MtDNA analysis of global populations support that major population expansions began before Neolithic Time

Hong-Xiang Zheng[1], Shi Yan[1,2,3], Zhen-Dong Qin[1] & Li Jin[1,2,3]

[1]MOE Key Laboratory of Contemporary Anthropology and Center for Evolutionary Biology, School of Life Sciences and Institutes of Biomedical Sciences, Fudan University, [2]Chinese Academy of Sciences and Max-Planck Society (CAS-MPG) Partner Institute for Computational Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China, [3]Key Laboratory of Computational Biology, CAS-MPG Partner Institute for Computational Biology, Chinese Academy of Sciences, Shanghai 200031, China.

Agriculture resulted in extensive population growths and human activities. However, whether major human expansions started after Neolithic Time still remained controversial. With the benefit of 1000 Genome Project, we were able to analyze a total of 910 samples from 11 populations in Africa, Europe and Americas. From these random samples, we identified the expansion lineages and reconstructed the historical demographic variations. In all the three continents, we found that most major lineage expansions (11 out of 15 star lineages in Africa, all autochthonous lineages in Europe and America) coalesced before the first appearance of agriculture. Furthermore, major population expansions were estimated after Last Glacial Maximum but before Neolithic Time, also corresponding to the result of major lineage expansions. Considering results in current and previous study, global mtDNA evidence showed that rising temperature after Last Glacial Maximum offered amiable environments and might be the most important factor for prehistorical human expansions.

Agriculture in modern society appeared first in the Fertile Crescent of West Asia about 11–12 thousand years ago (kya)[1–3]. During the subsequent several thousand years until ~4.5 kya[1], agriculture was developed independently in central China, West Africa, New Guinea highlands, Mesoamerica, central Andes, and eastern part of North America. From these origin homelands, farming was spread to the remaining of the world for its overwhelming advantage in food production compared to hunting and foraging, which was the main subsistence mode of human before Holocene[1–3]. The advent of agriculture, which demarcated the beginning of the Neolithic Time, revolutionarily impacted on the formation of modern society and shaped the distribution of modern human populations and language families[1]. As the primary consequences of the agriculture, Neolithic expansions were numerous, such as Bantu expansion in Africa (4−2 kya)[4], farmer influx into Europe (~10 kya)[5], Lapita expansion in Oceania (~5 kya)[6] and Northern Han Chinese expansion (~5−2 kya)[7,8]. The demographic growth during Neolithic Time was considered as population explosions, even continuing unabated to nowadays[9,10]. Lines of evidence in linguistics[1,9], molecular anthropology[11] and archaeology[2] supported rapid demographic, geographic and cultural expansions after the invention of agriculture. Thus, hypothesis was put forward that major population expansions began after the advent of agriculture, i.e. the Neolithic Time. To test such a hypothesis in the framework of population genetics requires a large-scale and random sampling strategy without ascertainment bias, so that major expansion lineages could be detected, the ages of the expansion lineages could be accurately estimated and compared with the dating of the beginning of agriculture.

Several studies on population expansions in worldwide populations were analyzed by mitochondrial DNA (mtDNA) variants. Atkinson et al. conducted a global Bayesian analysis on eight regions (Sub-Sahara Africa, Middle East, South Asia, Europe, North Asia, Australia and Americas), and found out that the main phase of prehistorical human population growth were approximately before 10 kya[12]. Gignoux et al. investigated global Neolithic expansions in three regions (Africa, Europe and Southeast Asia) by analyzing mitochondrial lineages associated with or without agriculture, and found some lineages associated with expansion in Holocene[11]. Unfortunately, these studies were based on limited but not randomly sampled individuals with whole mtDNA sequences then available by using the analytical methods in which random samples are required.

Samples in the 1000 Genome Project[13] were collected randomly without a priori strategy, therefore, provided an opportunity for investigating a large number of whole sequences of human mtDNA. Several tens of populations in Africa, Europe, East Asia and Americas were sequenced, far more than those in previous studies. Based on binary sequences alignment map (BAM) files, whole mtDNA sequences of high quality could be assembled and generated. Recently, using the mtDNA sequence data of East Asians and found that major lineage expansion and population expansion in East Asian began before the time that agriculture became a major food source, i.e. the advent of Neolithic Time[14]. We hypothesized that the rising temperature after Last Glacial Maximum (LGM) might have contributed to the population growth and the population expansion subsequently constituted a need for the introduction of agriculture. Furthermore, we speculated that the continuous growth of population size was likely one of the driving forces that led to the further development of agriculture and turned agriculture from a supplementary food source to a major one. With the sequence data from 1000 Genome Project, in this study, we extended the analysis to worldwide populations to examine whether the global patterns of population expansions were similar to East Asians.
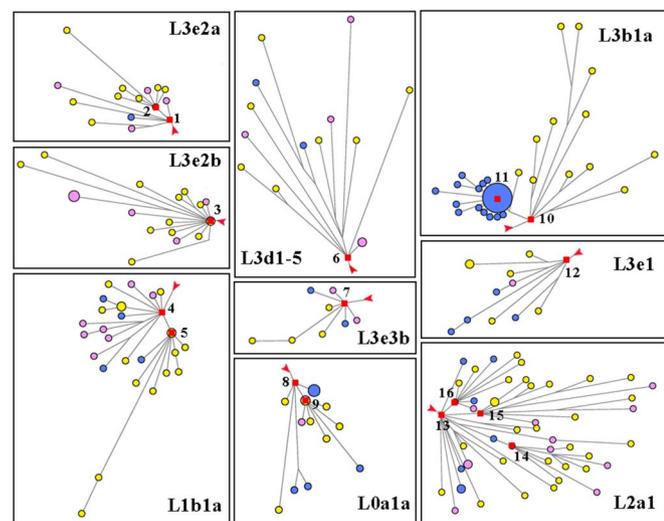
Results

Africans. Although agriculture developed independently in western part of Africa[1], Neolithic transition appeared in North Africa at the beginning of Holocene from the Middle East, and marked with the emergence of agriculture in the lower Nile Valley ~7 kya[15].

