Supplementary Materials

Sociocultural behavior, sex-biased admixture and effective population sizes in Central African Pygmies and non-Pygmies

Paul Verdu[1,2*], Noémie S.A. Becker[2], Alain Froment[3], Myriam Georges[2], Viola Grugni[4], Lluis Quintana-Murci[5], Jean-Marie Hombert[6], Lolke Van der Veen[6], Sylvie Le Bomin[2], Serge Bahuchet[2], Evelyne Heyer[2] and Frédéric Austerlitz[2]

[1] Department of Biology, Stanford University, Stanford CA 94305, USA

[2] CNRS, MNHN, Université Paris Diderot, Sorbonne Paris Cité, UMR 7206 Eco-Anthropologie et Ethnobiologie, F-75005 Paris, France

[3] IRD – MNHN, UMR 208 Patrimoines locaux, Paris, France

[4] Dipartimento di Biologia e Biotecnologie, Università di Pavia, 27100 Pavia, Italy

[5] CNRS-Institut Pasteur, URA3012 Human Evolutionary Genetics Unit, 75015 Paris, France

[6] ISH, UMR5596, Laboratoire Dynamique du Langage, 69363 Lyon, France

[*] *Corresponding author*: *Paul Verdu, CNRS-MNHN, UMR 7206 Eco-Anthropologie et Ethnobiologie, 47 rue Cuvier 75005 Paris CP 139 France ; verdu@mnhn.fr; Tel. +33 1 40 79 81 54 ; Fax. +33 1 40 79 32 31.*

Running title: Sex-biased admixture and population sizes in African Pygmies

Keywords:
Approximate Bayesian Computation;
African Pygmy;
Demography;
History;
Human population genetics;
Sex-specific.

Supplementary Materials and Methods:

Approximate Bayesian Computation (ABC)

Prior set

We simulated 1,000,000 data sets for each one of these 128 three-population ABC analyses separately. For all simulations, following Verdu et al. (2009), we considered the following prior distributions for the historical and demographic parameters (Figure S2). We used flat priors for the population sizes, divergence times and admixture rates. We considered a uniformly distributed prior between 1 and 10,000 for N_1, N_2, N_{ap} and N_A, and between 1 and 100,000 for N_{np}. Priors for the ancient admixture event, r_{aa}, and for the two recent admixture event between non-Pygmies and each one of the Pygmy population separately, r_{r1} and r_{r2}, were drawn from a uniform distribution between 0 and 1.

Following Verdu et al. (2009), since we lacked archaeological information in Central Africa, we used a uniformly distributed prior between 1 and 5,000 generations (25 and 125,000 years considering 25 years per generations), for the divergence times t_p, t_{pnp} and for the ancient admixture event t_{aa}. We further considered a log-uniform distribution for the prior of the time of recent admixture, t_{ra}, between 1 and 5,000 generations (Verdu et al. 2009). Only simulations satisfying $t_{ra} < t_p < t_{aa} < t_{pnp}$ were conducted to obtain 1,000,000 simulations for each separate ABC analysis. For the time, t_A, of the potential change of effective population size in the non-Pygmy lineage (from N_A to N_{np}), we considered a uniformly distributed prior between 1 and 10,000 generations.

Mutation model

Since all the microsatellites here considered were tetranucleotides, and included the possibility for insertions and deletions, we used the same mutation priors as the ones considered in Verdu et al. (2009), for the 28 autosomal, the eight X-chromosome and the six NRY markers. Namely, we considered a general stepwise mutation model (GSM, Estoup, Jarne, Cornuet 2002), with the possibility for 40 contiguous alleles. We considered a uniform prior between 10^{-3} and 10^{-4} for the mean mutation rate $\bar{\mu}$ (Zhivotovsky, Rosenberg, Feldman 2003), and drew the mutation rate for

each independent loci from a gamma distribution with mean $\bar{\mu}$ and Shape 2.0. The average parameter for the geometric distribution ($\bar{p}$) of the length of the number of repeats occurring for each mutation events was drawn from a uniform distribution between 0.1 and 0.3 (Dib et al. 1996; Estoup, Jarne, Cornuet 2002), and the specific value for each locus was drawn from a Gamma distribution with mean $\bar{p}$ and Shape 2.0. Finally, rates of insertion-deletion for each marker based on the observation of departure from a strictly tetranucleotide motif, were drawn from a Gamma distribution of mean = 2.5×10^{-8} and Shape = 2 (Pascual et al. 2007).

For the mutation rate of the 359 bp of the mitochondrial HVR-1, we considered the Hasegawa-Kishino-Yano model for substitutions (Hasegawa-Kishino-Yano Hasegawa, Kishino, Yano 1985), with all 359 sites possibly subject to mutation and with a shape of the Gamma distribution of mutations among sites of 0.26 (Endicott et al. 2009). We considered a uniform prior between 10^{-8} and 10^{-5} for the mean mutation rate $\bar{\mu}$ for the mitochondrial HVR-1 (Endicott et al. 2009), and drew the mutation rate for each independent loci from a Gamma distribution with mean $\bar{\mu}$ and Shape 0 (this is not a limiting case of the Gamma distribution but an argument of DIY ABC, which gives the same value of the mutation rate to all loci separately as recommended by the manual p.9) since we only consider mitochondrial HVR-1 sequences. The average coefficient *kappa* ($\bar{\kappa}$) for the transition/transversion ratio was drawn from a uniform distribution between 0.05 and 20 and from a Gamma distribution with mean $\bar{\kappa}$ and Shape 0 for the individual locus coefficient *kappa* since we only consider mitochondrial HVR-1 sequences.

Since effective population sizes and mutation rates are virtually undistinguishable, we also computed composite parameters as $\theta_i = 4N_i \bar{\mu}$ and $\tau_i = t_i \bar{\mu}$, with $\bar{\mu}$ the mean mutation rate of microsatellite loci or of mitochondrial sequences for the four chromosomal types here under study.

