doi:10.1038/nature11837

Supplemental Discussion accompanying Axelsson et al. "The genomic signature of dog domestication reveals adaptation to a starch-rich diet"

Supplementary Discussion, section 1. Experimental verification confirms high accuracy of SNP calling.

Stringent filtering criteria (see Methods) were used to identify 3,786,655 SNPs in the combined dog and wolf data. Based on a high accuracy of SNP calls in a previous study (>95%) using a nearly identical SNP calling methodology we believe that the majority of these SNPs are true[38]. To verify this, we designed an iPLEX assay targeting 124 SNPs. We obtained reliable genotyping calls for 114 of these SNPs, out of which 113 (99.1%) were confirmed to be variable (Table S3) in a panel of 71 dogs, representing 38 diverse breeds, and 19 wolves of worldwide distribution (Table S14). The iPLEX result thus confirms a high accuracy of SNP calls. One position identified as a SNP by the SOLiD data (chr26: 27,981,169 at which 10 reads support C and 3 support T) was found to be invariable in the iPLEX assay (only C's called). Although it is possible that we failed to sample the variant allele in the iPLEX assay, this result may reflect SOLiD sequencing errors.

Supplementary Discussion, section 2. Detecting signatures of selection in 200 Kb windows.

Selection on a single advantageous mutation affects patterns of genetic variation at nearby loci causing (i) a reduction in heterozygosity, (ii) a skewed allele frequency

distribution and (iii) an excess of high frequency derived alleles across a region surrounding the selected allele[39]. We used two approaches to search for regions in the dog genome that bear patterns of genetic variation consistent with these signatures of selection. Using a 200 Kb sliding window we first calculated the average pooled heterozygosity, H_P, in dogs and wolfs separately (see Supplementary Discussion, section 8, for a detailed description of the wolf analysis), and, secondly, the average fixation index, F_{ST}, between the two taxa. Putatively selected regions were identified by extracting windows in the low end of the $Z(H_P)$ distribution, and high end of the $Z(F_{ST})$ distribution, by applying a threshold at 5 standard deviations away from the mean. By applying these thresholds we extracted 38 windows, representing 14 unique regions with extremely low levels of heterozygosity in dog and 82 windows, representing 35 unique regions, with strongly elevated F_{ST} values. We partitioned the F_{ST} regions into those that are likely to represent selection in dog (n=30), wolf (n=3) or both taxa simultaneously (n=2), based on the corresponding $Z(H_P)$ scores in the two taxa (Methods). A total of 36 unique autosomal CDRs were identified by the two approaches combined (Table S6, Fig. 1bc).

As the variance of H_P and F_{ST} will depend on the number of SNPs used for each calculation, spurious selection signals will be more likely in windows with few variable sites. To reduce the number of false positives we have therefore required that windows harbour a minimum of 10 variable sites to be included in the analysis. We tested a range of different window sizes (50, 100, 200, 500 and 1000 Kb) and note that a window size of 200 Kb results in a low number of windows with few SNPs; forty-nine out of 21,927 200 Kb windows contained <10 SNPs, in contrast to 3,758 out of 87,873 50 Kb windows. Hence, using a window size of 200 Kb allowed us to screen a large fraction of the genome at a false positive rate that likely is lower than if a smaller window size had been used.

