

Supplementary information

Adaptive introgression and the origin of dog starch digestion

Rina Okada, Jun Gojobori, and Yohey Terai*

Types	Sample	chr16:7,156,695
Europe	Chihuahua	C C
	Labrador Retriever	C C
	Border Collie	C C
	Standard Poodle	C C
	German Shepherd	C C
	Boston Terrier	C C
Africa	Basenji	C C
	Nigeria dog 1	C C
	Nigeria dog 2	C C
	Nigeria dog 3	C C
	Nigeria dog 4	C C
	Nigeria dog 5	C C
Japan / East Asia	Nigeria dog 6	C C
	Akita	C C
	Kishu	C C
	Chinese dog 1	C T
	Chinese dog 2	C C
	Vietnamese dog	C C
South Asia / Central Asia / Oceania	DongGyeong	C C
	Borneo village dog	T T
	Central Asian Shepherd	C C
	Dingo 1	C T
	Dingo 2	T T
	Dingo 3	C C
Arctic	Dingo 4	C C
	NGSD 1	T C
	NGSD 2	C C
	Greenland sled dog 1	T T
	Greenland sled dog 2	C T
	Greenland sled dog 3	T C
Gray wolves	Siberian Husky	T T
	East Siberian Laika	C C
	Samoyed	C C
	Himalayan wolf	T T
	Chinese wolf	T T
	Indian wolf	T T
Other canids	Israel wolf	T T
	Croatian wolf	T T
	Iberian wolf	T T
	Daneborg wolf	T T
	BaffinNorth wolf	T T
	Minnesota wolf	T T
Other canids	African golden wolf 1	C C
	African golden wolf 2	T T
	African golden wolf 3	T T
	African golden wolf 4	C C
	African golden wolf 5	T C
	African golden wolf 6	C C
	African golden wolf 7	C C
	Coyote	T C
	Dhole	T T
	Ethiopian wolf	C C
	Golden jackal	T T
	Andean fox	T T

Dogs

Figure S1 Genotypes of the *MGAM* nonsynonymous SNP (CanFam3.1:chr16:7,156,695). Ancient samples were excluded because DNA damage affected this transition site. The table shows the genotypes of 42 modern dogs and gray wolves, together with 12 individuals representing six other canid species. Colors indicate the genotypes: orange (C|C), blue (T|T), and white (heterozygous).

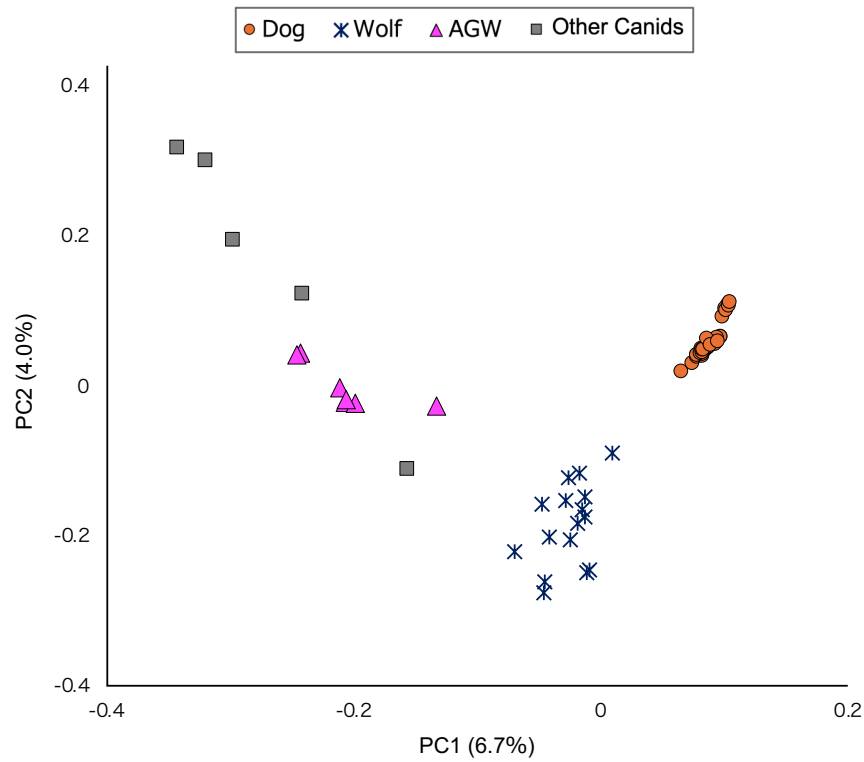


Figure S2 Principal component analysis (PCA) based on SNPs across chromosome 16. Samples include modern and ancient dogs, gray wolves, and other canid species. Consistent with the whole-genome data, dogs, gray wolves, and other canids formed clearly separated clusters.

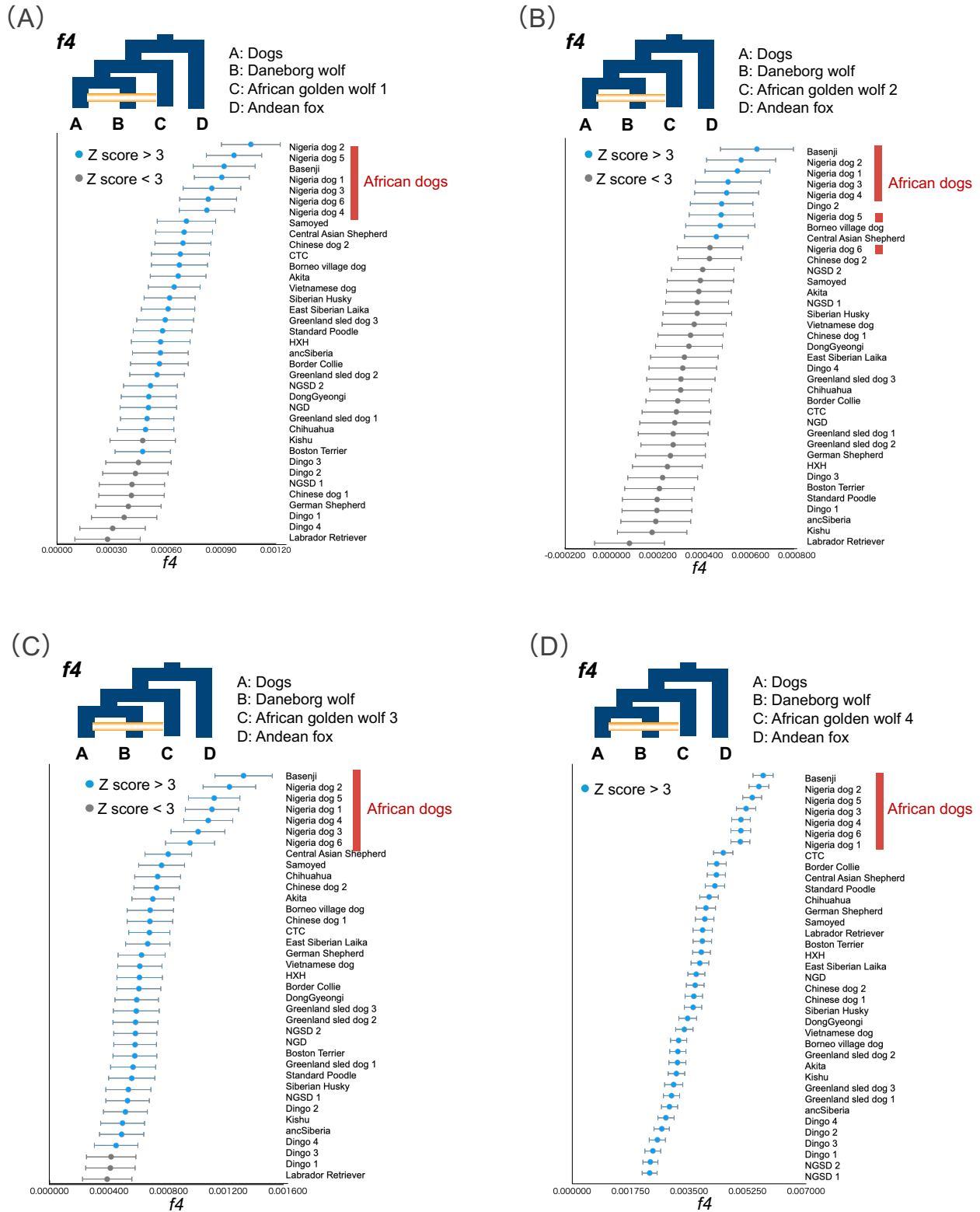
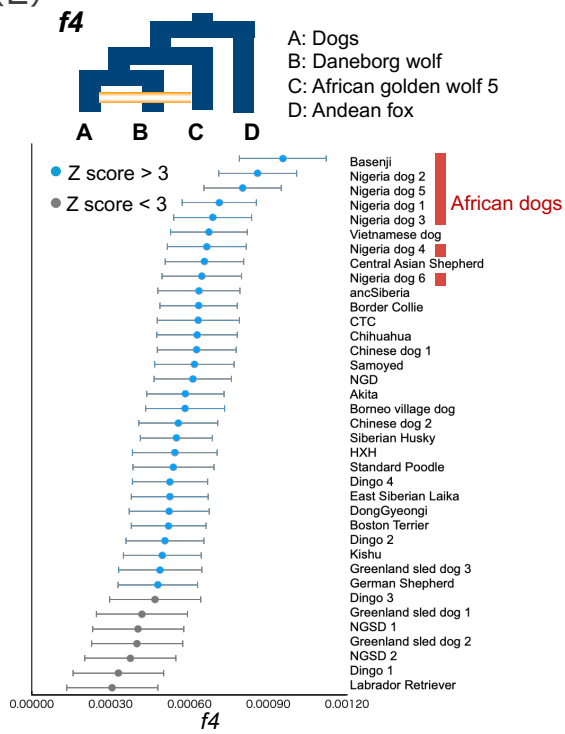
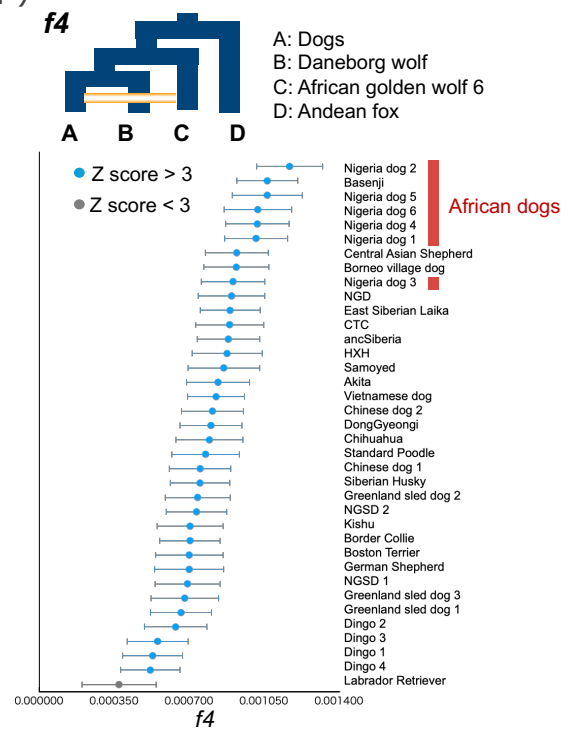


