Supplementary Information

The genomics of selection in dogs and the parallel evolution between dogs and humans

Guo-dong Wang[*], Weiwei Zhai[*], He-chuan Yang, Ruo-xi Fan, Xue Cao, Li Zhong, Lu Wang, Fei Liu, Hong Wu, Lu-guang Cheng, Andrei D. Poyarkov, Nikolai A. Poyarkov JR, Shu-sheng Tang, Wen-ming Zhao, Yun Gao, Xue-mei Lv, David M. Irwin, Peter Savolainen, Chung-I Wu[¶], Ya-ping Zhang [¶]

[*] These authors contributed equally to this work.

[¶] **Correspondence:** zhangyp@mail.kiz.ac.cn and ciwu@uchicago.edu

This file includes:

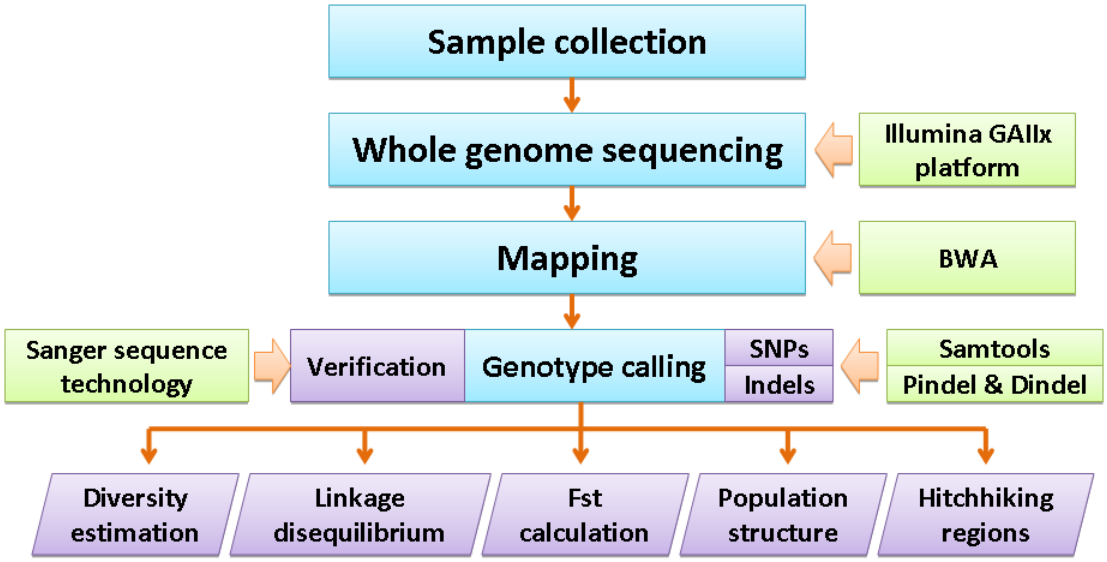
Supplementary Figure S1 to S11

Supplementary Tables S1 to S7

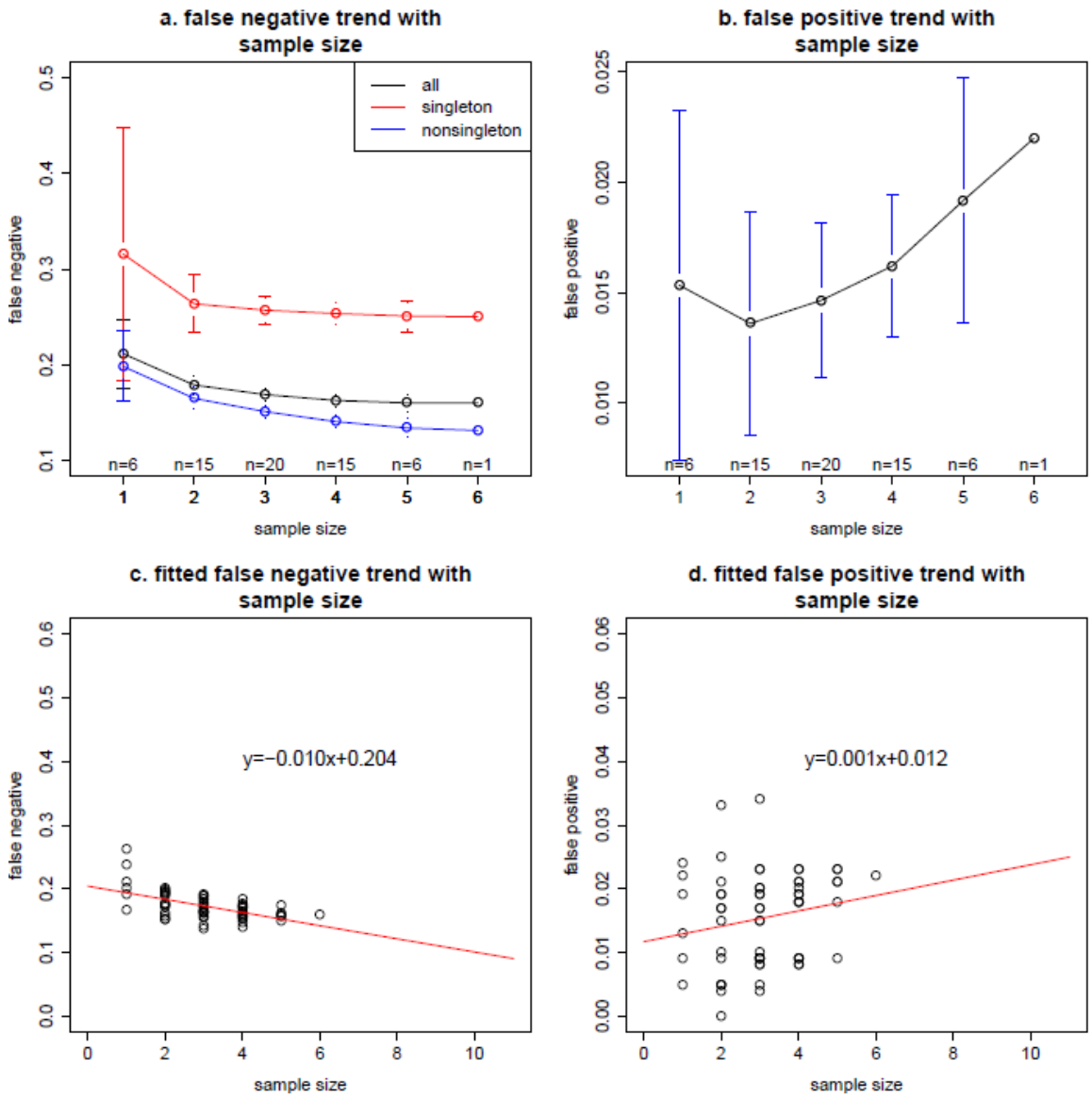
Supplementary Note 1-5

Supplementary References

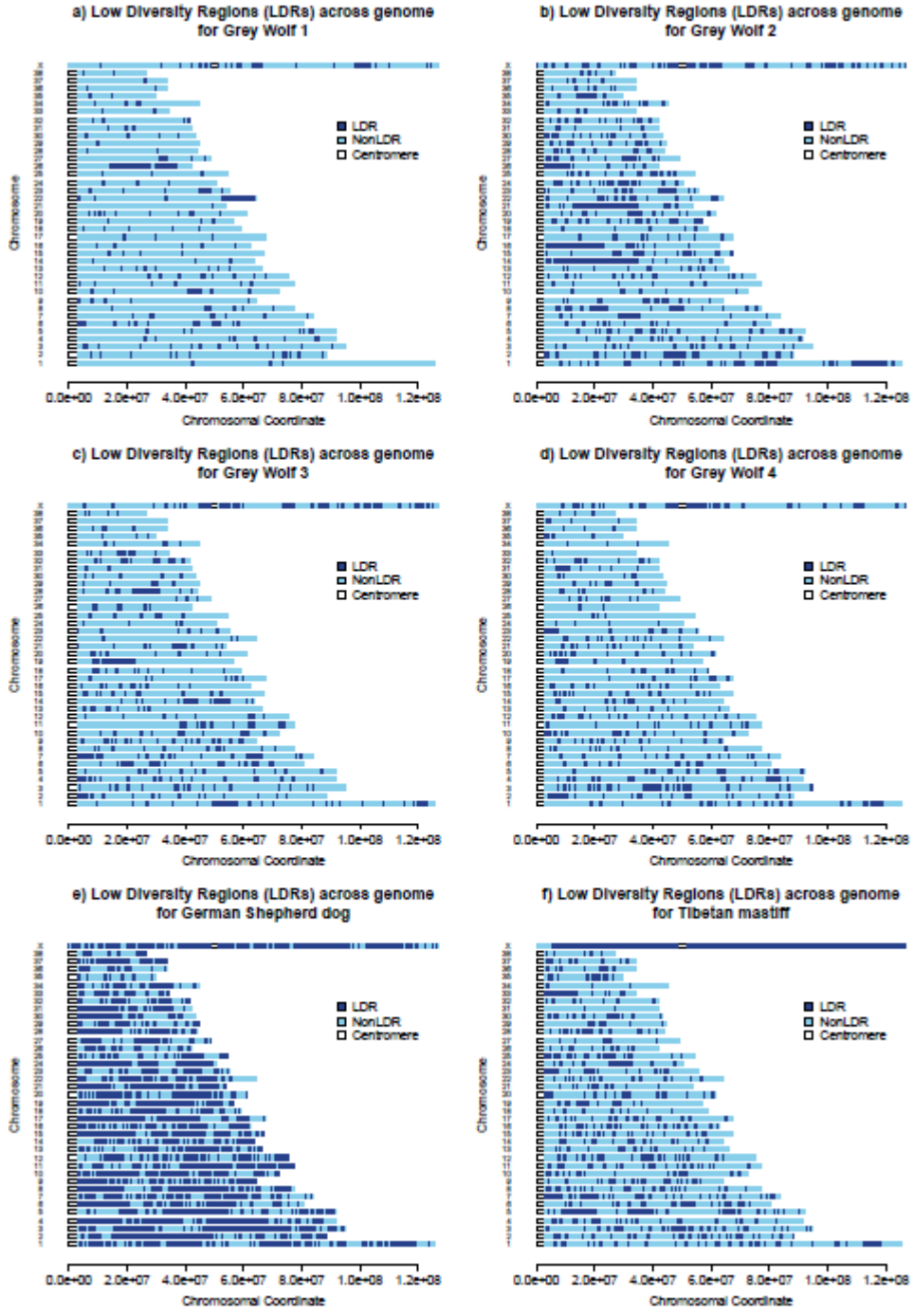
Supplementary Figures

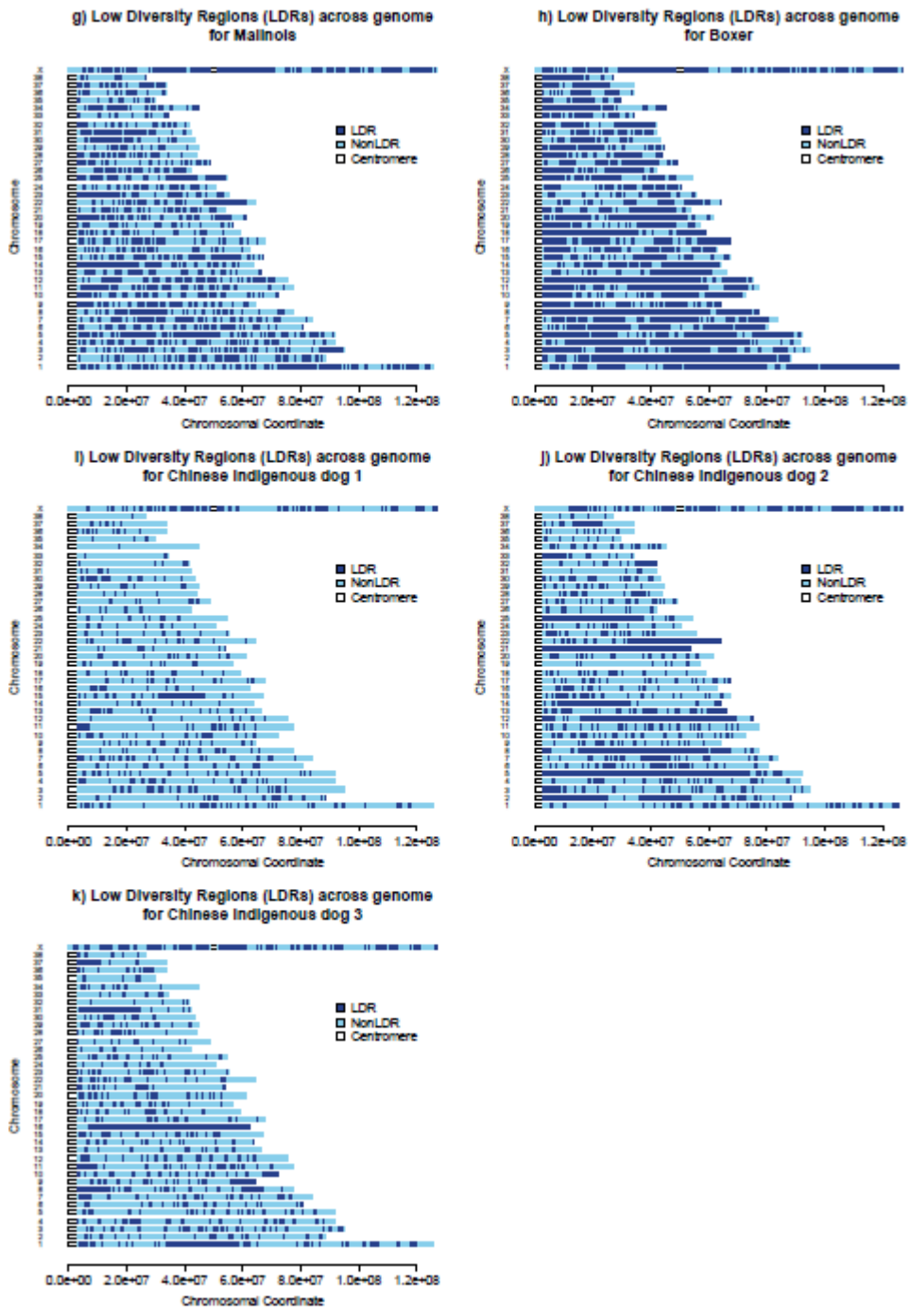


Supplementary Figure S1: Data flow of our sequencing and analysis.