Even if windows harbour sufficient amounts of variable sites it may still be challenging to differentiate between signals caused by selection and genetic drift, respectively. In this regard two parameters - window size and Z-score threshold, can be tuned to reduce the risk of including false positives that are due to drift. Determining an optimal window size, in terms of maximising the sensitivity towards detecting selection events at a low cost of false positives, is however complicated given the complex and partly unknown demographic history of dogs. By comparing the distribution of H_P in analyses using window sizes of 50 and 200 Kb (Fig 1 and Fig S7), respectively, we note that the larger window size is associated with a small standard deviation (sd) (sd_{200Kb} = 0.054 vs. sd_{50Kb} = 0.071) that may indicate that the outliers we have detected are enriched for rare and strong selective sweeps compared to if smaller window sizes had been used. Determining a Z-score threshold that maximises sensitivity at a low false positive rate is equally complicated, given the circumstances discussed above. We chose to set the thresholds at $Z(H_P)$<-5 and $Z(F_{ST})$>5 as they represent the extreme tails of the distributions and hence should be enriched for true signals of selection. It is however possible that windows with less extreme $Z(H_P)$ and $Z(F_{ST})$ values may merit further investigation too, as they may have contributed to dog domestication as well. We suggest that the best way to evaluate the CDRs detected here is to seek confirmation of the observed signatures of selection in additional dog and wolf individuals, as well as test if there is evidence for altered gene activity or protein function that is associated with the genetic signal.

Supplementary Discussion, section 3. Confirming signatures of selection.

High sequence coverage and mapping quality across CDRs.

During mapping and variant calling we have applied a number of filtering criteria (see Methods) to ensure that the data used for selection analyses is of high general quality.

Nevertheless, to test if CDRs may represent regions that are affected by experimental or technical artefacts more than other regions we compared CDRs and the genome as a whole, with regards to sequence coverage and mapping quality.

The average pooled heterozygosity (H_P) is negatively correlated with sequence coverage (rho=-0.53, p<0.0001, (Spearman), Fig. S8) and the average sequence coverage is higher in regions of low heterozygosity compared to the genomic average (24.5 ($Z(H_P)$<-5) vs. 19.1 (genomic average), p<0.0001 (wilcoxon)). In line with these observations, F_{ST} is positively correlated with sequence coverage (rho=0.17, p<0.0001, (Spearman), Fig. S9) and the average sequence coverage is higher in regions of high fixation index compared to the genomic average (24.5 ($Z(H_P)$<-5) vs. 19.1 (genomic average), p<0.0001 (wilcoxon)). The relatively high sequence coverage in CDRs reflects an inverse relationship between the theoretical minimum of allele frequencies and sequence coverage. Our method may thus be biased towards detecting outliers in regions of slightly higher sequence coverage, however this ensures that the evidence for putatively selected regions is strong.

We also note that the root mean square mapping quality (RMS) of reads mapping within CDRs (RMS_{CDR} = 42) is slightly higher than the genomic average (RMS_{GENOME} = 40.1, p<0.0001 (wilcoxon)). In summary, these results indicate that CDRs represent real outliers in terms of genetic varation.

Unequal sample size of dog and wolf.

To reduce the risk of introducing biases due to unequal sample size in dog and wolf we have used a method to estimate F_{ST}[40] that takes differences in sample size into account.

Nevertheless, to test if the unequal dog and wolf representation may have contributed to variance of statistics across the genome, we redid the F_{ST} and H_P analyses based on a random sub sample of the dog data such that the average coverage in dog equalled that in wolf (6.2x) (Fig. S10 and S11).

First, analysing the sub sampled data we detect 17 regions with $Z(H_P)<-5$, out of which 12 overlap the 14 regions from the original analysis. The 2 regions from the original analysis that did not overlap a region in the sub sampling analysis still show a clear reduction in heterozygosity ($Z(H_P)<-4$) in the sub sampled data.

Secondly, we find 20 regions with $Z(F_{ST})>5$ in the sub sampled data, out of which 18 overlap the 35 regions detected in the original analysis. The 17 regions from the original analysis that do not overlap a region in the sub sampling analysis still show a clear increase in F_{ST} in the sub sampling analysis (8 regions have $Z(F_{ST})>4$ and the remaining 9 regions have $Z(F_{ST})>3$).

The majority of the original CDRs were thus identified as CDRs in the sub sampled data, and the remainder still stand out, although not significantly so. These results suggest that the unequal sample sizes of dogs and wolves likely have little effect on the overall results of our selection analyses.