Figure S3 Genetic affinity between each African golden wolf individual ($n = 7$; A–G) and dog individuals shown by Patterson's f_4 -statistics. Error bars indicate standard errors. Blue dots indicate significant allele sharing between dogs and African golden wolves (Z score > 3).

(E)



(F)



(G)

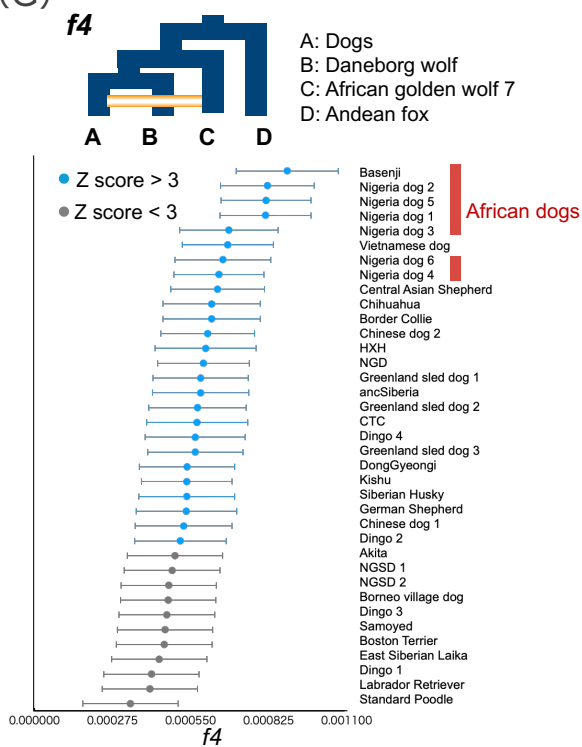


Figure S3 continued.

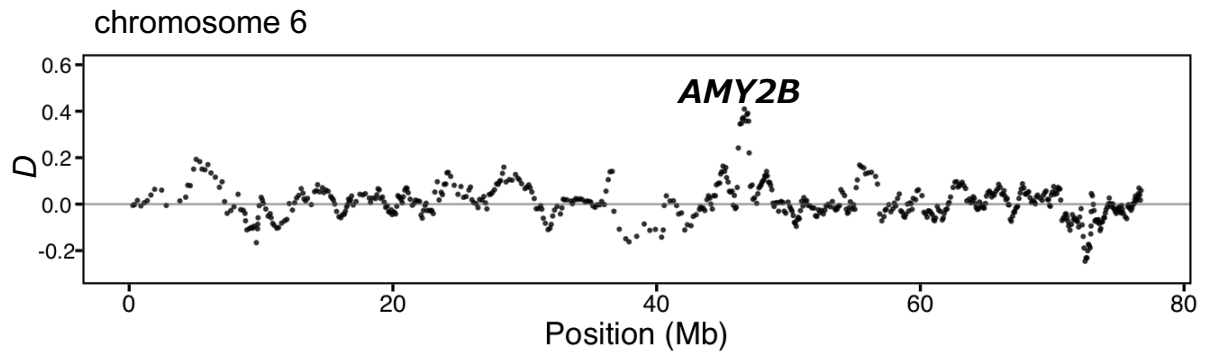


Figure S4 Sliding-window results of Patterson's D -statistics on chromosome 6. To eliminate potential bias due to copy-number variation, we used dog individuals with two copies per diploid genome that carry the dog-type haplogroup. A high- D -value peak was observed within the *AMY2B* region.

