

## Supplementary Materials for

### **Whole-genome sequence analysis shows that two endemic species of North American wolf are admixtures of the coyote and gray wolf**

Bridgett M. vonHoldt, James A. Cahill, Zhenxin Fan, Ilan Gronau, Jacqueline Robinson, John P. Pollinger, Beth Shapiro, Jeff Wall, Robert K. Wayne

Published 27 July 2016, *Sci. Adv.* **2**, e1501714 (2016)  
DOI: 10.1126/sciadv.1501714

#### **The PDF file includes:**

- Supplementary Results
- Legend for table S1
- table S2.  $\hat{f}$  estimates for wolf and coyote ancestry proportions in each sample across the 38 autosomes.
- table S3. Estimates of total migration rates and migration proportions.
- table S4. Fraction of nonwolf alleles in the full-sequence data and in sequences downsampled to six times (in parenthesis).
- fig. S1. Profile likelihood curve for  $T$ , the time of split between gray wolves and coyotes under a simple isolation model (blue).
- fig. S2. PCA of downsampled genomes.
- fig. S3. Results of all possible  $D$  statistic tests for Eurasian wolf introgression into North American canids, following  $D(P1, P2, \text{Eurasian wolf, fox})$ , where P1 and P2 are all possible combinations of North American canids.
- fig. S4. Estimates of total migration rates under four different models assumed in four separate *G-PhoCS* runs.
- fig. S5. Estimates of population divergence times and effective population sizes in the four separate *G-PhoCS* runs.
- fig. S6. Estimates of divergence time obtained in eight separate runs of *G-PhoCS*.

#### **Other Supplementary Material for this manuscript includes the following:**

(available at [advances.sciencemag.org/cgi/content/full/2/7/e1501714/DC1](http://advances.sciencemag.org/cgi/content/full/2/7/e1501714/DC1))

- table S1 (Microsoft Excel format). The  $D$  statistic test for admixture in coyotes and North American hybrids.

## Supplementary Materials

### Supplementary Results

**table S1. The *D* statistic test for admixture in coyotes and North American hybrids.** Results are calculated separately for the *X* chromosome and the autosomes. We test for significance using the Weighted Block Jackknife (WBJ) test with 5 Mb non-overlapping blocks. Tests with Z-scores ( $D\text{-statistic}/\text{WBJ}$ )  $>3$  are considered significantly different from the non-admixed expectation of  $D=0$ . All tests utilize a Channel Island fox as the outgroup. **(A)** Wolves are designated as the introgressor population. **(B)** Coyotes are designated as the introgressor population. **(C)** Hybrid genomes are designated as the candidate introgressor. These are included for completeness but are not useful for assessment of introgression. Table S1 is included as an excel file due to its large size.

**table S2.  $\hat{f}$  estimates for wolf and coyote ancestry proportions in each sample across the 38**

**autosomes.** These estimates are lower bounds on the amount of ancestry and are not expected to sum to 100%. (Abbreviations: CAC, California coyote; Fox, Channel Island fox)

Sample	Coyote Ancestry $\hat{f}$ (India, Sample Tested, CAC, Alabama, Fox)	Wolf Ancestry $\hat{f}$ (CAC, Sample Tested, Iran, India, Fox)
GShep	0.00%	58.11%
Iran	0.00%	Not Calculated
Basenji	0.10%	61.11%
Qinghai	1.00%	56.87%
Alaska	8.00%	56.91%
Yellowstone1	8.40%	54.85%
Yellowstone2	8.10%	54.90%
Yellowstone3	8.50%	55.21%
Mexican	10.80%	52.87%
Wisconsin	21.60%	44.33%
Minnesota	22.30%	43.91%
IRNP	24.40%	43.27%
Algonquin1	36.80%	33.04%
Algonquin2	34.80%	36.00%
Redwolf1	53.10%	19.47%
Redwolf2	52.20%	20.32%
QuebecCoy	60.10%	15.78%
Ohio	68.50%	9.51%
Florida	69.20%	8.93%
Illinois	73.70%	5.96%
QuebecWolf	76.10%	-1.44%
Alabama	Not Calculated	0.14%