Confirming signatures of selection.

We used several independent means and datasets to confirm that our methodology identifies regions that truly represent outliers in terms of genetic variation.

First, as strong selection for an advantageous allele will lead to a sharp rise in frequency of liked genetic variation, it is expected that a significant fraction of variable

sites surrounding the selected allele will become lost due to fixation. Genetic signatures of selection are thus expected to show a reduction in the number of segregating sites. In line with this expectation we note that H_P is positively correlated with the average number of segregating sites in 200 Kb windows (rho=0.23, p<0.0001, (Spearman), Fig. S12) and that the average number of segregating sites is markedly lower in regions of low heterozygosity compared to the genomic average (71.2 ($Z(H_P)$<-5) vs. 253.3 (genomic average), p<0.0001 (wilcoxon)). Similarly, we also observe that the average F_{ST} is negatively correlated with the corresponding average number of segregating sites in 200 Kb windows (rho=-0.25, p<0.0001, (Spearman), Fig. S13) and that the average number of segregating sites is markedly lower in regions of high fixation index compared to the genomic average (97.8 ($Z(F_{ST})$>5) vs. 244.9 (genomic average), p<0.0001 (wilcoxon)).

We then genotyped 47 and 48 randomly selected SNPs across the *MGAM* and *SGLT1* regions, respectively in 71 dogs representing 38 diverse breeds and 19 wolves of worldwide distribution (the reference panel, Table S14). We noted a significant reduction in heterozygosity for dogs, as well as a significant increase in F_{ST} across these regions, consistent with the signals observed in the resequencing data (Main text, Fig. 3 and S4). We also genotyping 17 additional diagnostic SNPs representing 13 more CDRs in the same panel of dogs and wolves as described above. The F_{ST} for these SNPs averaged 0.63, which is clearly above the genomic average of 0.22. This difference provides additional support that CDRs represent real outliers in terms of genetic differentiation between dog and wolf.

We finally compared patterns of genetic variation within and outside CDRs in a panel of 507 dogs and 15 wolves that we recently genotyped using a high density SNP array (Illumina 170K HD canine array) (containing 173, 622 SNPs with an average spacing of 13 Kb)[41]. We note a clear drop in SNP density (0.04/Kb in CDRs vs. 0.09/Kb

genome wide) and dog minor allele frequency (0.08 in CDRs vs. 0.24 genome wide), but only a slight decrease in wolf minor allele frequency (0.19 in CDRs vs. 0.29 genome wide) in CDRs compared to the genome wide average (Table S22). We also see a clear increase in F_{ST} in CDRs (0.27) compared to the genome wide average (0.13) (Table S22). These patterns are consistent with selection acting specifically in the dog lineage. We also studied the detailed patterns of genetic variation near the *MGAM* and *SGLT1* regions using the HD-array data. Within the *MGAM* region we note four consecutive SNPs spanning the *MGAM* and *TAS2R38* genes (chr.16:10,153,271; 10,171,680; 10,182,601 and 10,212,709) that are completely fixed in all 507 dogs (Fig. S14), again corroborating the results of the resequencing data (Fig. 3). Within the SGLT1 region we note a single SNP (chr. 26: 27,964,669) that is completely fixed in dogs (Fig. S15). This SNP is located just downstream (558 bp) of the candidate causative amino acid substitution (chr 26: 27,964,111) (Main text).

The consistency of the signals observed in the resequencing data and the two genotyping assays provide strong support for the ability of our method to identify potential signatures of selection affecting all, or nearly all dogs.

Comparison of CDRs and signatures of selection detected in a previous study.