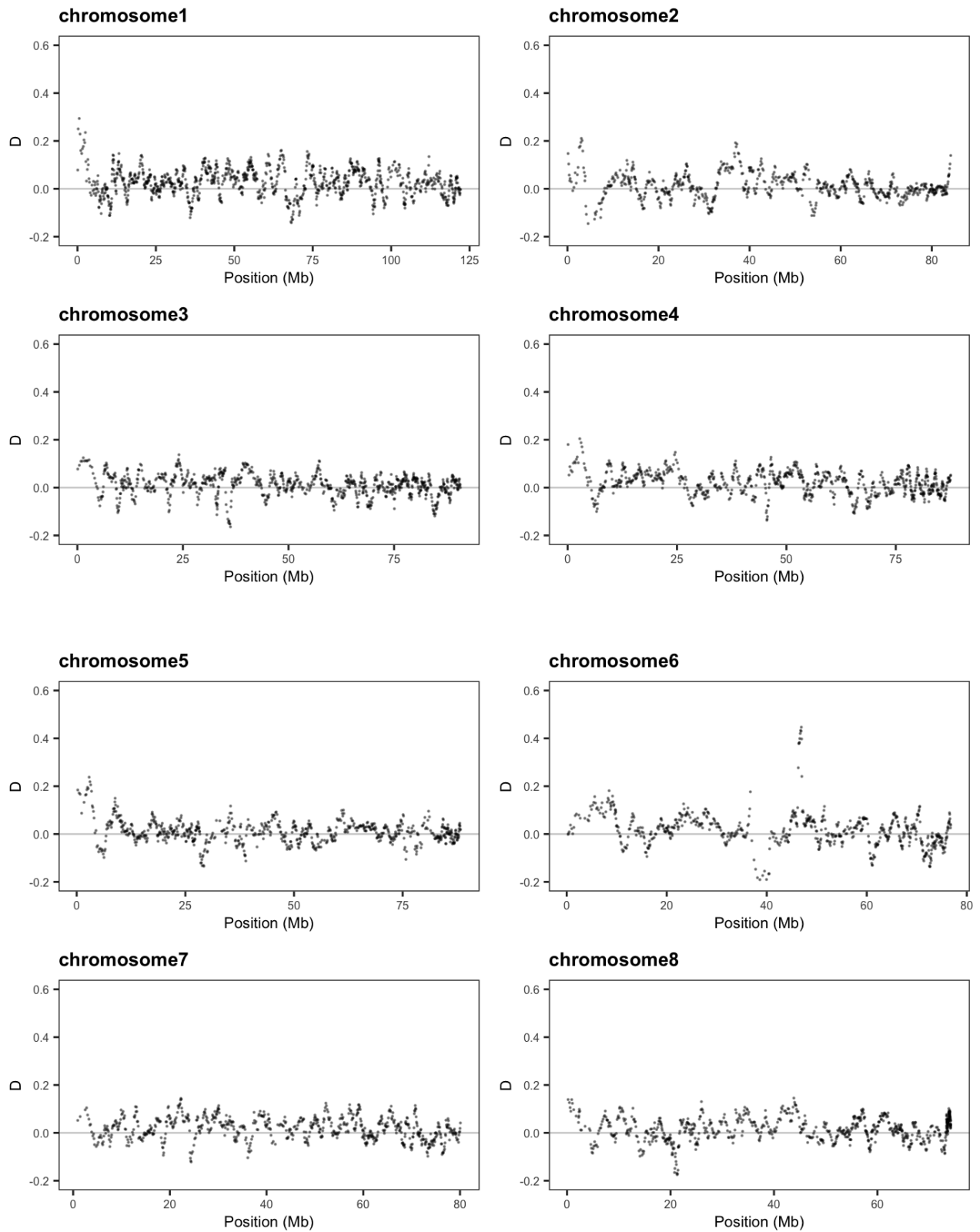


Figure S5 Sliding-window results of Patterson's D -statistics across all 38 autosomes.

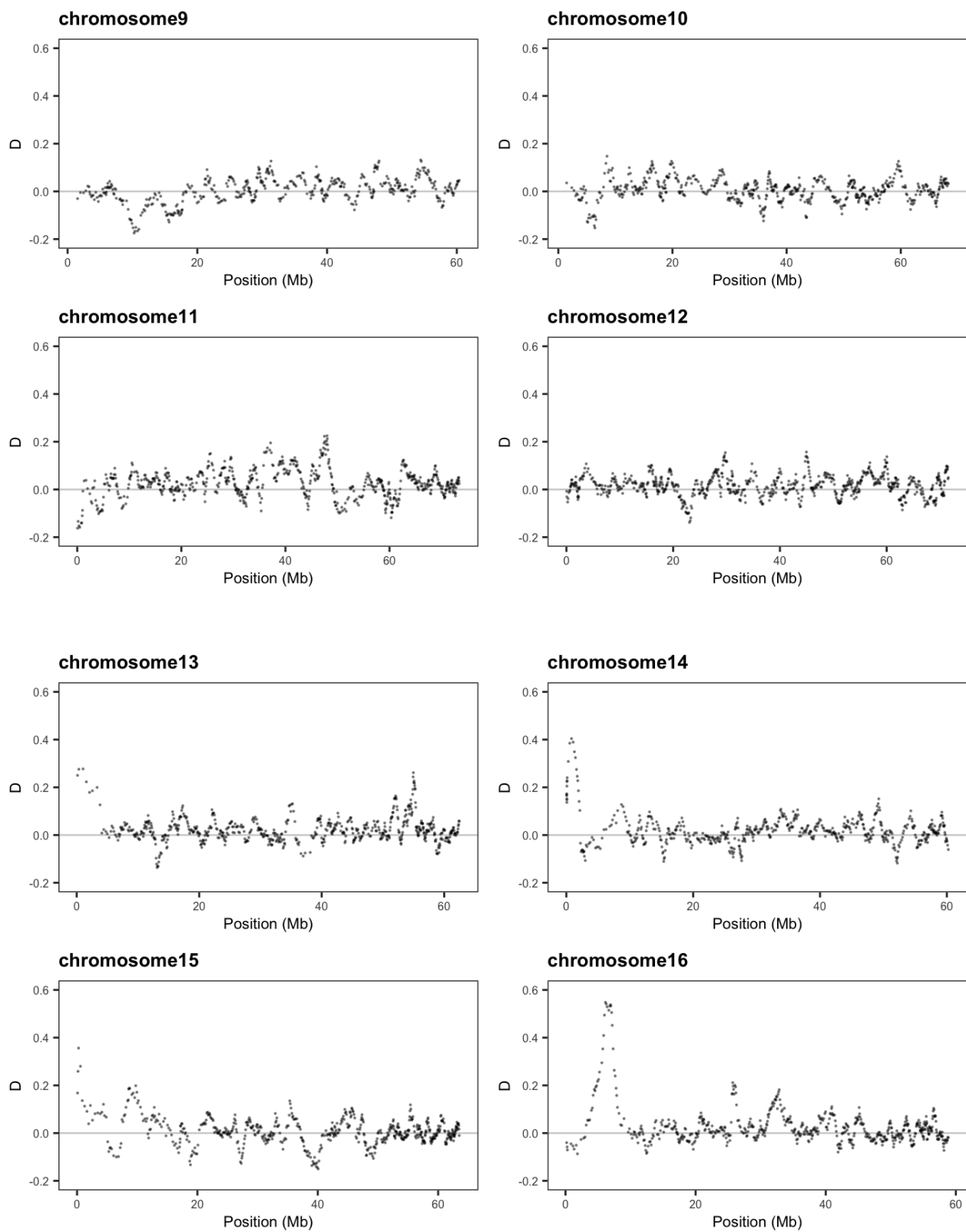


Figure S5 continued.