Summary Statistics

Autosomal, X-chromosome and NRY

We used 15 within- and among-populations summary statistics for the three chromosomal types microsatellite data sets respectively. Within population, we considered the mean number of alleles per locus, the mean expected heterozygosity H_e (Nei 1978) and the mean allele size

variance expressed in base pairs. Among populations, we considered pairwise allele sharing distances and genetic distances $(\delta\mu)^2$ (Goldstein et al. 1995).

mtDNA HVR-1 sequences

We used 21 within- and among-population summary statistics for the mtDNA HVR-1 sequences. Within population, we considered four statistics: the number of segregating sites, the mean number of pairwise difference, the variance of the number of pairwise differences and Tajima's *D* (Tajima 1989). Among populations, we considered three pairwise statistics: the mean of within sample pairwise differences, the mean between sample pairwise differences and the F_{ST} between two samples (Hudson 1992).

Supplementary Tables

Table S1. Posterior estimates (mode, median and 90% CI) of composite model parameters for the effective population sizes ($N\bar{\mu}$) in Pygmy populations estimated using Approximate Bayesian Computations.

Posterior estimates of composite model parameters for the effective population sizes $N\bar{\mu}$ obtained considering one-population models described in Materials and Methods. *"nd"* indicates that no data was available for this population (see Table 1). "90% CI" indicates the 90% Credibility Interval for the posterior estimates of the composite parameter. For the four chromosomal types separately, we simulated one million data sets for each Pygmy population. For each Pygmy population and each chromosomal type separately, posterior distributions of $N\bar{\mu}$ were estimated based on the 10,000 (1%) top simulations providing summary statistics closest to the observed data. See Supplementary Materials for prior distributions of the original parameters N and for the mean mutation rate $\bar{\mu}$. See Table 1 for population codes.

Table S1.

	Autosomes			X-chromosome			NRY			mtDNA HVR-1		
Population	**Mode**	**Median**	**90% CI**	**Mode**	**Median**	**90% CI**	**Mode**	**Median**	**90% CI**	**Mode**	**Median**	**90% CI**
EBK	2.601	2.751	[0.893 ; 5.472]	0.886	1.055	[0.412 ; 2.769]	3.878	4.277	[1.889 ; 7.775]	*nd*	*nd*	*nd*
CBK	2.556	2.580	[0.889 ; 5.043]	1.308	1.648	[0.696 ; 4.173]	*nd*	*nd*	*nd*	0.014546	0.014390	[0.006640 ; 0.030111]
GBK	2.654	2.612	[0.886 ; 4.939]	1.033	1.183	[0.449 ; 2.982]	3.489	3.910	[1.640 ; 7.480]	0.014000	0.013840	[0.006850 ; 0.027612]
SBK	2.304	2.352	[0.791 ; 4.663]	1.379	1.752	[0.717 ; 4.512]	2.860	3.710	[1.440 ; 7.390]	0.030902	0.031670	[0.018280 ; 0.055390]
BEZ	1.849	1.987	[0.654 ; 4.097]	1.139	1.375	[0.519 ; 3.597]	2.594	3.227	[1.325 ; 6.811]	0.005460	0.005070	[0.001840 ; 0.011541]
EBG	2.514	2.628	[0.820 ; 5.258]	1.242	1.637	[0.641 ; 3.957]	2.579	2.860	[1.110 ; 6.460]	0.017256	0.017370	[0.008370 ; 0.035081]
CBG	2.137	2.237	[0.730 ; 5.014]	1.141	1.437	[0.584 ; 4.285]	2.664	3.278	[1.276 ; 7.076]	*nd*	*nd*	*nd*
SBG	2.519	2.526	[0.888 ; 5.452]	1.290	1.706	[0.700 ; 4.003]	2.617	3.190	[1.208 ; 6.838]	*nd*	*nd*	*nd*
KOL	1.951	2.066	[0.697 ; 4.251]	1.009	1.242	[0.441 ; 3.382]	1.527	1.930	[0.700 ; 5.020]	0.005613	0.005300	[0.001970 ; 0.010892]
KOY	2.393	2.505	[0.831 ; 5.083]	1.301	1.504	[0.612 ; 4.018]	0.874	1.089	[0.377 ; 3.101]	0.010387	0.010120	[0.004010 ; 0.027592]

Table S2. Posterior estimates (mode and 90% CI) of original and composite model parameters for the effective population sizes in non-Pygmy populations estimated using Approximate Bayesian Computations.

Posterior estimates of composite model parameters for the effective population sizes $N\bar{\mu}$ obtained considering one-population models described in Materials and Methods. "*nd*" indicates that no data was available for this population (see Table 1). "90% CI" indicates the 90% Credibility Interval for the posterior estimates of the composite parameter. For the four chromosomal types separately, we simulated one million data sets for each non-Pygmy population. For each non-Pygmy population and each chromosomal type separately, posterior distributions of $N\bar{\mu}$ were estimated based on the 10,000 (1%) top simulations providing summary statistics closest to the observed data. See Supplementary Materials for prior distributions of the original parameters *N* and for the mutation rate $\bar{\mu}$. See Table 1 for population codes.

Table S2.