In the 1000 Genome Project, 313 African samples from 4 populations were collected, most of which (97.9%) were from Macrohaplogroup L excluding M and N under L3 (Table S1), and were also confirmed as autochthonous in African. Detailed information for the populations was annotated in Methods. Besides Macrohaplogroup L, we could find some Native American components which belong to A2, C1 and D1 (3 individuals in ASW), while there were also low frequent M32 which was common in Southeast of Africa (1 individual in ASW) and U6 which had the North African origin (1 individual in ASW and 1 in YRI). From the median-joining network

constructed by 313 African samples, 16 expansion lineages were identified, not including the very old expansion L3 lineage (see Figure 1). Nearly all the lineages with star-like structure were shared by at least 2 populations except for L3b1a1 that is LWK-specific, indicating that most of these expansions might have occurred in the African ancestral populations before the divergence of these populations and these star lineages were a representative of the African maternal evolution. Among the 16 expansion haplogroups, 5 lineages (L0a1a2, L2a1f, L3b1a1, L3e2a1b, and L3e3b) showed coalescence time less than 10 kya at least by three of five estimates mentioned in Methods, while the remaining 11 lineages (L0a1a, L1b1a3, L1b1a, L2a1a, L2a1c, L2a1, L3b1a, L3e1, L3e2a, L3e2b and L3d1–5) expanded before 10 kya (see Table S2). Specially, the expansion of lineage L3d1–5 took place before the LGM. From the ages estimated above, we found that most lineages shared among populations (11/15) expanded before 10 kya, i.e. the first occurrence of farming in the land of Africa. Thus, the result of lineage expansions showed that maternal African growth could be mainly attributed to pre-Neolithic expansion.

After removing the M and N lineages, we constructed Bayesian Skyline Plots for the 3 populations respectively and jointly, to describe the historical maternal effective variation trends. From the Bayesian Skyline Plots (BSP) (Figure 2) constructed for each population of African ancestry, two populations (ASW and YRI) showed pre-Neolithic population expansions. ASW showed a distinct trend of growth about 20 kya and extend to 5 kya, while YRI also showed a pre-Neolithic expansion after LGM about 15 kya. From the African BSP (Figure 3B), all the African random samples also showed a 5-fold growth at ~15−11 kya, corresponding to expansion haplogroups L0a1a, L1b1a, L1b1a3, L2a1a, L3b1a, L3e1, L3e2a and L3e2b, and subsequently a 2-fold growth ~5−4kya, which might be driven by the Neolithic Revolution. The time of some expansion linages estimated were compared to previous studies[16–18], showing little difference. In a recent published paper, African L3 lineage was also proved to have a growth peak before 10 kya, which was identical to our observation[18].
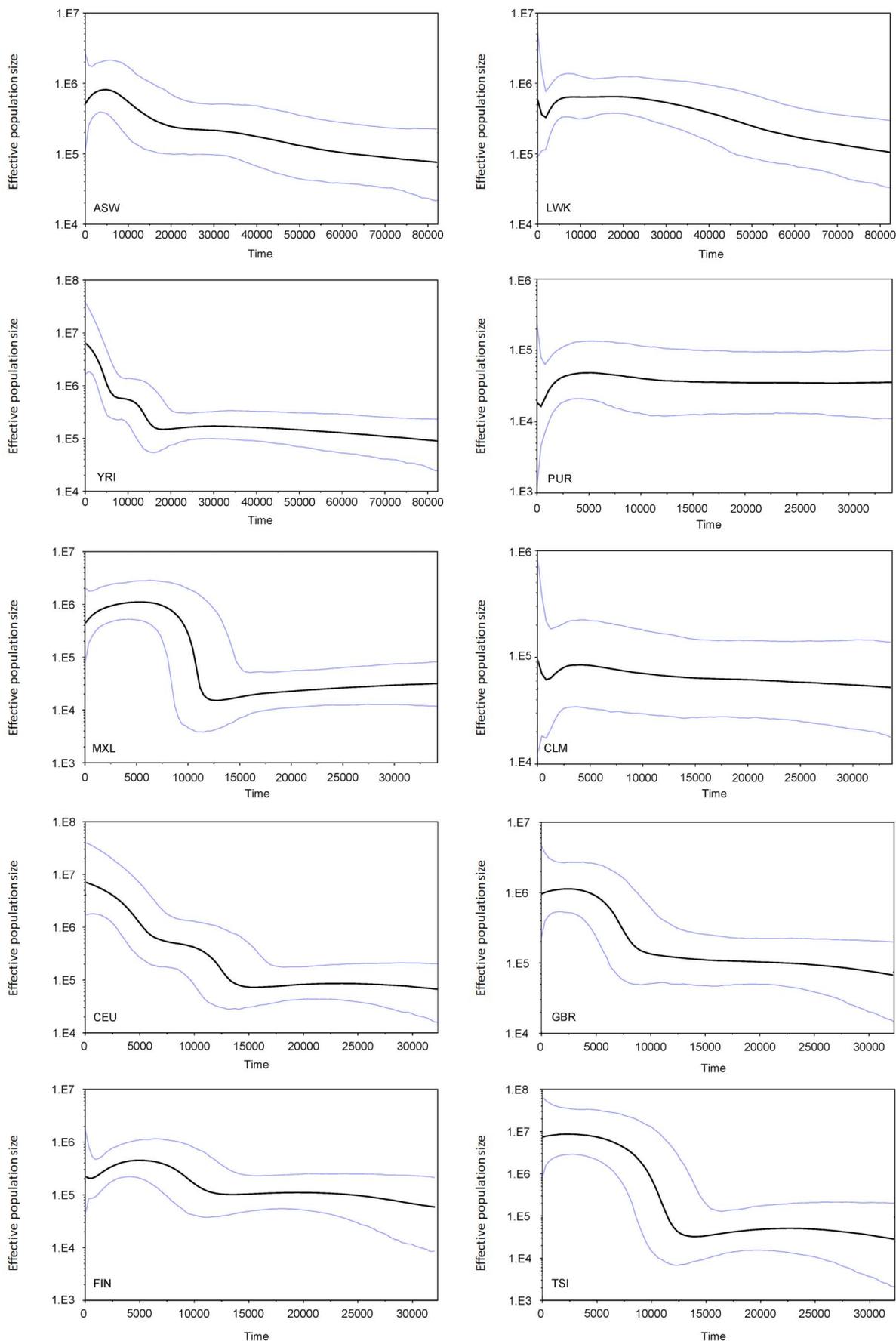
To summarize, both lineage expansions and population expansions in Africa suggested major pre-Neolithic expansion(s).