Using a combination of F_{ST} and haplotype based analyses VonHoldt and colleagues[42] identified a set of 14 regions that they argued may have been affected by selection during the initial phase of dog domestication. A comparison of the results of our study and those of VonHoldt et al. could thus represent an additional means to validate our results. We do however note that none of the regions detected by VonHoldt and colleagues overlap any of the CDRs detected here. We investigated this discrepancy in more detail by extracting the maximum $Z(F_{ST})$-score, as well as the minimum $Z(H_P)$-score, recorded in our data among all 200 Kb windows that overlap each of the 14

VonHoldt regions, respectively. We find that the average maximum $Z(F_{ST})$ of the 14 VonHoldt regions (avg. $Z(F_{ST})$=1.83, range: -0.32 – 3.66) is only moderately elevated. Similarly, we note that the average minimum $Z(H_P)$ of the 14 VonHoldt regions (avg $Z(H_P)$= -1.55, range: 0.53 - (-2.73)) represents a moderate drop in heterozygosity. How can this discrepancy between the results of the studies be explained?

First, VonHoldt et al. based their calculations on genotyping data, in which SNPs were ascertained almost exclusively in dog (this is unlike our study where wolves have contributed significantly to the SNP discovery). Since selection affecting the ancestors of all modern dogs results in a reduction in the number of segregating sites across the affected region, it follows that regions affected by strong selection are unlikely to be represented on the genotyping array. Based on this bias, most of the regions affected by strong selection during early dog domestication would have been missed in the analysis of VonHoldt et al. Weak selection or partial sweeps (such as selection in a subset of dog breeds) may still have been detected using the genotype data (Fig. S16).

Secondly, due to the low marker density many VonHoldt regions were represented by a relatively limited number of variable sites, which may have led to a large variance associated with the test statistic. Similarly, in a particular VonHoldt region, spanning ~250 Kb (Fig. S17), the wide spacing of markers may have lead the analysis of VonHoldt et al. to erroneously assume that markers from two separate regions with skewed allele frequencies represented one single signal. This may have led to a falsely inflated test statistic for this region.

Finally, since we have used a window size of 200 Kb it is possible that we could have missed smaller regions. However, among the 14 VonHoldt regions, only four span just under 200 Kb and in none of these cases do we see any clear indications of selection spanning a narrow region in the resequencing data.

To summarise, it thus appears as if the regions detected by VonHoldt et al. represents weak selection events. Based on these conclusions we see no reason to question the validity of the CDRs detected here due to the divergent results of the two studies.

Large overlap between H_P and F_{ST} regions.

Due to the way the fixation index and pooled heterozygosity are estimated, they are expected to be highly correlated. Yet, a number of CDRs are only detected as significant outliers using one of the statistics. To understand why this is the case we studied these regions in more detail.

23 F_{ST} regions do not overlap regions of significantly reduced heterozygosity in either dog (($Z(H_P)_{DOG}$)<-5) or wolf (($Z(H_P)_{WOLF}$)<-5). All of these regions do however still show clear reductions in heterozygosity, with an average $Z(H_P)$ = -4.41 (range: -3.87 - (-4.94)) (calculated based on the lowest of the values in dog and wolf) (see Fig. S18). These regions thus all appear to be regions that nearly reach the threshold for significance in the H_P analysis.

Four regions of reduced H_P in dog (($Z(H_P)_{DOG}$)<-5) do not overlap an F_{ST} region ($Z(F_{ST})$>5). Again, all of these regions do however still show clear increases in F_{ST}, with an average $Z(F_{ST})$ = 4.56 (range: 4.48 – 4.64). This shows that the non-overlapping regions are border line cases that almost reach the threshold of significance in the F_{ST}-analysis. We note that in two of these regions, there are only 12 and 17 segregating sites (see Fig. S19), respectively (average number of segregating sites in regions of $Z(F_{ST})$>5 is 97.8). This suggests that estimates for these particular borderline cases likely have higher variances that in turn could add to the explanation why they reach significance in the H_P-, but not the F_{ST} analysis.

Supplementary Discussion, section 4. Candidate domestication regions on chromosome X.