	Autosomes			X-chromosome			NRY			mtDNA HVR-1		
Population	**Mode**	**Median**	**90% CI**	**Mode**	**Median**	**90% CI**	**Mode**	**Median**	**90% CI**	**Mode**	**Median**	**90% CI**
AKL	4.279	7.020	[1.860 ; 31.043]	3.752	6.910	[1.330 ; 34.730]	2.530	4.220	[1.000 ; 27.632]	0.120750	0.136100	[0.077295 ; 0.328715]
BGD	2.600	4.430	[1.050 ; 24.141]	2.135	3.570	[0.820 ; 22.660]	1.649	3.160	[0.810 ; 22.750]	*nd*	*nd*	*nd*
CFG	3.737	5.130	[1.330 ; 23.321]	4.142	6.260	[1.520 ; 29.493]	3.298	6.190	[1.730 ; 30.001]	0.066252	0.079900	[0.043200 ; 0.268725]
EWD	6.371	7.940	[2.210 ; 31.183]	2.037	3.410	[0.760 ; 23.313]	1.800	3.100	[0.500 ; 27.900]	0.049119	0.059500	[0.030400 ; 0.249245]
GFG	3.160	5.540	[1.370 ; 26.264]	6.389	8.100	[2.260 ; 33.910]	1.446	2.130	[0.700 ; 13.930]	0.090634	0.102800	[0.060695 ; 0.234915]
KOT	6.822	7.890	[2.110 ; 31.450]	2.884	4.940	[1.160 ; 29.152]	5.333	7.010	[2.180 ; 27.271]	0.112430	0.130600	[0.075000 ; 0.318325]
NZE	6.684	7.480	[2.060 ; 29.981]	5.207	6.560	[1.550 ; 31.713]	7.000	2.700	[0.900 ; 16.400]	0.124586	0.142200	[0.083795 ; 0.322105]
NZI	6.376	7.830	[2.010 ; 29.695]	2.168	3.570	[0.840 ; 22.581]	1.467	5.065	[0.660 ; 35.891]	*nd*	*nd*	*nd*
TEK	6.665	7.575	[2.010 ; 29.742]	2.886	4.710	[1.050 ; 27.671]	3.339	4.550	[1.360 ; 23.621]	0.094350	0.104100	[0.062495 ; 0.251905]
TIK	2.942	3.920	[1.010 ; 20.291]	3.310	5.500	[1.280 ; 27.502]	5.312	7.290	[2.100 ; 33.632]	*nd*	*nd*	*nd*
TSG	3.139	4.350	[1.070 ; 21.333]	3.923	6.520	[1.570 ; 32.816]	3.327	4.470	[1.320 ; 21.761]	0.101085	0.115200	[0.066300 ; 0.274750]
NGB	4.293	5.210	[1.300 ; 25.233]	2.955	5.830	[1.400 ; 28.850]	*nd*	*nd*	*nd*	0.061831	0.067200	[0.041100 ; 0.159300]

Table S3. Posterior estimates (mode and 90% CI) of divergence times t_p among pairs of Pygmy populations estimated using Approximate Bayesian Computations for autosomal data.

Posterior estimates of the divergence time t_p were obtained considering three-population models described in Figure S2 and Materials and Methods. Lower-triangle: mode estimates; Upper-triangle: 90% Credibility Interval. "*NA*" indicates not applicable (see Materials and Methods). For autosomal microsatellites, we simulated one million data sets for all the procedures involving a different pair of Pygmy populations (together with the corresponding non-Pygmy neighbouring population). For each Pygmy population separately, posterior distributions of t_p were estimated based on the 10,000 (1%) top simulations providing summary statistics closest to the observed data. To obtain a single posterior distribution for a given Pygmy population, we merged all posterior distributions obtained for the ABC procedures considering that particular Pygmy population (see Materials and Methods). The prior distribution for t_p was drawn from a uniform distribution between 1 and 5,000 generations, with simulation constraints such that $tr_r < t_p < tr_a < t_{pnp}$ (see Supplementary Materials and Figure S2). See Table 1 for population codes.

Table S3.

	CBK	EBK	GBK	SBK	BEZ	CBG	SBG	EBG	KOY	KOL
CBK	--	*NA*	*NA*	*NA*	[85 ; 2094]	[95 ; 2287]	[87 ; 2114]	[98 ; 2341]	[88 ; 2259]	[90 ; 2128]
EBK	*NA*	--	*NA*	*NA*	[83 ; 2015]	[178 ; 3031]	[84 ; 2217]	[132 ; 2646]	[118 ; 2461]	[69 ; 1876]
GBK	*NA*	*NA*	--	*NA*	[98 ; 2168]	[152 ; 2766]	[103 ; 2323]	[127 ; 2576]	[93 ; 2110]	[92 ; 2196]
SBK	*NA*	*NA*	*NA*	--	[83 ; 2030]	[203 ; 3192]	[95 ; 2336]	[105 ; 2352]	[118 ; 2497]	[88 ; 2170]
BEZ	334	313	373	314	--	[114 ; 2369]	[113 ; 2509]	[122 ; 2535]	[90 ; 2102]	[104 ; 2372]
CBG	387	752	551	920	450	--	[122 ; 2596]	[143 ; 2774]	[168 ; 2876]	[126 ; 2609]
SBG	353	373	406	380	489	478	--	[97 ; 2358]	[102 ; 2363]	[74 ; 2042]
EBG	373	520	520	426	482	652	387	--	[134 ; 2699]	[107 ; 2448]
KOY	354	470	362	517	334	680	444	541	--	[78 ; 1898]
KOL	353	285	373	351	411	528	306	428	294	--

Table S4. Posterior estimates (mode and 90% CI) of divergence times t_p among pairs of Pygmy populations estimated using Approximate Bayesian Computations for X-chromosome data.

Posterior estimates of the divergence time t_p were obtained considering three-population models described in Figure S2 and Materials and Methods. Lower-triangle: mode estimates; Upper-triangle: 90% Credibility Interval. *"NA"* indicates not applicable (see Materials and Methods). For X-chromosome microsatellites, we simulated one million data sets for all the procedures involving a different pair of Pygmy populations (together with the corresponding non-Pygmy neighbouring population). For each Pygmy population separately, posterior distributions of t_p were estimated based on the 10,000 (1%) top simulations providing summary statistics closest to the observed data. To obtain a single posterior distribution for a given Pygmy population, we merged all posterior distributions obtained for the ABC procedures considering that particular Pygmy population (see Materials and Methods). The prior distribution for t_p was drawn from a uniform distribution between 1 and 5,000 generations, with simulation constraints such that $tr_r < t_p < tr_a < t_{pnp}$ (see Supplementary Materials and Figure S2). See Table 1 for population codes.

Table S4.