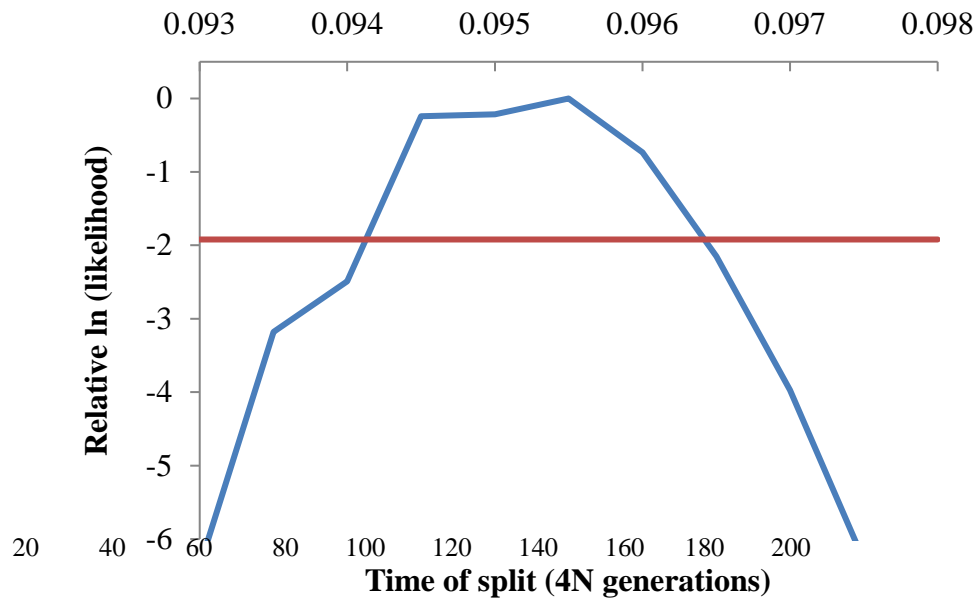
**table S3. Estimates of total migration rates and migration proportions.** See Fig. 4 for population phylogeny and ancestral population labels. Total migration rates ( $m$ ) measure the per-generation rate times the number of generations that gene flow was assumed to act. Total migration rates ( $m$ ) were converted to proportions ( $p$ ) in [0,1] using the following formula:  $p = 1 - e^{-m}$ . Migration proportions are shown in %. Significant gene flow is highlighted in bold (95% credible interval not overlapping 0).

Parameter	Total migration rates ( $m$ )			Probability of gene flow ( $p$ ) %		
	Posterior mean	95% Bayesian credible intervals		Posterior mean	95% Bayesian credible intervals	
		Lower	Upper		Lower	Upper
Great Lakes wolf → Red wolf	0.336	0.237	0.434	28.5	21.1	35.2
Coyote → Red wolf	1.244	0.647	2.101	71.2	47.7	87.8
Mexican wolf → Red wolf	0.064	0.050	0.077	6.2	4.9	7.4
Yellowstone wolf → Red wolf	0.152	0.111	0.191	14.1	10.5	17.4
Croatian wolf → Red wolf	0.001	0.000	0.008	0.1	0.0	0.8
Mongolian wolf → Red wolf	0.007	0.000	0.049	0.7	0.0	4.8
Basenji → Red wolf	0.002	0.000	0.017	0.2	0.0	1.6
Jackal → Red wolf	0.000	0.000	0.001	0.0	0.0	0.1
Red wolf → Great Lakes wolf	0.080	0.014	0.137	7.7	1.4	12.8
Coyote → Great Lakes wolf	0.343	0.280	0.413	29.0	24.5	33.8
Mexican wolf → Great Lakes wolf	0.073	0.049	0.096	7.1	4.8	9.2
Yellowstone wolf → Great Lakes wolf	0.565	0.450	0.643	43.1	36.2	47.4
Croatian wolf →	0.001	0.000	0.010	0.1	0.0	1.0