Europeans. The Neolithic transition in Europe has been debated for decades[5,19,20]. Agriculture in Europe was not developed independently, but was brought from farmer influx in Middle East. In the Eastern Europe, agriculture appeared in Greece about 9 kya[15], which could be considered as the first farming in Europe. It is still controversial whether the farmers replaced the majority of the original Paleolithic European residents or they only had limited contribution to the gene pool of modern Europeans when bringing agriculture to the land of Europe. According to previous mtDNA evidence which are mainly based on the analysis of hypervariable regions, the farmer influx only account for a very little proportion, about 20%[5]. The 1000 Genome Project provides 413 Europeans (103 CEU, 97 FIN, 94 GBR, 14 IBS and 105 TSI) in this study. As shown in Table S2, most of the European samples belonged to the Macrohaplogroup N, with 2 exceptions found in TSI (1 L1 and 1 D4). HV accounted for about a half of the gene pool, and frequencies of lineage U and JT were the next highest.

According to the median-joining network analysis, 15 star lineages were observed in Figure 4. Most of them (HV, H, H1, H3, J1c, T1, T2, U5a1, U5a, K1, V, W, U2'3'4'7'8'9) coalesced before 10 kya although 2 lineages (J1c3 and T2b) might expand in 10 kya. Except HV and U2'3'4'7'8'9, other lineages expanded about after LGM. A very distinct and major expansion in Figure 4 is the H lineage, and subsequent expansions of haplogroups H1 and H3 were also important in Europe. About 44.5% of European samples in current analysis belonged to the H expansion, which happened right after LGM according to our calculation (Table S3).


Figure 1 | African expansion lineages from median-joining network. Star expansions were selected from median-joining network of 313 African mtDNA coding region sequences corresponding to rCRS positions 577–16023. Blue, LWK; purple, ASW; yellow, YRI. Red squares and numbers indicate clusters with distinct expansions. 1, L3e2a; 2, L3e2a1b; 3, L3e2b; 4, L1b1a; 5, L1b1a3; 6, L3d1–5; 7, L3e3b; 8, L0a1a; 9, L0a1a2; 10, L3b1a; 11, L3b1a1; 12, L3e1; 13, L2a1; 14, L2a1a; 15, L2a1c; 16, L2a1f. Branch length is proportional to steps of mutations and red arrows indicate the root.

Figure 2 | mtDNA Bayesian skyline plot showing the size trend of 10 global populations. The y-axis is the product of maternal effective size and generation time. The x-axis is the time from present in units of years. The thick solid line is the median estimate and the thin lines (blue) show the 95% highest posterior density limits. Detailed settings refer to Methods.
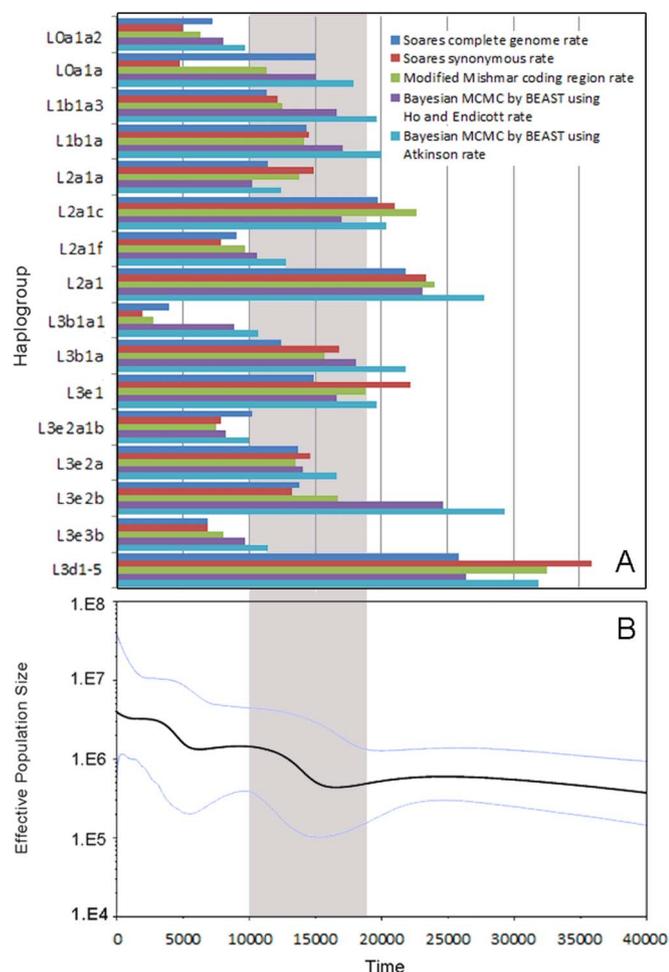
Figure 3 | Time estimation of each observed expansion in 3 African populations and Bayesian skyline plot for 3 African populations together. (A)Time estimation of each observed expansion in 3 African populations: ρ statistic-based method with Soares complete genome rate, Soares synonymous rate and modified Mishmar coding region rate, Bayesian MCMC method by BEAST using Ho and Endicott rate, and Atkinson rate respectively. (B) mtDNA Bayesian skyline plot showing size trend of the 3 African populations together, detailed settings refer to Methods. The grey area shows the time after LGM and before Neolithic Time.

To further verify that the lineage expansions indeed occurred in Europe, we extended the analysis to the populations in Middle East. In the network of Middle East which were mainly based on the data from Schönberg et al.[21] and whole mtDNA sequence data on Pakistan and Israel individuals from CEPH-HGDP (HX Zheng, unpublished data), we found 13 expansions (HV, H, I, L2a1, M4'67, M, N, R, T2b, T2, U2'3'4'7'8'9, U7 and X2) in Middle East, five of which were identical to European expansion lineages (HV, H, T2b, T2, U2'3'4'7'8'9). HV and U2'3'4'7'8'9 were too old for the discussion in this context.