	CBK	EBK	GBK	SBK	BEZ	CBG	SBG	EBG	KOY	KOL
CBK	--	*NA*	*NA*	*NA*	[122 ; 2714]	[174 ; 2886]	[93 ; 2387]	[100 ; 2624]	[179 ; 3008]	[117 ; 2596]
EBK	*NA*	--	*NA*	*NA*	[170 ; 3028]	[160 ; 2776]	[90 ; 2243]	[100 ; 2479]	[140 ; 2911]	[126 ; 2747]
GBK	*NA*	*NA*	--	*NA*	[82 ; 2180]	[228 ; 3115]	[87 ; 2182]	[118 ; 2633]	[203 ; 3111]	[81 ; 2249]
SBK	*NA*	*NA*	*NA*	--	[117 ; 2677]	[132 ; 2587]	[77 ; 2227]	[142 ; 3002]	[166 ; 3109]	[128 ; 2872]
BEZ	484	664	321	441	--	[259 ; 3319]	[89 ; 2380]	[91 ; 2432]	[264 ; 3553]	[102 ; 2624]
CBG	682	589	844	536	914	--	[172 ; 2940]	[276 ; 3379]	[283 ; 3382]	[282 ; 3444]
SBG	361	349	345	314	346	641	--	[120 ; 2830]	[182 ; 3160]	[102 ; 2449]
EBG	405	383	470	577	351	1062	495	--	[164 ; 3015]	[100 ; 2578]
KOY	661	552	726	655	1048	1019	663	661	--	[123 ; 2629]
KOL	437	501	310	525	446	1060	404	437	468	--

Table S5. Posterior estimates (mode and 90% CI) of divergence times t_p among pairs of Pygmy populations estimated using Approximate Bayesian Computations for NRY data.

Posterior estimates of the divergence time t_p were obtained considering three-population models described in Figure S2 and Materials and Methods. Lower-triangle: mode estimates; Upper-triangle: 90% Credibility Interval. *"NA"* indicates not applicable (see Materials and Methods). For NRY microsatellites, we simulated one million data sets for all the procedures involving a different pair of Pygmy populations (together with the corresponding non-Pygmy neighbouring population). For each Pygmy population separately, posterior distributions of t_p were estimated based on the 10,000 (1%) top simulations providing summary statistics closest to the observed data. To obtain a single posterior distribution for a given Pygmy population, we merged all posterior distributions obtained for the ABC procedures considering that particular Pygmy population (see Materials and Methods). The prior distribution for t_p was drawn from a uniform distribution between 1 and 5,000 generations, with simulation constraints such that $tr_r < t_p < tr_a < t_{pnp}$ (see Supplementary Materials and Figure S2). See Table 1 for population codes.

Table S5.

	EBK	GBK	SBK	BEZ	CBG	SBG	EBG	KOY	KOL
EBK	--	*NA*	*NA*	[145 ; 3048]	[113 ; 2854]	[113 ; 2905]	[129 ; 3076]	[164 ; 3156]	[136 ; 2953]
GBK	*NA*	--	*NA*	[172 ; 3312]	[130 ; 2955]	[108 ; 2907]	[129 ; 3014]	[158 ; 3060]	[114 ; 2913]
SBK	*NA*	*NA*	--	[137 ; 3176]	[92 ; 2751]	[95 ; 2784]	[209 ; 3508]	[153 ; 3066]	[135 ; 3055]
BEZ	592	761	617	--	[124 ; 2896]	[118 ; 2978]	[257 ; 3678]	[182 ; 3154]	[230 ; 3567]
CBG	478	523	428	495	--	[101 ; 2897]	[152 ; 3372]	[129 ; 3039]	[142 ; 3233]
SBG	461	457	441	543	448	--	[143 ; 3264]	[118 ; 3001]	[97 ; 2853]
EBG	549	535	926	1089	660	651	--	[139 ; 3068]	[125 ; 3016]
KOY	689	640	658	779	536	509	571	--	[107 ; 2684]
KOL	551	499	577	1029	618	432	538	446	--

Table S6. Posterior estimates (mode and 90% CI) of divergence times t_p among pairs of Pygmy populations estimated using Approximate Bayesian Computations for mtDNA HVR-1 data.

Posterior estimates of the divergence time t_p were obtained considering three-population models described in Figure S2 and Materials and Methods. Lower-triangle: mode estimates; Upper-triangle: 90% Credibility Interval. *"NA"* indicates not applicable (see Materials and Methods). For mtDNA HVR-1 sequences, we simulated one million data sets for all the procedures involving a different pair of Pygmy populations (together with the corresponding non-Pygmy neighbouring population). For each Pygmy population separately, posterior distributions of t_p were estimated based on the 10,000 (1%) top simulations providing summary statistics closest to the observed data. To obtain a single posterior distribution for a given Pygmy population, we merged all posterior distributions obtained for the ABC procedures considering that particular Pygmy population (see Materials and Methods). The prior distribution for t_p was drawn from a uniform distribution between 1 and 5,000 generations, with simulation constraints such that $tr_r < t_p < tr_a < t_{pnp}$ (see Supplementary Materials and Figure S2). See Table 1 for population codes.

Table S6.

	CBK	GBK	SBK	BEZ	EBG	KOY	KOL
CBK	--	*NA*	*NA*	*NA*	[289 ; 3131]	[99 ; 2016]	[157 ; 2564]
GBK	*NA*	--	*NA*	[64 ; 1600]	[255 ; 3033]	[126 ; 2309]	[89 ; 2076]
SBK	*NA*	*NA*	--	[102 ; 2188]	[200 ; 3053]	[95 ; 2411]	[106 ; 2151]
BEZ	*NA*	228	381	--	[345 ; 3199]	[67 ; 1574]	[118 ; 2207]
EBG	947	918	711	1105	--	[310 ; 3379.05]	[315 ; 3321]
KOY	359	352	417	239	1272	--	[110 ; 2219]
KOL	521	433	353	422	1025	392	--

Supplementary Figure Legends

Figure S1. Predicted ratio between the AMOVA F_{ST} at X-linked ($F_{ST}^{(X)}$) and autosomal loci ($F_{ST}^{(A)}$) as a function of ($F_{ST}^{(A)}$) and of the proportion of female migrants (m_f/m)

Using equation (4) in Ségurel et al. (2008), and assuming balanced sex ratio ($N_f/N = 1/2$), we plot the predicted AMOVA F_{ST} ratio of X-linked to autosomal-linked loci ($F_{ST}^{(X)}$ / $F_{ST}^{(A)}$) under an island model. Under this balanced sex ratio, the ratio $F_{ST}^{(X)}/F_{ST}^{(A)}$ is necessary comprised between 1.0 and 2.0. Since this equation predicts that $F_{ST}^{(X)}/F_{ST}^{(A)}$ is a decreasing function of N_f/N, this F_{ST} ratio can be above 2.0 only when the male effective population size is larger than the female effective population size, and below 1.0 only in the opposite case.