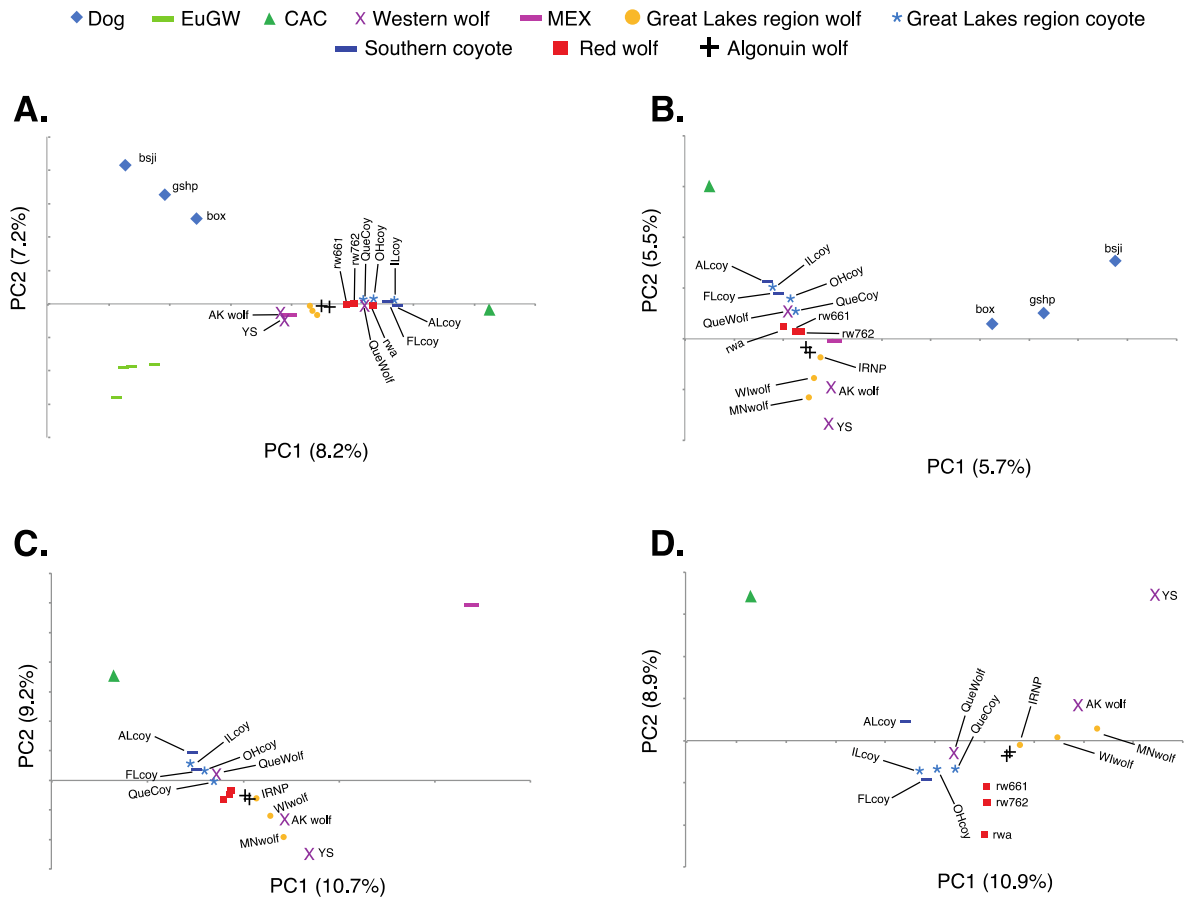
Great Lakes wolf						
Mongolian wolf	0.127	0.052	0.208	11.9	5.0	18.8
→						
Great Lakes wolf						
Basenji→	0.001	0.000	0.010	0.1	0.0	0.9
Great Lakes wolf						
Jackal→	0.001	0.000	0.002	0.0	0.0	0.2
Great Lakes wolf						
Jackal→	0.277	0.255	0.312	24.2	22.5	26.8
COY_ANC						