In the following, we focused on the analysis of the relatively younger lineages, including Haplogroup H. The H lineage in the Middle East was estimated ~15 kya, which was younger than European H (~18 kya). Although haplogroup H was thought to have a Middle East origin, previous work also supported that it expanded in Europe[22]. In Europe, H expanded 18–16 kya, which is definitely in Paleolithic Time. In addition, high frequency of H was observed in many European populations, almost about 40% or more[23–26]. Thus, the expansion of H lineage contributed greatly to current European gene pool. Another young lineage, T2b, was ~12−10 kya in Middle East, which is older than T2b in Europe
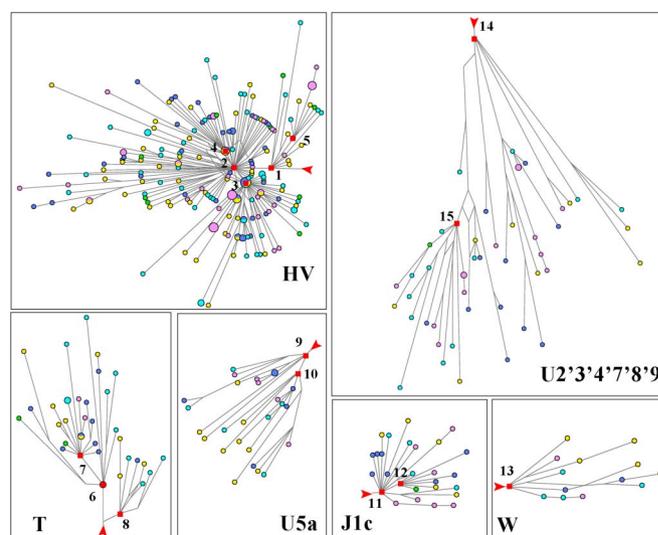

Figure 4 | European expansion lineages from median-joining network. Star expansions were selected from median-joining network of 413 European mtDNA coding region sequences corresponding to rCRS positions 577–16023. Blue, GBR; purple, FIN; yellow, CEU; Cyan, TSI; Green, IBS. Red squares and numbers indicate clusters with distinct expansions. 1, HV; 2, H; 3, H1; 4, H3; 5, V; 6, T2; 7, T2b; 8, T1; 9, U5a; 10, U5a1; 11, J1c; 12, J1c3; 13, W; 14, U2'3'4'7'8'9; 15, K1. Branch length is proportional to steps of mutations and red arrows indicate the root.

(10−9 kya), and T2b was previously suggested a Middle Eastern origin[26]. Although T1 and T2 were previously thought to associate with agriculture development[11], we did not find any expansion in T1 lineage while T2 lineage coalesced at 19 kya, which is much earlier than farming time, which was also concordant to a recent study considering T might in fact reflect dispersal from Near Eastern refugia in Post-LGM period[27]. Furthermore, H1, H3, and V were considered to expand northwards from the Southwestern European refuge right after LGM in former analysis[22,28]. Compared to the estimations of European lineages, our estimation might be lower in some lineages but still before the agriculture occurred in Europe[16,29]. Other expansions K1, W and J1c, were also ambiguous for their origins[26]. However, their ages (> 14 kya) indicated that these lineages had little chances for playing a role in agriculture transmission.

To conclude, H, H1, H3, J1c, K1, U5a, U5a1, V and W represent pre-Neolithic expansions, of which V, H1, H3, U5a and U5a1 were definitely autochthonous in Europe, indicating that main lineages in Europe began to expand before the agriculture while none of the lineages were found to expand in Europe after Neolithic Time.

From the European BSP plots (Figure 2), 3 populations (CEU, FIN and TSI) were found that they began distinct growth ~14−12 kya, which were concordant to previous analysis in Middle East[21]. As expected, from the BSP plot including all European samples in Macrohaplogroup N (Figure 5B), we found that the expansion began ~ 13 kya and showed a continuous trend to nowadays, which is very similar to the result of BSP analysis of East Asian[14]. The population expansion began ~ 13 kya might correlate to lineages H1, H3, J1c, J1c3, K1, T1, U5a1, V and W.

Americans. For Native Americans, we observed six expansions of lineages, of which four (A2, B2, C1 and D1) were shared by different populations. The four lineages were also the main constitutions and founding lineages in American gene pool[30–32]. The remaining 2 lineages, B2d and A2w were CLM-specific (see Figure 6). As expected, we found some African L lineages and European N lineages which might be admixed from recent contacts with immigrants. For example, in MXL, African component (U6) was
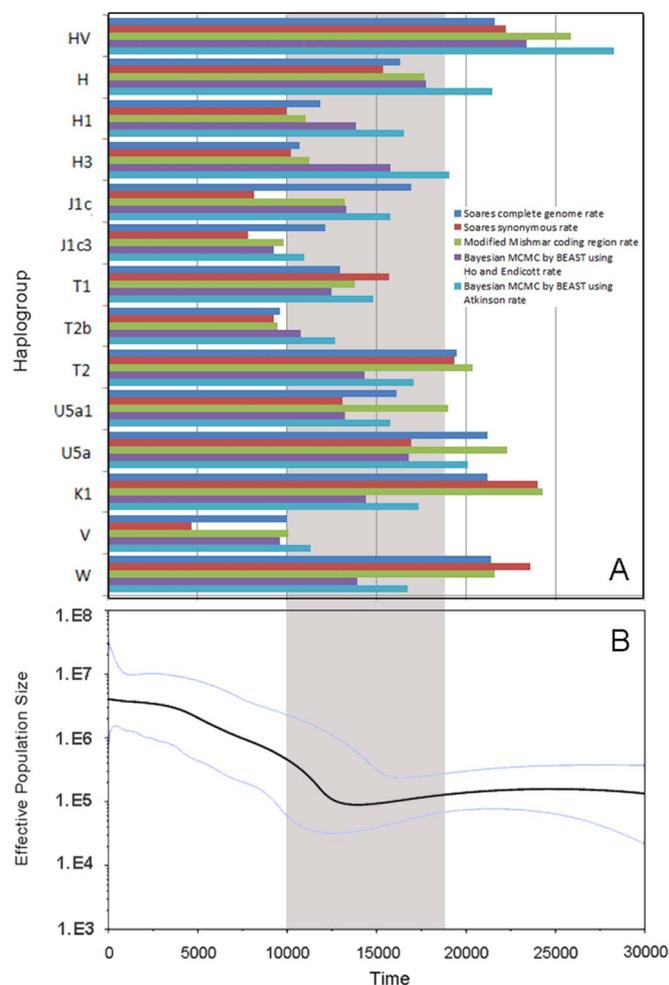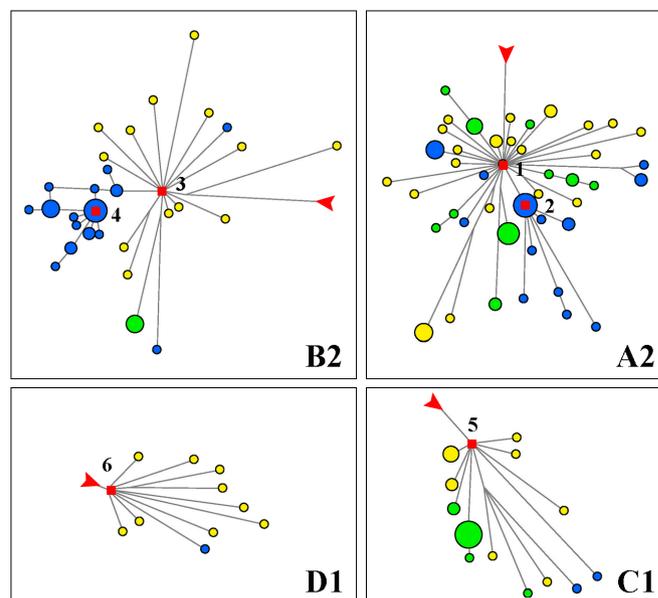
Figure 5 | Time estimation of each observed expansion in 5 European populations and Bayesian skyline plot for 5 European populations together. (A)Time estimation of each observed expansion in 5 European populations. (B) mtDNA Bayesian skyline plot showing size trend of the 5 European populations together, detailed setting refers to Methods. The grey area shows the time after LGM and before Neolithic Time.