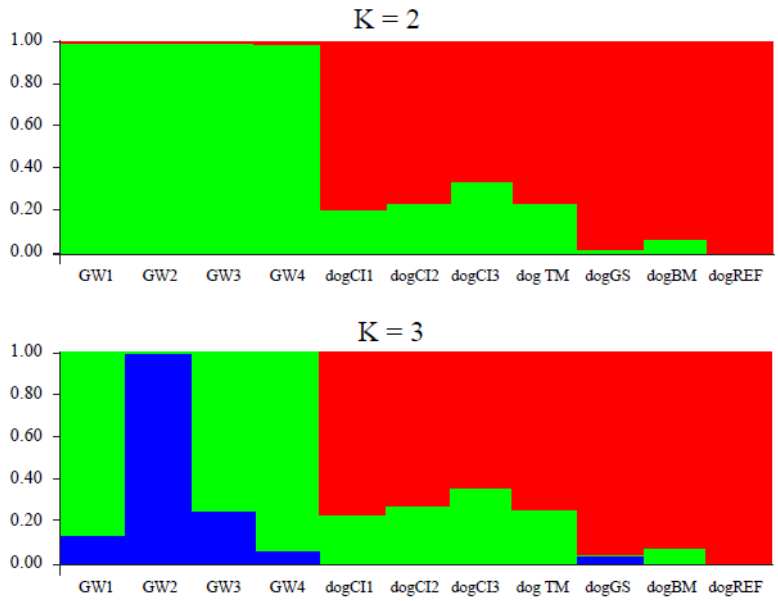


Supplementary Figure S2. False positives and false negatives in SNP calling. a) False negative for the SNP calling when there are x individuals (x axis). Singleton and non-singleton SNPs are also separately calculated. b) False positives for the SNP calling when there are x individuals (x axis). c) Fitted linear relationship between the sample size and the false negatives. d) Fitted linear relationship between the sample size and the false positives.

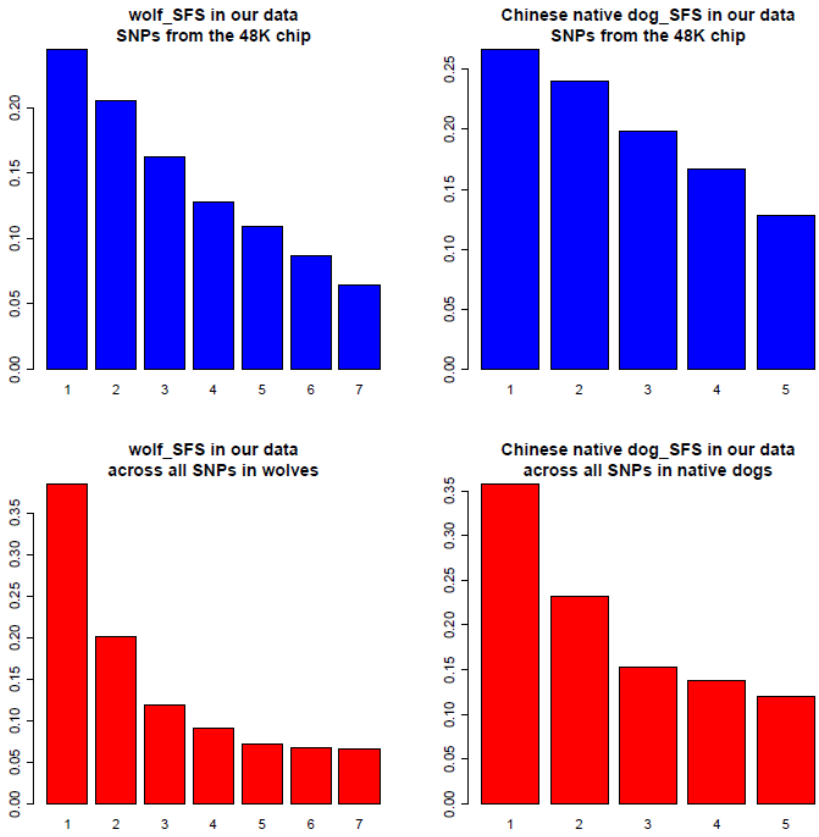

a) Low Diversity Regions (LDRs) across genome for Grey Wolf 1
b) Low Diversity Regions (LDRs) across genome for Grey Wolf 2
c) Low Diversity Regions (LDRs) across genome for Grey Wolf 3
d) Low Diversity Regions (LDRs) across genome for Grey Wolf 4
e) Low Diversity Regions (LDRs) across genome for German Shepherd dog
f) Low Diversity Regions (LDRs) across genome for Tibetan mastiff
LDR
NonLDR
Centromere
Chromosome
Chromosomal Coordinate
0.0e+00
2.0e+07
4.0e+07
6.0e+07
8.0e+07
1.0e+08
1.2e+08



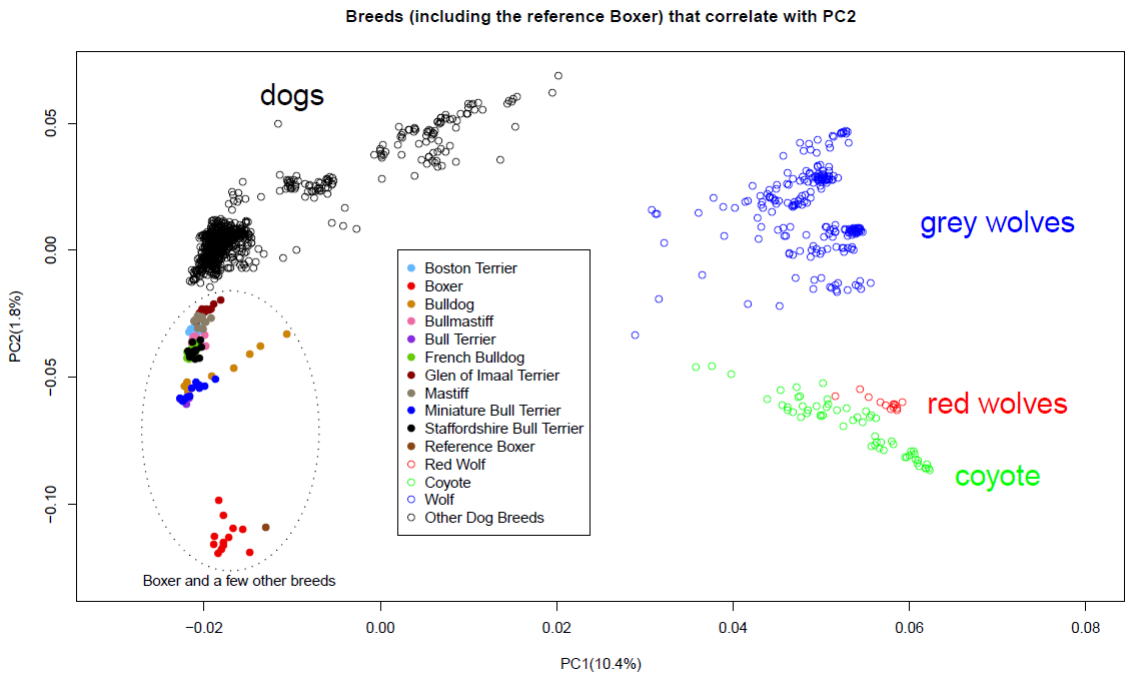
Supplementary Figure S3. LDR distributions across the 11 individual genomes. The genetic diversity across 38 autosomal and the sex chromosomes are plotted for each individual. The low diversity regions (cutoff is set as 0.00005) are plotted in dark blue regions.



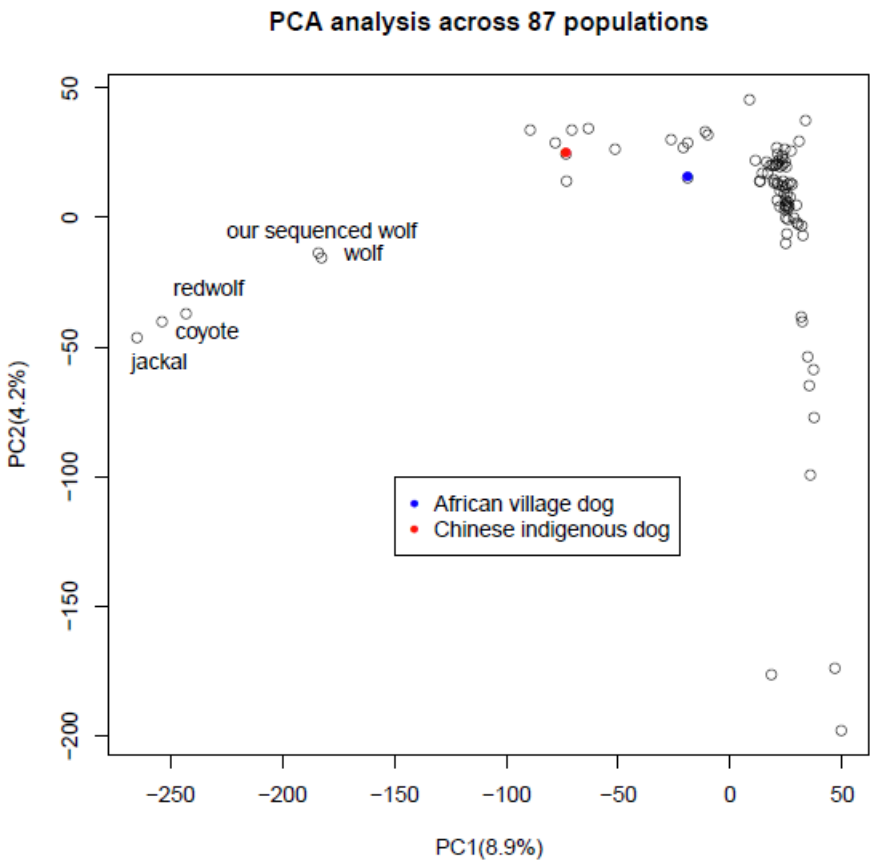
Supplementary Figure S4. Structure analysis with K=2 and K=3. The cluster of 11 individuals using STRUCTURE by partitioning the sample into 2 or 3 groups, respectively. Each vertical column represents an individual and each color represents a genetic component.



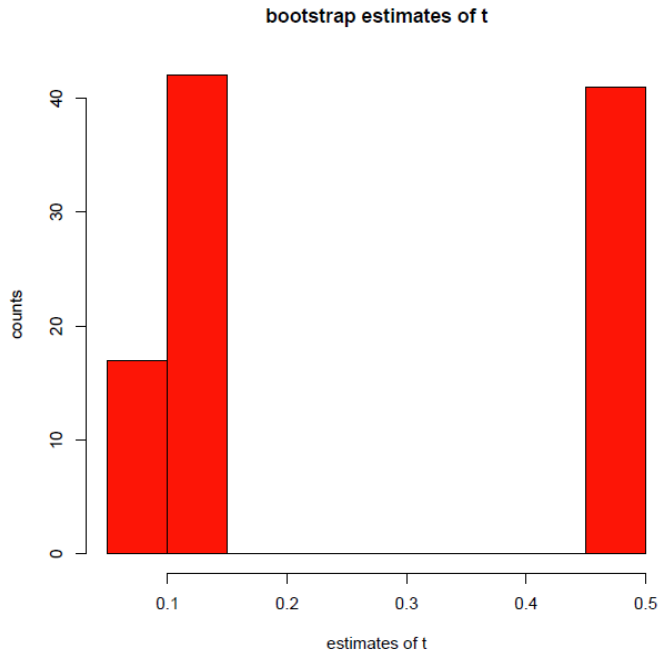
Supplementary Figure S5. SNP ascertainment biases in the 48K SNP chip. a) The site frequency spectra (polarized with an outgroup species) for these 48K SNPs in the wolves. b) The site frequency spectra for these 48K SNPs in the Chinese native dogs. c) The site frequency spectra for all the SNPs in the wolves. d) The site frequency spectra for all the SNPs in the Chinese native dogs.



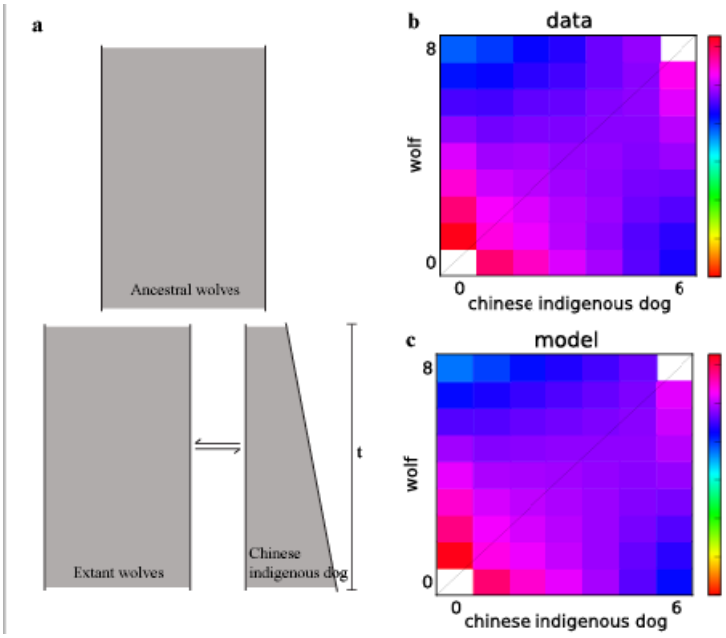
Supplementary Figure S6. Breeds that correlate with the PC2. This is the same plot as Figure 2d, but highlighting the dog breeds that is further away from the rest of the groups in the second principle component (PC2).



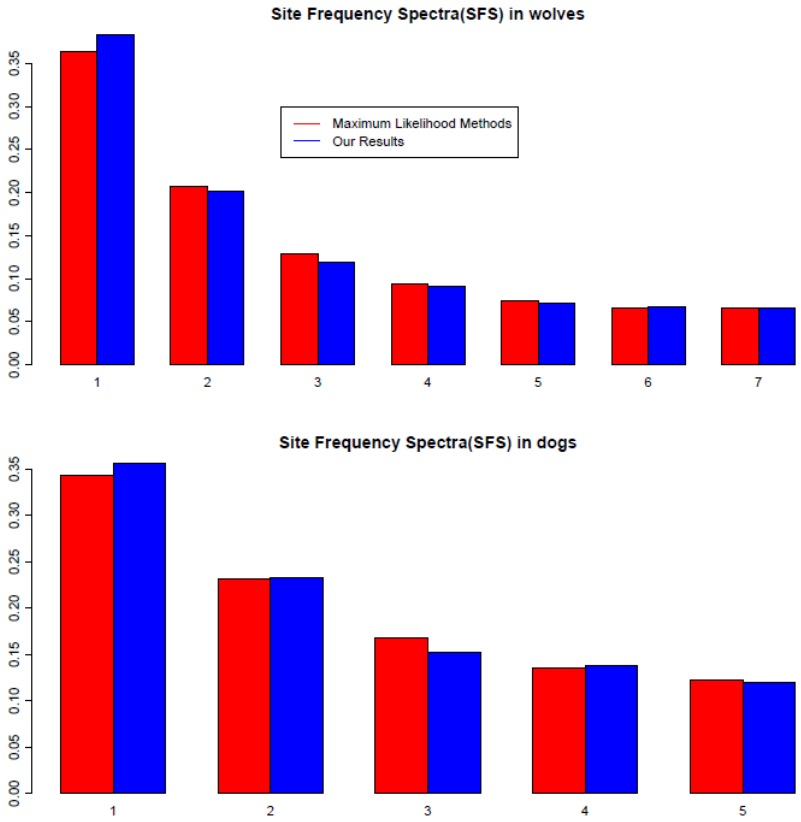
Supplementary Figure S7. PCA plot across 87 populations. The PCA plot for 87 populations using frequency spectra (including wolves and dogs sequenced in this study) are plotted in a two dimensional plot for the first two principle components.