**table S4. Fraction of nonwolf alleles in the full-sequence data and in sequences downsampled to six times (in parenthesis).**

Sample	No. of sites	No. of heterozygous sites	No. of fixed sites	New SNP %	Coverage
Basenji*	8591507	266285	294571	4.98	21.93
GShep	8652651 (7918361)	405320 (230757)	243654 (40703)	5.16 (1.97)	23.86
California	8434329 (7436797)	1459150 (854898)	557970 (108485)	15.27 (7.21)	24.25
Minnesota	8555543 (7868660)	814069 (510339)	121391 (22313)	6.18 (3.53)	24.90
Alabama	6737143 (6753601)	698210 (742879)	81643 (82484)	6.39 (6.72)	5.10
Florida	7448132 (7472412)	891439 (946337)	99361 (101162)	7.32 (7.69)	7.10
Illinois	7150060 (7330514)	928826 (1002926)	111931 (113637)	8.06 (8.58)	6.21
Ohio	7307270 (7330514)	780494 (826622)	69941 (71547)	6.30 (6.61)	5.92
QuebecCoy	7624898 (7644138)	851674 (904197)	77979 (79504)	6.61 (6.95)	6.41
Redwolf1	8459689 (7513456)	1188107 (731333)	402646 (74659)	11.78 (5.86)	28.30
Redwolf2	7347258 (7381204)	601483 (619489)	62940 (64188)	4.95 (5.07)	5.57
Redwolf3	7540732 (584256)	724595 (739681)	89423 (90709)	5.99 (6.07)	6.29
Algonquin2	7618704 (7274856)	620387 (530482)	25094 (15651)	4.40 (3.86)	4.48
Algonquin1	7258357 (7626507)	497372 (661591)	14914 (26060)	3.63 (4.68)	5.60
IRNP	7254632 (7268550)	254076 (269571)	16420 (17117)	1.98 (2.09)	4.62
QuebecWolf	5672095 (5696353)	409620 (437227)	47487 (48569)	4.45 (4.69)	3.88
Wisconsin	7973893 (7994621)	535295 (573171)	38864 (41261)	3.84 (4.10)	7.49

