006a). This Paleolithic demographic expansion could be linked to a rapid environmental change toward a dryer climate (Partridge et al. 1997) and/or to the emergence of new hunting technologies (Mellars 2006a).

According to Mellars (2006a), this period corresponds to a major increase in the complexity of the technological, economic, social, and cognitive behavior of certain African groups. It corresponds in particular to the emergence of projectile technologies (Shea 2009), which was probably part of a broader pattern of ecological diversification of early *Homo sapiens* populations. These changes could have been decisive for the human spread “Out of Africa” during the same period and could have ultimately also led to the sedentarization of the remaining populations. This inference is consistent with Sauer’s (1952) demographic theory, which stated that late Paleolithic demographic expansions could have favored the sedentarization and the emergence of agriculture in some human populations. In the case of Central Africa, the period of 60,000 YBP corresponds to the separation between hunter-gatherers and farmers ancestors (Patin et al. 2009; Verdu et al. 2009). Thus, these two groups may have presented contrasting demographic patterns since their divergence. Much later, higher expansion rates and larger population sizes among farmers’ ancestors may have induced the emergence of agriculture and sedentarization.

With respect to Eurasia, the expansion profiles inferred with HVS-I for all populations and with autosomes for the Han Chinese population also seem to have begun during the Paleolithic, thus before the Neolithic transition. Some genetic studies already reported pre-Neolithic expansions in Asia and Europe (e.g., Chaix et al. 2008). Notably, using mismatch and intermatch distributions, Chaix et al. (2008) showed an east-to-west Paleolithic expansion wave in Eurasia. We found a similar pattern here, as the inferred expansions of East-Asian populations were earlier than those of Central Asian populations, themselves earlier than those of European populations.

Moreover, we found this pattern in both sedentary farmer and nomadic herder populations. Thus, the ancestors of currently nomadic herder populations also experienced these Paleolithic expansions. However, Paleolithic expansion signals in nomadic populations seem lower than in sedentary populations. This is again compatible with the demographic theory of the Neolithic sedentarization (Sauer 1952): some populations may have experienced more intense Paleolithic expansions, which may have led ultimately to their sedentarization.

The inferred Paleolithic expansion signals might result partly from spatial expansions out of some refuge areas after the Last Glacial Maximum (LGM, 26,500–19,500 YBP; Clark et al. 2009), as this time interval matches with our inferred dating for expansion onsets in East Asia with HVS-I using the pedigree-based mutation rate and in Europe and Middle East using the transitional mutation rate. Some of the earlier date estimates might also be consistent with the out-of-Africa expansion of *H. sapiens*. However, the time radiocarbon-based estimates of the spread of *H. sapiens* in Eurasia are generally more ancient than our inferred expansion onset timings. For instance, Mellars (2006b) dated the colonization of Middle East by *H. sapiens* at 47,000–49,000 YBP and of Europe at 41,000–42,000 YBP. Pavlov et al. (2001) report traces of modern human occupation nearly 40,000 years old in Siberia. Finally, Liu et al. (2010) described modern human fossils from South China, dated to at least 60,000 YBP. Moreover out-of-Africa or post-LGM expansions would not explain our finding of an east-to-west gradient of expansion onset timing, which rather supports the hypothesis of a demographic expansion diffused from east to west in Eurasia in a demic (i.e., migrations of individuals) or cultural (favored by the diffusion of new technologies).

Possible Confounding Factors

Our approach makes the assumption that populations are isolated and panmictic, which is questionable for human populations. However, we analyzed a large set of populations sampled in very distant geographical regions (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 most of these areas, and it seems unlikely that processes such as admixture could have biased the estimates similarly everywhere. Moreover, in Central Africa, several studies have shown that hunter-gatherer populations show signals of admixture, whereas it is not the case for farmer populations (Patin et al. 2009; Verdu et al. 2009, 2013). If this introgression had been strong enough, this may have yielded a spurious expansion signal in the hunter-gatherer populations, which is not what we observed here. In Europe, spatial expansion processes during the Neolithic may have led to admixture with Paleolithic populations. As pointed out by a simulation study (Arenas et al. 2013), this may lead to a predominance of the Paleolithic gene pool. This may be one of the factors explaining why we observed mostly Paleolithic expansions here.