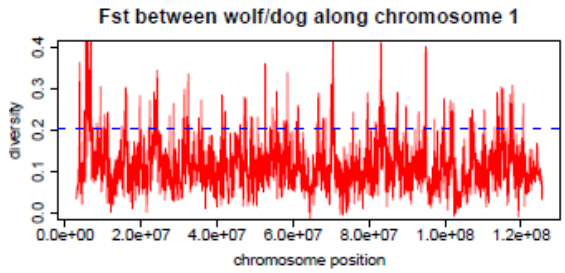
Supplementary Figure S8. The maximum likelihood estimates of divergence time over the bootstrap samples. The upper bound of t is set to be 0.5. The x axis is the point estimate for the divergence time (t), the y axis is the count (over 100 replicates).

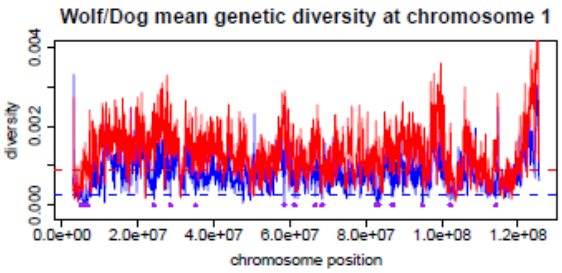


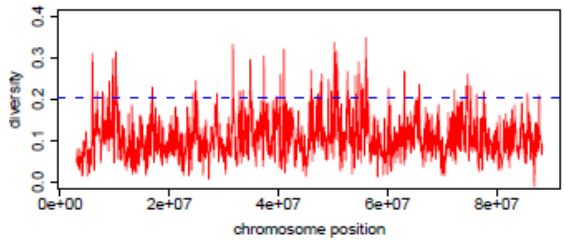
Supplementary Figure S9. Isolation and migration model. a) the cartoon illustrating of the IM(Isolation Migration) model we fitted to the data. b) the joint 2D site frequency spectra observed in the data (heatmap). c) the predicted joint 2D site frequency spectra from the best fitted model

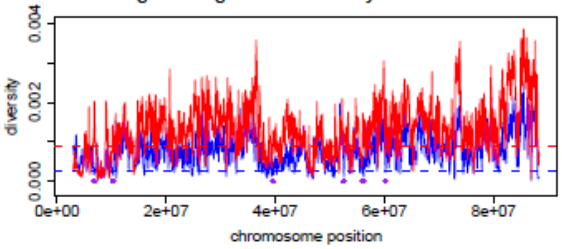


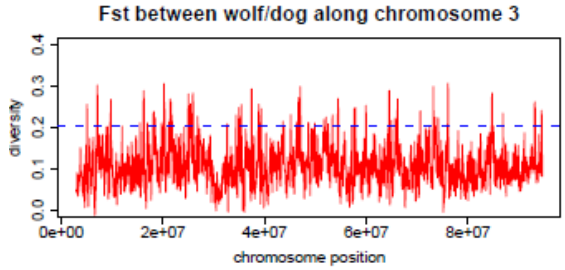
Supplementary Figure S10. Site frequency spectra estimated using two different methods. The site frequency spectra (SFS) inferred using two methods. The maximum likelihood is extracted by using a program from Kim et al 2011. Top panels are for wolves, bottom panels are from dogs.

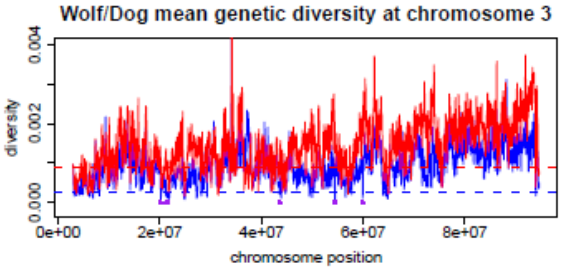

Fst between wolf/dog along chromosome 1
diversity
0.0
0.1
0.2
0.3
0.4
0.0e+00
2.0e+07
4.0e+07
6.0e+07
8.0e+07
1.0e+08
1.2e+08
chromosome position

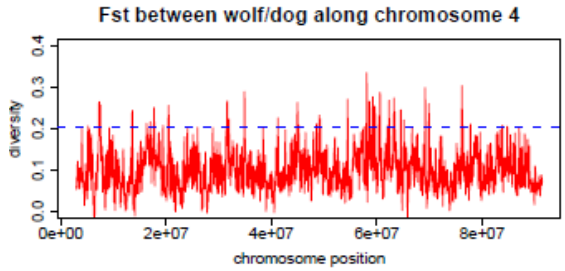

Wolf/Dog mean genetic diversity at chromosome 1
diversity
0.000
0.002
0.004
0.0e+00
2.0e+07
4.0e+07
6.0e+07
8.0e+07
1.0e+08
1.2e+08
chromosome position

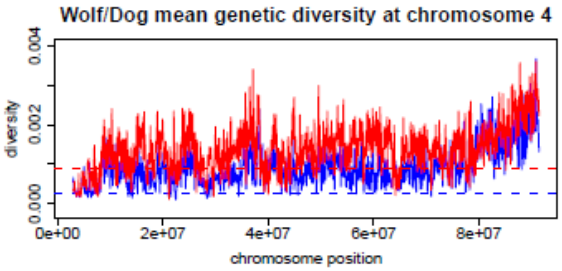

Fst between wolf/dog along chromosome 2
diversity
0.0
0.1
0.2
0.3
0.4
0e+00
2e+07
4e+07
6e+07
8e+07
chromosome position

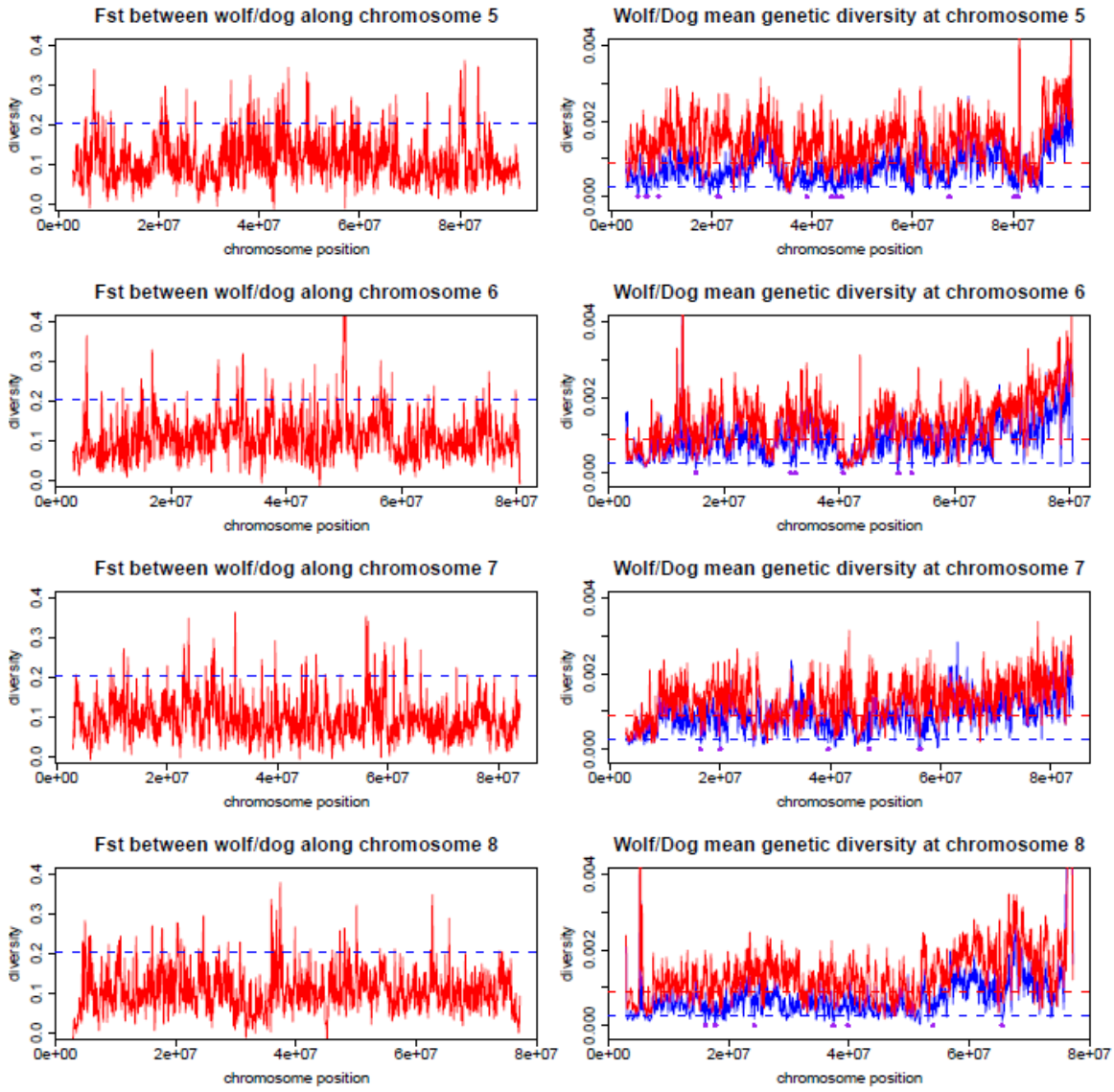

Wolf/Dog mean genetic diversity at chromosome 2
diversity
0.000
0.002
0.004
0e+00
2e+07
4e+07
6e+07
8e+07
chromosome position

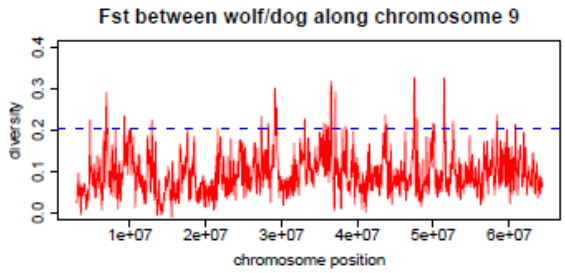

Fst between wolf/dog along chromosome 3
diversity
0.0
0.1
0.2
0.3
0.4
0e+00
2e+07
4e+07
6e+07
8e+07
chromosome position

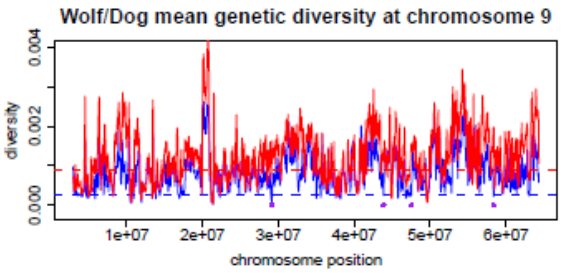

Wolf/Dog mean genetic diversity at chromosome 3
diversity
0.000
0.002
0.004
0e+00
2e+07
4e+07
6e+07
8e+07
chromosome position

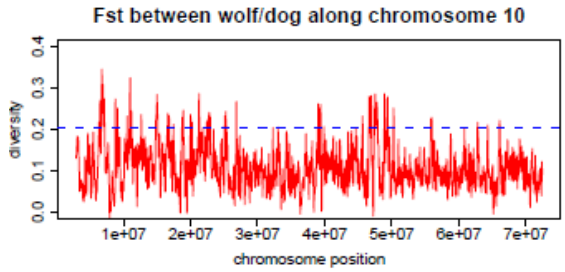

Fst between wolf/dog along chromosome 4
diversity
0.0
0.1
0.2
0.3
0.4
0e+00
2e+07
4e+07
6e+07
8e+07
chromosome position

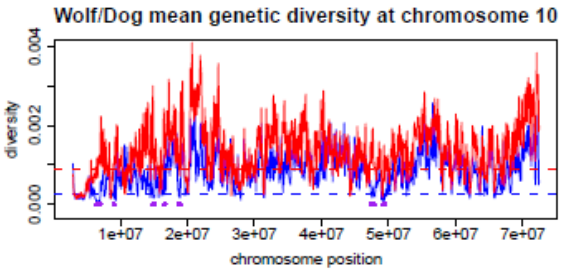

Wolf/Dog mean genetic diversity at chromosome 4
diversity
0.000
0.002
0.004
0e+00
2e+07
4e+07
6e+07
8e+07
chromosome position

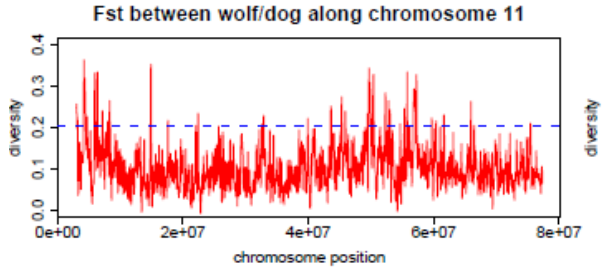

Fst between wolf/dog along chromosome 5
Wolf/Dog mean genetic diversity at chromosome 5
Fst between wolf/dog along chromosome 6
Wolf/Dog mean genetic diversity at chromosome 6
Fst between wolf/dog along chromosome 7
Wolf/Dog mean genetic diversity at chromosome 7
Fst between wolf/dog along chromosome 8
Wolf/Dog mean genetic diversity at chromosome 8
diversity
chromosome position
0e+00
2e+07
4e+07
6e+07
8e+07
0.0
0.1
0.2
0.3
0.4
0.000
0.002
0.004