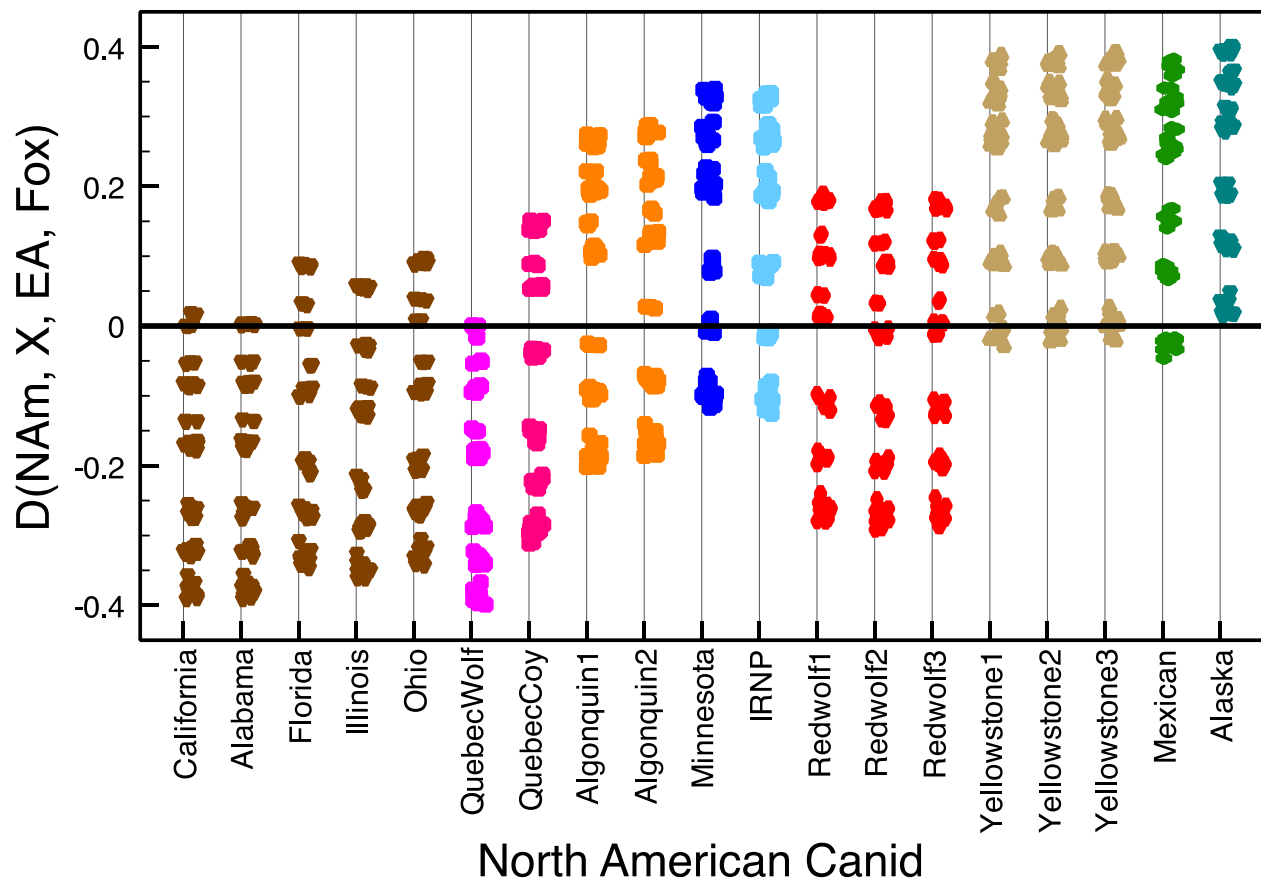


**fig. S1. Profile likelihood curve for  $T$ , the time of split between gray wolves and coyotes under a simple isolation model (blue). The cutoff for the approximate 95% confidence interval is shown in red.  $T$  is in units of  $4N$  generations, where  $N$  is the effective population size.**

