

1           **Human paternal and maternal demographic histories: insights from high-**  
2                               **resolution Y chromosome and mtDNA sequences**

3

4           Sebastian Lippold<sup>1</sup>, Hongyang Xu<sup>1,2</sup>, Albert Ko<sup>1</sup>, Mingkun Li<sup>1,3</sup>, Gabriel Renaud<sup>1</sup>,  
5                               Anne Butthof<sup>1,4</sup>, Roland Schröder<sup>1</sup>, Mark Stoneking<sup>1</sup>

6

7           <sup>1</sup> Department of Evolutionary Genetics, Max Planck Institute for Evolutionary  
8           Anthropology, Leipzig, Germany

9           <sup>2</sup> Department of Computational Genetics, CAS-MPG Partner Institute for  
10          Computational Biology, Shanghai, China

11          <sup>3</sup> Present address: Fondation Mérieux, 17 rue Bourgelat, Lyon, France

12          <sup>4</sup> Present address: Institute of Biochemistry, Faculty of Medicine, University of  
13          Leipzig, Leipzig, Germany

14

15

16          Correspondence should be addressed to M.S. ([stoneking@eva.mpg.de](mailto:stoneking@eva.mpg.de))

1 **Abstract**

2 To investigate in detail the paternal and maternal demographic histories of humans,  
3 we obtained ~500 kb of non-recombining Y chromosome (NRY) sequences and  
4 complete mtDNA genome sequences from 623 males from 51 populations in the  
5 CEPH Human Genome Diversity Panel (HGDP). Our results: confirm the  
6 controversial assertion that genetic differences between human populations on a  
7 global scale are bigger for the NRY than for mtDNA; suggest very small ancestral  
8 effective population sizes (<100) for the out-of-Africa migration as well as for many  
9 human populations; and indicate that the ratio of female effective population size to  
10 male effective population size ( $N_f/N_m$ ) has been greater than one throughout the  
11 history of modern humans, and has recently increased due to faster growth in  $N_f$ .  
12 However, we also find substantial differences in patterns of mtDNA vs. NRY  
13 variation in different regional groups; thus, global patterns of variation are not  
14 necessarily representative of specific geographic regions.

1 Comparisons of mtDNA and NRY variation have provided numerous important  
2 insights into the maternal and paternal histories of human populations<sup>1-3</sup>. However,  
3 such comparisons are limited by methodological differences in how mtDNA and  
4 NRY variation have been typically assayed. MtDNA variation is usually investigated  
5 by sequencing hypervariable segments of the control region, (or, increasingly, via  
6 complete mtDNA genome sequences), while human NRY variation is routinely  
7 assayed by genotyping SNPs of interest, often in combination with Y-STR loci.  
8 Nevertheless, NRY SNP-typing has several drawbacks due to the ascertainment bias  
9 inherent in the selection of SNPs<sup>1,4,5</sup>. This ascertainment bias precludes many  
10 analyses of interest, such as dating the age of the NRY ancestor or particular  
11 divergence events in the NRY phylogeny, as well as demographic inferences such as  
12 population size changes. Moreover, the difference in molecular methods used to assay  
13 NRY vs. mtDNA variation can complicate the interpretation of differences between  
14 patterns of NRY and mtDNA variation. For example, the seminal finding that NRY  
15 differences are bigger than mtDNA differences among global populations of humans,  
16 and that this is due to a higher rate of female than male migration due to  
17 patrilocality<sup>6</sup>, may instead reflect methodological differences in how mtDNA vs. NRY  
18 variation is assayed<sup>7</sup>. Another fundamental question concerns whether or not male  
19 and female effective population sizes have been the same over time. Attempts to  
20 address this question using the ratio of X chromosome to autosomal DNA diversity  
21 have come up with conflicting answers<sup>8,9</sup>, which may in part reflect the use of  
22 different methods that capture information about effective population size at different  
23 times in the past<sup>10</sup>. Moreover, the ratio of X to autosome diversity varies along the X  
24 chromosome, depending how far polymorphic sites are from genes<sup>11</sup>, indicating a  
25 potential role for selection in distorting effective population size estimates from

1 comparisons of X chromosome to autosomal DNA diversity. These questions –  
2 namely, are genetic differences between populations and effective population sizes  
3 the same for males and females –as well as other fundamental aspects of human  
4 maternal and paternal demographic history remain unanswered.

5 Recently, analyses have been carried out of NRY sequences obtained as part of  
6 whole genome sequencing projects<sup>12-14</sup>. While these studies provide very detailed  
7 insights into NRY diversity, they are nonetheless limited by the expense of whole  
8 genome sequencing, which precludes comprehensive global sampling. To allow for  
9 more accurate comparisons between mtDNA and NRY variation and to permit  
10 demographic inferences based on the NRY, we developed a capture-based array to  
11 enrich Illumina sequencing libraries for ~500kb of NRY sequence (**Supplementary**  
12 **Table 1**). We used this approach to obtain NRY sequences from 623 males in the  
13 HGDP<sup>15</sup>, from 51 globally-distributed populations. We also obtained complete  
14 mtDNA genome sequences from the entire HGDP, allowing us to investigate and  
15 directly compare the paternal and maternal relationships of global human populations  
16 in unprecedented detail.

17 The average coverage of the NRY sequences was 14.5X (range 5X-37.5X,  
18 **Supplementary Figure 1**), while for the mtDNA genome sequences the average  
19 coverage was 640X (range 46X-4123X, **Supplementary Figure 1**). After quality-  
20 filtering, imputation, and removal of sites with a high number of recurrent mutations,  
21 there remained 2228 SNPs in the NRY sequences. The mtDNA analyses here are  
22 restricted to the 623 males for which NRY sequences were obtained, for which there  
23 were 2163 SNPs; results based on the mtDNA genome sequences from the entire  
24 HGDP (952 individuals) did not differ from those based on the subset of 623 males  
25 (**Supplementary Figure 2**). More details about the results from each individual,

1 including mtDNA and NRY haplogroups, are provided in **Supplementary Table**  
2 **2**. The mtDNA sequences have been deposited in Genbank with accession numbers  
3 KF450814 -KF451871. The NRY raw data are in the European Nucleotide Archive  
4 (ENA) (<http://www.ebi.ac.uk/ena/home>) with the study accession number  
5 PRJEB4417 (sample accession numbers ERS333252-ERS333873).

6 Basic summary statistics for the mtDNA and NRY diversity in each population are  
7 provided in **Supplementary Table 3**. As the sample sizes for many of the individual  
8 populations are quite small, for most subsequent analyses we grouped the populations  
9 into the following regions (based on analyses of genome-wide SNP data<sup>16,17</sup>): Africa,  
10 America, Central Asia, East Asia, Europe, Middle East/North Africa (ME/NA), and  
11 Oceania (the regional affiliation for each population is in **Supplementary Table 2**).  
12 The Adygei, Hazara, and Uygur were excluded from these groupings due to extensive  
13 admixture evident in the genome-wide SNP data<sup>16,17</sup>. We stress that the use of  
14 regional names is a convenience to refer to these groupings of these specific  
15 populations, and should not be taken to represent the entirety of the regions (e.g.,  
16 “Africa” refers to the results based on the analysis of the combined African HGDP  
17 samples, not to Africa in general).