Figure 6 | American expansion lineages from median-joining network. Star expansions were selected from median-joining network of 184 Ameican mtDNA coding region sequences corresponding to rCRS positions 577–16023. Blue, CLM; yellow, MXL; Green, PUR. Red squares and numbers indicate clusters with distinct expansions. 1, A2; 2, A2w; 3, B2; 4, B2d; 5, C1; 6, D1. Branch length is proportional to steps of mutations and red arrows indicate the root.

about 3% and European component (H, V and W) was about 12% (Table S4), similar to the recent analysis in random Mexican sample (3.1% and 13.6% respectively)[33]. To analyze the lineages autochthonous to the New World, we focused on the classical Native American haplogroups A2, B2, C1, D1 and D4h3, of which the former 4 haplogroups showed star-like patterns. Time estimates were generated according to different methods and rates (Table S2), and the ages of 4 main clusters (A2, B2, C1 and D1) were between the LGM and 13 kya. According to the previous model, these lineages expanded rightly after the LGM via a coastal (Pacific) route from Northern refuge (Beringia) towards the south. The dispersal to the whole America continent was accomplished in a very short time, probably in just several thousand years[31,32,34–37].

The BSP plots (Figure 7B) including all Native American samples also showed a huge expansion about 100 folds at 12 kya, which is virtually identical to the former analysis[38]. In addition, all BSPs of Americans (Figure 2 and 7B) showed recent bottlenecks, which might be the impact of European contact[39].

In Americas, agriculture originated independently in central Mexico and Northern part of South America about 5−4 kya[2], while some researchers thought that the earliest agriculture could be traced to Valdivia Valley in Chile ~ 6.4 kya[15]. Whenever the first farming occurred, the expansions in America seemed have occurred much earlier than the first appearance of agriculture.

Discussion

This study showed that major population expansions in 3 continents began before Neolithic Time, i.e., 15−11 kya in Africa, 13 kya till now in Europe and 12−8 kya in America. All the expansions began at post-LGM as the temperature started to rise, i.e. before Neolithic time and the advent of agriculture. Considering the mtDNA evidences from Africa, Europe, Americas (current analysis), East Asia[14], South Asia[40], Southeast Asia[41], North Africa[42], and Middle East[21], we proposed that the post-LGM mild climate constituted an important factor for maternal expansion before Neolithic Time, and the increase of population size was likely one of the driving forces that led to the advent of agriculture. Climate change and technology development were believed to have played major roles on the archaic human demography, such as dispersals, expansions and bottlenecks[10,43]. LGM was the last extensively cold and arid period to modern human beings, when most of the human retreated to warmer regions in lower latitude. After LGM, the temperature rose and human beings re-occupied the remaining of the planet and flourished again. Rising temperature no longer confined human beings to limited regions and offered great opportunity for geographic expansions. Furthermore, mild climate not only benefited to the hunters and gatherers for more abundant food source, but also for farmers for crop cultivation, offering chances for demographic expansions. Thus, it is not surprising that the rising temperature after LGM resulted in the commencement of modern human major expansion. Although many former studies pointed out the importance of climate factor in human mtDNA evolution, this is the first global analysis that a large-scale random sample was used to ascertain the expansion lineages and construct historical demographic variations.

The star-like phylogeny is always interpreted as a signal of rapid population expansions[18,42]. Simulation results showed that under rapid population growth, most of the coalescent events occurred at
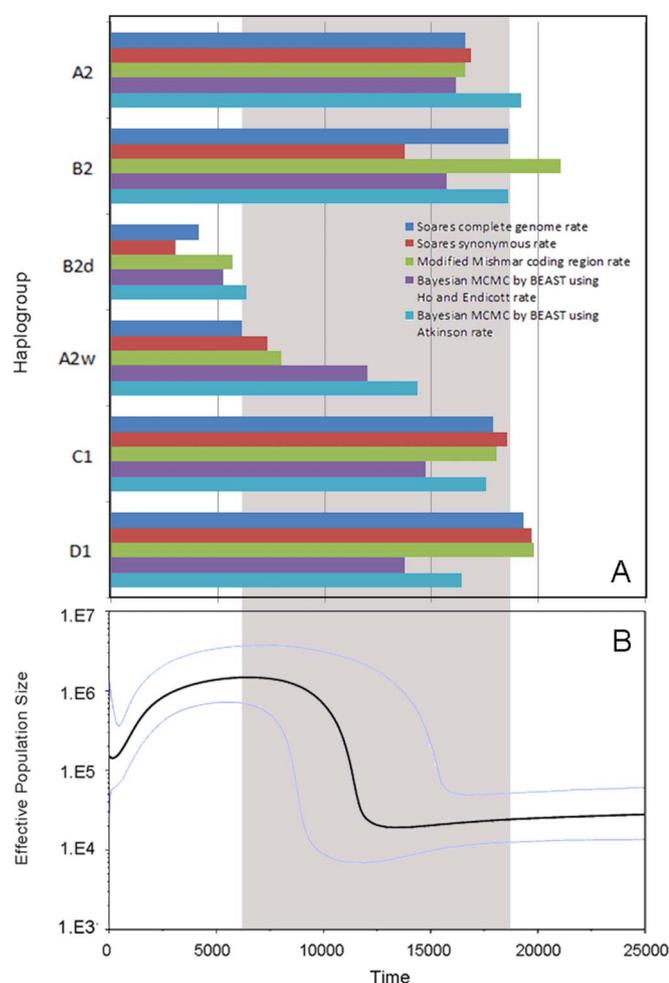
Figure 7 | Time estimation of each observed expansion in 3 American populations and Bayesian skyline plot for 3 American populations together. (A)Time estimation of each observed expansion in 3 American populations. (B) mtDNA Bayesian skyline plot showing size trend of the 3 American populations together, detailed settings refer to Methods. The grey area shows the time after LGM and before Neolithic Time.