Similarly, potential selection occurring on the whole mitochondrial genome (e.g., Pakendorf and Stoneking 2005) seems unlikely to have impacted in the same way all the studied populations within each group (e.g., stronger positive selection on sedentary than on nomadic populations), as we analyzed different nomadic and sedentary populations living near each other, in several geographically distant areas.

Regarding the potential effects of recombination on the inferences from autosomal data, we found that neutrality tests gave similar results on the whole sequences, when using a simulation procedure that was taking the known recombination rate of each sequence into account (table 1), and on the largest non-recombining blocks as inferred with IMgc, without taking recombination into account in the simulation process (supplementary table S12, Supplementary Material online). It thus appears unlikely that the BEAST analyses that can only handle the largest inferred non-recombining blocks are biased because of this.

Finally, note that the effective population sizes inferred using BEAST correspond to the N_e of the populations during their recent history, rather than a value of N_e averaged over the history of the population. It explains the finding that, for most populations, we inferred N_e estimates much higher than generally assumed for humans by population geneticists (about 10,000).

Material and Methods

Genetic Markers

Autosomal Sequences

We used data from 20 noncoding, a priori neutral, and unlinked autosomal regions selected by Patin et al. (2009) to be at least 200 kb away from any known or predicted gene, to not be in linkage disequilibrium (LD) neither with each other nor with any known or predicted gene, and to have a region of homology with the chimpanzee genome. These regions are on average 1,253 bp long. Using the four-gamete test (Hudson and Kaplan 1985) as implemented in IMgc online (Woerner et al. 2007), we identified recombination events for 6 of these 20 regions. As some methods used in this study cannot handle recombination, we retained for these six sequences the largest non-recombining block inferred by IMgc. Because of this reduction, the 20 regions used were on average 1,228 bp long. To identify potential bias related to this method (e.g., some recombination events may not be detected using the four-gamete test; larger blocks of non-recombining sequence may select for gene trees that are shorter than expected), we computed the summary statistics and performed neutrality tests (discussed later) both on the whole sequences (table 1) and on the largest non-recombining blocks (supplementary table S12, Supplementary Material online).

Mitochondrial Sequences

We used the first hypervariable segment of the mitochondrial control region (HVS-I), sequenced between positions 16067 and 16383, excluding the hypervariable poly-C region (sites 16179–16195). The total length of the sequence was thus of 300 bp.

Population Panel

For Africa, we used the autosomal sequences data set of Patin et al. (2009), which consists of five farmer populations ($N = 118$ individuals) and five Pygmy hunter-gatherer populations ($N = 95$). In addition, we used the HVS-I data set from Quintana-Murci et al. (2008), which consists of nine Central African farmer populations ($N = 486$) and seven Central African hunter-gatherer populations ($N = 318$) (supplementary table S1, Supplementary Material online).

For Eurasia, we used the autosomal sequences data set of Laval et al. (2010), consisting of 48 individuals from two East-Asian populations (Han Chinese and Japanese) and 47 individuals from two European populations (Chuvash and Danes). We also used the data from 48 individuals from one sedentary Central Asian population (Tajik farmers) and 48 individuals from one nomadic Central Asian population (Kyrgyz herders) of Ségurel et al. (2013). For HVS-I, we analyzed data from 17 Eurasian populations ($N = 494$ in total) located from Eastern to Western Eurasia, belonging to several published data sets (Derenko et al. 2000; Richards et al. 2000; Bermisheva et al. 2002; Imaizumi et al. 2002; Yao et al. 2002; Kong et al. 2003; Quintana-Murci et al. 2004; supplementary table S1, Supplementary Material online).