18 Some basic summary statistics concerning mtDNA and NRY diversity for the  
19 regions are provided in **Table 1**. Notably, there is substantial variation among regions  
20 in amounts of mtDNA vs. NRY diversity; this is shown further in the comparison of  
21 the mean number of pairwise differences (mpd) for mtDNA and the NRY (**Figure**  
22 **1a**). The mtDNA mpd for Africa is about twice that for other regions, in keeping with  
23 previous observations of substantially greater mtDNA diversity in Africa than outside  
24 Africa<sup>18,19</sup>. However, the NRY mpd is greatest in the Middle East/North Africa  
25 region, and only slightly greater in Africa than in the other regions (with the exception

1 of the Americas, which show substantially lower NRY diversity). The greater NRY  
2 diversity in the Middle East/North Africa region can be attributed to male-biased  
3 admixture involving substantially-diverged NRY haplogroups. Nevertheless, there are  
4 striking differences in the ratio of NRY:mtDNA mdp (**Table 1**), with Africa, Central  
5 Asia, and the Americas having significantly less NRY diversity relative to mtDNA  
6 diversity, compared to the other regional groups. These results indicate substantial  
7 regional variation in the maternal vs. paternal demographic history of human  
8 populations.

9 An outstanding question is whether or not there are differences in the relative  
10 amounts of between-population vs. within-population diversity for mtDNA vs. the  
11 NRY. Some studies have found much larger between-population differences for the  
12 NRY than for mtDNA<sup>6</sup>, while others have not<sup>7</sup>. To address this question, we carried  
13 out an AMOVA; the results (**Figure 1b**) show that in the entire worldwide dataset, the  
14 between-population differences are indeed bigger for the NRY (~36% of the variance)  
15 than for mtDNA (~25% of the variance). However, there are substantial differences  
16 among the regional groups. The ME/NA, East Asia, and Europe regional groups  
17 follow the worldwide pattern in having bigger between-population differences for the  
18 NRY than for mtDNA. In contrast, Africa, Oceania, and the Americas have  
19 substantially bigger between-population differences for mtDNA than for the NRY,  
20 while for central Asia the between-population variation is virtually identical for the  
21 NRY and mtDNA. These regional differences likely reflect the influence of sex-  
22 biased migrations and admixture, and moreover indicate that focussing exclusively on  
23 the worldwide pattern of mtDNA vs. NRY variation misses these important regional  
24 differences.

1 Multidimension-scaling (MDS) plots based on  $\Phi_{ST}$  distances further indicate  
2 regional differences in mtDNA vs. NRY variation (**Figure 2**). Nonetheless, and  
3 despite the small sample sizes at the population level, both mtDNA and NRY  $\Phi_{ST}$   
4 distances are significantly correlated with geographic distances between populations  
5 (Mantel tests with 1000 replications: mtDNA,  $r=0.41$ ,  $p<0.001$ ; NRY,  $r=0.36$ ,  
6  $p=0.002$ ) as well as with each other ( $r=0.23$ ,  $p=0.025$ ). Thus, NRY and mtDNA  
7 diversity are both highly associated with geographic distances among populations.

8 We used a Bayesian method to estimate the phylogeny and divergence times for  
9 both mtDNA and the NRY (**Supplementary Figure 2**); for the latter, we used both a  
10 “fast” mutation rate of  $1 \times 10^{-9}$ /bp/year and a “slow” mutation rate of  $0.62 \times 10^{-9}$   
11 /bp/year as there is currently much uncertainty regarding mutation rates<sup>20-22</sup>. The  
12 resulting phylogenies are quite consistent with the existing mtDNA and NRY  
13 phylogenies<sup>23,24</sup>, with some small discrepancies involving lineages that are not well-  
14 resolved. The age of the mtDNA ancestor is estimated to be about 160 thousand years  
15 (ky), and the ages of the non-African mtDNA lineages M and N are about 65-70 ky,  
16 in good agreement with previous estimates<sup>18</sup>. Our estimate for the age of the  
17 NRY ancestor is 103 ky based on the fast rate, and 165 ky based on the slow rate;  
18 however these estimates do not include the recently-discovered “A00” lineage<sup>20</sup>,  
19 which would result in much older ages for the NRY ancestor. The close agreement  
20 between the slow NRY ancestor age (165 ky) and the mtDNA ancestor age (160 ky)  
21 might be taken as evidence in favor of the slow NRY mutation rate. However, the  
22 slow NRY mutation rate gives an estimated age for the initial out-of-Africa divergence  
23 of about 100 ky, and an age for the divergence of Amerindian-specific haplogroup Q  
24 lineages of about 20 ky, while the fast rate gives corresponding estimates of about 60  
25 ky for out-of-Africa and about 12.5 ky for Amerindian haplogroup Q lineages, in

1 better agreement with the mtDNA and other evidence for these events<sup>18,25-27</sup>. Given  
2 the current uncertainty over mutation rate estimates, we have chosen to use either  
3 both estimates in further analyses (e.g., Bayesian skyline plots) or an average of the  
4 fast and slow rates (e.g., in simulation-based analyses).

5 NRY and mtDNA haplogroup frequencies per population are shown in  
6 **Supplementary Table 4**. NRY haplogroups for the CEPH-HGDP males were  
7 previously determined by SNP-genotyping<sup>28</sup>, and the phylogenetic relationships for  
8 the NRY sequences are generally concordant with the SNP-genotyping results (with  
9 some exceptions, discussed in the legends to **Supplementary Figures 4-13**). The  
10 haplogroup frequencies provide further insights into some of the different regional  
11 patterns of mtDNA vs. NRY diversity noted previously, and these are discussed in the  
12 legend to **Supplementary Table 4**. We note some additional features in the  
13 individual NRY haplogroup phylogenies provided in **Supplementary Figures 4-13**,  
14 while the full mtDNA phylogeny is provided in **Supplementary Figure 14**.

15 Sequence-based analysis of NRY variation permits demographic analyses that  
16 cannot be carried out with ascertained SNP genotype data. As an example, we  
17 estimated the history of population size changes via Bayesian skyline plots (BSPs) for  
18 the NRY and mtDNA genome sequences for each region (**Figure 3**). These results  
19 should be interpreted cautiously, both because of the small sample sizes for some of  
20 the regions (in particular, America and Oceania), and because grouping populations  
21 with different histories can produce spurious signals of population growth<sup>29</sup>.  
22 Nevertheless, both the mtDNA and NRY BSPs indicate overall population growth in  
23 almost all groups, but for mtDNA there is a more pronounced signal of growth at  
24 around 15,000-20,000 years ago than there is for the NRY, and during much of the  
25 past it appears as if the effective size for females was larger than that for males.



1 To further investigate female and male demographic history, we used a simulation-  
2 based approach (**Supplementary Figure 15**) to estimate the current and ancestral  
3 effective population size for females ( $N_f$ ) and males ( $N_m$ ) for Africa, Europe, East  
4 Asia, Central Asia, Oceania, and the Americas (excluding ME/NA populations  
5 because of their admixed history). We also estimated the ancestral  $N_f$  and  $N_m$  for the  
6 out-of-Africa migration. The detailed results are provided in **Supplementary Figures**  
7 **16-19** and **Supplementary Tables 5-10**; a pictorial depiction of the results is in  
8 **Figure 4**. These results indicate a small founding size in Africa of about 60 females  
9 and 30 males (all population sizes are effective population sizes); migration out of  
10 Africa about 75 ky ago (kya) associated with a bottleneck of around 25 females and  
11 15 males; migrations from this non-African founding population to Oceania 61 kya, to  
12 Europe 49 kya, to Central and East Asia 37 kya, and from East Asia to the Americas  
13 about 15 kya. There was concomitant population growth in all regions (with the most  
14 growth in East Asia); however, throughout history the mtDNA and NRY results  
15 indicate consistently larger effective population sizes for females than for males  
16 (except, possibly, in the ancestors of East Asians).

