Journal of Biogeography

SUPPORTING INFORMATION

Whole mitochondrial genomes illuminate ancient intercontinental dispersals of grey wolves (*Canis lupus*)

Stephan Koblmüller, Carles Vilà, Belen Lorente-Galdos, Marc Dabad, Oscar Ramirez, Tomas Marques-Bonet, Robert K. Wayne, Jennifer A. Leonard

Appendix S1 Details of the approximate Bayesian computation (ABC) model testing and power analyses.

Table S1. Prior distributions for demographic parameters and mutation rate used for ABC analyses.

Demographic parameter	Type	Prior
N_1	N	UN~[2 x 10^3, 2 x 10^5]
N_2 ($\leq N_1$)	N	UN~[1 x 10^2, 5 x 10^4]
N_3 ($< N_2$)	N	UN~[1 x 10^2, 5 x 10^4]
N_4 ($\geq N_1$)	N	UN~[1 x 10^4, 2 x 10^5]
N_A	N	UN~[1 x 10^4, 2.5 x 10^5]
t_{old}	T	UN~[8 x 10^3, 1 x 10^4]
t_{recent}	T	UN~[2.6 x 10^3, 5 x 10^3]
r_1	A	UN~[0.51, 0.95]
r_2	A	UN~[0.01, 0.5]
Mutation model		TN93 (α = 0.114)
μ		UN~[1.85 x 10^{-7}, 3.00 x 10^{-7}]

Type of parameter: N, effective population size; T, time of the event in generations; A, admixture rate (see Fig. 2 for information on the demographic parameters). Priors were set as uniform distribution between minimum and maximum values. See Fig. S1 for the demographic parameters of each scenario tested. The mutation model used was TN93 (Tamura & Nei 1993) with the gamma shape parameter α = 0.114, as selected by JMODELTEST 0.1 (Posada 2008). μ, mutation rate.

Table S2. Parameter estimates for the best fitting scenario (SC1): mean, mode, median (Q_{50}) and quantiles.

Parameter	Mean	Mode	$Q_{2.5}$	$Q_{5.0}$	$Q_{25.0}$	Median	$Q_{75.0}$	$Q_{95.0}$	$Q_{97.5}$
N_1	31,000	**24,800**	17,100	18,700	24,100	28,900	36,700	48,100	52,700
N_2	7,750	**5,960**	3,960	4,520	5,960	7,460	9,210	12,200	13,300
N_A	61,900	**17,800**	12,300	14,900	29,800	50,800	81,000	152,000	171,000
t_{old}*	8,960	**9,440**	8,000	8,060	8,480	8,980	9,440	9,830	9,890
μ	2.34E-7	**2.03E-7**	1.85E-7	1.88E-7	2.06E-7	2.31E-7	2.59E-7	2.92E-7	2.98E-7

*time in number of generations (3 years per generation in *C. lupus*).
See Fig. 2 for the demographic parameters tested

Table S3. Model checking results for the best fitting scenario (SC1) based on 1,000 simulated datasets.

Summary statistics	Observed value	p-value ($S_{simul.} < S_{obs.}$)
NHA_1	33	0.1580
NHA_2	17	0.2435
NSS_1	338	0.5265
NSS_2	84	0.7425
MPD_1	43.717	0.1730
MPD_2	20.380	0.7290
VPD_1	330.001	0.0740
VPD_2	169.960	0.6730
DTA_1	-1.569	**0.0310***
DTA_2	0.071	0.5495
PSS_1	322	0.5510
PSS_2	68	0.7630
MNS_2	3.484	**0.0330***
MNS_2	6.274	0.5980
VNS_1	11.615	**0.0500***
VNS_2	28.032	0.6870
NH2_1&2	50	0.1425
NS2_1&2	406	0.4400
MP2_1&2	34.571	0.1840