Figure S2. Three-population models used for the inference of demographic and admixture posterior parameters with Approximate Bayesian Computation.

Pop1 and Pop2 correspond to two different Pygmy populations with effective population sizes N_1 and N_2 respectively. Pop3 corresponds to the two respective neighboring non-Pygmy populations pooled together with effective population size N_{np}. We consider in turn each pair of Pygmy population in our data set and the corresponding non-Pygmy immediate neighbors. Following Verdu et al. (2009), we considered a common origin of the two Pygmy populations at time t_p and a more ancient divergence between the ancestral Pygmy and non-Pygmy populations at time t_{pnp}. N_{ap} denotes the effective population size of the ancestral Pygmy population. We considered recent genetic introgression events from the non-Pygmy population into each Pygmy population, occurring at time tr_r with potentially variable intensity (r_{r1} and r_{r2} respectively), and a more ancient admixture event between the ancestral Pygmy and non-Pygmy populations occurring at time tr_a with intensity r_a. We considered a potential increase in the non-Pygmy effective population size (from N_A to N_{np}) occurring at any time (t_A) during the history of the non-Pygmy lineage. We simulated 1,000,000 data sets for each chromosomal type separately (28 autosomal, eight X-chromosome, six NRY microsatellites and mtDNA HVR-1 359 bp sequences). See Supplementary Materials and Methods for detailed parameter prior sets used in the simulations.

Figure S3. Posterior distributions of the effective population size of Pygmy populations using three-population models with Approximate Bayesian Computations.

Posterior distributions were obtained considering three-population models (Supplementary Figure S2). For each one of the four chromosomal types, we simulated separately one million data sets for all the procedures involving a different pair of Pygmy populations (together with the corresponding non-Pygmy neighbouring population). For each chromosomal type separately, posterior distributions of the effective population size were estimated based on the 10,000 (1%) top simulations providing summary statistics closest to the observed data. To obtain a single posterior distribution for a given Pygmy population, we merged all posterior distributions obtained for the ABC procedures considering that particular Pygmy population (see Materials and Methods). Results in black correspond to the 28 autosomal microsatellites; in orange: eight X-chromosome microsatellites; in blue: six NRY microsatellites; and in red: mtDNA HVR-1 359 bp sequences. See Materials and Methods and Supplementary Materials for further description of the simulation priors and ABC analyses. See Table 1 for population codes. Note that the effective population size posterior distributions (90% CI, 1st and 4th quartile, rank order of medians and of most modes within population across chromosomal types) estimated with three-population ABC analyses are very consistent with posterior distributions obtained with one-population ABC analyses (Figure 4).

Figure S4. Posterior distributions of divergence times t_p among pairs of Pygmy populations estimated using Approximate Bayesian Computations.

Density of the posterior distributions of the divergence time t_p were obtained considering three-population scenarios described in Figure S2 and Materials and Methods. For each one of the four chromosomal types separately, we simulated one million data sets for all the procedures involving a different pair of Pygmy populations (together with the corresponding non-Pygmy neighbouring population). For each Pygmy population and each chromosomal types separately, posterior distributions of t_p were estimated based on the 10,000 (1%) top simulations providing summary statistics closest to the observed data. To obtain a single posterior distribution for a given Pygmy population, we merged all posterior distributions obtained for the ABC procedures

considering that particular Pygmy population (see Materials and Methods). The prior distribution for t_p is indicated by the thick red line and was drawn from a uniform distribution between 1 and 5,000 generations, with simulation constraints such that $tr_r < t_p < tr_a < t_{pnp}$ (see Supplementary Materials and Figure S2). The upper-left plot corresponds to the 28 autosomal microsatellites; Upper-right plot: eight X-chromosome microsatellites; Lower-left plot: six NRY microsatellites; and Lower-right plot: mtDNA HVR-1 359 bp sequences. See Materials and Methods and Supplementary Materials for further description of the simulation and ABC procedures. See Table 1 for population codes.

LITTERATURE CITED

Dib C, Faure S, Fizames C, et al. 1996. A comprehensive genetic map of the human genome based on 5,264 microsatellites. Nature 380:152-154.

Endicott P, Ho SY, Metspalu M, Stringer C. 2009. Evaluating the mitochondrial timescale of human evolution. Trends Ecol Evol 24:515-521.

Estoup A, Jarne P, Cornuet JM. 2002. Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. Mol Ecol. 11:1591-1604.

Goldstein DB, Ruiz Linares A, Cavalli-Sforza LL, Feldman MW. 1995. An evaluation of genetic distances for use with microsatellite loci. Genetics. 139:463-471.

Hasegawa M, Kishino H, Yano T. 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J Mol Evol 22:160-174.

Hudson RR. 1992. Gene trees, species trees and the segregation of ancestral alleles. Genetics. 131:509-513.

Pascual M, Chapuis MP, Mestres F, Balanya J, Huey RB, Gilchrist GW, Serra L, Estoup A. 2007. Introduction history of Drosophila subobscura in the New World: a microsatellite-based survey using ABC methods. Mol Ecol. 16:3069-3083.

Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123:585-596.

Zhivotovsky LA, Rosenberg NA, Feldman MW. 2003. Features of evolution and expansion of modern humans, inferred from genomewide microsatellite markers. Am J Hum Genet. 72:1171-1186.