The X-chromosome differs from autosomes in several population genetic aspects, including a reduction in effective population size and recombination rate. Via the effects of genetic drift this is expected to lead to a significant reduction in levels of genetic variation on the X-chromosome compared to the autosomes. In addition, a reduced mutation rate will further add to the depravation of genetic variation on chromosome X. Furthermore, as a result of the influence of drift it is also expected that the X-chromosome will show increased genetic differentiation between dog and wolf relative to on autosomes. To not confound the results of the main analyses due to these circumstances we decided to analyse chromosome X separately. We used the same methods for these analyses as described previously for the autosomes. As expected, we find that the average pooled heterozygosity, H_P, is lower (H_PX: 0.30 vs. H_PA: 0.33) and the average fixation index, F_{ST}, is higher (F_{ST}X 0.31 vs. F_{ST}A 0.20) on X compared to on autosomes (Fig. S20). We also note that the standard deviations of the H_P (σX: 0.096 vs. σA: 0.054) and F_{ST} (σX: 0.18 vs. σA: 0.091) distributions are larger on the X-chromosome relative to the autosomes, and that a relatively large proportion of the windows reside in the tails of the distributions (Fig. S20). Although this could indicate massive selection on the X-chromosome during dog domestication, it is likely that this mainly reflects a more prominent role of genetic drift on X compared to the autosomes. No windows passed the thresholds of significance ($Z(H_P)$<-5 or $Z(F_{ST})$>5) used for the autosomal analyses, again emphasising the difficulty of separating selection from drift on the X-chromosome. With this difficulty in mind, we nevertheless decided to extract the most extreme windows for further analyses as they represent clear candidates for selection on this chromosome. We did this by applying a threshold of 3 standard deviations from the mean of both the H_P and F_{ST} distributions. 13 windows,

representing 6 unique regions passed this cut-off in the F_{ST} analysis (Fig. S21, Table S23). No windows passed this threshold for the H_P analysis. A region spanning 200 Kb at chr.X: 42,905,187-43,105,187 represents the most extreme CDR in terms of absolute H_p (0.015) and F_{ST} (0.96) values recorded though out the entire dog genome (Table S23). This region harbours *CCNB3* that codes for the G2/mitotic-specific cyclin-B3, which is essential for oocyte maturation[43]. It is thus possible that selection for variants of this gene could be associated with the reproductive changes accompanying dog domestication, for example changes from one to two estrous cycles per year.

Supplementary Discussion, section 5. Amylase CNV break point suggests duplications arose via unequal crossing over using L1 LINE elements as template.

A CDR on chromosome 6 coincides with a sharp increase in coverage at the locus coding for amylase (Fig. 2). Two independent CNV detection methods[44,45] and real time quantitative PCR (Fig. 2c, Table S11) confirm a drastic increase in amylase copy numbers in dog relative to wolf. This copy number difference is further supported by a 4-8-fold increase in aligned read depth in dog relative to wolf across 5 additional *alpha-amylase* copies residing on three unmapped contigs (Fig. S22). Two of these copies reside on individual unplaced contigs while three are annotated in tandem on the same unplaced contig (chromosome Un) (Fig. S22). The close proximity of the three neighboring copies suggests that the alpha-amylase CNV arose via tandem duplications as a result of unequal crossing–over between neighboring homologous DNA sequences. To identify the likely target for the initial unequal crossing over event we sought to map the break point between individual copies by studying sequence coverage across the three chromosome Un contigs. A clear increase in dog relative to wolf coverage extended to the contig ends in all cases except the upstream end of the contig harbouring three amylase copies. We estimated the average sequence coverage in 1kb

windows and identified the first window in this contig for which the ratio of dog to wolf coverage exceeded 3. This window is located between positions 46,696,000 and 46,697,000 on chromosome Un. We compared sequence coverage at individual sites across a 10 Kb region spanning this window and noted a clear increase in coverage just downstream of position 46,696,000 (data not shown). This position harbours a L1 LINE element that reoccurs between all three amylase copies on this contig as well as upstream of the two additional annotated amylase copies. It is thus likely that L1 LINE elements located on opposite sides of the ancestral amylase copy served as templates for an unequal crossing over event that caused the initial duplication event from which several additional duplications followed. In support of this, we recently identified L1 LINE elements as the most highly enriched sequence feature near CNVs in the dog genome[46].