about the same time, forming the star lineage and corresponding to the time of major expansion[44]. Furthermore, the BSPs were constructed by all lineages (including star and non-star lineages) in a population with random data and reconstructed the general variation of population size. Thus, the correspondence between the coalescence age of most star lineages and the growth peak of BSPs showed the major population expansion time. Furthermore, the accuracy of time estimation on star-like haplogroups is critical to this study. Considering the fact that different approaches and rates could lead to varied results, we adopted a comprehensive strategy by comparing two different methods of time estimation, i.e. the method based on ρ statistics, and the Bayesian MCMC method. In addition, a total of 5 rates were also employed, including rates on the whole mtDNA genome, on the coding region only, or on synonymous site. We estimated and judged the time of specific linages considering the majority of methods used. Specifically, the BSP were constructed according to a relatively high rate 2.038×10^{-8} subs/site/year, making our results more reliable because higher rate would result in lower time estimates and the time of population expansion still predated agriculture. In this study, the coalescence time of each lineage estimated by the two aforementioned methods and five different rates showed some discrepancies, which might be caused by natural selection and random drift in different lineages, mutation rate heterogeneity among different

mtDNA regions, or different internal calibration points. However, the discrepancies were not that substantial and did not affect our conclusions. To confirm the age estimation of a lineage, we compared the results with the published literature[16–18,29] and showed little difference.

This study showed that lineage expansions and population expansions in 3 continents began before Neolithic Time. In Africa, 11 lineages (L0a1a, L1b1a3, L1b1a, L2a1a, L2a1c, L2a1, L3b1a, L3e1, L3e2a, L3e2b and L3d1–5) out of 15 star lineages shared by different populations were estimated to coalesce above 10 kya, and African samples also showed a 5-fold growth ~15−11 kya, while agriculture in Africa emerged ~7 kya. In Europe, all the autochthonous expansion lineages (H1, H3, U5a, U5a1, V) were older than 10 kya, and Europe witnessed a major population expansion from ~13 kya to nowadays, while the appearance of farming in Europe were after 10 kya. In Americas, the ages of 4 founding and expansion lineages (A2, B2, C1 and D1) were older than 13 kya. The American population also showed a demographic leap 12−8 kya. When the different regions entered the Neolithic Time in ~11−6 kya, agriculture offered the possibility of further population growth. Considering results in current and previous study, global mtDNA evidence showed that rising temperature after Last Glacial Maximum offered amiable environments and might be the most important factor for prehistorical human expansions.

Methods

Populations and samples. Three African, five European and three American populations sequenced in the 1000 Genome Project were included in the current analysis. For African populations, Southwest African individuals (ASW) are those of African ancestry residing in the southwest of the United States; Yoruba individuals were from Ibadan in Nigeria (YRI); Luhya individuals (LWK) were from Webuye in Kenya. For European populations, European Caucasians (CEU) were residents with northern and western European ancestry collected in Utah, USA; Finnish individuals (FIN) were from Finland; British individuals (GBR) were from England and Scotland; Tuscan individuals (TSI) were collected in a small town near Florence in the Tuscany region of Italy; Iberian Populations in Spain (IBS) were collected throughout the Spanish territory. For Native American populations, Mexican individuals (MXL) were from Los Angeles, California; Colombian individuals (CLM) were gathered in the Medellín, Colombia, metropolitan area; Puerto Ricans (PUR) were collected throughout Puerto Rico. More detailed population information could be found in the homepage of 1000 Genome Project[13](http://www.1000genomes.org). All mtDNA sequences in this analysis are maternally unrelated.

Whole mtDNA sequence assembly. The binary sequence alignment/map (BAM) files of mtDNA genomes in this study were obtained from NCBI ftp site (ftp://ftp.ncbi.nlm.nih.gov/1000genomes/). The duplicate reads were removed by MarkDuplicates, implemented in Picard v1.36 (http://picard.sourceforge.net) and the mtDNA sequences were locally realigned by GATK v1.2.59[45]. Pileup files were generated by SAMtools v1.0.16[46]. Consensus sequences were then obtained based on the pileup files and indels were checked manually afterwards. Variations for haploid and missing site were called according to the criteria used before[14]. Finally, we obtained sequences of 910 samples, of which 313 Africans (61 ASW, 116 LWK, and 136 YRI), 413 Europeans (103 CEU, 97 FIN, 94 GBR, 14 IBS and 105 TSI), and 184 Native Americans (62 CLM, 67 MXL, and 55 PUR). The average ambiguous sites were 0.54, and the average coverage of these 910 bams was 1269× and the minimum was 6.7×. All the variations to rCRS were attached as supplemental material (Table S5).

Haplogroup assignment. Complete sequences were aligned to rCRS by MUSCLE v3.8.31[47] and manually checked, then assigned to the haplogroups according to Phylotree.org Build 12[48]. As in Phylotree, positions 309.1C(C), 16182C, 16183C, 16193.1C(C), and 16519 were not used for haplogroup assignment since these were subject to highly recurrent mutations.

Data analysis. The median-joining network of complete mtDNA was constructed by Network v4.6[49] using the coding region (577–16023) in each continent. Each star cluster was identified with the pattern that 5 or more branches splitted out from one internal node, which was also considered as a distinct expansion. Then, to test the assumption of a molecular clock, a maximum likelihood phylogenetic tree was also reconstructed for the coding region using PhyML v3.0[50] under the HKY+G mutation model with an α parameter of 0.12[51]. In all the three continents, the null hypothesis of a molecular clock cannot be rejected (P > 0.05) using PAML package v4.4[52].