17 Previous studies of  $N_f$  and  $N_m$  have largely relied on comparisons of X  
18 chromosome vs. autosomal variation, and have come to contradictory conclusions  
19 concerning the historical  $N_f/N_m$  ratio<sup>8,9,30</sup>, because of methodological differences,  
20 difficulties in accounting for differences in male vs. female mutation rates, as well as  
21 the potentially greater effect of selection on the X chromosome<sup>10,11</sup>. Comparison of  
22 mtDNA vs. NRY variation offers a more direct assessment free of some of the issues  
23 concerning X:autosome comparisons, but requires unbiased estimates of NRY  
24 variation, which until our study were only available from either whole genome  
25 sequencing studies<sup>5,12-14</sup> or more limited targeted studies of NRY sequence

1 variation<sup>7,31</sup>. Our results support a consistent excess of  $N_f$  vs.  $N_m$  starting even before  
2 the out-of-Africa migration that has been carried through almost all subsequent  
3 migrations (with the possible exception of East Asia), and has become even more  
4 pronounced in recent times (**Supplementary Figure 16; Figure 4**) due to higher rates  
5 of growth in  $N_f$  than in  $N_m$  (**Figures 3 and 4**).

6 In conclusion, we have developed a rapid and cost-effective means of obtaining  
7 unbiased NRY sequence information at comparable resolution to that of complete  
8 mtDNA genome sequences. Application to the HGDP provides new insights into the  
9 comparative demographic history of males and females, including support for larger  
10 between-population differences for the NRY than for mtDNA (albeit with  
11 considerable regional variation), significant bottlenecks associated with the migration  
12 of modern humans out-of-Africa and with other migration events; and overall higher  
13 female than male effective population sizes. We anticipate that this approach should  
14 enable more detailed comparative analyses of the demographic history of males vs.  
15 females, and the influence of sex-specific processes during human evolution.

16

## 17 **METHODS**

18 Methods and any associated references are available in the online version of the  
19 paper.

20

21 *Note: Supplementary information is available in the online version of the paper.*

22

## 23 **ACKNOWLEDGMENTS**

24 We thank: Maanasa Raghavan and Martina Codina for technical assistance; the MPI-  
25 EVA Evolutionary Genetics Sequencing Group and Bioinformatics Group for

1 carrying out the sequencing and the initial processing of the sequence data; and  
2 Chiara Barbieri, Peter de Knijff, Brigitte Pakendorf, Kay Prüfer, and Udo Stenzel for  
3 helpful discussion. Research supported by the Max Planck Society; we acknowledge  
4 financial support from the Biotechnology and Biological Sciences Research Council  
5 (to Francois Balloux).

6

## 7 **AUTHOR CONTRIBUTIONS**

8 M.S. designed the study. S.L. designed the NRY capture microarray. A.B., S.L. and  
9 R.S. carried out the laboratory work. S.L., M.L. and G.R. processed the sequences.  
10 S.L., H.X., and A.K. analyzed the data. M.S. and S.L. wrote the paper, with input  
11 from all authors.

12

13

## 1 REFERENCES

- 2 1. Jobling, M.A. The impact of recent events on human genetic diversity. *Philos Trans R*  
3 *Soc Lond B Biol Sci* **367**, 793-9 (2012).
- 4 2. Heyer, E., Chaix, R., Pavard, S. & Austerlitz, F. Sex-specific demographic behaviours  
5 that shape human genomic variation. *Mol Ecol* **21**, 597-612 (2012).
- 6 3. Wilkins, J.F. Unraveling male and female histories from human genetic data. *Curr*  
7 *Opin Genet Dev* **16**, 611-7 (2006).
- 8 4. Dupanloup, I. *et al.* A recent shift from polygyny to monogamy in humans is  
9 suggested by the analysis of worldwide Y-chromosome diversity. *J Mol Evol* **57**, 85-  
10 97 (2003).
- 11 5. Wei, W., Ayub, Q., Xue, Y. & Tyler-Smith, C. A comparison of Y-chromosomal lineage  
12 dating using either resequencing or Y-SNP plus Y-STR genotyping. *Forensic Sci Int*  
13 *Genet* **7**, 568-72 (2013).
- 14 6. Seielstad, M.T., Minch, E. & Cavalli-Sforza, L.L. Genetic evidence for a higher female  
15 migration rate in humans. *Nat Genet* **20**, 278-80 (1998).
- 16 7. Wilder, J.A., Kingan, S.B., Mobasher, Z., Pilkington, M.M. & Hammer, M.F. Global  
17 patterns of human mitochondrial DNA and Y-chromosome structure are not  
18 influenced by higher migration rates of females versus males. *Nat Genet* **36**, 1122-5  
19 (2004).
- 20 8. Hammer, M.F., Mendez, F.L., Cox, M.P., Woerner, A.E. & Wall, J.D. Sex-biased  
21 evolutionary forces shape genomic patterns of human diversity. *PLoS Genet* **4**,  
22 e1000202 (2008).
- 23 9. Keinan, A., Mullikin, J.C., Patterson, N. & Reich, D. Accelerated genetic drift on  
24 chromosome X during the human dispersal out of Africa. *Nat Genet* **41**, 66-70 (2009).
- 25 10. Emery, L.S., Felsenstein, J. & Akey, J.M. Estimators of the human effective sex ratio  
26 detect sex biases on different timescales. *Am J Hum Genet* **87**, 848-56 (2010).
- 27 11. Hammer, M.F. *et al.* The ratio of human X chromosome to autosome diversity is  
28 positively correlated with genetic distance from genes. *Nat Genet* **42**, 830-1 (2010).
- 29 12. Francalacci, P. *et al.* Low-pass DNA sequencing of 1200 Sardinians reconstructs  
30 European Y-chromosome phylogeny. *Science* **341**, 565-9 (2013).
- 31 13. Poznik, G.D. *et al.* Sequencing Y chromosomes resolves discrepancy in time to  
32 common ancestor of males versus females. *Science* **341**, 562-5 (2013).
- 33 14. Wei, W. *et al.* A calibrated human Y-chromosomal phylogeny based on resequencing.  
34 *Genome Res* **23**, 388-95 (2013).
- 35 15. Cann, H.M. *et al.* A human genome diversity cell line panel. *Science* **296**, 261-2  
36 (2002).
- 37 16. Li, J.Z. *et al.* Worldwide human relationships inferred from genome-wide patterns of  
38 variation. *Science* **319**, 1100-4 (2008).
- 39 17. Lopez Herraez, D. *et al.* Genetic variation and recent positive selection in worldwide  
40 human populations: evidence from nearly 1 million SNPs. *PLoS One* **4**, e7888 (2009).
- 41 18. Behar, D.M. *et al.* The dawn of human matrilineal diversity. *Am J Hum Genet* **82**,  
42 1130-40 (2008).
- 43 19. Pakendorf, B. & Stoneking, M. Mitochondrial DNA and human evolution. *Annu Rev*  
44 *Genomics Hum Genet* **6**, 165-83 (2005).
- 45 20. Mendez, F.L. *et al.* An African American paternal lineage adds an extremely ancient  
46 root to the human Y chromosome phylogenetic tree. *Am J Hum Genet* **92**, 454-9  
47 (2013).
- 48 21. Scally, A. & Durbin, R. Revising the human mutation rate: implications for  
49 understanding human evolution. *Nat Rev Genet* **13**, 745-53 (2012).