For our detailed study of Central Asia, we used HVS-I sequences from 12 farmer populations ($N = 408$ in total) and 16 herder populations ($N = 567$ in total). These data come from the studies by Chaix et al. (2007) and Heyer et al. (2009) for 25 populations (supplementary table S1, Supplementary Material online). The other populations (KIB, TAB, and TKY) were sequenced for this study. As in Chaix et al. (2007) and Heyer et al. (2009), DNA was extracted from blood samples using standard protocols, and the sequence quality was ensured as follows: each base pair was determined once with a forward and once with a reverse primer; any ambiguous base call was checked by additional and independent PCR and sequencing reactions; all sequences were examined by two independent investigators. All sampled individuals were healthy donors from whom informed consent was obtained. The study was approved by appropriate Ethic Committees and scientific organizations in all countries where samples have been collected.

Demographic Inferences from Sequences Analysis

Summary Statistics and Neutrality Tests

We computed classical summary statistics (number of polymorphic sites S , number of haplotypes K) and four neutrality tests (Tajima's [1989] D , Fu and Li's [1993] D^* and F^* , and Fu's [1997] F_s) on both mitochondrial and autosomal sequences. Although neutrality tests were originally designed to detect selective events, they also give information about demographic processes, especially when applied to neutral markers. Indeed, expansion events lead to more negative values than expected in the absence of selective and demographic processes. Conversely, contraction events lead to more positive values of the neutrality tests. For HVS-I sequences, we computed all summary statistics and neutrality tests and tested

their departure from neutrality using the coalescent-based tests provided in DnaSP (Librado and Rozas 2009).

For the autosomal sequences, we used the procedure developed in Laval et al. (2010), which combines all autosomal sequences into a single test. This procedure consists in computing the mean value of each summary statistics across the 20 loci and in testing whether this mean value departs significantly from its expectation under neutrality in a constant-size population model using a simulation procedure. For a given population with sample size n , we produced 10^5 simulated samples of the same size n , under a constant population size model, using the generation-per-generation coalescent-based algorithm implemented in SIMCOAL v2 (Laval and Excoffier 2004). Each simulated individual was constituted by 20 independent sequences of 1,253 bp (the average sequence length for the real data). Then, we used ARLEQUIN v3 (Excoffier et al. 2005) as modified by Laval et al. (2010) to compute the summary statistics on these simulated samples. We assessed whether the observed statistics differed significantly from the constant population model under neutrality by comparing these statistics with their null distribution obtained from the simulated data. We used gamma distributed mutation rates with a mean value of 2.5×10^{-8} /generation/site (95% confidence interval: $1.476 \times 10^{-8} - 4.036 \times 10^{-8}$), in agreement with previous studies (Pluzhnikov et al. 2002; Voight et al. 2005). This procedure yielded a P value for the significance of the departure from a constant size model. We performed this procedure both on the whole sequences and on the largest non-recombining blocks. For the whole sequences (i.e., including recombination), we performed the data simulation under a coalescent model with recombination, using for each locus the recombination rate provided by the HapMap build GRCh37 genetic map (International HapMap Consortium 2003) (supplementary table S8, Supplementary Material online), whereas for the largest non-recombining blocks we used a coalescent model without recombination.

Both for autosomes and for HVS-I sequences, we adjusted the obtained P values for each neutrality test using a false discovery rate (FDR) correction (Benjamini and Hochberg 1995) in R v2.14.1 (R Development Core Team 2011), in order to take into account the increased error probability in the case of multiple testing.