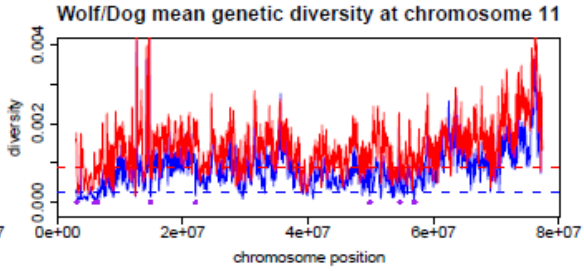

Fst between wolf/dog along chromosome 9
diversity
0.0
0.1
0.2
0.3
0.4
1e+07
2e+07
3e+07
4e+07
5e+07
6e+07
chromosome position

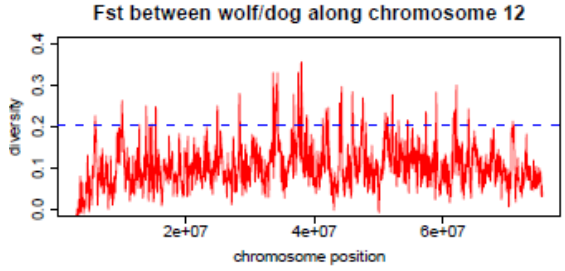

Wolf/Dog mean genetic diversity at chromosome 9
diversity
0.000
0.002
0.004
1e+07
2e+07
3e+07
4e+07
5e+07
6e+07
chromosome position

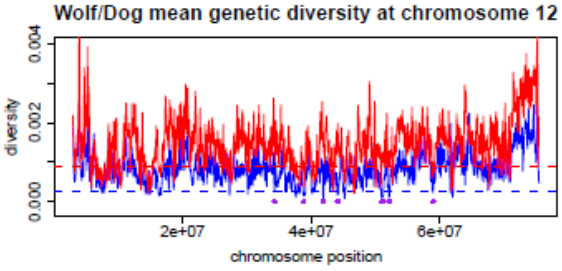

Fst between wolf/dog along chromosome 10
diversity
0.0
0.1
0.2
0.3
0.4
1e+07
2e+07
3e+07
4e+07
5e+07
6e+07
7e+07
chromosome position

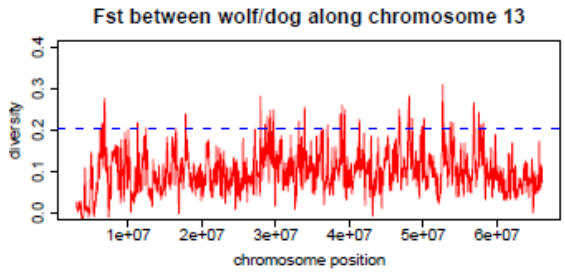

Wolf/Dog mean genetic diversity at chromosome 10
diversity
0.000
0.002
0.004
1e+07
2e+07
3e+07
4e+07
5e+07
6e+07
7e+07
chromosome position

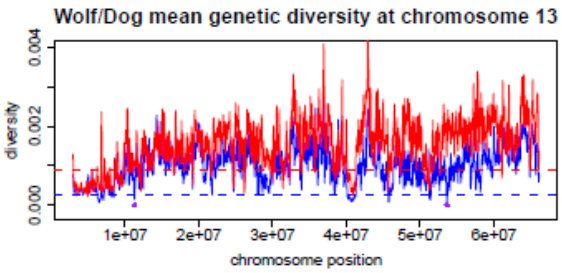

Fst between wolf/dog along chromosome 11
diversity
0.0
0.1
0.2
0.3
0.4
0e+00
2e+07
4e+07
6e+07
8e+07
chromosome position

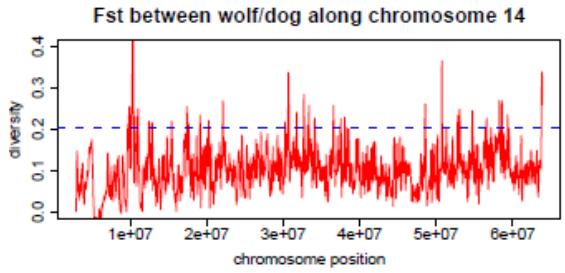

Wolf/Dog mean genetic diversity at chromosome 11
diversity
0.000
0.002
0.004
0e+00
2e+07
4e+07
6e+07
8e+07
chromosome position

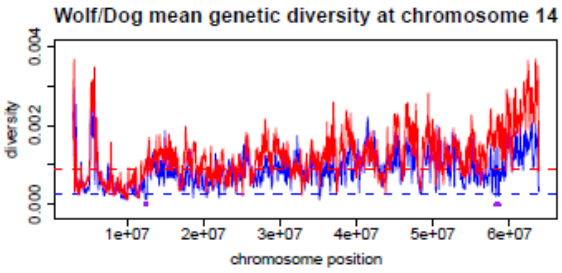

Fst between wolf/dog along chromosome 12
diversity
0.0
0.1
0.2
0.3
0.4
2e+07
4e+07
6e+07
chromosome position

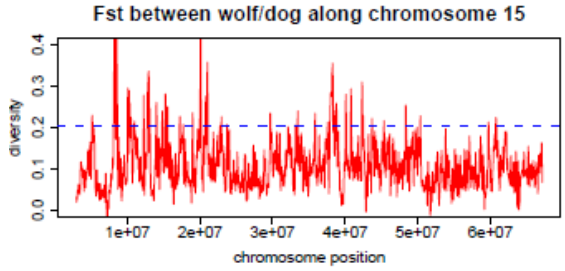

Wolf/Dog mean genetic diversity at chromosome 12
diversity
0.000
0.002
0.004
2e+07
4e+07
6e+07
chromosome position

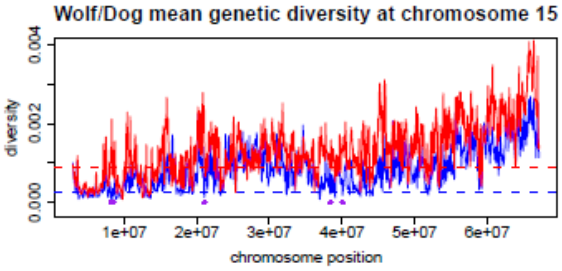

Fst between wolf/dog along chromosome 13
diversity
0.0
0.1
0.2
0.3
0.4
1e+07
2e+07
3e+07
4e+07
5e+07
6e+07
chromosome position

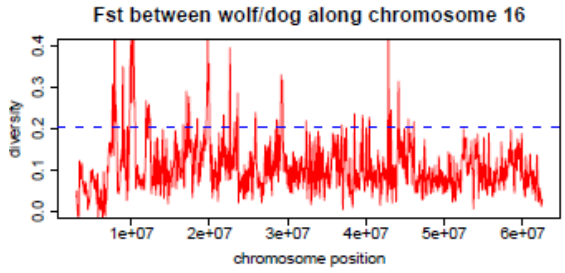

Wolf/Dog mean genetic diversity at chromosome 13
diversity
0.000
0.002
0.004
1e+07
2e+07
3e+07
4e+07
5e+07
6e+07
chromosome position

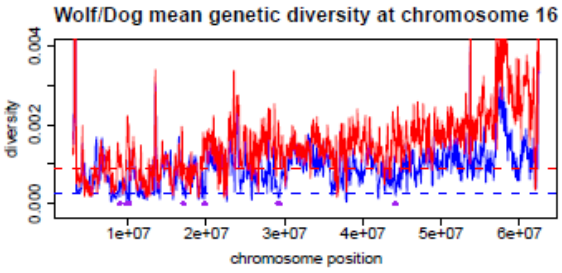

Fst between wolf/dog along chromosome 14
diversity
0.0
0.1
0.2
0.3
0.4
1e+07
2e+07
3e+07
4e+07
5e+07
6e+07
chromosome position

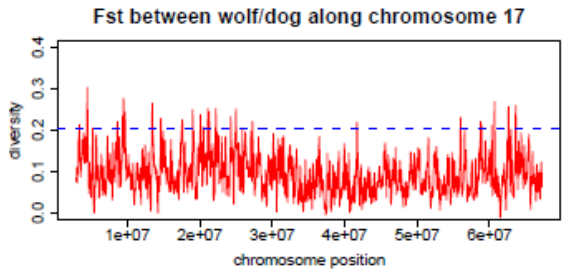

Wolf/Dog mean genetic diversity at chromosome 14
diversity
0.000
0.002
0.004
1e+07
2e+07
3e+07
4e+07
5e+07
6e+07
chromosome position

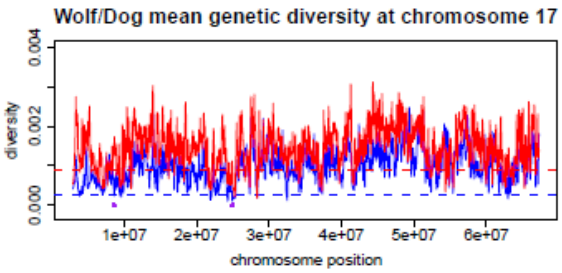

Fst between wolf/dog along chromosome 15
diversity
0.0
0.1
0.2
0.3
0.4
1e+07
2e+07
3e+07
4e+07
5e+07
6e+07
chromosome position

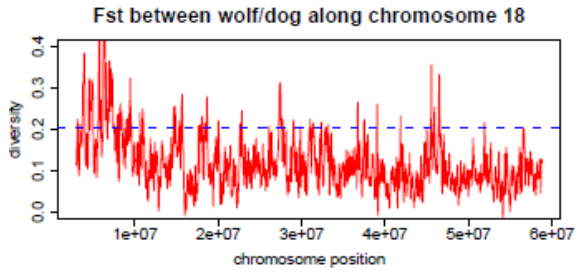

Wolf/Dog mean genetic diversity at chromosome 15
diversity
0.000
0.002
0.004
1e+07
2e+07
3e+07
4e+07
5e+07
6e+07
chromosome position

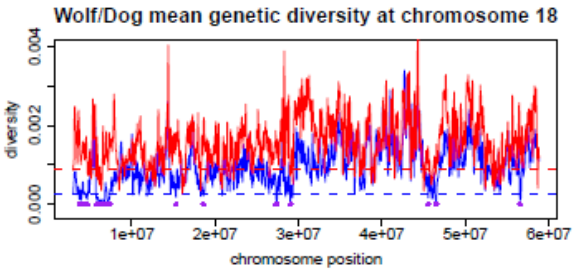

Fst between wolf/dog along chromosome 16
diversity
0.0
0.1
0.2
0.3
0.4
1e+07
2e+07
3e+07
4e+07
5e+07
6e+07
chromosome position

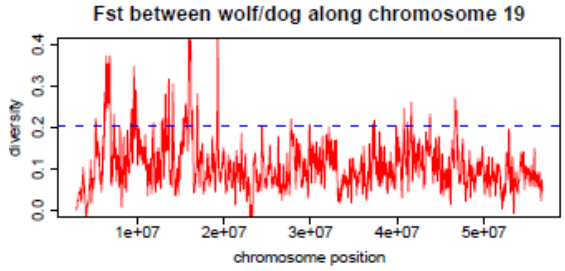

Wolf/Dog mean genetic diversity at chromosome 16
diversity
0.000
0.002
0.004
1e+07
2e+07
3e+07
4e+07
5e+07
6e+07
chromosome position

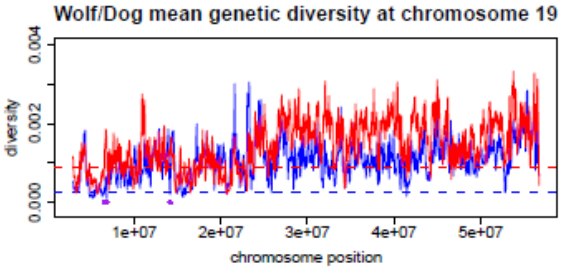

Fst between wolf/dog along chromosome 17
diversity
0.0
0.1
0.2
0.3
0.4
1e+07
2e+07
3e+07
4e+07
5e+07
6e+07
chromosome position

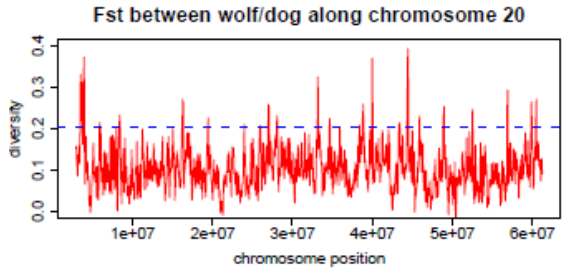

Wolf/Dog mean genetic diversity at chromosome 17
diversity
0.000
0.002
0.004
1e+07
2e+07
3e+07
4e+07
5e+07
6e+07
chromosome position

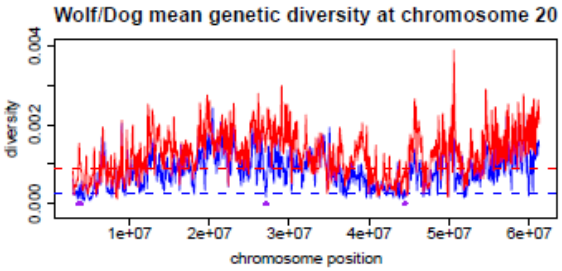

Fst between wolf/dog along chromosome 18
diversity
0.0
0.1
0.2
0.3
0.4
1e+07
2e+07
3e+07
4e+07
5e+07
6e+07
chromosome position

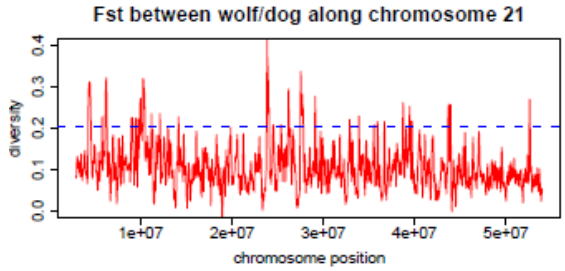

Wolf/Dog mean genetic diversity at chromosome 18
diversity
0.000
0.002
0.004
1e+07
2e+07
3e+07
4e+07
5e+07
6e+07
chromosome position

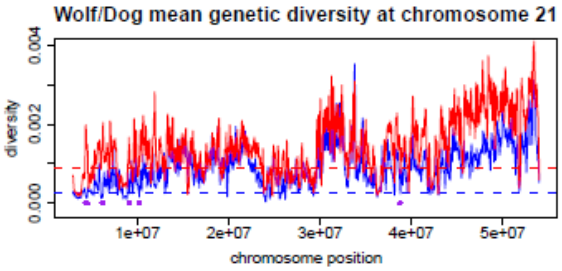

Fst between wolf/dog along chromosome 19
diversity
0.0
0.1
0.2
0.3
0.4
1e+07
2e+07
3e+07
4e+07
5e+07
chromosome position