Figure S1

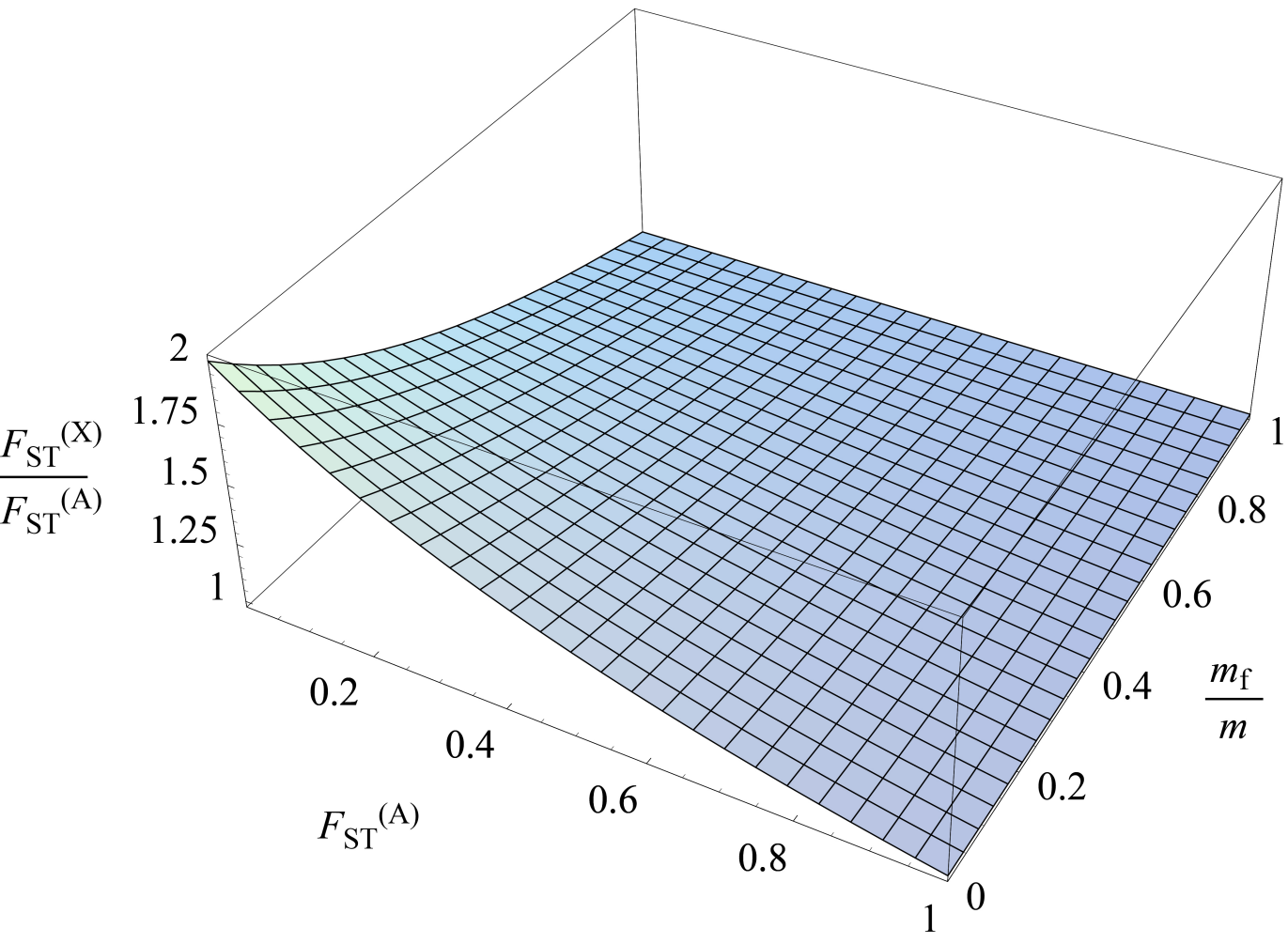

$\frac{F_{ST}^{(X)}}{F_{ST}^{(A)}}$
2
1.75
1.5
1.25
1
0.2
0.4
0.6
0.8
1
$F_{ST}^{(A)}$
0
0.2
0.4
0.6
0.8
1
$\frac{m_f}{m}$

Figure S2

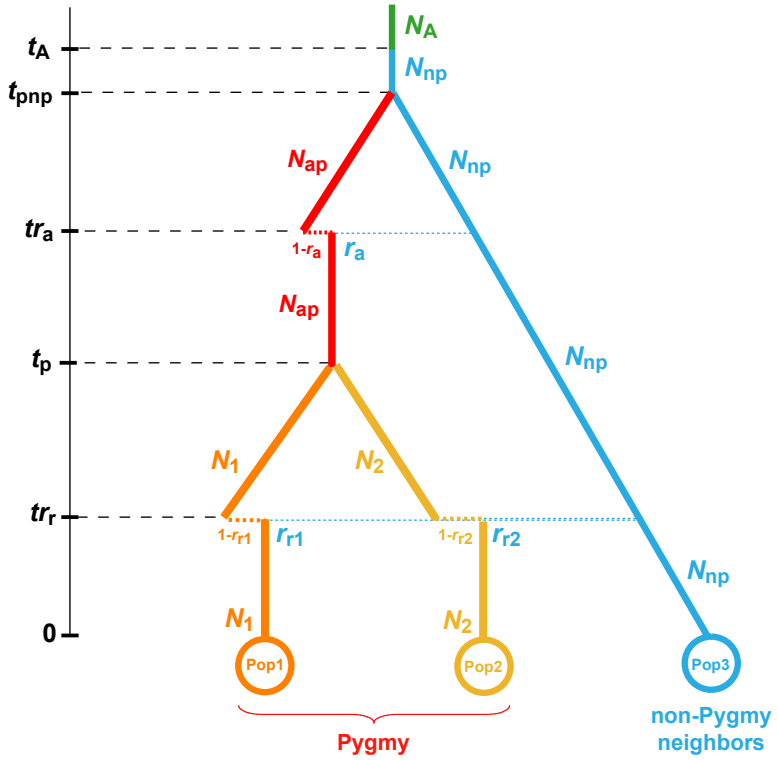

t_A
t_{pnp}
tr_a
t_p
tr_r
0
N_A
N_{np}
N_{ap}
N_{np}
1-r_a
r_a
N_{ap}
N_{np}
N_1
N_2
N_{np}
1-r_{r1}
r_{r1}
1-r_{r2}
r_{r2}
N_1
N_2
Pop1
Pop2
Pop3
Pygmy
non-Pygmy neighbors