MCMC Estimations of Demographic Parameters

We used the MCMC algorithm implemented in BEAST v1.6 (Drummond and Rambaut 2007). We tested the four demographic models implemented in this software: constant effective population size (N_0) (constant model), population expansion with an increasing growth rate (g) (exponential model), population expansion with an decreasing growth rate (g) (logistic model), and the expansion model, in which N_0 is the present day population size, N_1 the population size that the model asymptotes to going into the distant past, and g the exponential growth rate that determines how fast the transition is from near the N_1 population size to N_0 population size. In fact, BEAST estimates composite parameters for each model, namely $N_0\mu$ and g/μ , where N_0 is the current

effective population size, g the growth rate, and μ the mutation rate. In addition, for the expansion model, the ratio between the current (N_0) and ancestral (N_1) effective population size is also estimated. To infer N_0 and g from these composite parameters, we needed to assume a value for the mutation rate μ . However, there is no consensus for mutation rates in humans in the literature, as different methods lead to different estimations. For autosomes, the most commonly used value is the phylogenetic rate of $\mu = 2.5 \times 10^{-8}$ /generation/site (Pluzhnikov et al. 2002). However, recent studies based on the 1,000 genome project (1000 Genomes Project Consortium 2012) have found a 2-fold lower rate ($\mu = 1.2 \times 10^{-8}$ /generation/site) by directly comparing genome-wide sequences from children and their parents (Conrad et al. 2011; Scally and Durbin 2012). We used here both mutation rates. Similarly, for HVS-I, estimated mutation rates are highly dependent on methodologies and modes of calibration (Endicott et al. 2009). We used both the lower and the higher estimated mutation rates: the transitional changes mutation rate of $\mu = 5 \times 10^{-6}$ /generation/site (Forster et al. 1996) and the pedigree-based rate of $\mu = 10^{-5}$ /generation/site (Howell et al. 1996; Heyer et al. 2001). We used a general time-reversible substitution model (Rodriguez et al. 1990). We assumed a generation time of 25 years, permitting the comparison with previous human population genetics studies (e.g., Chaix et al. 2008; Patin et al. 2009; Laval et al. 2010). As BEAST cannot handle recombination events, we used the largest nonrecombining block within each sequence (discussed earlier).

We performed three runs of 10^7 steps per population and per demographic model for the HVS-I sequence and three runs of 2×10^8 steps (which corresponded to three runs of 10^7 steps per locus) for the autosomal sequences. We recorded one tree every 1,000 steps, which thus implied a total of 10^5 trees per locus and per run. We then removed the first 10% steps of each run (burn-in period) and combined the runs to obtain acceptable effective sample sizes (ESSs of 100 or above). The convergence of these runs was assessed using two methods: visual inspection of traces using Tracer v1.5 (Rambaut and Drummond 2007) to check for concordance between runs and computation of Gelman and Rubin's (1992) convergence diagnostic using R v2.14.1 (R Development Core Team 2011) with the function *gelman.diag* available in the package *coda* (Plummer et al. 2006).

To facilitate a large exploration of the parameter space, for the autosomal sequences, we chose uniform priors between 0 and 0.05 for $2N_0\mu$ and between -10^9 and 10^9 for g/μ . For HVS-I sequences, we chose uniform priors for $N_0\mu$ between 0 and 10 and for g/μ between -2.5×10^6 and 2.5×10^6 , resulting for the same priors on N_0 and g than for autosomal sequences if we assumed $\mu = 10^{-5}$ /generation/site (i.e., N_0 constrained between 0 and 10^6 and g constrained between -1 and 1 per year). Conversely, if we assumed $\mu = 5 \times 10^{-6}$ /generation/site, it meant that N_0 was constrained between 0 and 2×10^6 and g was constrained between -0.5 and 0.5 per year.

For each population and model, we obtained the mode and the 95% HPD of N_0 and g inferred from their

posterior distributions (supplementary tables S4 and S7, Supplementary Material online) using the add-on package Locfit (Loader 1999) in R v2.14.1. We selected the best-fitting model among the four tested demographic models by estimating marginal likelihoods using two methods: path sampling and stepping-stone sampling (Baele et al. 2012). The model with the greater marginal likelihood (supplementary table S3, Supplementary Material online) was considered as the best-fitting model.