- 1 22. Xue, Y.L. *et al.* Human Y Chromosome Base-Substitution Mutation Rate Measured by  
2 Direct Sequencing in a Deep-Rooting Pedigree. *Current Biology* **19**, 1453-1457 (2009).
- 3 23. Karafet, T.M. *et al.* New binary polymorphisms reshape and increase resolution of  
4 the human Y chromosomal haplogroup tree. *Genome Res* **18**, 830-8 (2008).
- 5 24. van Oven, M. & Kayser, M. Updated comprehensive phylogenetic tree of global  
6 human mitochondrial DNA variation. *Hum Mutat* **30**, E386-94 (2009).
- 7 25. Zegura, S.L., Karafet, T.M., Zhivotovsky, L.A. & Hammer, M.F. High-resolution SNPs  
8 and microsatellite haplotypes point to a single, recent entry of Native American Y  
9 chromosomes into the Americas. *Mol Biol Evol* **21**, 164-75 (2004).
- 10 26. O'Rourke, D.H. & Raff, J.A. The human genetic history of the Americas: the final  
11 frontier. *Curr Biol* **20**, R202-7 (2010).
- 12 27. Ray, N. *et al.* A statistical evaluation of models for the initial settlement of the  
13 american continent emphasizes the importance of gene flow with Asia. *Mol Biol Evol*  
14 **27**, 337-45 (2010).
- 15 28. Shi, W. *et al.* A worldwide survey of human male demographic history based on Y-  
16 SNP and Y-STR data from the HGDP-CEPH populations. *Mol Biol Evol* **27**, 385-93  
17 (2010).
- 18 29. Gunnarsdottir, E.D., Li, M., Bauchet, M., Finstermeier, K. & Stoneking, M. High-  
19 throughput sequencing of complete human mtDNA genomes from the Philippines.  
20 *Genome Res* **21**, 1-11 (2011).
- 21 30. Labuda, D., Lefebvre, J.F., Nadeau, P. & Roy-Gagnon, M.H. Female-to-male breeding  
22 ratio in modern humans-an analysis based on historical recombinations. *Am J Hum*  
23 *Genet* **86**, 353-63 (2010).
- 24 31. Wilder, J.A., Mobasher, Z. & Hammer, M.F. Genetic evidence for unequal effective  
25 population sizes of human females and males. *Mol Biol Evol* **21**, 2047-57 (2004).

26

27

1 **FIGURE LEGENDS**

2

3 **Figure 1** Diversity and AMOVA results. **(a)** Mean number of pairwise differences  
4 (and SE bars) for the NRY and mtDNA sequences from each regional group. **(b)**  
5 AMOVA results for the entire worldwide dataset, and for each regional group of  
6 populations. Two comparisons are shown for the entire dataset; the left comparison  
7 includes regional groups as an additional hierarchical level, while the right one does  
8 not. \* indicates that the among population component of diversity does not differ  
9 significantly from zero (after Bonferroni adjustment of the p-value for multiple  
10 comparisons).

11

12 **Figure 2** MDS plots based on  $\Phi_{ST}$  distances among regional groups. The stress values  
13 are 0.055 for the mtDNA plot and 0.031 for the NRY plot.

14

15 **Figure 3** Bayesian skyline plots for regional groups. Two curves are shown for the  
16 NRY data, based on “fast” and “slow” mutation rate estimates.

17

18 **Figure 4** Pictorial representation of the simulation results estimating divergence times  
19 and female and male effective population sizes. Red numbers reflect  $N_f$  (with  
20 ancestral  $N_f$  at the point of the triangle and current  $N_f$  at the base of the red triangle),  
21 blue numbers reflect ancestral and current  $N_m$ , the numbers in the black oval indicate  
22 the founding effective sizes for the initial out-of-Africa migration, and dates on arrows  
23 indicate divergence times based on the tree in **Supplementary Figure 14**. Arrows are  
24 meant to indicate the schematic direction of migrations and should not be taken as  
25 indicating literal migration pathways, e.g. the results indicate divergence of the

- 1 ancestors of Oceanians 61,000 years ago, but not the route(s) people took to get to
- 2 Oceania.
- 3

1 **Table 1** Summary statistics for regional groups. n, sample size; H, number of different haplotypes (sequences); S, number of polymorphic sites;  
 2 mpd  $\pm$  SE, mean number of pairwise differences  $\pm$  standard error;  $\pi$   $\pm$  SE, nucleotide diversity  $\pm$  standard error; mpd ratio, ratio of the  
 3 mpd<sub>NRY</sub>/mpd<sub>mtDNA</sub>. \* group ratios that differ significantly (p<0.05) from the overall average ratio for the entire HGDP, based on random  
 4 resampling of NRY and mtDNA sequences.

5

6

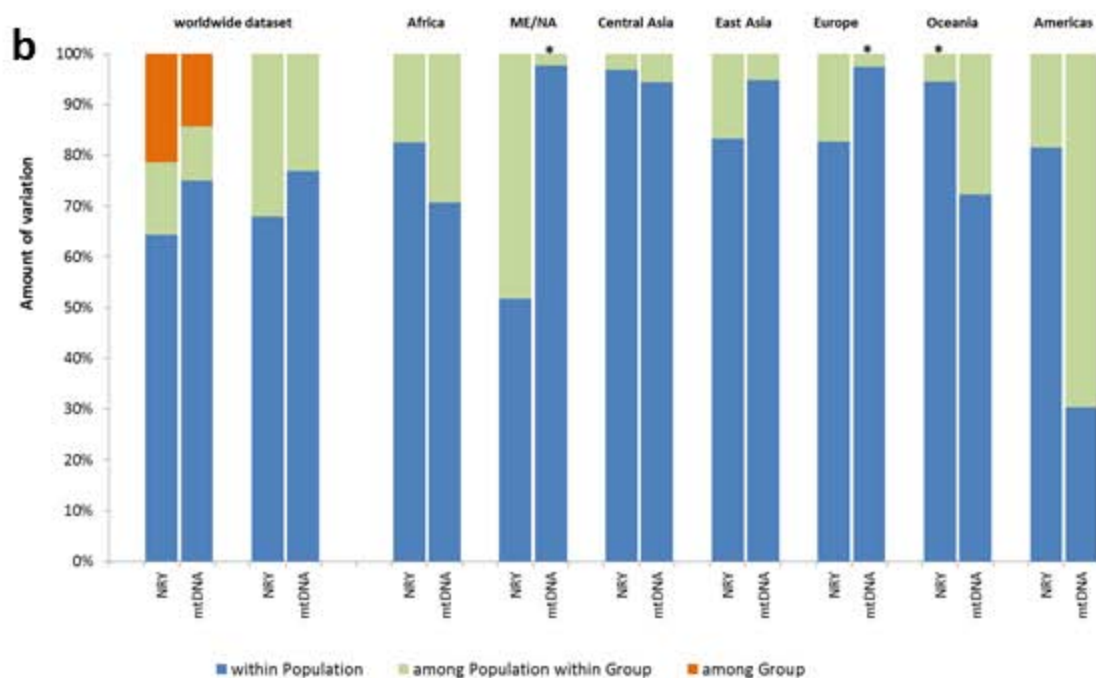
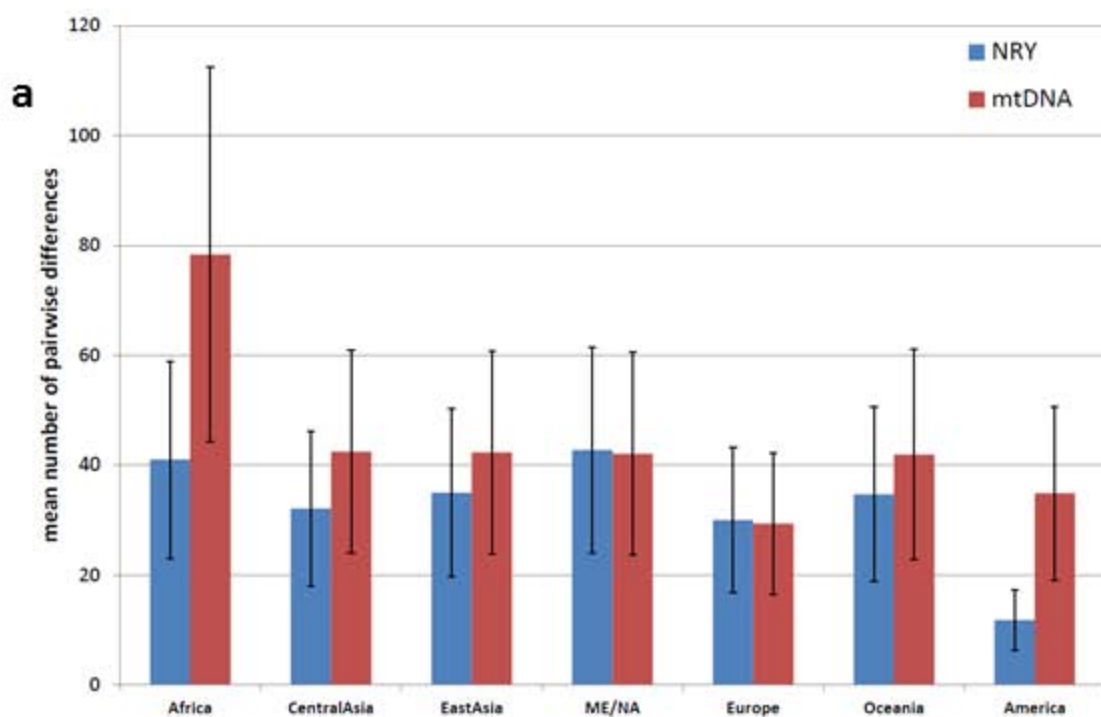
Group	-----NRY-----					-----mtDNA-----				mpd ratio
	n	H	S	mpd $\pm$ SE	$\pi$ $\pm$ SE <sup>a</sup>	H	S	mpd $\pm$ SE	$\pi$ $\pm$ SE <sup>b</sup>	
Africa	85	71	545	41.0 $\pm$ 18.0	80 $\pm$ 40	70	617	78.3 $\pm$ 34.0	47 $\pm$ 23	0.52*
CentralAsia	146	106	524	32.1 $\pm$ 14.1	62 $\pm$ 31	131	833	42.4 $\pm$ 18.5	26 $\pm$ 12	0.76*
EastAsia	162	141	709	35.0 $\pm$ 15.3	71 $\pm$ 36	156	899	42.3 $\pm$ 18.5	26 $\pm$ 12	0.83
ME/NA	75	47	301	42.7 $\pm$ 18.7	85 $\pm$ 40	71	618	42.0 $\pm$ 18.4	25 $\pm$ 12	1.02
Europe	79	68	350	30.0 $\pm$ 13.2	58 $\pm$ 31	78	432	29.3 $\pm$ 12.9	18 $\pm$ 9	1.02
Oceania	17	16	147	34.7 $\pm$ 15.9	71 $\pm$ 36	16	175	41.9 $\pm$ 19.2	25 $\pm$ 13	0.83
America	22	19	96	11.8 $\pm$ 5.5	22 $\pm$ 13	15	148	34.9 $\pm$ 15.8	21 $\pm$ 11	0.39*