The coalescence time of each distinct expansion was estimated using ρ statistic-based method and Bayesian MCMC method. For ρ statistic-based method, standard deviation was calculated following Saillard et al.[53]. Then the time to TMRCA of each expansion was estimated using Soares rate for synonymous mutations, for complete mitochondrial genomes (all the substitutions excluding the 16519 mutation and the 16182C, 16183C, and 16194C)[16], and a corrected rate of Mishmar's rate for coding

6

regions respectively[34]. For Bayesian MCMC analysis, the time of each distinct expansion was estimated using BEAST v1.6.1[54]. Each MCMC sample of each cluster with distinct expansion was based on a run of 40 million generations sampled every 1,000 steps with the first 4 million generations regarded as burn-in. For African and European data, we combined 3 independent runs together for adequate effective sample size (>200). We used the HKY+G model of nucleotide substitution without partitioning the coding region. A strict clock was used and prior substitution rate was assumed to be normally distributed, with a mean of 2.038×10^{-8} subs/site/year and an SD of 2.064×10^{-9} subs/site/year[38]. To confirm our result, another rate 1.691×10^{-8} subs/site/year calibrated with Q lineage in New Guinea was also employed[12]. Each run was subsequently analyzed using Tracer v1.5.1.

Bayesian skyline plots for each population and each continent together were also generated by BEAST v1.6.1 and Tracer v1.5.1, using the similar settings as above and allowing 10 discrete changes (for each individual population and Americans) and 30 discrete changes (for Africans and Europeans) in the population history regarding that population size grows or declines linearly between changing points. Φ_{ST} distances between populations in current study or previous analysis[14,21,33,55–61] were calculated in Arlequin 3.11 also via coding regions, and plotted in PAST 1.85[62] with a non-metric multidimensional scaling method (see Figure S1), showing that populations in each continent were clustered together.