Extended Bayesian Skyline Plots

EBSPs (Heled and Drummond 2010), also implemented in BEAST, estimate demographic changes occurring continuously through time in a population, using the time intervals between successive coalescent events. This method allows a visualization of the evolution of N_e through time. As above, we combined three runs of 10^7 steps for mitochondrial sequences and three runs of 2×10^8 steps for autosomal sequences to obtain acceptable ESS values. We assumed the same mutation rates as above and a generation time of 25 years. Outputs were analyzed with Tracer v1.5 to visually check for convergence and ESS, also to obtain the 95% HPD interval for the number of demographic changes that occurred in the population (supplementary table S5, Supplementary Material online). A constant population size could be rejected when the 95% HPD of the number of change points excluded 0 but included 1 (Heled and Drummond 2010). Then, we used R v2.14.1 to compute Gelman and Rubin's (1992) convergence diagnostic as above, as well as to compute skyline plots. Finally, we used the population growth curves generated from BEAST to assess the time at which populations began to expand. Each Skyline plot consisted of smoothed data points at ≈ 10 –20 generation intervals. We consider that the population increased (or decreased) when both the median and 95% HPD values for N_e increased (or decreased) between more than two successive data points. Although this method did not provide a 95% HPD interval for the inferred expansion timings, this conservative approach ensured that we considered only relevant expansion signals.

Correlation Tests of Inferred Growth Rates and Isolation/Immigration Patterns

To test how isolation degrees and migration patterns differences could have impacted our demographic inferences, we used ARLEQUIN v3.11 (Excoffier et al. 2005) to compute population-specific F_{ST} values (Weir and Hill 2002) from HVS-I data for Central Africa, Eurasia, and Central Asia (supplementary table S9, Supplementary Material online) and estimate immigration rates from mismatch distributions under a spatially explicit model (Excoffier 2004) (supplementary table S10, Supplementary Material online). We performed then Spearman tests using R v2.14.1 to investigate for each region how the inferred parametric growth rates were correlated with those F_{ST} values and immigration rates. We used a value of 0 for the growth rate when the constant model best fitted the data.

Supplementary Material

Supplementary figures S1 and S2 and tables S1–S12 are available at *Molecular Biology and Evolution* online (<http://www.mbe.oxfordjournals.org/>).

Acknowledgments

The authors would like to warmly thank all volunteer participants. They also thank Michael Fontaine for his help on some data analyses, Phillip Endicott and Julio Bendezu-Sarmiento for insightful discussions, Alexei Drummond for helpful discussions, and Friso Palstra for his help on English usage. They thank Laurent Excoffier, two anonymous reviewers, and the editor for helpful comments and suggestions. All computationally intensive analyses were run on the Linux cluster of the Museum National d'Histoire Naturelle (administrated by Julio Pedraza) and on the web-based portal "Bioportal" (Kumar et al. 2009). This work was supported by Actions Transversales du Muséum - Muséum National d'Histoire Naturelle (grant "Les relations Sociétés-Natures dans le long terme") and Agence Nationale de la Recherche (grants "Altérité culturelle" ANR-10-ESVS-0010 and "Demochips" ANR-12-BSV7-0012). C.A. was financed by a PhD grant from the Centre National de la Recherche Scientifique.

References

- 1000 Genomes Project Consortium. 2012. An integrated map of genetic variation from 1,092 human genomes. *Nature* 491:56–65.
- Ambrose SH. 2001. Paleolithic technology and human evolution. *Science* 291:1748–1753.
- Arenas M, François O, Currat M, Ray N, Excoffier L. 2013. Influence of admixture and Paleolithic range contractions on current European diversity gradients. *Mol Biol Evol.* 30: 57–61.
- Atkinson QD, Gray RD, Drummond AJ. 2009. Bayesian coalescent inference of major human mitochondrial DNA haplogroup expansions in Africa. *Proc R Soc Lond B Biol Sci.* 276:367–373.
- Baele G, Lemey P, Bedford T, Rambaut A, Suchard MA, Alekseyenko AV. 2012. Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. *Mol Biol Evol.* 29:2157–2167.
- Balaresque PL, Ballereau SJ, Jobling MA. 2007. Challenges in human genetic diversity: demographic history and adaptation. *Hum Mol Genet.* 16:R134–R139.
- 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.
- Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate—a practical and powerful approach to multiple testing. *J R Stat Soc Ser B.* 57:289–300.
- Bermisheva M, Salimova A, Korshunova T, Svyatova G, Berezina G, Vilems R, Khusnutdinova E. 2002. Mitochondrial DNA diversity in the populations of Middle Asia and Northern Caucasus. *Eur J Hum Genet.* 10:179–180.
- Bocquet-Appel JP. 2011. When the world's population took off: the springboard of the Neolithic Demographic Transition. *Science* 333: 560–561.
- Bocquet-Appel J-P, Bar-Yosef O. 2008. The Neolithic demographic transition and its consequences. Dordrecht (The Netherlands): Springer.