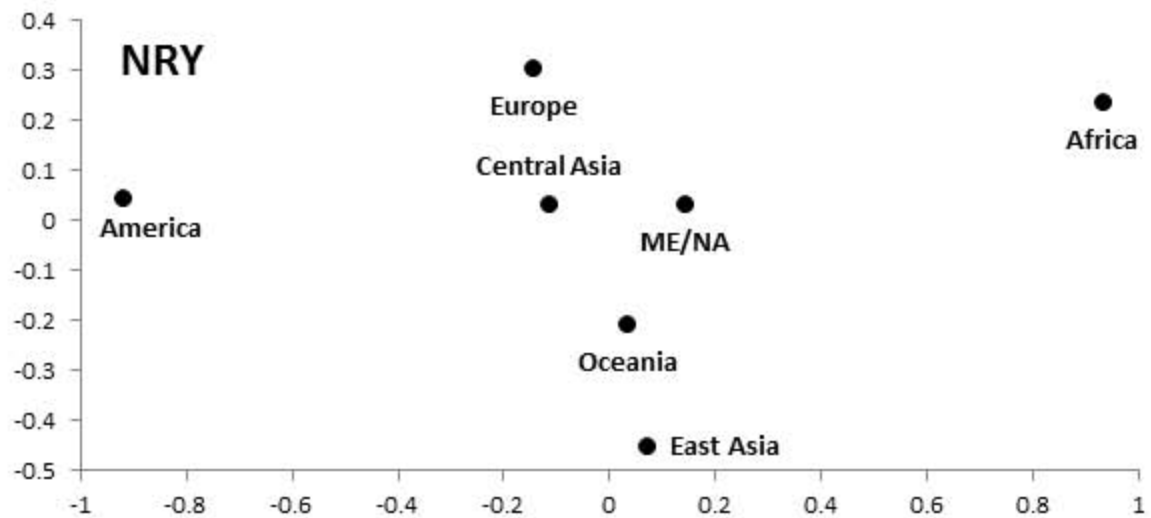
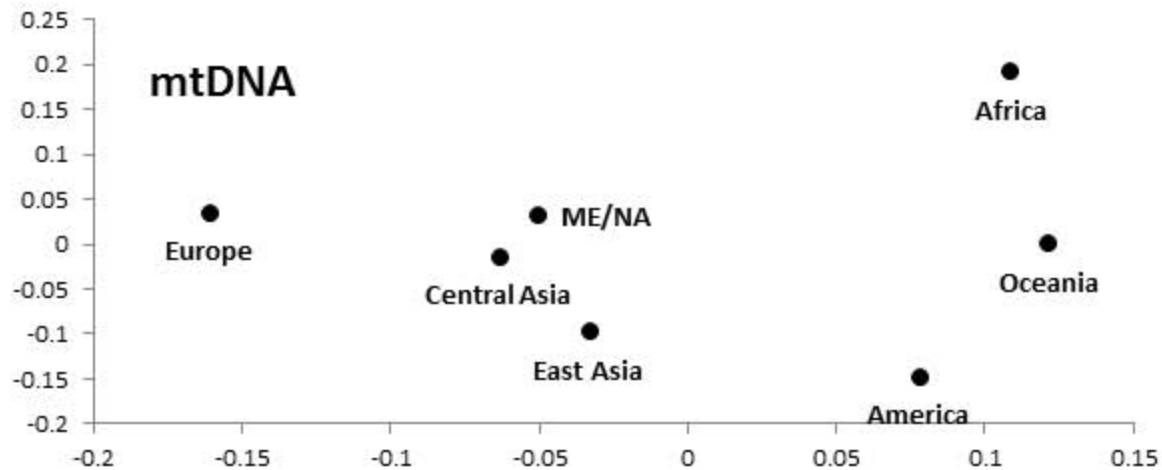
7