Supplementary Discussion, section 6. Chromosome 16 sweep likely represents two independent selection events.

A single 200 Kb window (chr. 16: 10,107,390-10,307,390) spanning the *MGAM* locus has a significant reduction in average pooled heterozygosity (H_P). This putatively selected region is however extended by the F_{ST} analysis to include 300 Kb upstream sequence, thus in total spanning a 500 Kb region (chr. 16: 9,807,391-10,307,390). By studying heterozygosity and F_{ST} estimates for individual sites across this large region (Fig. 3b) we note a short, highly variable section (spanning approximately chr. 16: 10,055,000-10,095,000) that separates this region into two smaller regions. We believe that this reflects a true increase in genetic variation rather than alignment artefacts for several reasons. First, the region spanning this highly variable 40 Kb region is syntenic in the human genome, arguing against a miss assembly in the dog reference genome. Secondly, we see no average difference in sequence coverage between wolf and dog,

nor any other evidence indicative of a CNV in this region. Finally, the frequency of alternatively oriented mate-pairs reads in this region does not differ from surrounding sequence. This argues that, rather than representing a single sweep, the 500 Kb region represents two independent selection events, one of which affects the *MGAM* locus (which is discussed in detail in the main text), and a second affecting the nearby T-cell receptor cluster. T-cell receptors play crucial roles in the immune system by recognizing antigens bound to class I or class II major histocompatibility proteins and given that immune system genes are frequent targets of selection[47] it is not unexpected that selection may have affected the T-cell receptor cluster in dog. We searched for candidate mutations that may have been targeted by this sweep and observed a non-conservative amino-acid substitution leading to a shift from glycine in wolf, to glutamic acid in dog, at residue 61 (chr. 16: 9,852,935) of the T cell receptor beta variable 2 (TRBV2). The resequencing data indicate that this represents a fixed difference between dog and wolf as all canine reads (n=36) support a C, while all wolf reads (n=6) support a T at this position. We confirmed a high degree of differentiation by genotyping this mutation in 72 dogs representing 38 diverse breeds and 21 wolves of worldwide origin (Table S14). Among dogs genotyped, all except two West Highland White Terriers carried at least one copy of the C allele: 62 dogs were homozygous and 8 heterozygous for this allele. 17 out of 21 wolves were homozygous and 2 were heterozygous for the T allele while two wolves were homozygous for the C allele. In addition, we note that three consecutive SNPs (chr. 16: 9,874,039; 9,890,639 and 9,967,827) on the Illumina 170 K Canine HD array located just downstream of the candidate mutation are fixed in 502 dogs (Fig. S14).

The high degree of dog–wolf differentiation in combination with the non-conservative nature of this amino-acid substitution argues that this represents the target of the selective sweep. The fact that the mutation affects a receptor responsible for

antigen binding suggest that this change may reflect exposure to new pathogens as ancestral dogs shifted to a life in close proximity to humans.

Supplementary Discussion, section 7. CDRs are not enriched for CNVs.

To test if CNVs might have played a major role in forming the molecular basis for adaptations during dog domestication we compared the relative abundance of CNVs that are present in all dog breeds, but lacking in wolf, in CDRs (n=8) and the genome as a whole (n=881). Based on the genome wide estimates of CNV abundance, the expected number of CNVs per 200 Kb window is 0.038 and the observed number in 200 Kb windows in CDRs is 0.032. We conclude that there is no evidence for an enrichment of these CNVs in CDRs (p=0.66, Chi square test).