- 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, 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.
- Clark PU, Dyke AS, Shakun JD, Carlson AE, Clark J, Wohlfarth B, Mitrovica JX, Hostetler SW, McCabe AM. 2009. The Last Glacial Maximum. *Science* 325:710–714.
- Conrad DF, Keebler JEM, DePristo MA, et al. (17 co-authors). 2011. Variation in genome-wide mutation rates within and between human families. *Nat Genet.* 43:712–714.
- Derenko MV, Malyarchuk BA, Dambueva IK, Shaikhaev GO, Dorzhu CM, Nimaev DD, Zakharov IA. 2000. Mitochondrial DNA variation in two South Siberian aboriginal populations: implications for the genetic history of North Asia. *Hum Biol.* 72:945–973.
- Dirksen VG, van Geel B. 2004. Mid to late Holocene climate change and its influence on cultural development in South Central Siberia. *Nato Sci S Ss IV Ear.* 42:291–307.
- Drummond AJ, Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol.* 7:214.
- Endicott P, Ho SYW, Metspalu M, Stringer C. 2009. Evaluating the mitochondrial timescale of human evolution. *Trends Ecol Evol.* 24: 515–521.
- 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. *Nat Rev Genet.* 7:745–758.
- Excoffier L, Laval G, Schneider S. 2005. Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evol Bioinform Online.* 1:47–50.
- Fontaine MC, Snirc A, Frantzis A, Koutrakis E, Ozturk B, Ozturk AA, Austerlitz F. 2012. History of expansion and anthropogenic collapse in a top marine predator of the Black Sea estimated from genetic data. *Proc Natl Acad Sci U S A.* 109:E2569–E2576.
- Forster P, Harding R, Torroni A, Bandelt HJ. 1996. Origin and evolution of native American mtDNA variation: a reappraisal. *Am J Hum Genet.* 59:935–945.
- Fu YX. 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* 147: 915–925.
- Fu YX, Li WH. 1993. Statistical tests of neutrality of mutations. *Genetics* 133:693–709.
- Gelman A, Rubin DB. 1992. Inference from iterative simulation using multiple sequences. *Stat Sci.* 7:457–511.
- Heled J, Drummond AJ. 2010. Bayesian inference of species trees from multilocus data. *Mol Biol Evol.* 27:570–580.
- Heyer E, Balaresque P, Jobling MA, Quintana-Murci L, Chaix R, Segurel 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, Zietkiewicz E, Rochowski A, Yotova V, Puymirat J, Labuda D. 2001. Phylogenetic and familial estimates of mitochondrial substitution rates: study of control region mutations in deep-rooting pedigrees. *Am J Hum Genet.* 63:1113–1126.
- Ho SYW, Shapiro B. 2011. Skyline-plot methods for estimating demographic history from nucleotide sequences. *Mol Ecol Res.* 11: 423–434.
- Howell N, Kubacka I, Mackey DA. 1996. How rapidly does the human mitochondrial genome evolve? *Am J Hum Genet.* 59:501–509.
- Hudson RR, Kaplan NL. 1985. Statistical properties of the number of recombination events in the history of a sample of DNA sequences. *Genetics* 111:147–164.
- Imaizumi K, Parsons TJ, Yoshino M, Holland MM. 2002. A new database of mitochondrial DNA hypervariable regions I and II sequences from 162 Japanese individuals. *Int J Legal Med.* 116:68–73.
- International HapMap Consortium. 2003. The International HapMap Project. *Nature* 426:789–796.
- Kingman JFC. 1982. The coalescent. *Stoch Proc Appl.* 13:235–248.
- Kong QP, Yao YG, Sun C, Bandelt HJ, Zhu CL, Zhang YP. 2003. Phylogeny of East Asian mitochondrial DNA lineages inferred from complete sequences. *Am J Hum Genet.* 73:671–676.