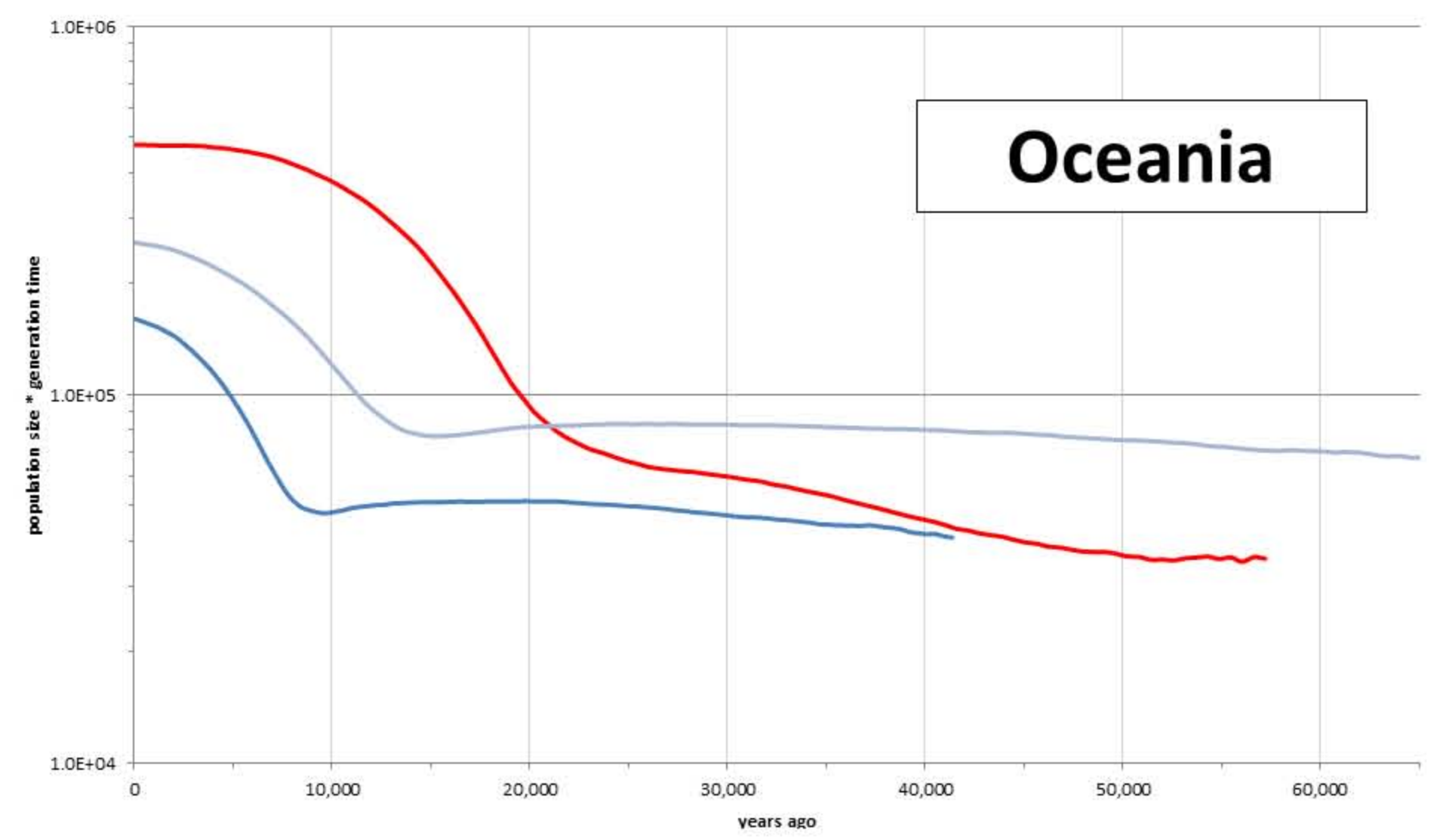
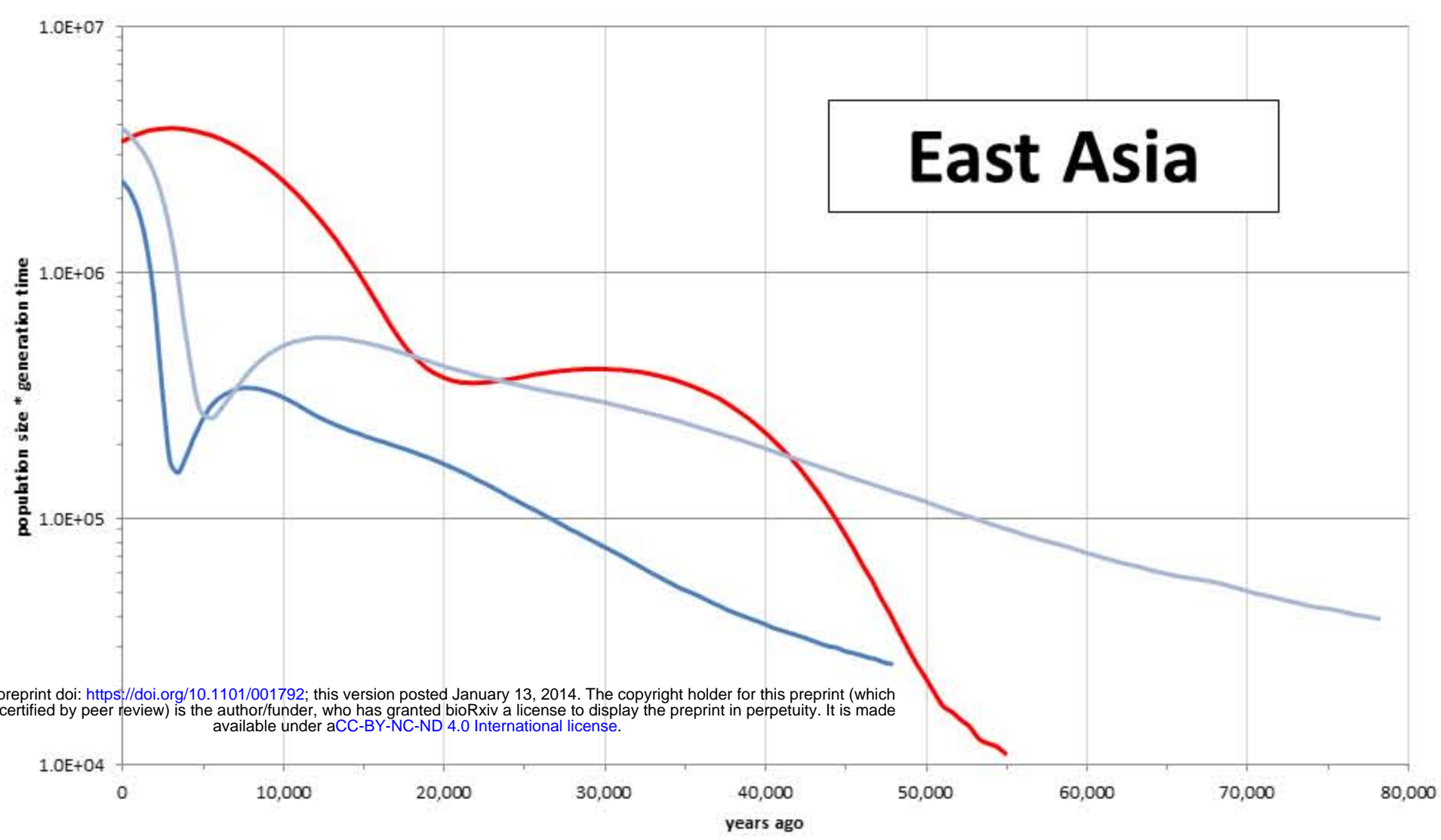
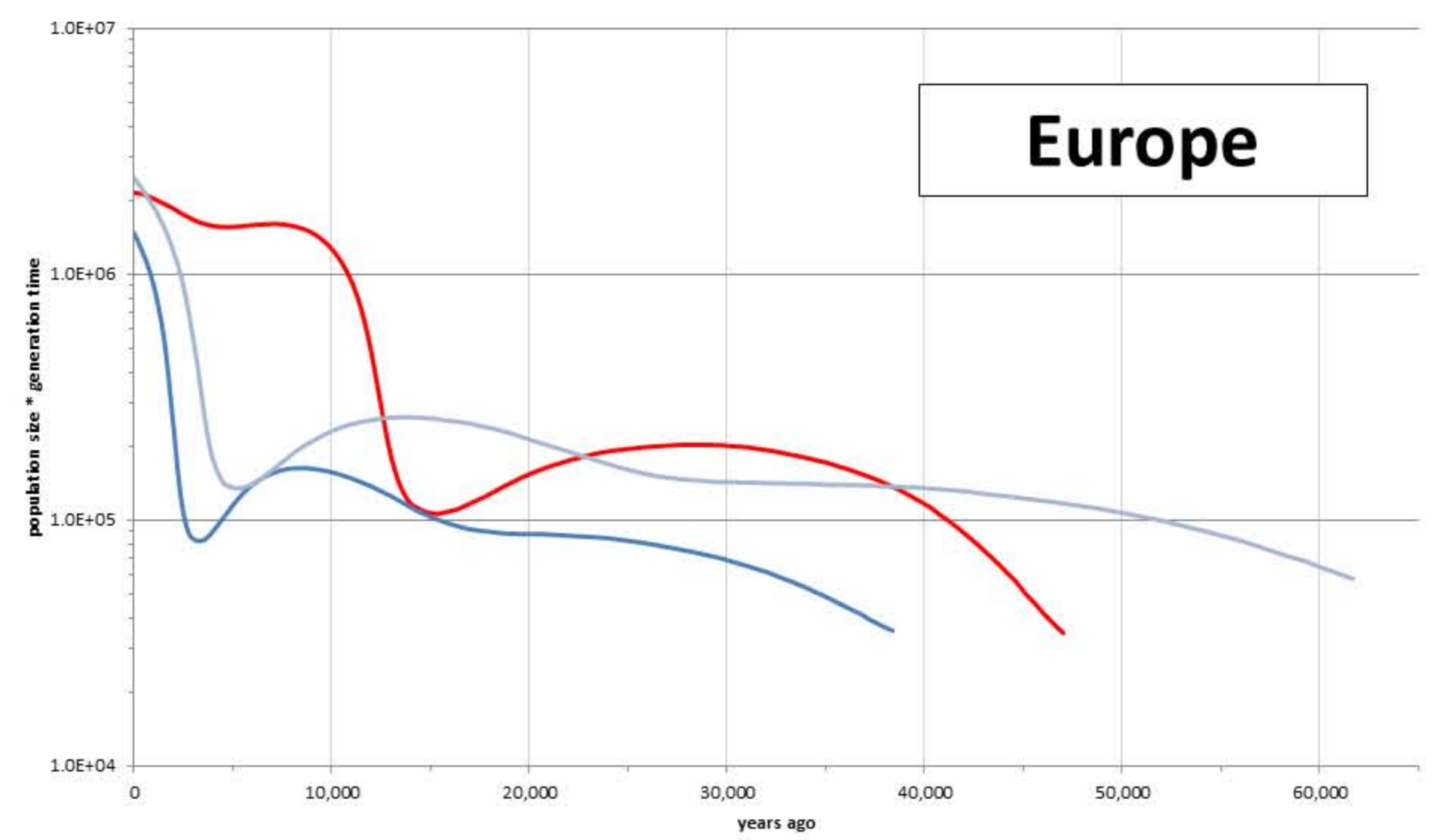
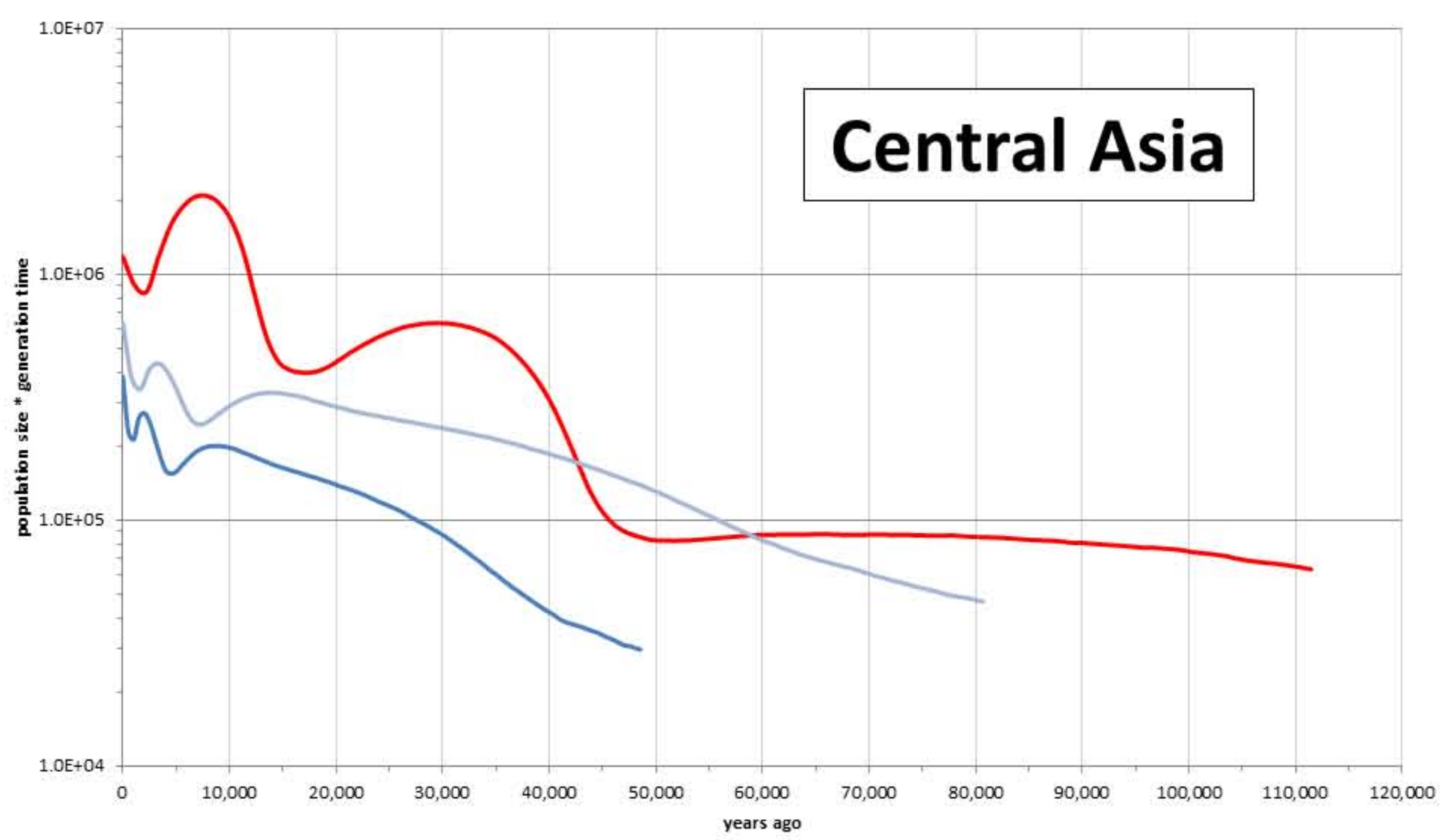