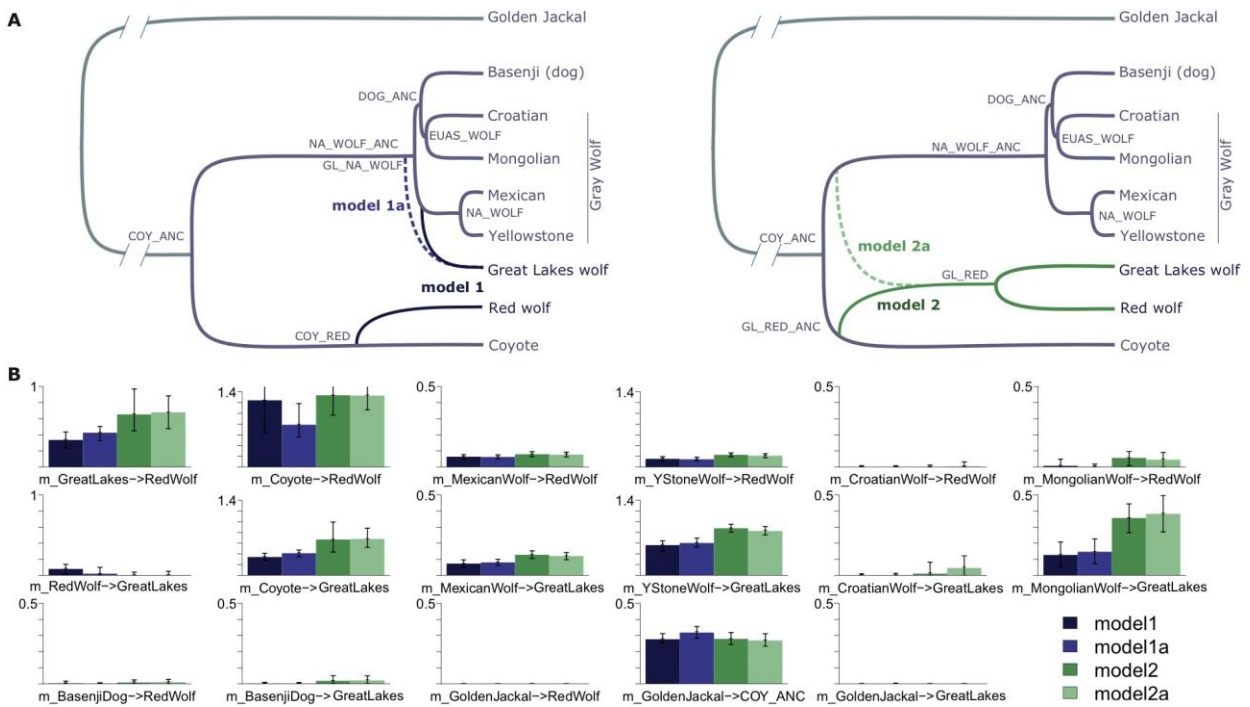


**fig. S2. PCA of downsampled genomes.** Each population or geographic representative is labeled in an analysis that excludes one population or species at a time. **(A)** Excludes the golden jackal; **(B)** Excludes the golden jackal and Eurasian gray wolf (EuGW); **(C)** Excludes golden jackal, EuGW, and dogs; **(D)** Excludes golden jackal, EuGW, dogs, and Mexican wolf.