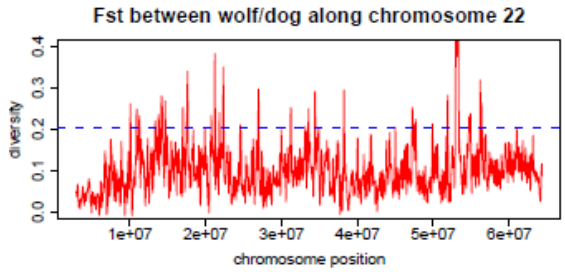

Wolf/Dog mean genetic diversity at chromosome 19
diversity
0.000
0.002
0.004
1e+07
2e+07
3e+07
4e+07
5e+07
chromosome position

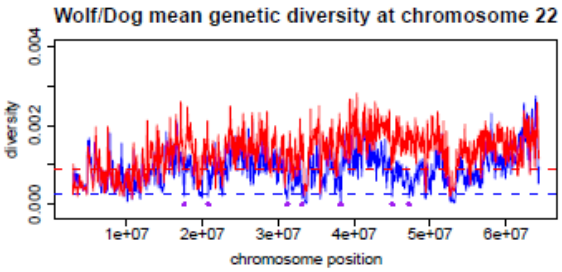

Fst between wolf/dog along chromosome 20
diversity
0.0
0.1
0.2
0.3
0.4
1e+07
2e+07
3e+07
4e+07
5e+07
6e+07
chromosome position

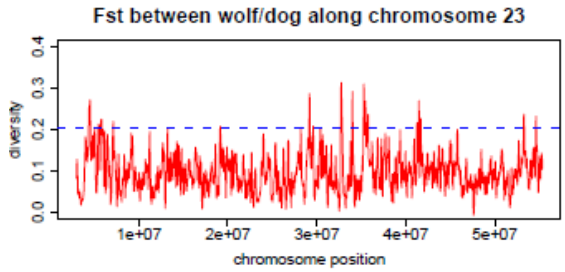

Wolf/Dog mean genetic diversity at chromosome 20
diversity
0.000
0.002
0.004
1e+07
2e+07
3e+07
4e+07
5e+07
6e+07
chromosome position

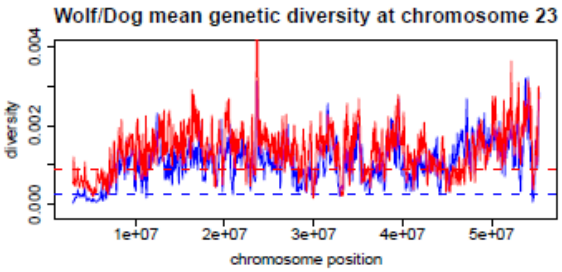

Fst between wolf/dog along chromosome 21
diversity
0.0
0.1
0.2
0.3
0.4
1e+07
2e+07
3e+07
4e+07
5e+07
chromosome position

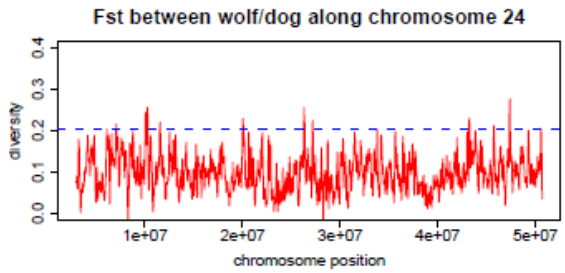

Wolf/Dog mean genetic diversity at chromosome 21
diversity
0.000
0.002
0.004
1e+07
2e+07
3e+07
4e+07
5e+07
chromosome position

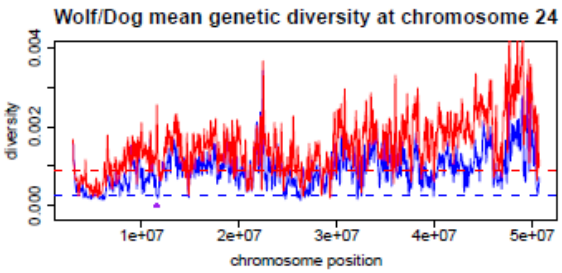

Fst between wolf/dog along chromosome 22
diversity
0.0
0.1
0.2
0.3
0.4
1e+07
2e+07
3e+07
4e+07
5e+07
6e+07
chromosome position

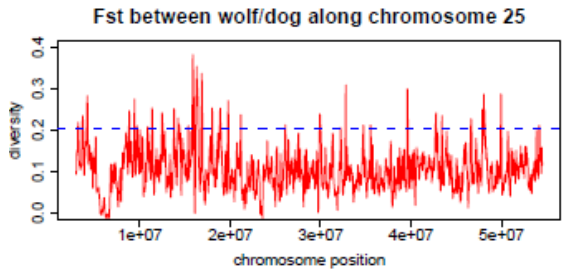

Wolf/Dog mean genetic diversity at chromosome 22
diversity
0.000
0.002
0.004
1e+07
2e+07
3e+07
4e+07
5e+07
6e+07
chromosome position

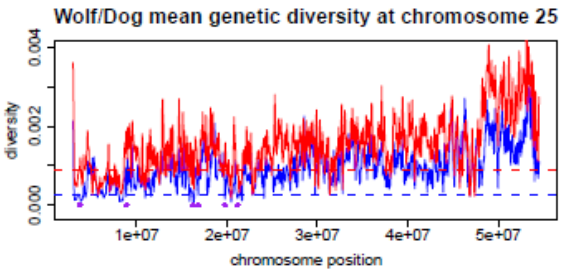

Fst between wolf/dog along chromosome 23
diversity
0.0
0.1
0.2
0.3
0.4
1e+07
2e+07
3e+07
4e+07
5e+07
chromosome position

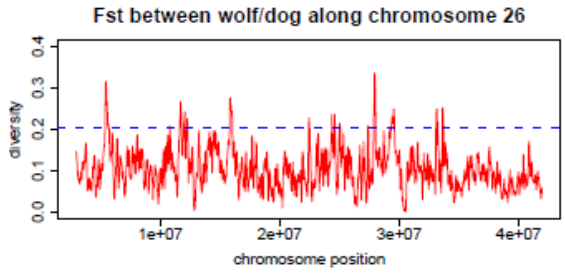

Wolf/Dog mean genetic diversity at chromosome 23
diversity
0.000
0.002
0.004
1e+07
2e+07
3e+07
4e+07
5e+07
chromosome position

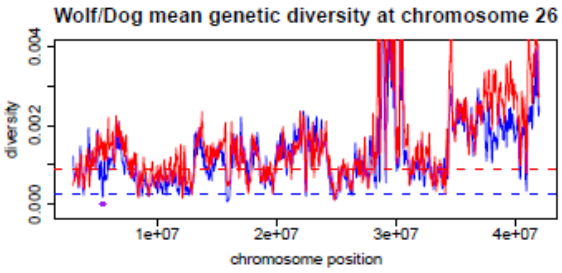

Fst between wolf/dog along chromosome 24
diversity
0.0
0.1
0.2
0.3
0.4
1e+07
2e+07
3e+07
4e+07
5e+07
chromosome position

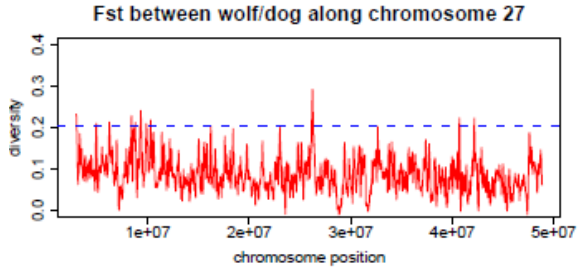

Wolf/Dog mean genetic diversity at chromosome 24
diversity
0.000
0.002
0.004
1e+07
2e+07
3e+07
4e+07
5e+07
chromosome position

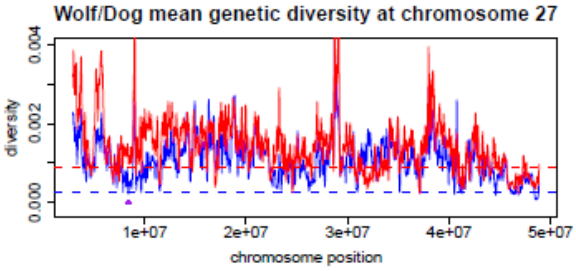

Fst between wolf/dog along chromosome 25
diversity
0.0
0.1
0.2
0.3
0.4
1e+07
2e+07
3e+07
4e+07
5e+07
chromosome position

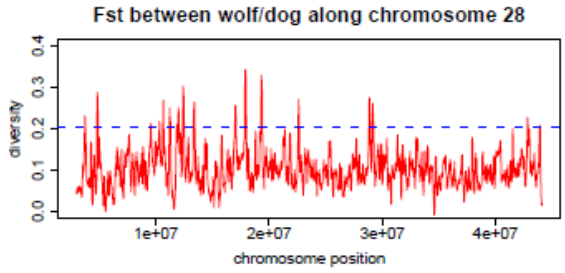

Wolf/Dog mean genetic diversity at chromosome 25
diversity
0.000
0.002
0.004
1e+07
2e+07
3e+07
4e+07
5e+07
chromosome position

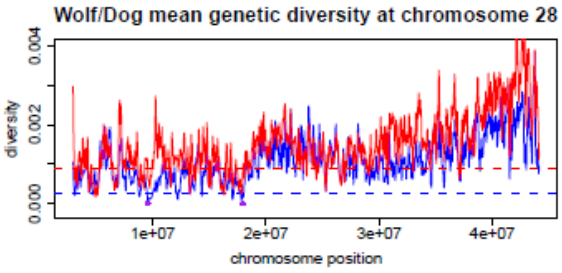

Fst between wolf/dog along chromosome 26
diversity
0.0
0.1
0.2
0.3
0.4
1e+07
2e+07
3e+07
4e+07
chromosome position

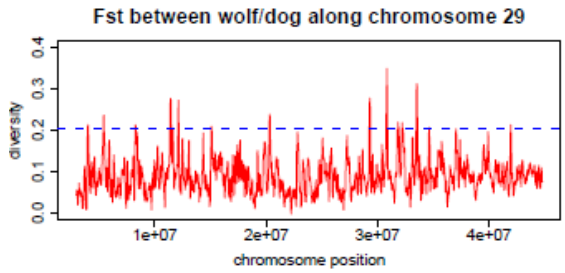

Wolf/Dog mean genetic diversity at chromosome 26
diversity
0.000
0.002
0.004
1e+07
2e+07
3e+07
4e+07
chromosome position

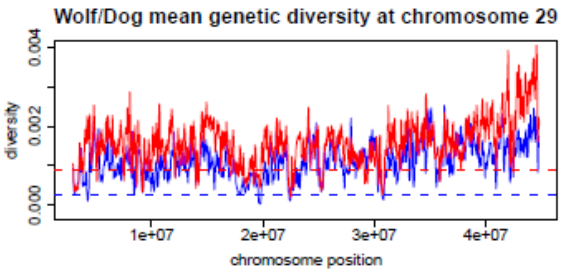

Fst between wolf/dog along chromosome 27
diversity
0.0
0.1
0.2
0.3
0.4
1e+07
2e+07
3e+07
4e+07
5e+07
chromosome position

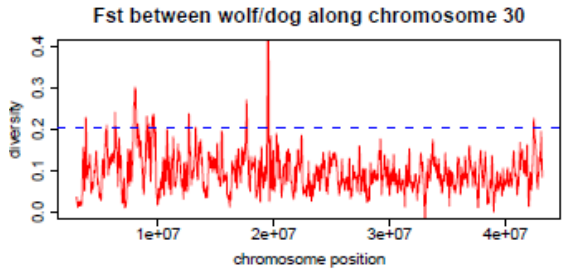

Wolf/Dog mean genetic diversity at chromosome 27
diversity
0.000
0.002
0.004
1e+07
2e+07
3e+07
4e+07
5e+07
chromosome position

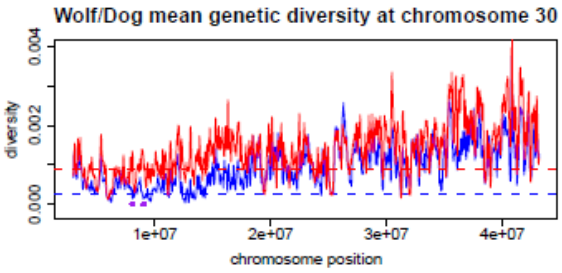

Fst between wolf/dog along chromosome 28
diversity
0.0
0.1
0.2
0.3
0.4
1e+07
2e+07
3e+07
4e+07
chromosome position

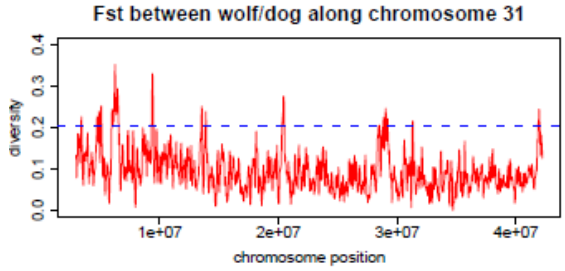

Wolf/Dog mean genetic diversity at chromosome 28
diversity
0.000
0.002
0.004
1e+07
2e+07
3e+07
4e+07
chromosome position

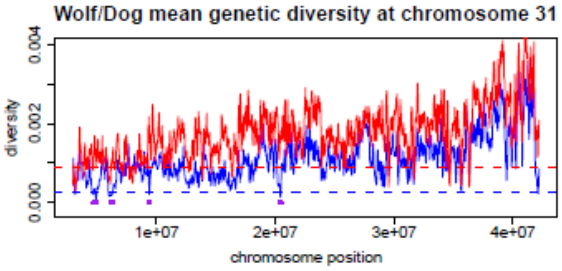

Fst between wolf/dog along chromosome 29
diversity
0.0
0.1
0.2
0.3
0.4
1e+07
2e+07
3e+07
4e+07
chromosome position

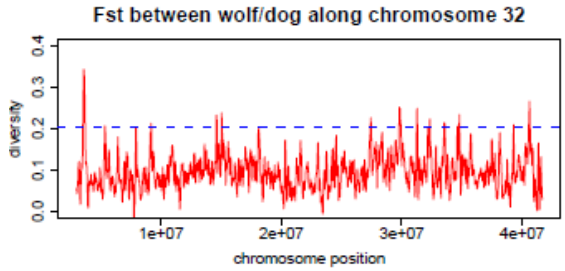

Wolf/Dog mean genetic diversity at chromosome 29
diversity
0.000
0.002
0.004
1e+07
2e+07
3e+07
4e+07
chromosome position