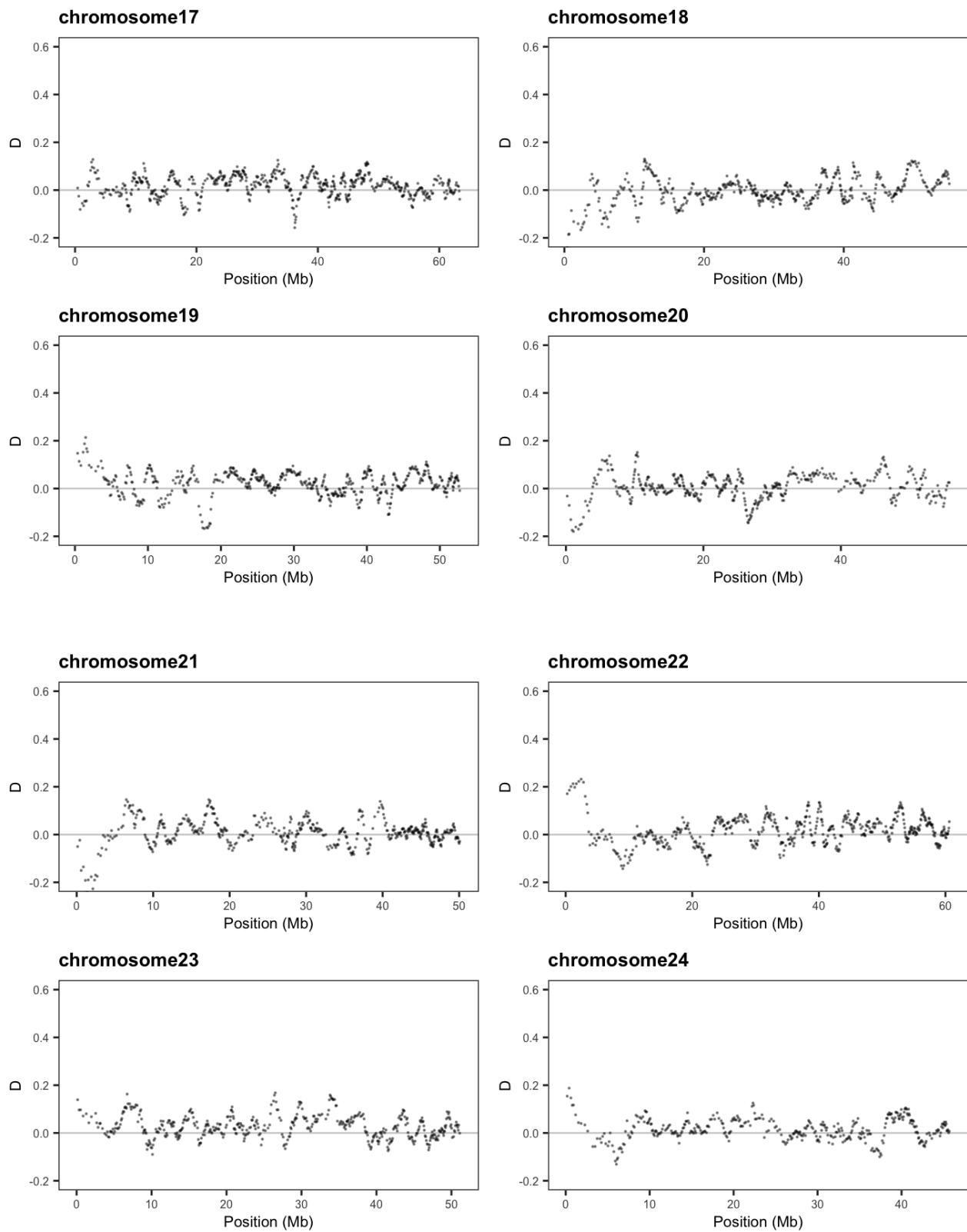


Figure S5 continued.

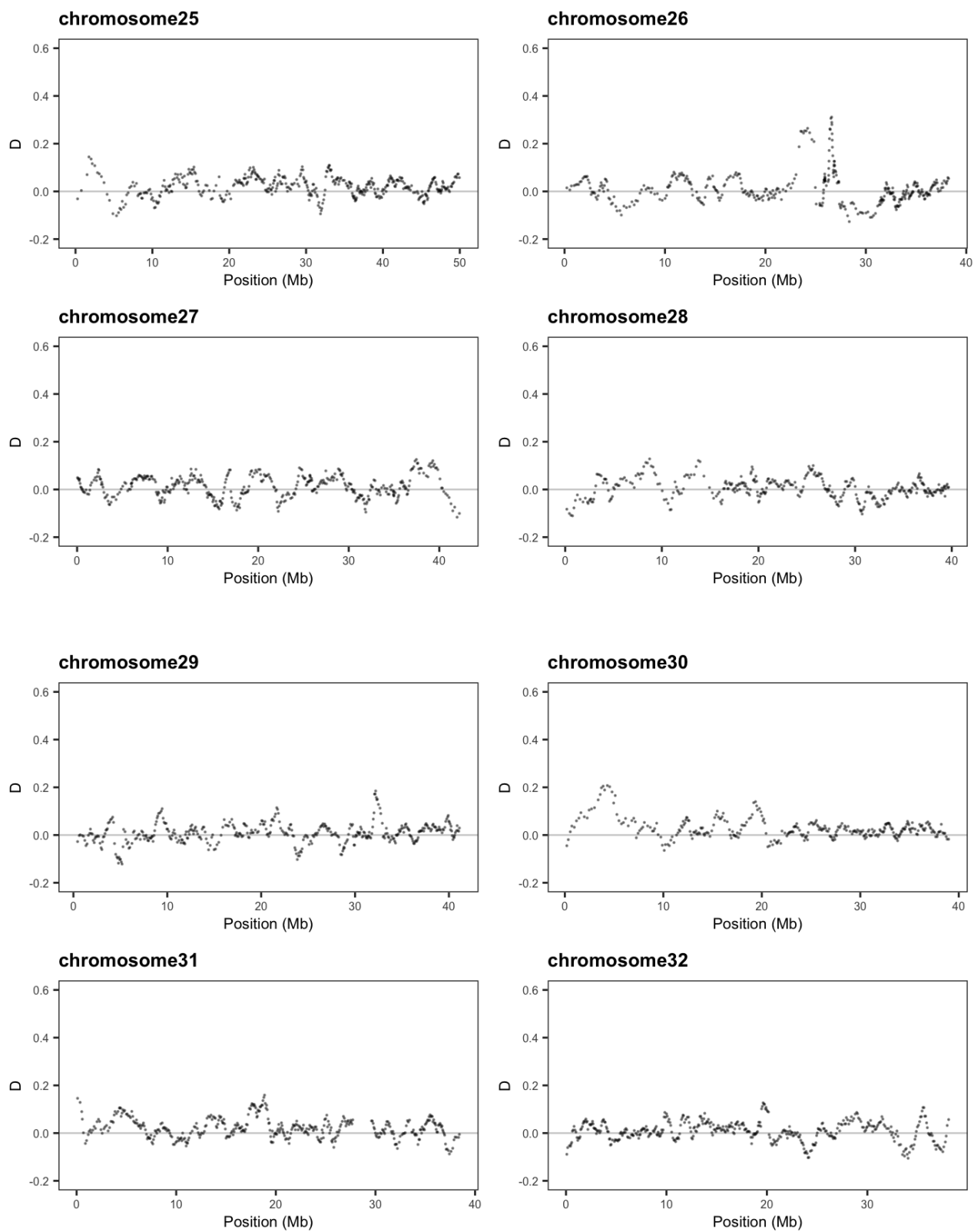


Figure S5 continued.