- Kumar V, Langstieh BT, Madhavi KV, Naidu VM, Singh HP, Biswas S, Thangaraj K, Singh L, Reddy BM. 2006. Global patterns in human mitochondrial DNA and Y-chromosome variation caused by spatial instability of the local cultural processes. *PLoS Genet.* 2:e53.
- Kumar S, Skjaeveland A, Orr RJS, Enger P, Ruden T, Mevik BH, Burki F, Botnen A, Shalchian-Tabrizi K. 2009. AIR: a batch-oriented web program package for construction of supermatrices ready for phylogenomic analyses. *BMC Bioinformatics* 10:357.
- Laval G, Excoffier L. 2004. SIMCOAL 2.0: a program to simulate genomic diversity over large recombining regions in a subdivided population with a complex history. *Bioinformatics* 20:2485–2487.
- Laval G, Patin E, Barreiro LB, Quintana-Murci L. 2010. Formulating a historical and demographic model of recent human evolution based on resequencing data from noncoding regions. *PLoS One* 5: e10284.
- Librado P, Rozas J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25:1451–1452.
- Liu W, Jin CZ, Zhang YQ, et al. (13 co-authors). 2010. Human remains from Zhirendong, South China, and modern human emergence in East Asia. *Proc Natl Acad Sci U S A.* 107:19201–19206.
- Loader C. 1999. Local regression and likelihood. New York: Springer.
- Mellars P. 2006a. Why did human populations disperse from Africa ca. 60,000 years ago? A new model. *Proc Natl Acad Sci U S A.* 103: 9381–9386.
- Mellars P. 2006b. A new radiocarbon revolution and the dispersal of modern humans in Eurasia. *Nature* 439:931–935.
- Nei M. 1995. The origins of human populations: genetic, linguistic, and archeological data. In: Brenner S, Hanihara K, editors. The origin and past of modern humans as viewed from DNA. Singapore: World Scientific. p. 71–91.
- Nei M, Roychoudhury AK. 1993. Evolutionary relationships of human populations on a global scale. *Mol Biol Evol.* 10:927–943.
- Nelson SM. 1993. The archaeology of Korea. Cambridge: Cambridge University Press. p. 307.
- Oota H, Settheetham-Ishida W, Tiwawech D, Ishida T, Stoneking M. 2001. Human mtDNA and Y-chromosome variation is correlated with matrilineal versus patrilineal residence. *Nat Genet.* 29: 20–21.
- Pakendorf B, Stoneking M. 2005. Mitochondrial DNA and human evolution. *Annu Rev Genomics Hum Genet.* 6:165–183.
- Partridge TC, Demenocal PB, Lorentz SA, Paiker MJ, Vogel JC. 1997. Orbital forcing of climate over South Africa: a 200,000-year rainfall record from the Pretoria Saltpan. *Quat Sci Rev.* 16:1125–1133.
- Patin E, Laval G, Barreiro LB, et al. (15 co-authors). 2009. Inferring the demographic history of African farmers and pygmy hunter-gatherers using a multilocus resequencing data set. *PLoS Genet.* 5: e1000448.
- Pavlov P, Svendsen J, Indrelid SI. 2001. Human presence in the European Arctic nearly 40,000 years ago. *Nature* 413:64–67.
- Phillipson DW. 1993. African archaeology. Cambridge: Cambridge University Press.
- Plummer M, Best N, Cowles K, Vines K. 2006. CODA: Convergence diagnosis and output analysis for MCMC. *R. News* 6:7–11.
- Pluzhnikov A, Di Rienzo A, Hudson RR. 2002. Inferences about human demography based on multilocus analyses of noncoding sequences. *Genetics* 161:1209–1218.
- Quintana-Murci L, Chaix R, Wells RS, et al. (17 co-authors). 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, Quach H, Harmant C, et al. (23 co-authors). 2008. Maternal traces of deep common ancestry and asymmetric gene flow between Pygmy hunter-gatherers and Bantu-speaking farmers. *Proc Natl Acad Sci U S A.* 105:1596–1601.