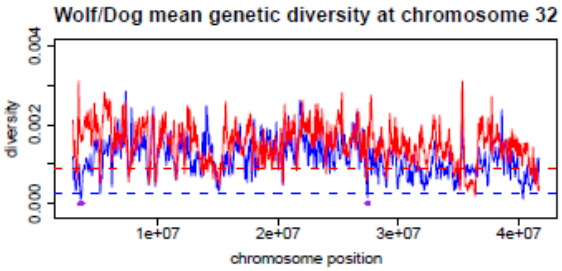

Fst between wolf/dog along chromosome 30
diversity
0.0
0.1
0.2
0.3
0.4
1e+07
2e+07
3e+07
4e+07
chromosome position

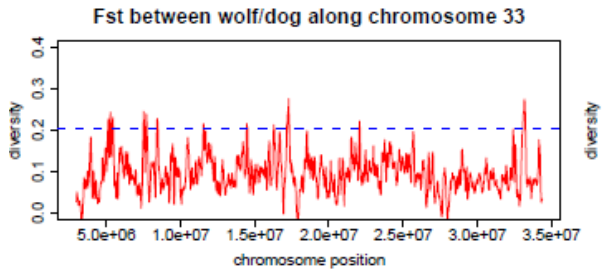

Wolf/Dog mean genetic diversity at chromosome 30
diversity
0.000
0.002
0.004
1e+07
2e+07
3e+07
4e+07
chromosome position

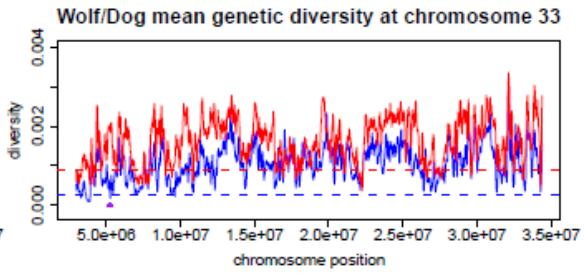

Fst between wolf/dog along chromosome 31
diversity
0.0
0.1
0.2
0.3
0.4
1e+07
2e+07
3e+07
4e+07
chromosome position

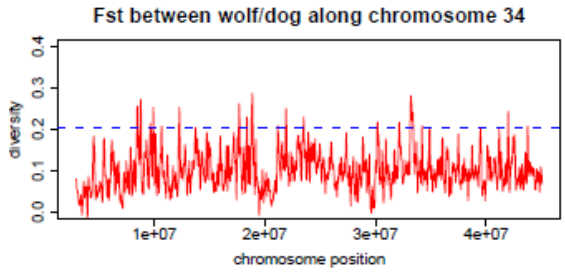

Wolf/Dog mean genetic diversity at chromosome 31
diversity
0.000
0.002
0.004
1e+07
2e+07
3e+07
4e+07
chromosome position

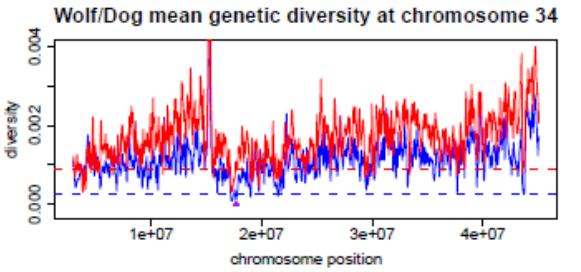

Fst between wolf/dog along chromosome 32
diversity
0.0
0.1
0.2
0.3
0.4
1e+07
2e+07
3e+07
4e+07
chromosome position

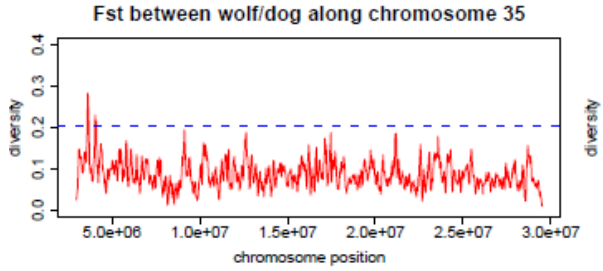

Wolf/Dog mean genetic diversity at chromosome 32
diversity
0.000
0.002
0.004
1e+07
2e+07
3e+07
4e+07
chromosome position

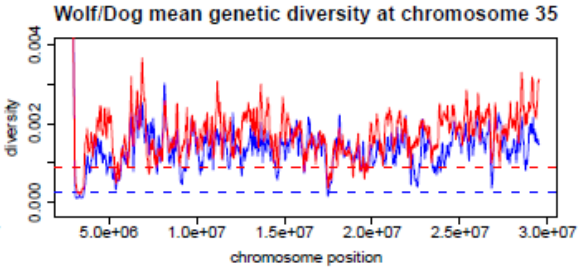

Fst between wolf/dog along chromosome 33
diversity
0.0
0.1
0.2
0.3
0.4
5.0e+06
1.0e+07
1.5e+07
2.0e+07
2.5e+07
3.0e+07
3.5e+07
chromosome position

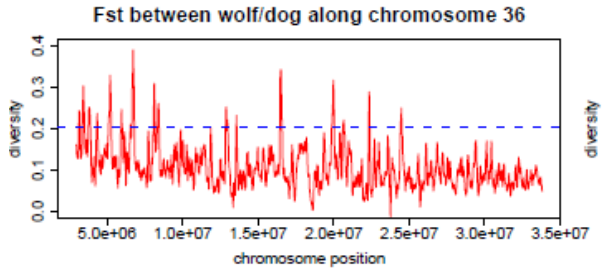

Wolf/Dog mean genetic diversity at chromosome 33
diversity
0.000
0.002
0.004
5.0e+06
1.0e+07
1.5e+07
2.0e+07
2.5e+07
3.0e+07
3.5e+07
chromosome position

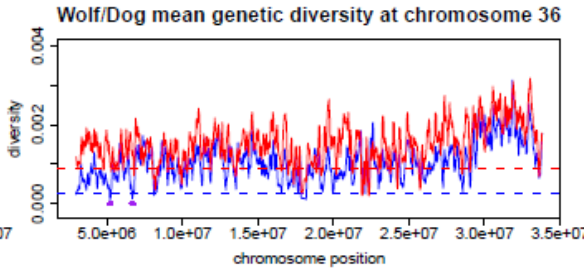

Fst between wolf/dog along chromosome 34
diversity
0.0
0.1
0.2
0.3
0.4
1e+07
2e+07
3e+07
4e+07
chromosome position

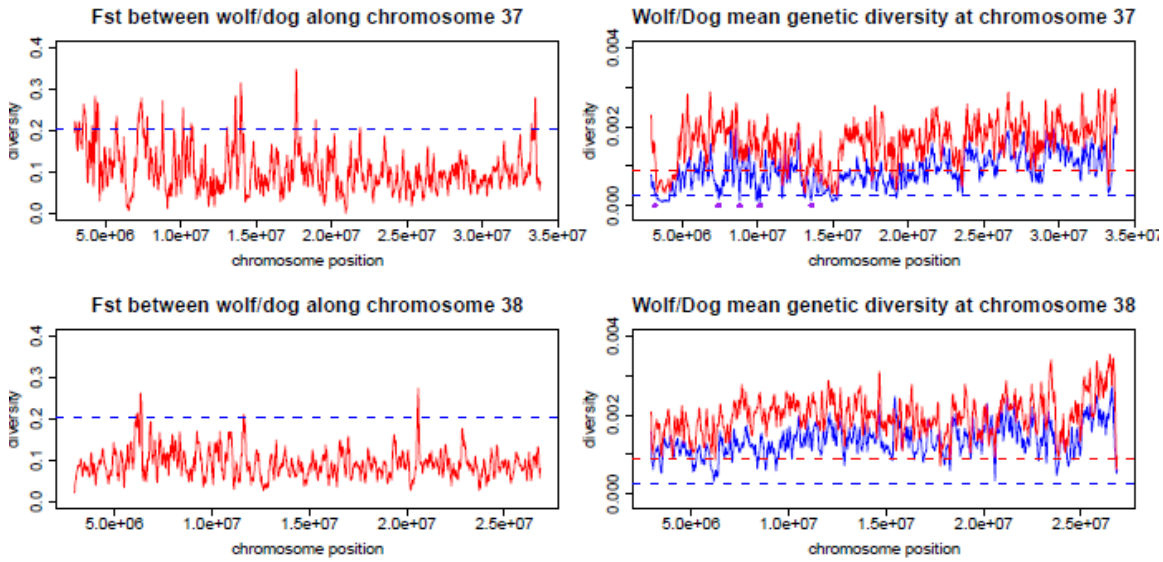

Wolf/Dog mean genetic diversity at chromosome 34
diversity
0.000
0.002
0.004
1e+07
2e+07
3e+07
4e+07
chromosome position


Fst between wolf/dog along chromosome 35
diversity
0.0
0.1
0.2
0.3
0.4
5.0e+06
1.0e+07
1.5e+07
2.0e+07
2.5e+07
3.0e+07
chromosome position


Wolf/Dog mean genetic diversity at chromosome 35
diversity
0.000
0.002
0.004
5.0e+06
1.0e+07
1.5e+07
2.0e+07
2.5e+07
3.0e+07
chromosome position


Fst between wolf/dog along chromosome 36
diversity
0.0
0.1
0.2
0.3
0.4
5.0e+06
1.0e+07
1.5e+07
2.0e+07
2.5e+07
3.0e+07
3.5e+07
chromosome position


Wolf/Dog mean genetic diversity at chromosome 36
diversity
0.000
0.002
0.004
5.0e+06
1.0e+07
1.5e+07
2.0e+07
2.5e+07
3.0e+07
3.5e+07
chromosome position



Supplementary Figure S11. Fst and mean diversity across the dog and wolf genome. The left panels are the mean Fst across the genome. The right panels are the mean genetic diversity for wolves (in red) and dogs (in blue). Window size is set to be 100kb and step size is 20kb. Dashed lines are the cutoffs for picking the potentially positively selected regions, which are also plotted with the purple horizontal bars.

Supplementary Tables

Supplementary Table S1. Library sizes and throughput for different individuals

Individual	Location	InsertSize	Std	Reads Length	Matching Bps (Gbp)
GW1	Altai, Russia	409.38	124.03	111	5.51
		340.76	39.47	120	16.91
		217.64	91.49	111	0.09
		339.82	43.87	81	4.29
GW2	Chukotka, Russia	333.82	25.76	120	18.79
		297.69	129.2	111	1.31
		286.46	105.46	111	1.18
GW3	Bryansk, Russia	329.62	77.74	101	26.48
GW4	Inner Mongolia, China	2257.61	337.58	44	2.96
		177.79	11.84	44	1.72
		521.58	14.24	44	3.46
		491.09	12.37	44	0.90
		176.91	12	75	3.85
		520.09	14.53	75	2.18
		490.66	12.71	75	5.08
		-	-	44	2.77
dogCI1	Xi'an, China	371.95	36.35	101	15.31
		421.37	51.38	101	12.38
		485.13	101.3	101	4.67
dogCI2	Simao, China	313.04	19.3	171	23.44
dogCI3	Ya'an, China	332.45	21.67	121	24.44
dogTM	Lijiang, China	314.36	82.51	101	24.73
dogGS	Germany	478.97	17.17	44	5.71
		373.11	116.52	65	17.10
dogBM	France	488.61	16.17	44	5.32
		472.56	22.47	90	18.78

Supplementary Table S2. The Sanger read coverage distribution across six individuals

Coverage	1	2	3	4	5	6
region_length	4923	2243	2792	5688	8806	29866

Supplementary Table S3. False positive and false negative for SNP calling

Sample	Sanger_snp	GAIIx	overlap	false_positive	false_negative[a]
GW3	206	170	166	0.024	0.194
GW4	251	200	199	0.005	0.207
dogCI1	211	180	176	0.022	0.166
dogGS	152	113	112	0.009	0.263
dogTM	194	158	155	0.019	0.201
dogBM	210	159	157	0.013	0.252
Sum	1224	980	965	0.015	0.212

[a]A large proportion of the false negatives are due to low coverage in the sequencing

Supplementary Table S4. False positive and false negative rate for indel calling

sanger_indels	solexa_indels	Overlap	false_positive	false_negative
222	107	100	0.065	0.55

Supplementary Table S5. Demographic inference with different upper bound

Parameter	Estimated value	nonparametric bootstrap confidence interval[a]	nonparametric bootstrap confidence interval[b]
N_{anc}	53,000	-	-
N_{wolf}	50,000	(49,543, 50,163)	(49,522,50,190)
$N_{dog\_bottleneck}$	8,500	(8,161, 8,961)	(8,089, 8956)
$N_{dog\_current}$	17,000	(16,552, 17,585)	(16,573, 17,692)
T	32,000 (years)	(31,300, 33,191)	(31,158, 33,137)
M_{wd}	1.31	(1.24, 1.40)	(1.22, 1.41)
M_{dw}	1.71	(1.64, 1.83)	(1.63, 1.82)

[a]: bootstrap with a confined parameter bound, [b]: bootstrap with a large parameter bound, but bootstrap estimates with parameter values hitting the right boundary are removed.

$M_{wd} = 2N_{ref} \times m_{wd}$, m_{wd} is the fraction of the wolf population that is migrants from dogs.

N_{ref} is the population size for the reference population. M_{dw} is defined similarly.