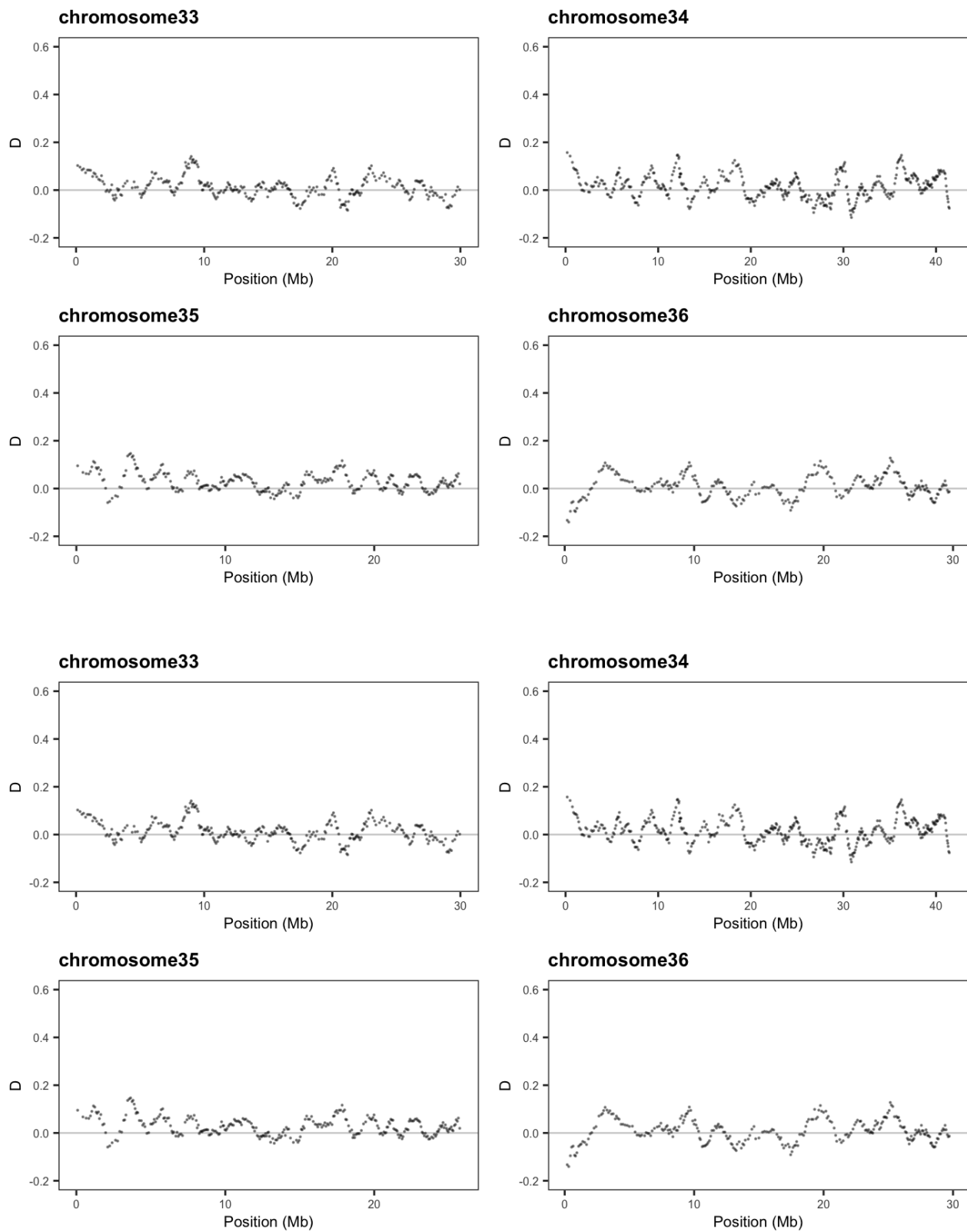


Figure S5 continued.

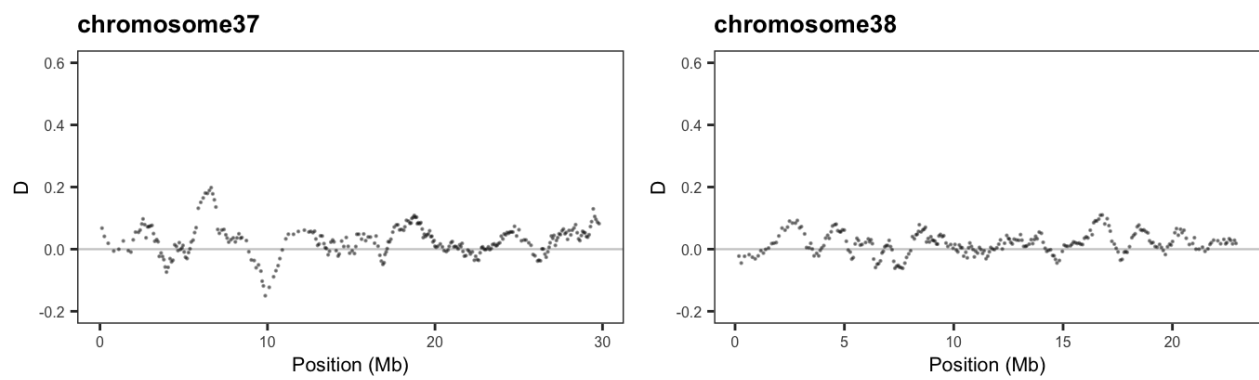


Figure S5 continued.

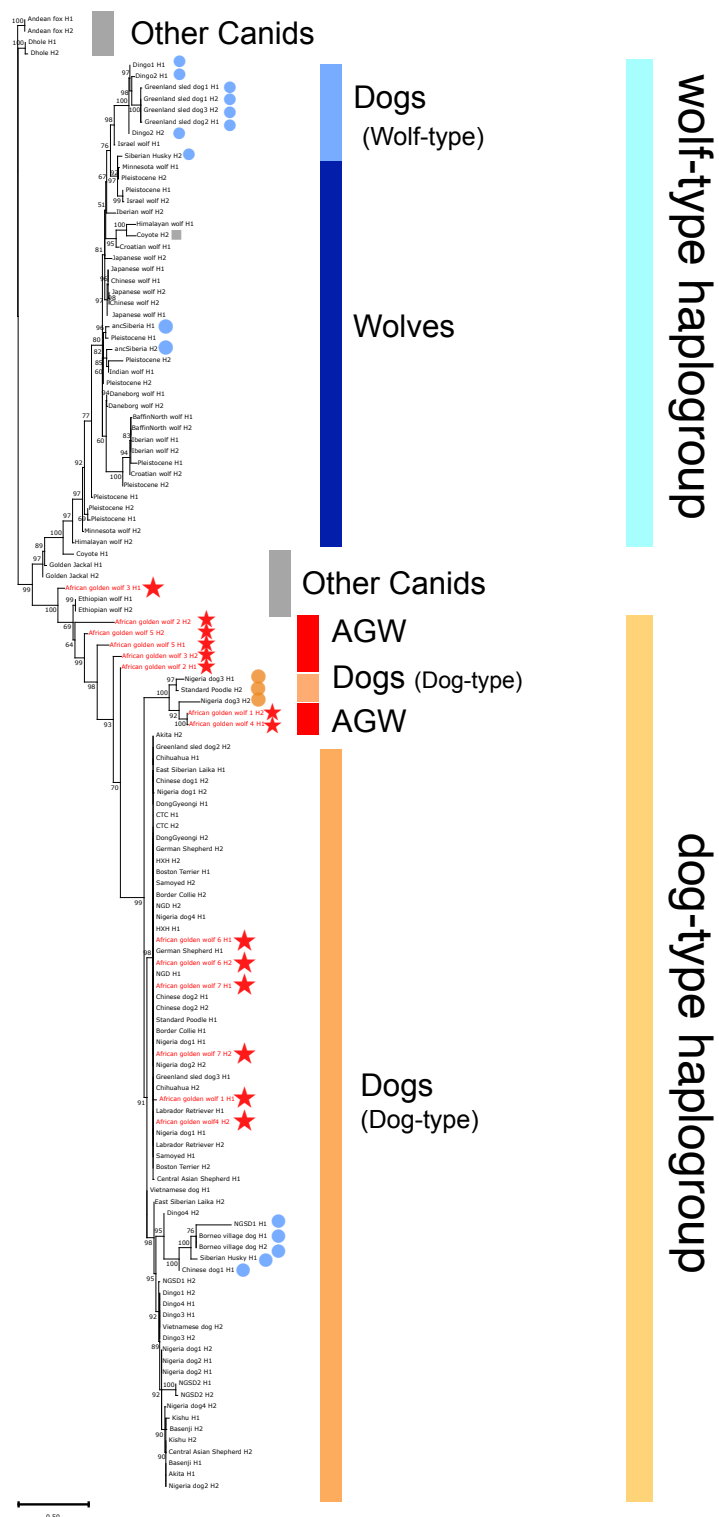


Figure S6 Maximum-likelihood (ML) phylogenetic tree for the *MGAM* region (chr16:7,125,960–7,303,209).

Stars (★) and circles (●) indicate African golden wolves and dogs, respectively. Blue and orange circles indicate dogs that carry the T and C alleles at the *MGAM* nonsynonymous SNP (CanFam3.1:chr16:7,156,695), respectively.