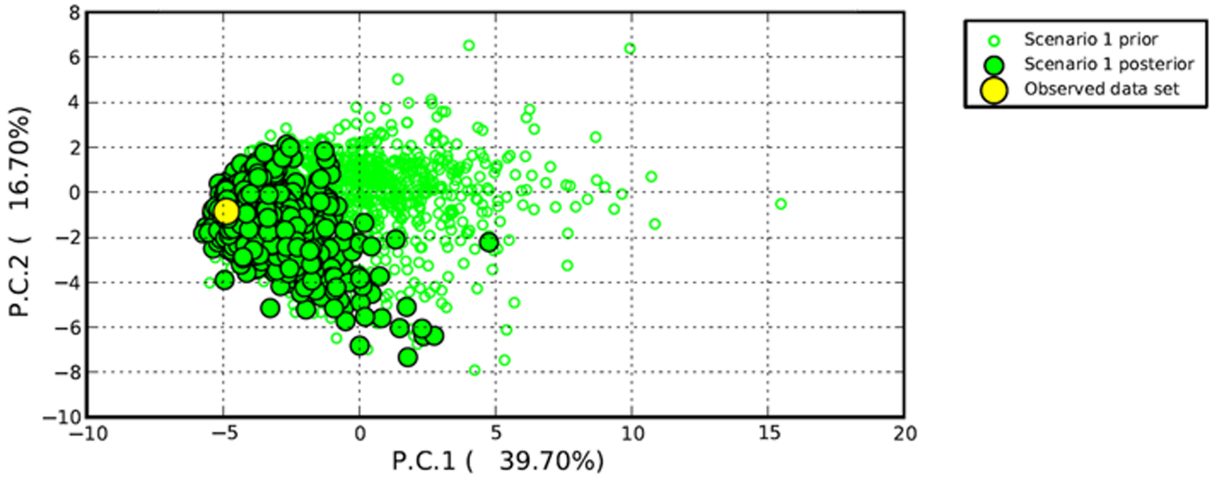
*all p-values were non-significant after applying the Benjamini & Hochberg (1995) correction for multiple testing.
Abbreviations for the summary statistics are as follows: number of haplotypes (NHA); number of segregating sites (NSS); mean pairwise distance (MPD); variance of pairwise distance (VPD); Tajima's D (DTA); private segregating sites (PSS); mean number of the rarest nucleotide at segregating sites (MNS); variance of numbers of the rarest nucleotide at segregating sites (VNS); number of distinct haplotypes in the pooled samples (NH2); number of segregating sites in the pooled samples (NS2); and the mean of pairwise differences between the two samples (MP2).

Table S4. ABC power analysis. Percentage of time a scenario (SC 1-4) was selected as best based on data simulated under the different scenarios (D 1-4).

Performance	SC1	SC2	SC3	SC4
D1	**60.0%**	14.0%[b]	11.6%[b]	26.0%[b]
D2	10.0%[a]	51.4%	22.8%	11.4%
D3	9.4%[a]	27.8%	48.0%	29.4%
D4	20.6%[a]	6.4%	17.6%	37.2%

D, proportion of cases in which the simulation-based model choice procedure was able to select a scenario as the most probable one.
[a]type-I error rate
[b]type-II-error rate

Figure S1. Fit of the selected scenario (SC1) with the observed data.



Appendix S2 Data for mitogenomes reconstructed from whole-genome shotgun sequencing data.

Table S5. Data for mitogenomes reconstructed from whole-genome shotgun sequencing data. Total number of reads in dataset, and whole genome coverage (X). Out of those reads, the number of reads used to reconstruct the mitochondrial genome using standard and modified assemblies, and the coverage that that yields (X). Length: length of the final mitochondrial genomes; Length woDloop: length excluding the repetitive d-loop.

Sample	Whole genome		Mitochondria Standard Assembly		Mitochondria Modified origin Assembly		Mitochondria	
	Total Reads	X*	Total Reads	X#	Total Reads	X#	Length	Length woDloop
MexicanWolfB	189,807,096	7.48	31,608	170	31,642	170	16,520	15,461
Italy2	203,474,928	8.02	952,846	5,126	954,906	5,137	16,520	15,460
YellowstoneC	226,260,028	8.92	755,396	4,064	754,536	4,059	16,528	15,460
MexicanWolfA	618,720,272	25.66	344,780	2,061	344,392	2,059	16,559	15,461
YellowstoneB	631,577,114	26.20	157,284	940	156,400	935	16,562	15,460
India2	661,696,716	27.45	358,338	2,142	357,874	2,139	16,529	15,460
YellowstoneA	680,714,876	28.23	249,002	1,488	248,222	1,484	16,599	15,460
Portugal	686,719,836	28.48	77,328	462	76,822	459	16,520	15,460
China5	694,048,646	28.79	286,646	1,713	286,722	1,714	16,579	15,460
Iberian	696,825,370	28.90	3,430,014	20,502	3,467,400	20,726	16,580	15,460
Iran2	727,480,320	30.17	200,158	1,196	198,310	1,185	16,541	15,460

*Coverage relative to the 2.4 Gbp of the dog assembly.

#Coverage relative to the length of the corresponding mitochondrial reference.