Figure S3

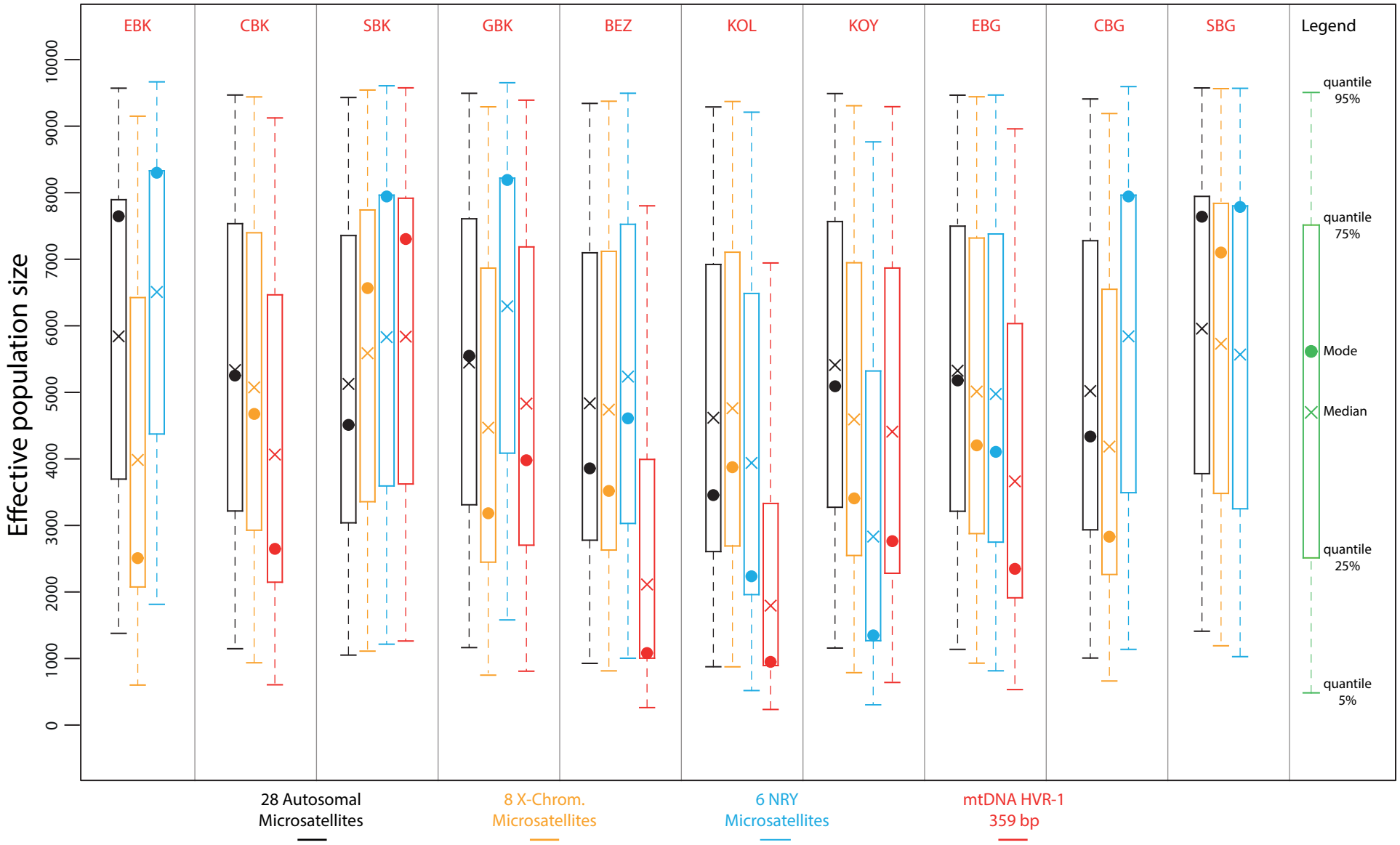
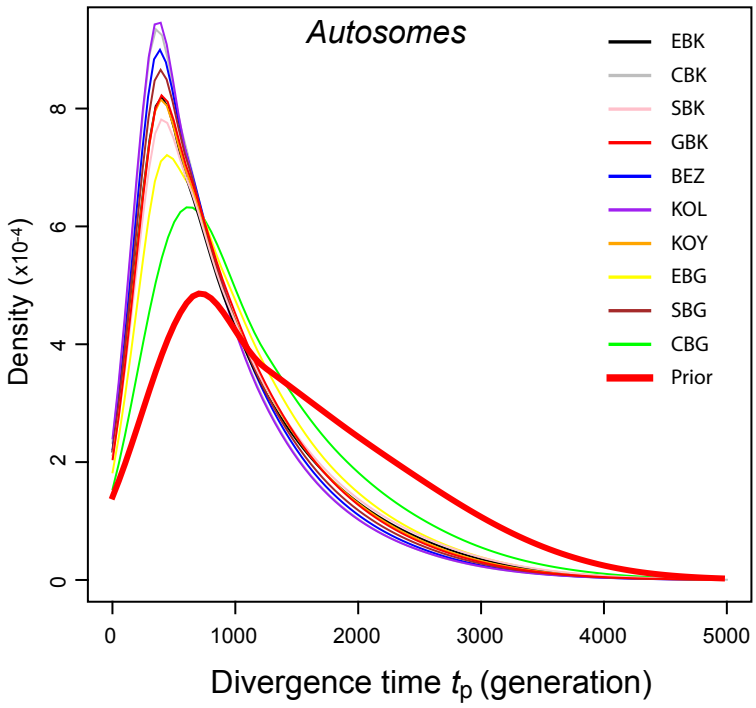
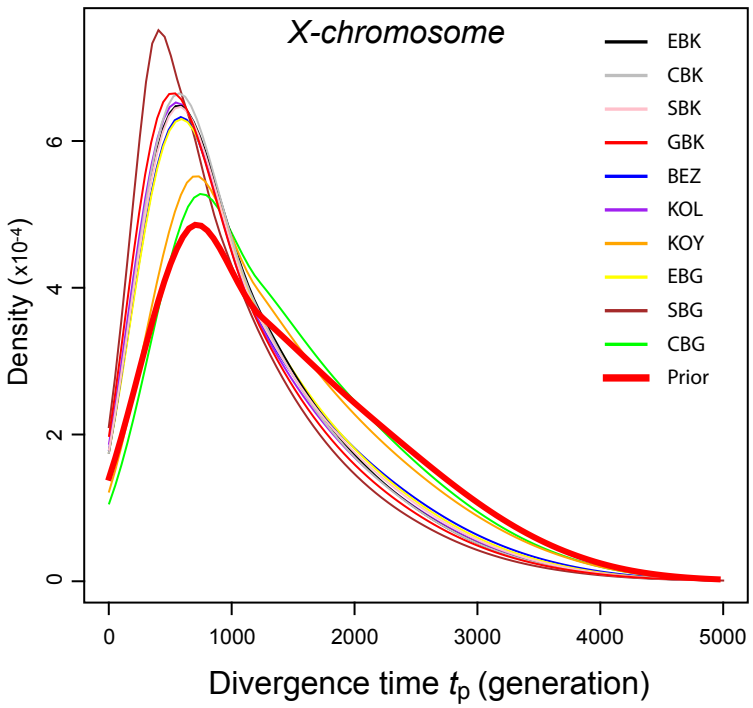
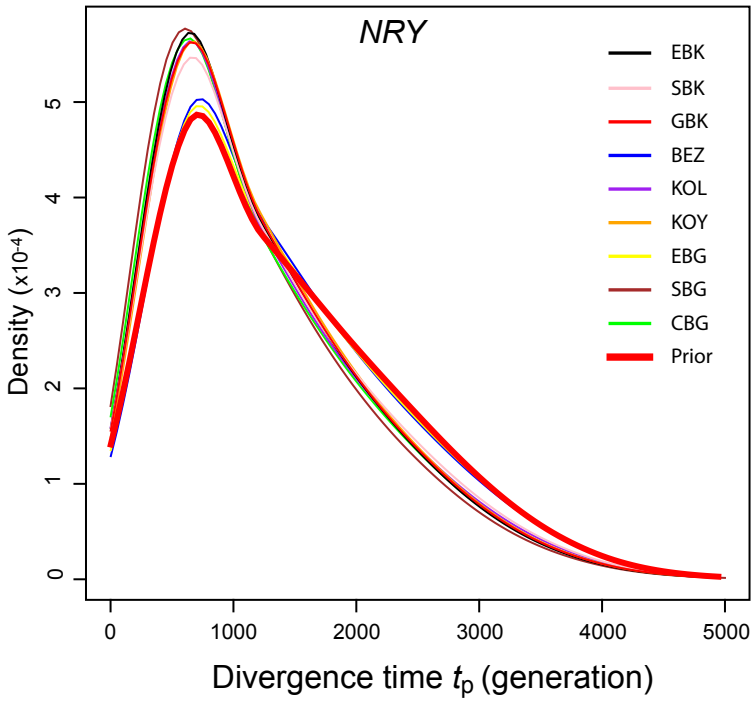
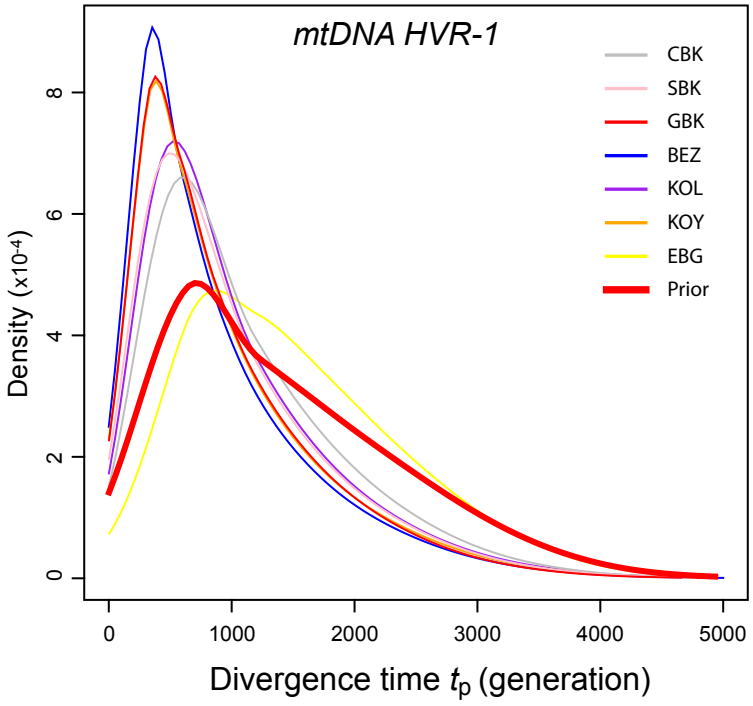