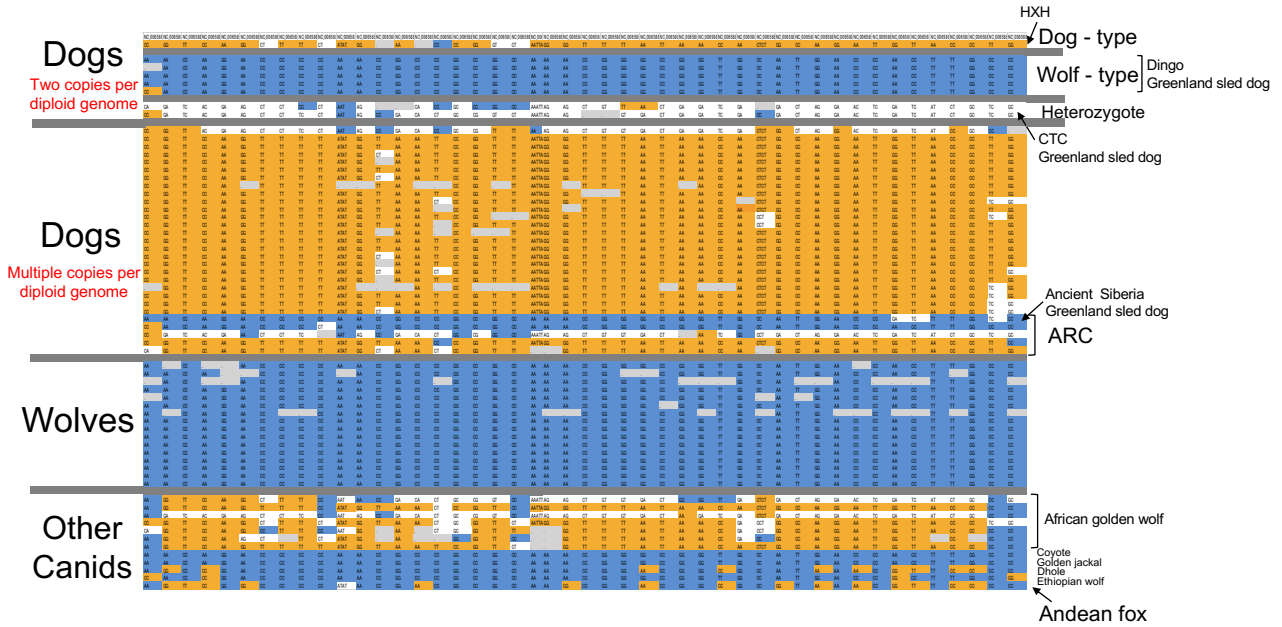


Figure S7 Genotype distribution in the *AMY2B* region. The genotype at the genomic region containing the *AMY2B* gene (chr6:46,854,109–47,454,177). Rows represent individuals, and columns represent positions of SNPs, respectively. SNPs that are dominant in gray wolves and dogs are shown in blue and orange backgrounds, respectively. A white background indicates a heterozygous site, and a gray background indicates missing data. Since most dogs exhibit an increased copy number in the *AMY2B* region, these genotype patterns represent the major haplotypes in each dog.

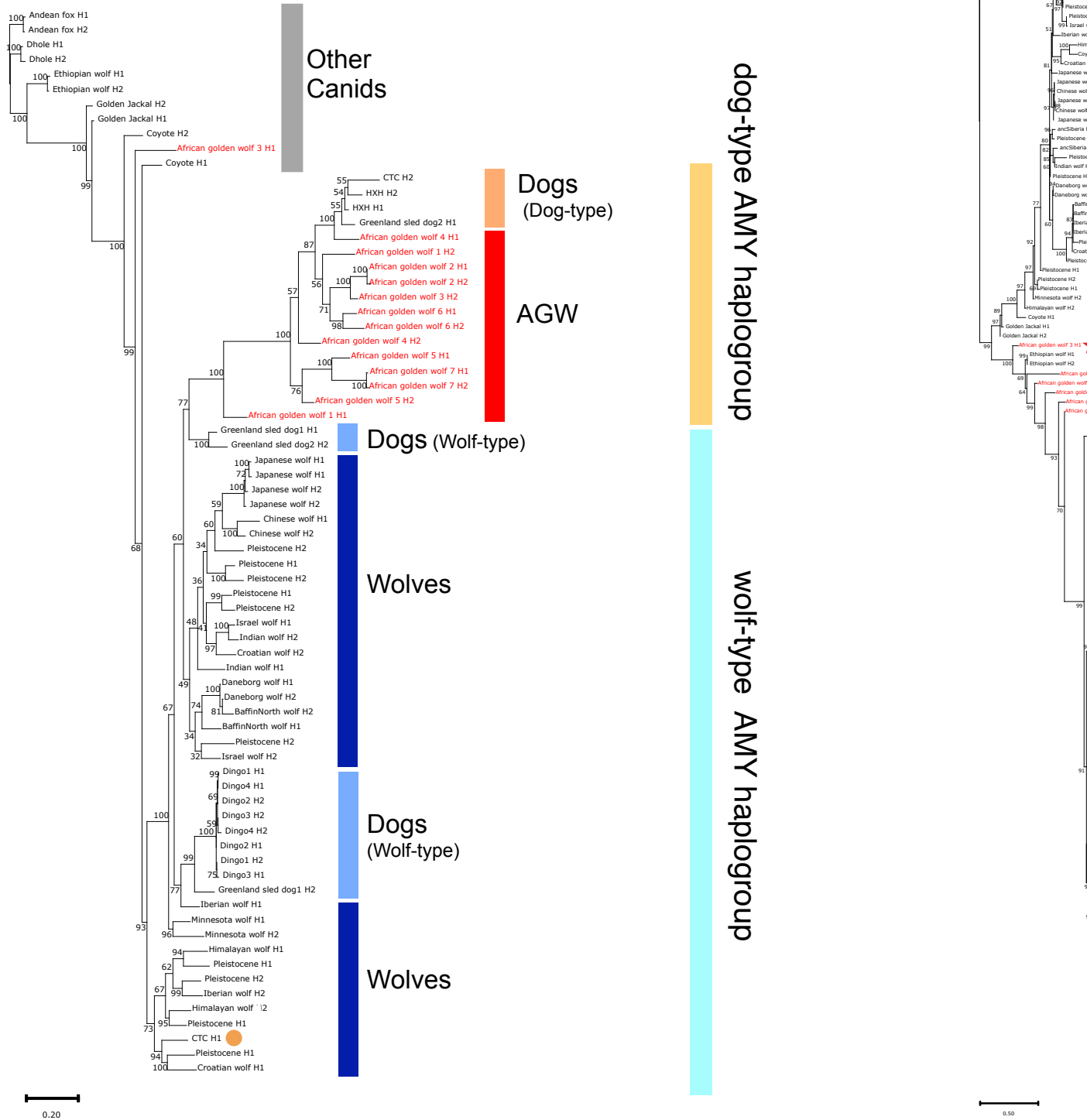


Figure S8 Maximum-likelihood (ML) phylogenetic tree for the *AMY2B* region (chr6:46,854,109–47,454,177). To eliminate potential bias from copy number variation (CNV), we used dogs with two copies per diploid genome. Circles (●) indicate an ancient dog within wolves.