Here, we set the reference population to be the ancestral population. Confidence interval is calculated as the $\overline{\theta^*} \pm 1.96 \times \sigma(\theta^*)$

Supplementary Table S6: Regions show strong signal of positive selection

Chrom	Start	End	Chrom	Start	End
chr1	4920000	5020000	chr12	34240000	34380000
chr1	5540000	5640000	chr12	38760000	38900000
chr1	5820000	6100000	chr12	41800000	41960000
chr1	6060000	6200000	chr12	44020000	44160000
chr1	6360000	6500000	chr12	44200000	44320000
chr1	6580000	6760000	chr12	50920000	51180000
chr1	6740000	6920000	chr12	51200000	51300000
chr1	24200000	24340000	chr12	52020000	52260000
chr1	28480000	28600000	chr12	58920000	59040000
chr1	35140000	35240000	chr13	11240000	11340000
chr1	58440000	58580000	chr13	53620000	53760000
chr1	61160000	61260000	chr14	12400000	12540000
chr1	66700000	66800000	chr14	12480000	12580000
chr1	68440000	68580000	chr14	58220000	58320000
chr1	82620000	82880000	chr14	58560000	58680000
chr1	82920000	83160000	chr15	8120000	8400000
chr1	83240000	83400000	chr15	8440000	8580000
chr1	86840000	86940000	chr15	20960000	21140000
chr1	87140000	87260000	chr15	38360000	38500000
chr1	94840000	94980000	chr15	40020000	40180000
chr1	102080000	102280000	chr16	8900000	9040000
chr1	114160000	114280000	chr16	9740000	10040000
chr2	6700000	6800000	chr16	10100000	10200000
chr2	6740000	6880000	chr16	17020000	17120000
chr2	10140000	10320000	chr16	19680000	19860000
chr2	10260000	10380000	chr16	29100000	29420000
chr2	39520000	39640000	chr16	44080000	44200000
chr2	52500000	52620000	chr16	44140000	44280000
chr2	52580000	52680000	chr17	8580000	8680000
chr2	55940000	56080000	chr17	24920000	25020000
chr2	56200000	56300000	chr18	3660000	3840000
chr2	60160000	60260000	chr18	3840000	4060000
chr3	20160000	20280000	chr18	4420000	4600000
chr3	21160000	21260000	chr18	4580000	4720000
chr3	21520000	21660000	chr18	5700000	5860000
chr3	43580000	43680000	chr18	5800000	5960000
chr3	54400000	54580000	chr18	6140000	6280000
chr3	59900000	60040000	chr18	6440000	6720000
chr5	5260000	5360000	chr18	6700000	6820000
chr5	6840000	7260000	chr18	6980000	7200000
chr5	9400000	9500000	chr18	7160000	7420000

chr5	21080000	21200000	chr18	15240000	15360000
chr5	21220000	21400000	chr18	18520000	18660000
chr5	21440000	21540000	chr18	27240000	27480000
chr5	38880000	38980000	chr18	28980000	29080000
chr5	43780000	43900000	chr18	45500000	45600000
chr5	44740000	44900000	chr18	46420000	46580000
chr5	45740000	45900000	chr18	56500000	56600000
chr5	67180000	67280000	chr19	6520000	6620000
chr5	67220000	67340000	chr19	6760000	6920000
chr5	67420000	67540000	chr19	14100000	14220000
chr5	80120000	80280000	chr20	3600000	3700000
chr5	80900000	81080000	chr20	3800000	3980000
chr6	15040000	15160000	chr20	27020000	27160000
chr6	15100000	15200000	chr20	44440000	44540000
chr6	31520000	31700000	chr21	4280000	4580000
chr6	32380000	32540000	chr21	6100000	6300000
chr6	40640000	40780000	chr21	9080000	9180000
chr6	50200000	50420000	chr21	10160000	10260000
chr6	52560000	52680000	chr21	38740000	38840000
chr7	16400000	16520000	chr22	17580000	17680000
chr7	19980000	20100000	chr22	20720000	20820000
chr7	39540000	39680000	chr22	31160000	31260000
chr7	46940000	47080000	chr22	33080000	33180000
chr7	56040000	56160000	chr22	38160000	38340000
chr7	56300000	56400000	chr22	44960000	45060000
chr8	16060000	16220000	chr22	47180000	47300000
chr8	17640000	17880000	chr24	11460000	11560000
chr8	24240000	24340000	chr24	11540000	11640000
chr8	37280000	37500000	chr25	3620000	3800000
chr8	37460000	37560000	chr25	8800000	8920000
chr8	39800000	39920000	chr25	16160000	16300000
chr8	53960000	54060000	chr25	16760000	16860000
chr8	65440000	65600000	chr25	19700000	19840000
chr9	29060000	29240000	chr25	21100000	21240000
chr9	43820000	43920000	chr26	5380000	5540000
chr9	47480000	47580000	chr27	8360000	8480000
chr9	58340000	58460000	chr28	9520000	9620000
chr10	6420000	6600000	chr28	17920000	18060000
chr10	6680000	7040000	chr30	8020000	8160000
chr10	9060000	9160000	chr30	8980000	9100000
chr10	14800000	15120000	chr30	9040000	9160000
chr10	16580000	16680000	chr31	4660000	4780000
chr10	18700000	18820000	chr31	4840000	4960000
chr10	18840000	18980000	chr31	6140000	6340000

chr10	47340000	47440000	chr31	9340000	9440000
chr10	47420000	47540000	chr31	20380000	20520000
chr10	47480000	47620000	chr32	3520000	3720000
chr10	47560000	47700000	chr32	27380000	27500000
chr10	47740000	47960000	chr33	5200000	5320000
chr10	49220000	49480000	chr34	17620000	17780000
chr11	3020000	3140000	chr36	5140000	5300000
chr11	5820000	6000000	chr36	6620000	6800000
chr11	6340000	6440000	chr37	3160000	3260000
chr11	14800000	14960000	chr37	7380000	7500000
chr11	22000000	22100000	chr37	8780000	8880000
chr11	49780000	49940000	chr37	10100000	10240000
chr11	54640000	54740000	chr37	13560000	13660000
chr11	56820000	57080000			
chr11	57020000	57120000			

Supplementary Table S7. GO analysis of positively selected genes

Go_term	Pvalue	Fold Enrichment
Biological Process		
GO:0048609~reproductive process in a multicellular organism	0.002	2.68
GO:0032504~multicellular organism reproduction	0.002	2.68
GO:0044058~regulation of digestive system process	0.006	25.44
GO:0007276~gamete generation	0.010	2.60
GO:0019953~sexual reproduction	0.010	2.44
GO:0009057~macromolecule catabolic process	0.019	1.91
GO:0008037~cell recognition	0.021	6.79
GO:0060457~negative regulation of digestive system process	0.021	93.30
GO:0010949~negative regulation of intestinal phytosterol absorption	0.021	93.30
GO:0045796~negative regulation of intestinal cholesterol absorption	0.021	93.30
GO:0044265~cellular macromolecule catabolic process	0.022	1.93
GO:0016337~cell-cell adhesion	0.028	2.70
GO:0009063~cellular amino acid catabolic process	0.036	5.49
GO:0032368~regulation of lipid transport	0.040	9.33
GO:0033631~cell-cell adhesion mediated by integrin	0.042	46.65
GO:0009310~amine catabolic process	0.050	4.78
GO:0046777~protein amino acid autophosphorylation	0.062	4.39
GO:0030300~regulation of intestinal cholesterol absorption	0.062	31.10
GO:0033627~cell adhesion mediated by integrin	0.072	26.66
GO:0032372~negative regulation of sterol transport	0.072	26.66
GO:0007158~neuron adhesion	0.072	26.66
GO:0032375~negative regulation of cholesterol transport	0.072	26.66
GO:0010604~positive regulation of macromolecule metabolic process	0.072	1.63
GO:0007586~digestion	0.073	4.10
GO:0007368~determination of left/right symmetry	0.073	6.66
GO:0009799~determination of symmetry	0.077	6.51
GO:0009855~determination of bilateral symmetry	0.077	6.51
GO:0030299~intestinal cholesterol absorption	0.082	23.32
GO:0006281~DNA repair	0.083	2.30
GO:0006259~DNA metabolic process	0.091	1.84
GO:0009077~histidine family amino acid catabolic process	0.092	20.73
GO:0006548~histidine catabolic process	0.092	20.73
GO:0048610~reproductive cellular process	0.095	2.88
GO:0006508~proteolysis	0.095	1.50
Cellular Component		

GO:0042995~cell projection	5.14E-04	2.43
GO:0031981~nuclear lumen	0.010	1.63
GO:0030424~axon	0.011	3.73
GO:0070013~intracellular organelle lumen	0.014	1.52
GO:0043233~organelle lumen	0.019	1.49
GO:0043005~neuron projection	0.019	2.48
GO:0031974~membrane-enclosed lumen	0.025	1.46
GO:0031594~neuromuscular junction	0.027	11.54
GO:0005654~nucleoplasm	0.054	1.63
GO:0044463~cell projection part	0.058	2.53
GO:0005929~cilium	0.065	3.28
GO:0005730~nucleolus	0.067	1.70
GO:0005911~cell-cell junction	0.073	2.67
Molecular Function		
GO:0016894~endonuclease activity, active with either ribo- or deoxyribonucleic acids and producing 3'-phosphomonoesters	1.94E-07	26.15
GO:0004522~pancreatic ribonuclease activity	5.02E-07	35.03
GO:0016892~endoribonuclease activity, producing 3'-phosphomonoesters	1.30E-06	29.50
GO:0004540~ribonuclease activity	6.08E-06	11.32
GO:0004518~nuclease activity	4.69E-05	5.91
GO:0004519~endonuclease activity	9.26E-05	7.47
GO:0004521~endoribonuclease activity	1.35E-04	11.92
GO:0005518~collagen binding	0.056	7.78
GO:0001640~adenylate cyclase inhibiting metabotropic glutamate receptor activity	0.062	31.13
GO:0070742~C2H2 zinc finger domain binding	0.062	31.13
GO:0046982~protein heterodimerization activity	0.071	2.69
GO:0016888~endodeoxyribonuclease activity, producing 5'-phosphomonoesters	0.082	23.35
GO:0003684~damaged DNA binding	0.099	5.60

(cutoff pvalue at 0.1 level)

Supplementary Table S8. Genes found in human genome scans for positive selection

Gene	NS	Description	Reference
ABCG5	4	ATP-binding cassette, sub-family G (WHITE), member 5	[58-60],*
ABCG8	4	ATP-binding cassette, sub-family G (WHITE), member 8	[58-60],*
ALS2CR11	2	amyotrophic lateral sclerosis 2 (juvenile) chromosome region, candidate 11	[61],*
BFAR	3	bifunctional apoptosis regulator	[62-64]
BRE	2	brain and reproductive organ-expressed (TNFRSF1A modulator)	[60],*
C11orf49	2	chromosome 11 open reading frame 49	[60],*
CENPP	2	centromere protein P	[60],*
CPEB3	2	cytoplasmic polyadenylation element binding protein 3	[59,60]
DYNC2LI1	4	dynein, cytoplasmic 2, light intermediate chain 1	[58-60],*
GRM8	2	glutamate receptor, metabotropic 8	[60,61]
HECW2	5	HECT, C2 and WW domain containing E3 ubiquitin protein ligase 2	[58-61,64]
ITGB1	2	integrin, beta 1	[59,60]
LRPPRC	4	leucine-rich PPR-motif containing	[58-60],*
MET	2	met proto-oncogene (hepatocyte growth factor receptor)	[60,61]
MIR423	4	microRNA 423	[58-60],*
MOXD1	2	monooxygenase, DBH-like 1	[58,61]
MRPL46	3	mitochondrial ribosomal protein L46	[61,63,65]
MRPS11	3	mitochondrial ribosomal protein S11	[61,63,65]
NSRP1	4	nuclear speckle splicing regulatory protein 1	[58-60],*
PARN	3	poly(A)-specific ribonuclease	[62-64]
PLA2G10	3	phospholipase A2, group X	[62-64]
PLEKHH2	4	pleckstrin homology domain containing, family H (with MyTH4 domain) member 2	[58-60],*
PRSS1	6	protease, serine, 1 (trypsin 1)	[58,61-64],*
RBKS	2	ribokinase	[60],*
SLC6A4	4	solute carrier family 6 (neurotransmitter transporter, serotonin), member 4	[58-60],*
SPDYE2	4	speedy homolog E2 (Xenopus laevis)	[60,61,65],*
SPDYE6	4	speedy homolog E6 (Xenopus laevis)	[60,61,65],*
STK17B	5	serine/threonine kinase 17b	[58-61,64]
WDR75	2	WD repeat domain 75	[58,60]
ZMYM2	6	zinc finger, MYM-type 2	[60-63,65],*
ZMYM5	6	zinc finger, MYM-type 5	[60-63,65],*
ZNF786	2	zinc finger protein 786	[61],*

*, Fst based calculation (unpublished) in Akey 2009 Genome Research

Supplementary Note 1. Sanger verification of Single Nucleotide Variants (SNVs)

Our genetic information (e.g. SNP calling) was extracted aggregating across all 11 individuals including the boxer reference. DNA for Sanger verification from all individuals, however, is not available. Here, we performed the experimental verification in a small dataset and extracted the trend for the larger sample. Thus, PCR primers were designed to sequence genome segments across the dog genome for six individuals during our whole genome sequencing. False positives and false negatives in SNP calling within a single individual were examined first, followed by across all the 11 individuals.

After quality control, we were able to sequence, using Sanger methods, 614 segments with a total length of 263,763bp (across all six individuals). For most amplicons we were able to sequence all six individuals (Supplementary Table S2).

The levels of false positive and false negative SNP calling within each individual is shown in Supplementary Table S3. False positives are quite low and false negatives are around 0.2-0.3. A large proportion of the false negatives are due to low coverage in the genome sequencing. If we restrict our analysis to parts of the genome that are sequenced at higher coverage, the false negatives are much fewer.

Since errors are mostly random and non-overlapping, when multiple individuals are sequenced, the false positives will increase, but false negatives decrease as germline SNPs will be shared between individuals. Therefore, we randomly sampled 2-6 of the Sanger sequenced individuals and examined the trend of false positives and false negatives by changing the number of sampled individuals (Supplementary Figure S2). As shown, the observed trends match our expectations. It should be emphasized that the decrease in false negative is rapid while the increase in false positive is quite slow. If we fit a line through the observed points, the predicted false negative should be less than 10%, while false positive should be no larger than 5% across 11 individuals (Supplementary Figure S2).

A set of similar calculations was done for Indels. The calculated false positive is 6.5% while false negative is 55% (Supplementary Table S4). The high false negative is mostly due to the fact that we used very stringent criteria and only picked high quality indel mutations.

Supplementary Note 2. PCA analysis

a) PCA analysis over all the canids

Genotype data was obtained from previous studies[66,67] for 912 dogs, 209 grey wolves, 58 coyotes and 12 red wolves and combined with data obtained from the genomes reported here for 11 individuals and one additional outgroup species (a red wolf that we sequenced) for a total of 1203 individuals surveyed across the 48k SNP markers.

SmartPCA was employed to perform the PCA analysis across the 1203 individuals. From Fig.2d we see that the split between the domesticated dogs and wolves is clear. The first principle component, which accounts for 10.4% of the total variation, directly separates these two groups. Interestingly, there is a group of dogs (we call this group 1) that are closer to the wolves.

Our Chinese indigenous dogs are located within group 1. In addition to the Chinese indigenous dogs, other breeds that are known to originate in China/Southeast Asia (e.g., Tibetan Mastiff (that we sequenced here), Chow-chow, Chinese Shar-Pei) are in group 1. A few other breeds, for example, Dingo and New Guinea Singing Dog, are suggested to have South-east Asian origin[68,69], while a few other breeds from Siberia (Siberian Husky and Alaskan Malamute) and Japan (Akita) might also have a Southeast Asian origin[70]. All of these breeds are in group 1. The Basenji, an African dog that is often classified as an ancient breed, is the only other breed found in group 1, and a previous mtDNA study also suggested a closer relationship of this breed to many Chinese breeds[71]. The exact history of the Basenji, especially whether it has an Asian root is not very clear at this moment.