1. Diamond, J. & Bellwood, P. Farmers and their languages: The first expansions. *Science* **300**, 597–603 (2003).
2. Bellwood, P. & Oxenham, M. The expansions of farming societies and the role of the Neolithic Demographic Transition. In *The Neolithic Demographic Transition and its consequences.* J.-P. Bocquet-Appel & O. Bar-Yosef, eds.,13 34, (Springer Netherlands, 2008).
3. Gupta, A. K. Origin of agriculture and domestication of plants and animals linked to early Holocene climate amelioration. *Curr Sci India* **87**, 54–59 (2004).
4. Salas, A. *et al.* The making of the African mtDNA landscape. *Am. J. Hum. Genet.* **71**, 1082–1111 (2002).
5. Richards, M. *et al.* Tracing European founder lineages in the near eastern mtDNA pool. *Am. J. Hum. Genet.* **67**, 1251–1276 (2000).
6. Gray, R. D., Drummond, A. J. & Greenhill, S. J. Language phylogenies reveal expansion pulses and pauses in Pacific settlement. *Science* **323**, 479–483 (2009).
7. Su, B. *et al.* Y chromosome haplotypes reveal prehistorical migrations to the Himalayas. *Hum Genet* **107**, 582–590 (2000).
8. Wen, B. *et al.* Genetic evidence supports demic diffusion of Han culture. *Nature* **431**, 302–305 (2004).
9. Diamond, J. Evolution, consequences and future of plant and animal domestication. *Nature* **418**, 700–707 (2002).
10. Jobling, M., Hurles, M. & Tyler-Smith, C. *Human evolutionary genetics: Origins, peoples and disease.* (Garland Publishing, 2003).
11. Gignoux, C. R., Henn, B. M. & Mountain, J. L. Rapid, global demographic expansions after the origins of agriculture. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 6044–6049 (2011).
12. Atkinson, Q. D., Gray, R. D. & Drummond, A. J. MtDNA variation predicts population size in humans and reveals a major southern Asian chapter in human prehistory. *Mol. Biol. Evol.* **25**, 468–474 (2008).
13. The 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061–1073 (2010).
14. Zheng, H. X. *et al.* Major population expansion of East Asians began before Neolithic Time: Evidence of mtDNA genomes. *PLoS One* **6**, e25835 (2011).
15. Bandy, M. Global patterns of early village development. In *The Neolithic Demographic Transition and its consequences.* J.-P. Bocquet-Appel and O. Bar-Yosef, eds. 333–357, (Springer Netherlands, 2008).
16. Soares, P. *et al.* Correcting for purifying selection: An improved human mitochondrial molecular clock. *Am. J. Hum. Genet.* **84**, 740–759 (2009).
17. Behar, D. M. *et al.* The dawn of human matrilineal diversity. *Am. J. Hum. Genet.* **82**, 1130–1140 (2008).
18. Soares, P. *et al.* The expansion of mtDNA haplogroup L3 within and out of Africa. *Mol Biol Evol* **25**, 915–917 (2011).
19. Richards, M. B., Macaulay, V. A., Bandelt, H. J. & Sykes, B. C. Phylogeography of mitochondrial DNA in western Europe. *Ann Hum Genet* **62**, 241–260 (1998).
20. Richards, M. The neolithic invasion of Europe. *Annu Rev Anthropol* **32**, 135–162 (2003).
21. Schönberg, A., Theunert, C., Li, M., Stoneking, M. & Nasidze, I. High-throughput sequencing of complete human mtDNA genomes from the Caucasus and West Asia: high diversity and demographic inferences. *Eur J Hum Genet* **19**, 988–994 (2011).
22. Pereira, L. *et al.* High-resolution mtDNA evidence for the late-glacial resettlement of Europe from an Iberian refugium. *Genome Res* **15**, 19–24 (2005).
23. Torroni, A. *et al.* mtDNA analysis reveals a major late Paleolithic population expansion from southwestern to northeastern Europe. *Am J Hum Genet* **62**, 1137–1152 (1998).
24. Achilli, A. *et al.* The molecular dissection of mtDNA haplogroup H confirms that the Franco-Cantabrian glacial refuge was a major source for the European gene pool. *Am. J. Hum. Genet.* **75**, 910–918 (2004).
25. Roostalu, U. *et al.* Origin and expansion of haplogroup H, the dominant human mitochondrial DNA lineage in West Eurasia: The near eastern and Caucasian perspective. *Mol. Biol. Evol.* **24**, 436–448 (2007).
26. Soares, P. *et al.* The archaeogenetics of Europe. *Curr Biol* **20**, R174–183 (2010).
27. Pala, M. *et al.* Mitochondrial DNA signals of late glacial recolonization of europe from near eastern refugia. *Am J Hum Genet* **90**, 915–924 (2012).
28. Torroni, A., Achilli, A., Macaulay, V., Richards, M. & Bandelt, H. J. Harvesting the fruit of the human mtDNA tree. *Trends Genet* **22**, 339–345 (2006).
29. Pierron, D. *et al.* Mutation rate switch inside Eurasian mitochondrial haplogroups: Impact of selection and consequences for dating settlement in Europe. *PLoS One* **6**, e21543 (2011).
30. Torroni, A. *et al.* Asian affinities and continental radiation of the four founding Native American mtDNAs. *Am J Hum Genet* **53**, 563–590 (1993).
31. Fagundes, N. J. R. *et al.* Mitochondrial population genomics supports a single pre-Clovis origin with a coastal route for the peopling of the Americas. *Am. J. Hum. Genet.* **82**, 583–592 (2008).
32. Perego, U. A. *et al.* The initial peopling of the Americas: A growing number of founding mitochondrial genomes from Beringia. *Genome Res.* **20**, 1174–1179 (2010).
33. Kumar, S. *et al.* Large scale mitochondrial sequencing in Mexican Americans suggests a reappraisal of Native American origins. *Bmc Evol Biol* **11**, 293 (2011).
34. Perego, U. A. *et al.* Distinctive Paleo-Indian migration routes from Beringia marked by two rare mtDNA haplogroups. *Curr Biol* **19**, 1–8 (2009).
35. Achilli, A. *et al.* The Phylogeny of the four Pan-American mtDNA haplogroups: Implications for evolutionary and disease studies. *PLoS One* **3**, e1764 (2008).
36. Tamm, E. *et al.* Beringian standstill and spread of Native American founders. *PLoS One* **2**, e829 (2007).
37. Mulligan, C. J., Kitchen, A. & Miyamoto, M. M. Updated three-stage model for the peopling of the Americas. *PLoS One* **3**, e3199 (2008).
38. Ho, S. Y. W. & EndiCott, P. The crucial role of calibration in molecular date estimates for the peopling of the Americas. *Am. J. Hum. Genet.* **83**, 142–146 (2008).
39. O'Fallon, B. D. & Fehren-Schmitz, L. Native Americans experienced a strong population bottleneck coincident with European contact. *Proc Natl Acad Sci U S A* **108**, 20444–20448 (2011).
40. Kumar, S. *et al.* The earliest settlers' antiquity and evolutionary history of Indian populations: evidence from M2 mtDNA lineage. *Bmc Evol Biol* **8**, 230 (2008).
41. Soares, P. *et al.* Climate change and postglacial human dispersals in Southeast Asia. *Mol. Biol. Evol.* **25**, 1209–1218 (2008).
42. Pereira, L. *et al.* Population expansion in the North African late Pleistocene signalled by mitochondrial DNA haplogroup U6. *Bmc Evol Biol* **10**, 390 (2010).
43. Forster, P. Ice Ages and the mitochondrial DNA chronology of human dispersals: a review. *Philos Trans R Soc Lond B Biol Sci* **359**, 255–264; discussion 264 (2004).
44. Relethford, J. H. *Human Population Genetics.* (Wiley, 2012).
45. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**, 1297–1303 (2010).
46. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
47. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
48. van Oven, M. & Kayser, M. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum Mutat* **30**, E386–394 (2009).
49. Bandelt, H. J., Forster, P. & Rohl, A. Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.* **16**, 37–48 (1999).
50. Guindon, S. & Gascuel, O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* **52**, 696–704 (2003).
51. Macaulay, V. *et al.* Single, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes. *Science* **308**, 1034–1036 (2005).
52. Yang, Z. H. PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).
53. Saillard, J., Forster, P., Lynnerup, N., Bandelt, H. J. & Norby, S. mtDNA variation among Greenland Eskimos: The edge of the Beringian expansion. *Am. J. Hum. Genet.* **67**, 718–726 (2000).
54. Drummond, A. J. & Rambaut, A. BEAST: Bayesian evolutionary analysis by sampling trees. *Bmc Evol Biol* **7**, 214 (2007).
55. Barbieri, C. *et al.* Contrasting maternal and paternal histories in the linguistic context of Burkina Faso. *Mol Biol Evol* **29**, 1213–1223 (2012).
56. Malyarchuk, B., Derenko, M., Denisova, G. & Kravtsova, O. Mitogenomic diversity in Tatars from the Volga-Ural Region of Russia. *Mol. Biol. Evol.* **27**, 2220–2226 (2010).
57. Finnila, S., Lehtonen, M. S. & Majamaa, K. Phylogenetic network for European mtDNA. *Am. J. Hum. Genet.* **68**, 1475–1484 (2001).
58. Gasparre, G. *et al.* Disruptive mitochondrial DNA mutations in complex I subunits are markers of oncocytic phenotype in thyroid tumors. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 9001–9006 (2007).
59. Howell, N. *et al.* Sequence analysis of the mitochondrial genomes from Dutch pedigrees with Leber hereditary optic neuropathy. *Am J Hum Genet* **72**, 1460–1469 (2003).
60. Pichler, I. *et al.* Drawing the history of the Hutterite population on a genetic landscape: inference from Y-chromosome and mtDNA genotypes. *Eur J Hum Genet* **18**, 463–470 (2010).
61. Just, R. S., Diegoli, T. M., Saunier, J. L., Irwin, J. A. & Parsons, T. J. Complete mitochondrial genome sequences for 265 African American and U.S. "Hispanic" individuals. *Forensic Sci Int Genet* **2**, e45–48 (2008).

62. Hammer, Ø., Harper, D. A. T. & Ryan, P. D. PAST: paleontological statistics software package for education and data analysis. *Palaeontologia Electronica* **4**, 9 (2001).

Author Contributions

Z.H.X. designed the study, carried out the molecular genetic studies and data analysis, and drafted the manuscript. Y.S. and Q.Z.D. participated in the genetic studies and helped in data analysis. J.L. participated in the design of the study, conceived of the study and drafted the manuscript. All authors read and approved the final manuscript.

Additional information

Supplementary Information accompanies this paper at http://www.nature.com/scientificreports

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Zheng, H., Yan, S., Qin, Z. & Jin, L. MtDNA analysis of global populations support that major population expansions began before Neolithic Time. *Sci. Rep.* **2**, 745; DOI:10.1038/srep00745 (2012).