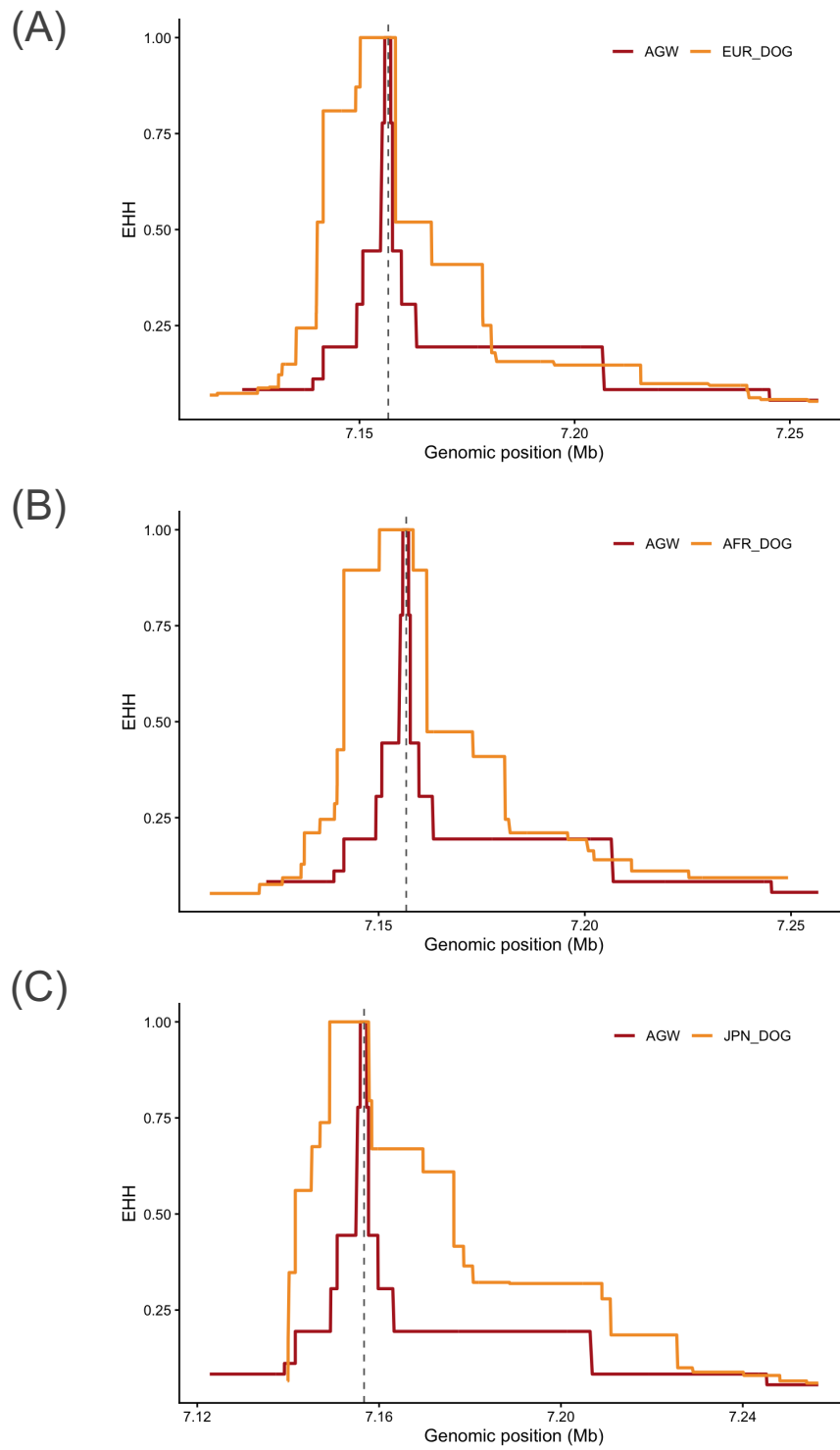


Figure S9 Extended haplotype homozygosity (EHH) analysis anchored at the nonsynonymous SNP (C allele) (CanFam3.1:chr16:7,156,695) within the *MGAM* gene region. All the dogs from Europe and the Americas (EUR_DOG)(A), Africa (AFR_DOG)(B), and Japan (JPN_DOG)(C) showed longer haplotypes around this SNP than African golden wolves (AGW; $n = 7$).

Table S1 Sample information, an attached file.

Table S2 Windows in the top 0.1% of Patterson's *D*-statistics

Chromosome	windowStart	windowEnd	<i>D</i>
6 (NC_006588.3)	46542534	47292691	0.383212
	46595286	47392051	0.39918
	46693308	47550298	0.420548
	46775875	47636220	0.433236
	46860834	47706241	0.430036
	46918327	47803981	0.446849
	46979457	47879629	0.397635
14 (NC_006596.3)	538336	2133765	0.384863
	820072	2225606	0.404132
	1077577	2338282	0.389319
16 (NC_006598.3)	5865837	7183327	0.409204
	5987840	7320597	0.494392
	6140225	7488285	0.548187
	6259392	7635903	0.541007
	6404447	7772777	0.529236
	6659724	7866524	0.514812
	6818813	7981867	0.532775
	6917471	8211925	0.538047
	6967255	8472024	0.533101
	7090014	8646614	0.505388
	7183400	8990857	0.45177

Table S3 Windows with $D \geq 0.35$ on chromosome 6

Chromosome	windowStart	windowEnd	D
6 (NC_006588.3)	46498567	47309406	0.367424
	46578411	47428235	0.372577
	46650810	47583996	0.409448
	46754791	47674811	0.358159
	46847108	47750551	0.385569
	46910903	47853017	0.391117
	46977830	47914866	0.35722

* Only individuals carrying two copies of the dog-type haplogroup per diploid genome were used for the calculations.

Table S4 Annotation of genes located within regions corresponding to the top 0.1% of *D*-statistics

Chromosome	Start	Stop	Gene symbol	NCBI Gene ID	UniProt Protein
6 (NC_006588.3)	46960033	47044272	<i>RNPC3</i>	490136	RNA-binding region-containing protein 3
	47425387	47622189	<i>COL11A1</i>	100685969	Collagen type XI alpha 1 chain
	-	-	<i>AMY2B</i>	-	-
14 (NC_006596.3)	604444	605584	<i>H2AC25</i>	482203	Histone H2A
	633767	639543	<i>TRIM17</i>	482204	E3 ubiquitin-protein ligase <i>TRIM17</i>
	640375	647427	<i>TRIM11</i>	482205	E3 ubiquitin-protein ligase <i>TRIM11</i>
	664404	782376	<i>OBSCN</i>	482206	-
	799540	804191	<i>IBA57</i>	482207	Iron-sulfur cluster assembly factor <i>IBA57</i> , mitochondrial
	810388	819248	<i>GJC2</i>	607294	Gap junction protein
	820048	826694	<i>GUK1</i>	475178	Guanylate kinase
	826731	830107	<i>MRPL55</i>	607317	Mitochondrial ribosomal protein L55
	832005	833681	<i>C14H1orf35</i>	607351	Multiple myeloma tumor-associated protein 2-like N-terminal domain-containing protein
	835476	836424	<i>ARF1</i>	100856517	ADP-ribosylation factor
	864393	899698	<i>WNT3A</i>	482208	Protein Wnt
	946843	959289	<i>WNT9A</i>	482209	Protein Wnt
	997712	1007499	<i>PRSS38</i>	475179	Trypsin
	1058891	1104568	<i>SNAP47</i>	475181	Synaptosomal-associated protein 47
	1100002	1109965	<i>JMJD4</i>	475182	2-oxoglutarate and iron-dependent oxygenase <i>JMJD4</i>
	1387375	1388310	<i>OR2L17</i>	482213	Olfactory receptor
	1582244	1599482	<i>TRIM58</i>	482216	Tripartite motif containing 58
	1661954	1662749	<i>OR6AB1P</i>	482218	-
	1896389	1897357	<i>OR2AV2</i>	482223	Olfactory receptor family 2 subfamily AG member 2
	2051227	2052165	<i>OR2L5</i>	482224	Olfactory receptor
	2104120	2105058	<i>OR2T24</i>	482227	Olfactory receptor
	2127466	2128404	<i>OR2T15</i>	482228	Olfactory receptor
	2146120	2148523	<i>OR2T13</i>	482229	Olfactory receptor
16 (NC_006598.3)	5884765	5885780	<i>OR6B1</i>	482725	Olfactory receptor
	5942879	5943832	<i>OR2F1D</i>	482726	Olfactory receptor
	5966896	5967849	<i>OR2F1</i>	482727	-
	5992676	5993629	<i>OR2F1C</i>	482728	-
	6109869	6132436	<i>TCAF2</i>	482731	TRPM8 channel associated factor 2
	6230851	6231777	<i>TAS2R41</i>	482734	Taste receptor type 2
	6255260	6257954	<i>TAS2R60</i>	106559803	-
	6287757	6304082	<i>EPHA1</i>	482735	Ephrin type-A receptor 1
	6304163	6313923	<i>ZYX</i>	482736	Zyxin
	6335862	6343235	<i>FAM131B</i>	608789	Family with sequence similarity 131 member B
	6342772	6373580	<i>CLCN1</i>	403723	Chloride channel protein 1
	6378447	6396251	<i>CASP2</i>	482737	Caspase-2
	6396724	6645221	<i>KEL</i>	100855519	Kell metallo-endopeptidase (Kell blood group)
	6397507	6401385	<i>TMEM139</i>	608825	Transmembrane protein 139
	6411678	6416135	<i>GSTK1</i>	475518	Glutathione S-transferase kappa
	6439097	6440017	<i>TAS2R40</i>	608842	Taste receptor type 2
	6465643	6466605	<i>TAS2R39</i>	100271735	Taste receptor type 2
	6491424	6498175	<i>PIP</i>	608856	-
	6556517	6560189	<i>OR6W1</i>	482740	Olfactory receptor
	6645433	6646953	<i>LLCFC1</i>	100855808	LLLL and CFNLAS motif containing 1
	6647070	6678080	<i>TRPV5</i>	100855561	Transient receptor potential cation channel subfamily V member 5
	6691521	6706205	<i>TRPV6</i>	119874628	Transient receptor potential cation channel subfamily V member 6
	6706302	6718566	<i>EPHB6</i>	119874627	Ephrin type-B receptor 6
	6778109	6781404	<i>PRSS2</i>	100686744	Trypsin
	7053433	7119396	<i>MGAM2</i>	482756	Maltase-glucoamylase 2 (putative)
	7124343	7218035	<i>MGAM</i>	475523	Maltase-glucoamylase
	7224077	7225027	<i>TAS2R38</i>	100271737	Taste receptor type 2
	7242424	7269986	<i>CLEC5A</i>	609161	C-type lectin domain family 5 member A
	7293612	7295559	<i>OR9A7</i>	482759	-
	7406855	7407742	<i>TAS2R5</i>	100271743	Taste receptor type 2
	7424277	7425302	<i>TAS2R4</i>	100688996	Taste receptor type 2
	7435147	7436097	<i>TAS2R3</i>	100271736	Taste receptor type 2
	7445351	7461513	<i>SSBP1</i>	475524	Single-stranded DNA-binding protein
	7466651	7485150	<i>WEE2</i>	475525	Wee1-like protein kinase 2
	7531025	7559156	<i>DENND11</i>	609262	DENN domain-containing protein 11
	7561084	7637017	<i>AGK</i>	482764	Acylglycerol kinase, mitochondrial
	7698154	8042210	<i>TMEM178B</i>	482765	Transmembrane protein 178B
	8083252	8092845	<i>MRPS33</i>	488358	Small ribosomal subunit protein mS33
	8222909	8318179	<i>BRAF</i>	475526	Non-specific serine/threonine protein kinase
	8329367	8341292	<i>NDUFB2</i>	100683846	NADH dehydrogenase [ubiquinone] 1 beta subcomplex subunit 2, mitochondrial
	8346159	8363835	<i>ADCK2</i>	475527	AarF domain containing kinase 2
	8377389	8479791	<i>DENND2A</i>	482766	DENN domain containing 2A
	8508036	8530009	<i>MKRN1</i>	475528	RING-type E3 ubiquitin transferase
	8544353	8562714	<i>RAB19</i>	482767	Ras-related protein Rab-19
	8574973	8615961	<i>SLC37A3</i>	482768	Sugar phosphate exchanger 3
	8737670	8820925	<i>KDM7A</i>	482769	Lysine-specific demethylase 7A
	8828399	8867610	<i>PARP12</i>	482770	Protein mono-ADP-ribosyltransferase <i>PARP12</i>
	8870283	9032105	<i>TBXAS1</i>	482771	Thromboxane-A synthase