Previous studies have argued for a Middle-Eastern origin of dog based on geographic patterns from wolves, especially the fact that Middle Eastern wolves, as a group, seem to be closer to dogs than wolves from other places using the 48K SNP chip data[66,67]. There are several potential confounding factors that might affect conclusions drawn from that study.

1) Wolf populations have been greatly affected by human activities in recent history. For example, the ancestral Chinese wolves for the domesticated dog may be extinct. It is difficult to use patterns from extant wolves to infer the domestication location of an ancestral population.

2) Several European wolves are found to be even closer to dogs, in spite of the fact that European wolves as a whole are slightly further away from dogs than Middle Eastern wolves (Fig.2d).

3) The Southwest Asian wolves are much further away from dogs, and are quite distinct from Middle Eastern wolves, even though geographically they are very close (Fig.2d). This suggests that there might be possible confounding factors contributing to the discrepancy between genetic relationship and geographic location.

4) SNP ascertainment biases seem to be quite strong in the 48K SNP chip data. For example, when we look at allele frequencies at the SNP positions in our resequenced data, these SNPs are strongly enriched towards high frequency (Supplementary Figure S5). In addition, SNP ascertainment biases are found to be correlated with the fact that these SNP markers were developed during the sequencing of the first dog genome (a boxer individual). In the PCA plot, the boxer/bull dogs group is distributed at some distance from the other dog breeds (Supplementary Figure S6). SNP ascertainment biases might have significantly contributed to PC2.

These observations indicate that demographic relationship of wolves is not yet resolved. The argument for a Middle Eastern origin using patterns from wolves might be confounded by these factors. These observations suggest that a careful examine of these factors is needed before we can better explore the question of domestication location using patterns from wolves.

b) Chinese indigenous dogs and native dogs from other geographic regions

Using mtDNA and Y chromosome data, several of the previous studies found that genetic diversity is highest in Southeast Asia/China and is generally higher than populations from the rest of the world [71-74]. One study that challenged this view is from Boyko et al 2009, where the authors found that African village dogs have comparable genetic diversity as those from Southeast Asia[75]. However, this conclusion was later contested and was found to be questionable[73,74]. A recent study with many native dogs across much of the old world also revealed a pattern with the highest genetic diversity in Southeast Asia [72]. In addition, Chinese indigenous dogs together with several dog breeds originated from Southeast Asia (often designated as ancient breeds) are found to be the most basal lineages linking to grey wolves [71,73,74].

Besides studies using only a few genetic markers, majority of the whole genome studies are based on SNP chips. For example, several work from Wayne and his colleagues have surveyed the genetic information among a global collection of more than 1000 canids [66,67,76]. In these studies, several dog breeds from Asia are found to be the first tier of groups that are closest to wolves [67]. Majority of the dogs surveyed were breed dogs except a few cases where the African village dogs were included[66,67,76].

When we extracted the genetic information, including all the dog breeds and the African dogs, we found that, the African village dogs are much further away from the wolves than the first tier of dogs, which include the Chinese native dogs sequenced here and the ancient breeds (Supplementary Figure S7). It supports that Chinese indigenous dogs together with a group of breed dogs that originated in Southeast Asia are the first tier of dogs that are closest to wolves.

(*data were extracted from http://genomemirror.bscb.cornell.edu/cgi-bin/hgGateway)

Supplementary Note 3. Demographic analysis

When performing demographic analysis using dadi, there are various options for the search space for the underlying parameters. In our analysis, we set the upper bound of t to be 0.3 (equivalence of ~100,000 years), the maximum likelihood estimates show a single sharp peak around the 0.1 (32,000 years). This parameter setting is equivalent for using a hard bound (a fixed interval for the possible divergence time) similar to what is typically conducted in molecular dating.

In the exploratory analysis, we noticed a phenomenon with the statistical model that is not discussed in the literature very extensively and worth a discussion here. The combination of parameters, in particular, migration rate (m) and divergence time (t) can have multiple peaks when inspecting within a large parameter domain (e.g. when setting the parameter boundary to be much wider). In particular, the likelihood function might take a form $f(x,y)=f(x+y, x*y)$, where x and y are two parameters. This form means that the x and y are somewhat "equivalent", in other words, a situation with a large x and a small y might be as "good" (in the sense of the likelihood) as those with a small x and a large y. The parameter t (divergence time) and m (migration rate) here seem to have a form similar to this effect. Thus, there are multiple peaks in the likelihood surface with different combinations of parameters.

Since this is purely a mathematical property, in practice, it can be treated in two ways. One is putting a prior and modeling things in the Bayesian framework (e.g. similar to the soft bound in the molecular dating literature). The other is to restrict the parameter bound within biologically reasonable ranges (e.g. hard bound). In our setting, we know that very deep divergence is not possible for domesticated species (e.g. through fossil and archeological records).

In our analysis, when we set the upper bound of t to be 0.3 (equivalence of ~100,000 years). The maximum likelihood estimates show a single sharp peak around the 0.1 (32,000 years). When we expanded the upper bound of t to be 0.5 or 0.8 (equivalence of 160,000 or 250,000 years), the second parameter peak "appears" (Supplementary Figure S8). For example, in the 100 bootstrap replicates, a significant proportion of the replicates have point estimate of t to be 0.5 (the right boundary, the mean of the right peak estimates is 0.499392).

The reality is that, in the parameter domain, there is a second peak at large t and the estimated value is hitting the parameter boundary. When we set the upper bound to 0.8, the observation is very similar. So, there are two peaks in the likelihood surface. One of the peaks for t is at 0.1, the other is at a very large t. Since the maximum likelihood optimizer (search function) is doing a local hill climbing. Which peak it finds (absorbing to) critically depends on the starting value. We found that, for a given bootstrap sample, if we run the optimizing a lot more times (corresponding to different starting values), the number of bootstrap samples that "absorb" to the right boundary will also decrease.

Through our limited explorations, we find that the peak might be much larger than 0.8. (*The estimates keep hitting the right bound when we extend the upper bound. However,

the program becomes very slow when the bound is set to be very large, say >2 due to the large search space). Since we know that very deep divergence is not possible for these domesticated species, we thus removed the estimates from these bootstrap replicates (Equivalence of implementing a hard bound). The maximum likelihood estimates for the rest of the bootstrap samples are very similar to the case when we set the upper bound to be 0.3 (Supplementary Table S5).

When we extracted site frequency spectra from the sequenced individuals, we restricted ourselves to the non-coding part of the genome that is sequenced to a higher coverage (See maintext). We are interested in quantify possible ascertainment biases in our extracted site frequencies in addition to Sanger sequencing verifications.

We employed methods developed in a previous study, which uses a maximum likelihood method to estimate allele frequencies in low and medium coverage next-generation sequencing data. This method is based on integrating over uncertainties in the data for each individual[77]. When we compared our extracted Site Frequency Spectra (SFS) with the inferred SFS using this maximum likelihood method, we found that our results match quite well with the inferred SFS as well as the best fitted model (Supplementary Figure S9 and S10).

Supplementary Note 4. Diversity estimates and genome wide plots

Since the dogs in our sample are quite diverse, and do not come from the same population, the heterogeneity of the individuals will prevent sophisticated population genetic methods from identifying traces of adaption. We thus looked for the footprints of adaption by looking for: 1) focal regions that show reduced genetic diversity in the dog population (we used the bottom 5% quantile from the dog mean genome wide distribution), 2) segments are not in a low diversity regions in wolves (we used the bottom 20% quantile from the wolf mean genome wide distribution), 3) there is a high divergence between the dog and wolf populations (we used the top 95% quantile in the Fst distribution as the cutoff).

A full display of the diversity and cutoffs are presented in Supplementary Figure S11 and the associated genomic regions are listed in the Supplementary Table S6.

Supplementary Note 5. Genes of interest in addition to those in the main text

Several genes (Supplementary Table S7), in particular genes for metabolic enzymes are in our list of overlapping genes. For example, *PLA2G10* is an important phospholipase involved in the biochemical pathways transforming phospholipids and other lipophilic molecules[78]. *PRSS1* gene is a trypsinogen gene[79] and *RBKS* is a ribokinase[80]. All of these proteins play very important and diverse biological functions in processes such as inflammation, cell growth, signaling and death and maintenance of membrane phospholipids.

In addition to genes such as *MET*, other cancer-related genes involved in the apoptosis pathway were found. For example, *ITGB1* is a member of integrin family involved in the metastatic diffusion of tumor cells [81]. *BFAR*, a gene that have been found in five human genome scans, is an important apoptosis regulator[82]. The *BRE* gene is part of a protein complex that is involved in DNA damage and may also act as a death receptor-associated anti-apoptotic protein[83]. *STK17B* is a positive regulator of apoptosis[84]. Natural selection on apoptosis related genes on the human lineage was noticed in previous studies, e.g. [85]. *ZMYM2* is a transcription factor and translocation of this gene with fibroblast growth factor receptor-1 gene (*FGFR1*) can result in a fusion gene that is associated with the stem cell leukemia lymphoma syndrome [86].

Supplementary References

58 Tang, K., Thornton, K. R. & Stoneking, M. A New Approach for Using Genome Scans to Detect Recent Positive Selection in the Human Genome. *PLoS biology* **5**, e171 (2007).

59 Voight, B. F., Kudaravalli, S., Wen, X. & Pritchard, J. K. A map of recent positive selection in the human genome. *PLoS biology* **4**, e72 (2006).

60 Wang, E. T., Kodama, G., Baldi, P. & Moyzis, R. K. Global landscape of recent inferred Darwinian selection for Homo sapiens. *Proc Natl Acad Sci U S A* **103**, 135-140 (2006).

61 Kimura, R., Fujimoto, A., Tokunaga, K. & Ohashi, J. A practical genome scan for population-specific strong selective sweeps that have reached fixation. *PLoS ONE* **2**, e286 (2007).

62 Carlson, C. S. *et al.* Genomic regions exhibiting positive selection identified from dense genotype data. *Genome research* **15**, 1553-1565 (2005).

63 Kelley, J. L., Madeoy, J., Calhoun, J. C., Swanson, W. & Akey, J. M. Genomic signatures of positive selection in humans and the limits of outlier approaches. *Genome research* **16**, 980-989 (2006).

64 Sabeti, P. C. *et al.* Genome-wide detection and characterization of positive selection in human populations. *Nature* **449**, 913-918 (2007).

65 Williamson, S. H. *et al.* Localizing recent adaptive evolution in the human genome. *PLoS genetics* **3**, e90 (2007).

66 vonHoldt, B. M. *et al.* A genome-wide perspective on the evolutionary history of enigmatic wolf-like canids. *Genome research* **21**, 1294-1305 (2011).

67 Vonholdt, B. M. *et al.* Genome-wide SNP and haplotype analyses reveal a rich history underlying dog domestication. *Nature* **464**, 898-902 (2010).

68 Oskarsson, M. C. *et al.* Mitochondrial DNA data indicate an introduction through Mainland Southeast Asia for Australian dingoes and Polynesian domestic dogs. *Proc Biol Sci* **279**, 967-974 (2012).

69 Savolainen, P., Leitner, T., Wilton, A. N., Matisoo-Smith, E. & Lundeberg, J. A detailed picture of the origin of the Australian dingo, obtained from the study of mitochondrial DNA. *Proc Natl Acad Sci U S A* **101**, 12387-12390 (2004).

70 Parker, H. G. *et al.* Genetic structure of the purebred domestic dog. *Science (New York, N.Y* **304**, 1160-1164 (2004).

71 Savolainen, P., Zhang, Y. P., Luo, J., Lundeberg, J. & Leitner, T. Genetic evidence for an East Asian origin of domestic dogs. *Science* **298**, 1610-1613 (2002).

72 Brown, S. K. *et al.* Phylogenetic distinctiveness of Middle Eastern and Southeast Asian village dog Y chromosomes illuminates dog origins. *PLoS One* **6**, e28496 (2011).

73 Ding, Z. L. *et al.* Origins of domestic dog in southern East Asia is supported by analysis of Y-chromosome DNA. *Heredity* **108**, 507-514 (2012).

74 Pang, J. F. *et al.* mtDNA data indicate a single origin for dogs south of Yangtze River, less than 16,300 years ago, from numerous wolves. *Molecular biology and evolution* **26**, 2849-2864 (2009).

75 Boyko, A. R. *et al.* Complex population structure in African village dogs and its implications for inferring dog domestication history. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 13903-13908 (2009).
76 Boyko, A. R. *et al.* A simple genetic architecture underlies morphological variation in dogs. *PLoS biology* **8**, e1000451 (2010).
77 Kim, S. Y. *et al.* Estimation of allele frequency and association mapping using next-generation sequencing data. *BMC Bioinformatics* **12**, 231 (2011).
78 Cupillard, L., Koumanov, K., Mattei, M. G., Lazdunski, M. & Lambeau, G. Cloning, chromosomal mapping, and expression of a novel human secretory phospholipase A2. *J Biol Chem* **272**, 15745-15752 (1997).
79 Emi, M. *et al.* Cloning, characterization and nucleotide sequences of two cDNAs encoding human pancreatic trypsinogens. *Gene* **41**, 305-310 (1986).
80 Park, J., van Koeverden, P., Singh, B. & Gupta, R. S. Identification and characterization of human ribokinase and comparison of its properties with E. coli ribokinase and human adenosine kinase. *FEBS Lett* **581**, 3211-3216 (2007).
81 Garmy-Susini, B. *et al.* Integrin alpha4beta1 signaling is required for lymphangiogenesis and tumor metastasis. *Cancer Res* **70**, 3042-3051 (2010).
82 Prat, M. *et al.* The receptor encoded by the human c-MET oncogene is expressed in hepatocytes, epithelial cells and solid tumors. *Int J Cancer* **49**, 323-328 (1991).
83 Li, Q. *et al.* A death receptor-associated anti-apoptotic protein, BRE, inhibits mitochondrial apoptotic pathway. *J Biol Chem* **279**, 52106-52116 (2004).
84 Sanjo, H., Kawai, T. & Akira, S. DRAKs, novel serine/threonine kinases related to death-associated protein kinase that trigger apoptosis. *J Biol Chem* **273**, 29066-29071 (1998).
85 da Fonseca, R. R., Kosiol, C., Vinar, T., Siepel, A. & Nielsen, R. Positive selection on apoptosis related genes. *FEBS Lett* **584**, 469-476 (2010).
86 Lierman, E. & Cools, J. Recent breakthroughs in the understanding and management of chronic eosinophilic leukemia. *Expert Rev Anticancer Ther* **9**, 1295-1304 (2009).