- R Development Core Team. 2011. R: a language and environment for statistical computing. Vienna (Austria): R Foundation for Statistical Computing.
- Rambaut A, Drummond AJ. 2007. Tracer, ver. 1.5. [cited 2013 Sep 24]. Available from: <http://tree.bio.ed.ac.uk/software/tracer/>
- Ramos-Onsins SE, Rozas J. 2002. Statistical properties of new neutrality tests against population growth. *Mol Biol Evol.* 19:2092–2100.
- Ray N, Currat M, Excoffier L. 2003. Intra-deme molecular diversity in spatially expanding populations. *Mol Biol Evol.* 20:76–86.
- Richards M, Macaulay V, Hickey E, et al. (37 co-authors). 2000. Tracing European founder lineages in the Near Eastern mtDNA pool. *Am J Hum Genet.* 67:1251–1276.
- Rodriguez F, Oliver JL, Marin A, Medina JR. 1990. The general stochastic model of nucleotide substitution. *J Theor Biol.* 142:485–501.
- Sauer CO. 1952. Agricultural origins and dispersals. Cambridge (MA): American Geographical Society.
- Scally A, Durbin R. 2012. Revising the human mutation rate: implications for understanding human evolution. *Nat Rev Genet.* 13:745–53.
- Ségurel L, Austerlitz F, Toupance B, et al. (16 co-authors). 2013. 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.
- Shea JJ. 2009. The impact of projectile weaponry on Late Pleistocene hominin evolution. In: Hublin J-J, Richards MP, editors. Evolution of hominin diets. Dordrecht (The Netherlands): Springer. p. 189–199.
- Short R. 1982. The biological basis for the contraceptive effects of breast feeding. *Int J Gynecol Obstet.* 25:207–217.
- Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123:585–595.
- 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.
- Valeggia C, Ellison PT. 2009. Interactions between metabolic and reproductive functions in the resumption of postpartum fecundity. *Am J Hum Biol.* 21:559–566.
- Verdu P, Austerlitz F, Estoup A, et al. (14 co-authors). 2009. Origins and genetic diversity of pygmy hunter-gatherers from Western Central Africa. *Curr Biol.* 19:312–318.
- Verdu P, Becker NS, Froment A, et al. (12 co-authors). 2013. Sociocultural behavior, sex-biased admixture, and effective population sizes in Central African Pygmies and non-Pygmies. *Mol Biol Evol.* 30:918–937.
- Voight BF, Adams AM, Frisse LA, Qian YD, Hudson RR, Di Rienzo A. 2005. Interrogating multiple aspects of variation in a full resequencing data set to infer human population size changes. *Proc Natl Acad Sci U S A.* 102:18508–18513.
- Weir BS, Hill WG. 2002. Estimating F-statistics. *Ann Rev Genet.* 36:721–750.
- Woerner AE, Cox MP, Hammer MF. 2007. Recombination-filtered genomic datasets by information maximization. *Bioinformatics* 23:1851–1853.
- Yao YG, Kong QP, Bandelt HJ, Kivisild T, Zhang YP. 2002. Phylogeographic differentiation of mitochondrial DNA in Han Chinese. *Am J Hum Genet.* 70:635–651.