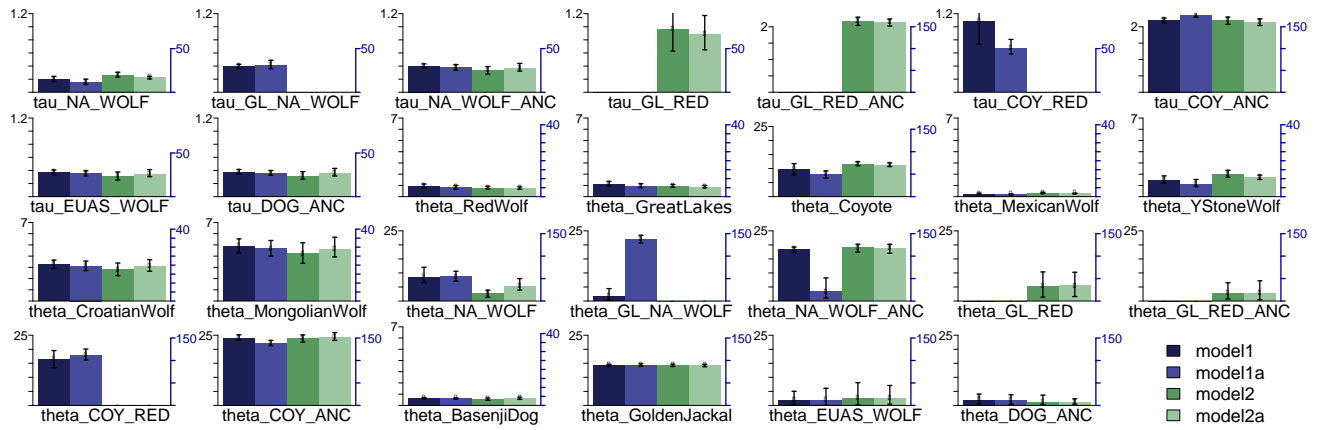




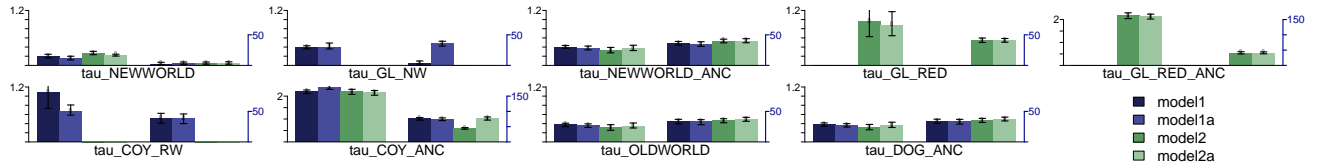
**fig. S3. Results of all possible  $D$  statistic tests for Eurasian wolf introgression into North American canids, following  $D(P1, P2, \text{Eurasian wolf}, \text{fox})$ , where  $P1$  and  $P2$  are all possible combinations of North American canids.** Each symbol represents a single  $D$ -statistic comparison. The individual indicated on the x-axis is  $P2$ . Individuals that always produce  $D$ -statistics less than or equal to zero have no detectable wolf ancestry. Apart from the California and Alabama coyotes and the Quebec wolf, all other individuals have Eurasian wolf ancestry either via descent or introgression.



**fig. S4. Estimates of total migration rates under four different models assumed in four separate *G-PhoCS* runs. (A)** The four different topologies assumed with labels representing the ancestral populations. In two, the red wolf and Great Lakes region wolf are not sister species (model 1 and 1a; left), and in the other two they are sister species (model 2 and 2a; right). Model 1 was the one used in the main analysis (Fig. 4). In all four models, the population phylogeny was augmented with 17 migration bands (see Methods). **(B)** Total migration rates inferred for each of the 17 migration bands in the model in each of the four separate runs. Each migration band is specified by the source and target populations. Bars represent posterior means and error bars represent 95% Bayesian credible intervals. The total migration rate equals the per-generation rate times the number of generations that migration is allowed. Estimates are fairly compatible across the different models, with overall slightly lower rates inferred for model 1 used in the main analysis.



**fig. S5. Estimates of population divergence times and effective population sizes in the four separate *G-PhoCS* runs.** Each parameter is specified as tau (divergence time) or theta (effective population size) of a branch in the population phylogeny. See fig. S3A for the four different topologies assumed and the names of ancestral populations. Divergence times and effective population sizes are given in their raw mutation-scaled values scaled up by a factor of 10,000 (left Y axis), and in their calibrated form (in thousands of years or thousands of individuals; right Y axis). Calibration is done by assuming an average generation time of three years and an average per-generation mutation rate of  $\mu=4 \times 10^{-9}$  / base pair. Bars represent posterior means and error bars represent 95% Bayesian credible intervals. Estimates in all four models are consistent with previous analysis of Eurasian wolves and dogs (26). Divergence of red wolves and Great Lakes region wolves in these models likely captures the divergence of the main populations (coyotes and North American gray wolves) contributing to these two admixed lineages.



**fig. S6. Estimates of divergence time obtained in eight separate runs of *G-PhoCS*.** The models used in each run correspond to the four different topologies (fig. S3), with gene flow (left four bars) and without gene flow (right four bars). Labels of ancestral populations are as indicated in fig. S3A. Estimates of divergence time are given in their raw mutation-scaled values scaled up by a factor of 10,000 (left Y axis), and in their calibrated form (in thousands of years or thousands of individuals; right Y axis). Calibration is done by assuming an average generation time of three years and an average per-generation mutation rate of  $m=4 \times 10^{-9}$  / base pair. Bars represent posterior means and error bars represent 95% Bayesian credible intervals (same as in fig. S4).