*Gene IDs were obtained from the NCBI Gene database based on chromosome numbers and genomic coordinates (CanFam3.1), and gene symbols were updated to current names. Predicted or uncharacterized loci beginning with “LOC” were excluded. Protein annotations were assigned using the UniProt database; genes without matching entries are shown as –. *AMY2B* was included in chromosome 6 CanFam3.1:chr6:46,854,109–47,454,177.

Table S5 Region-wide average nucleotide diversity (π) in the *MGAM* region and flanking regions

Region (NC_006598.3)	Group	pi_per_bp
6725960–6925959	AGW	1.749505×10^{-3}
	DOG	1.583898×10^{-3}
	WOLF	1.542811×10^{-3}
6925960–7125959	AGW	1.793790×10^{-3}
	DOG	1.921790×10^{-3}
	WOLF	2.093987×10^{-3}
7125960–7303209	AGW	1.340813×10^{-3}
	DOG	3.504510×10^{-4}
	WOLF	9.350560×10^{-4}
7303210–7503209	AGW	1.212911×10^{-3}
	DOG	1.021862×10^{-3}
	WOLF	9.706860×10^{-4}
7503210–7703209	AGW	1.204175×10^{-3}
	DOG	1.436954×10^{-3}
	WOLF	1.332582×10^{-3}

*Values are shown for five genomic regions spanning the *MGAM* region (chr16:7,125,960–7,303,209; 177 kb) and the 400 kb upstream and downstream flanking regions in dogs, gray wolves, and African golden wolves.

Table S6 Selection coefficients estimated using CLUES2 (1–5 epochs)

Epoch	logLR	-log ¹⁰ P	AIC	Cutoffs (kyr BP)	5.0k	7.5k	10.0k	12.5k	15.0k	17.5k	20.0k	>20.0k
1	3.40	2.04	- 4.79	-	0.00167							
2	11.59	5.03	- 19.18	5.0k	0.00659	- 0.00013						
2	10.67	4.63	- 17.34	7.5k	0.00533		- 0.00060					
2	7.48	3.25	- 10.96	10.0k	0.00452			- 0.00074				
2	4.69	2.04	- 5.38	12.5k	0.00363				- 0.00067			
2	3.41	1.48	- 2.81	15.0k	0.00162					0.00207		
3	11.68	4.47	- 17.36	5.0k, 7.5k	0.00647	0.00116	- 0.00033					
3	12.80	4.94	- 19.60	5.0k, 15.0k	0.00685	- 0.00187				0.00267		
3	13.60	5.27	- 21.19	7.5k, 10.0k	0.00567		- 0.01319	0.00209				
3	13.79	5.35	- 21.58	7.5k, 15.0k	0.00554		- 0.00472			0.00422		
3	13.93	5.41	- 21.86	7.5k, 17.5k	0.00550		- 0.00353					0.00650
3	14.30	5.57	- 22.61	7.5k, 20.0k	0.00548		- 0.00283					0.01245
3	11.64	4.45	- 17.27	7.5k, 22.5k	0.00572		- 0.00082					0.00755
3	11.19	4.26	- 16.38	10.0k, 20.0k	0.00463			-0.00369				0.01255
4	13.93	4.87	- 19.85	5.0k, 7.5k, 10.0k	0.00635	0.00307	- 0.01236	0.00210				
4	14.16	4.97	- 20.32	5.0k, 7.5k, 12.5k	0.00637	0.00307	- 0.00643		0.00289			
4	14.30	5.02	- 20.5	5.0k, 7.5k, 15.0k	0.00636	0.00276	- 0.00431			0.00400		
4	14.94	5.29	- 21.88	5.0k, 7.5k, 20.0k	0.00638	0.00221	- 0.00244					0.01205
4	13.98	4.90	- 19.9	7.5k, 15.0k, 22.5k	0.00527		- 0.00397			0.00322 / 0.00663		
5	14.72	4.72	- 19.45	5.0k, 7.5k, 10.0k, 17.5k	0.00624	0.00322	- 0.00761	- 0.00148			0.00550	
5	14.61	4.68	- 19.23	5.0k, 7.5k, 12.5k, 17.5k	0.00630	0.00281	- 0.00451		- 0.00143		0.00567	

* Models with 1–5 epochs were compared using CLUES2.

* Bold red text indicates the best-fitting model (lowest AIC).

* Cutoffs (kyr BP): Time boundaries (in thousands of years before present) used to define epochs in the multi-epoch model, assuming a generation time of 3 years.

* Tmax = 30 kyr BP; values in the “>20.0 kyr BP” epoch correspond to 20.0–30 kyr BP.

* AIC: Akaike Information Criterion ($2k - 2\ln L$), where k is the number of model parameters and $\ln L$ is the log-likelihood ratio compared with a neutral model.

* For multi-epoch models, values are listed from most recent to oldest.