EBK
CBK
SBK
GBK
BEZ
KOL
KOY
EBG
CBG
SBG
Legend
quantile 95%
quantile 75%
Mode
Median
quantile 25%
quantile 5%
Effective population size
0
1000
2000
3000
4000
5000
6000
7000
8000
9000
10000
28 Autosomal Microsatellites
8 X-Chrom. Microsatellites
6 NRY Microsatellites
mtDNA HVR-1 359 bp

Figure S4


Autosomes
EBK
CBK
SBK
GBK
BEZ
KOL
KOY
EBG
SBG
CBG
Prior
Density (x10-4)
0
2
4
6
8
0
1000
2000
3000
4000
5000
Divergence time tp (generation)


X-chromosome
EBK
CBK
SBK
GBK
BEZ
KOL
KOY
EBG
SBG
CBG
Prior
Density (x10-4)
0
2
4
6
0
1000
2000
3000
4000
5000
Divergence time tp (generation)


NRY
EBK
SBK
GBK
BEZ
KOL
KOY
EBG
SBG
CBG
Prior
Density (x10-4)
0
1
2
3
4
5
0
1000
2000
3000
4000
5000
Divergence time tp (generation)


mtDNA HVR-1
CBK
SBK
GBK
BEZ
KOL
KOY
EBG
Prior
Density (x10-4)
0
2
4
6
8
0
1000
2000
3000
4000
5000
Divergence time tp (generation)