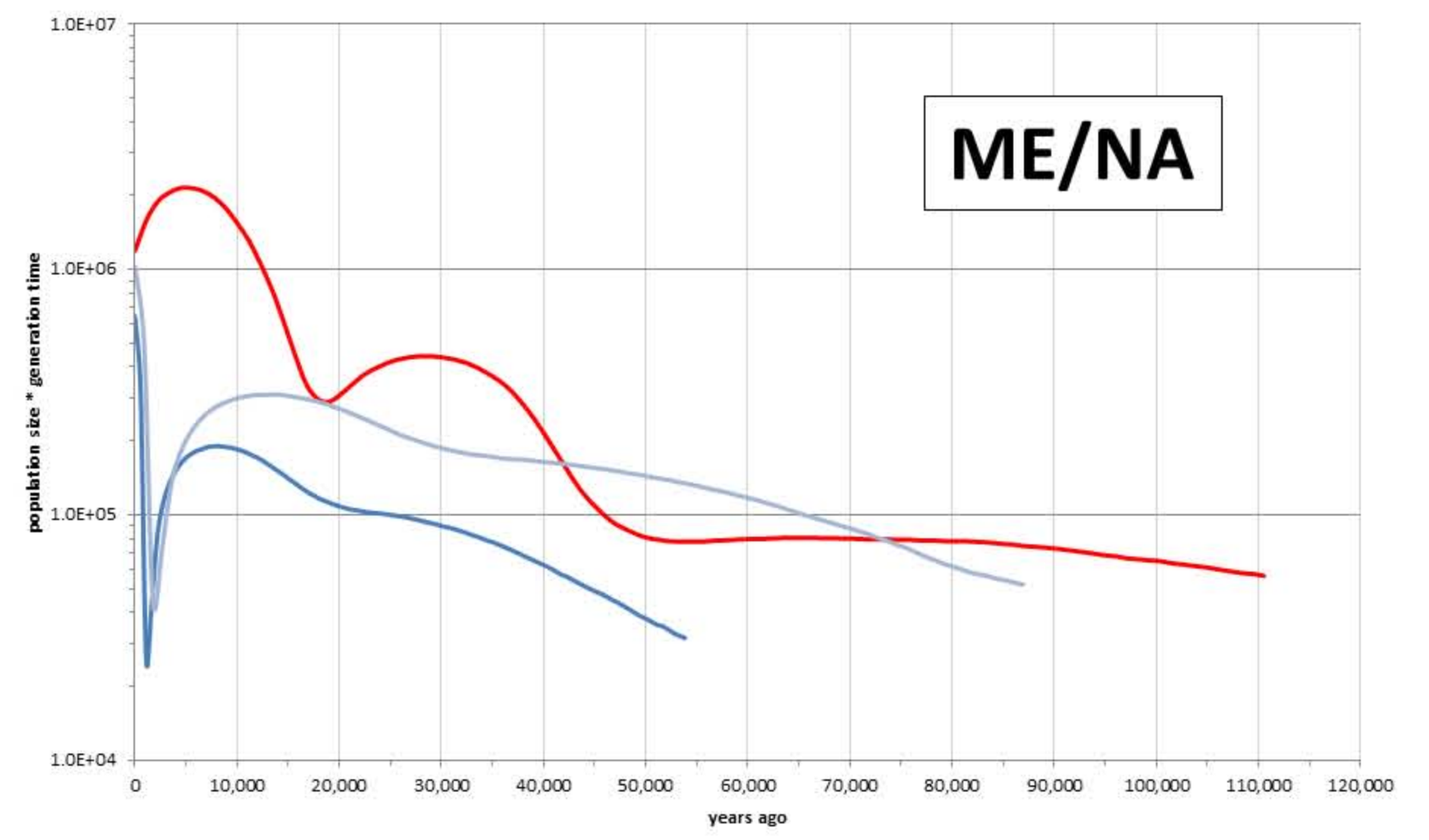
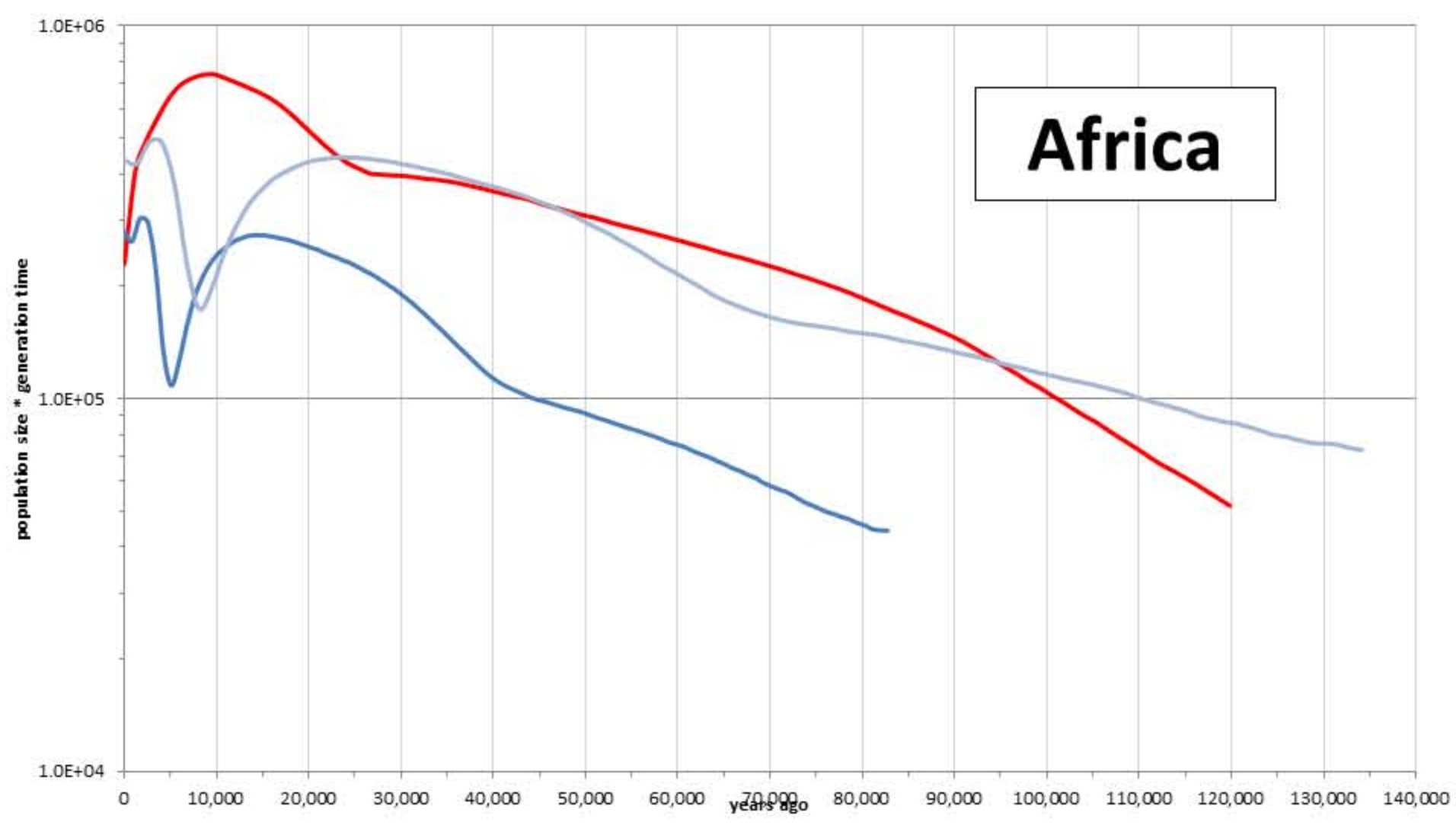
8 <sup>a</sup> multiply values by 10<sup>-6</sup>

9 <sup>b</sup> multiply values by 10<sup>-4</sup>









bioRxiv preprint doi: <https://doi.org/10.1101/001792>; this version posted January 13, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