Supplementary Discussion, section 8. Detecting signatures of selection in wolf.

Although it is possible that some wolf populations have experienced severe bottlenecks recently, the significantly larger effective population size in wolf relative to in the ancestral pre-breed dog population[48], suggests that wolves suffered to a less extent than dogs from historical bottlenecks. Selection and drift should thus be more easily teased apart in wolf relative to in dog, arguing that our data, despite including only a single wolf pool, represents a first opportunity to detect signatures of recent selection through out the wolf genome. In agreement with this prediction the distribution of average pooled heterozygosity (H_P) in 200 Kb windows across the autosomal part of the wolf genome forms a narrow peak centred around the mean (avg. autosomal H_P: 0.280), with a few clear outliers harbouring very low levels of genetic variation. We applied the same threshold for identifying outliers ($Z(H_P)<-5$) as in the dog analysis and extracted 23 windows, representing 18 unique regions with extremely low levels of heterozygosity in the wolf genome (avg. length: 272 Kb, avg. H_P: 0.049 (0.013-0.067),

(Table S24, Fig. 1, S2 and S23). We combined these results with those of the previously described F_{ST} analyses (see Supplementary Discussion, section 2) to detect 21 unique wolf candidate selection regions (wolf CSR) containing 113 annotated genes. To formally characterise the function of genes in wolf CSRs we searched for overrepresented GO-terms associated to these genes and find several terms related to intracellular signalling and protein kinase cascades (Table S25). For example, in a wolf CSR on chromosome 13 we note *STK3* that encodes the stress activated Serine/threonine-protein kinase 3 that mediates apoptotic signals[49]. It is unclear what may have triggered the potential selection for altered signal transduction, although stress related to pathogens or other environmental stimuli are potential candidates.

Two wolf CSRs encompass single genes, which may hence represent the target of the putative selection; *ADCY10* encodes adenylyl cyclase 10 that may play a role in sperm maturation[50,51], and Neurobeachin codes for a protein that is involved in body weight control, presumably via an effect on feeding behaviours[52].

Apart from casting light on the recent evolutionary history of the wolf itself, studying selection in wolves serves to contrast the results of the dog analyses. By comparing the results of the GO-analyses of regions under the potential influence of selection in dog and wolf, respectively, we find no apparent functional overlap, indicating that distinct selective pressures have been operating in the two lineages. The result of the wolf analysis thus confirms that selection for efficient starch digestion and a potentially altered nervous system development is unique to the dog lineage. Based on the divergent results of the dog and wolf analyses, it is furthermore unlikely that methodological problems related to the GO-analysis have biased the results of the dog analysis significantly.

Supplementary Discussion, section 9. The genetic relationship of dog and wolf pools.

We used the allele frequencies at the 3,786,655 SNPs in individual pools to construct a phylogenetic tree (Fig. S24) that summarises the genetic relationship of wolves and the five dog pools[53]. This analysis shows that all dog pools are more closely related to each other than to the wolf pool and that no dog pool can be considered more ancestral (or ancient) than other pools. This agrees with earlier results suggesting the wolf is the ancestor of all dogs and that the formation of most modern breeds happened in a short time frame starting with the same relatively limited breeding stock.

References

38 Rubin, C. J. *et al.* Whole-genome resequencing reveals loci under selection during chicken domestication. *Nature* **464**, 587-U145, doi:Doi 10.1038/Nature08832 (2010).

39 Stephan, W. Genetic hitchhiking versus background selection: the controversy and its implications. *Philos T R Soc B* **365**, 1245-1253, doi:Doi 10.1098/Rstb.2009.0278 (2010).

40 Weir, B. S. & Cockerham, C. C. Estimating F-Statistics for the Analysis of Population-Structure. *Evolution* **38**, 1358-1370 (1984).

41 Vaysse, A. *et al.* Identification of Genomic Regions Associated with Phenotypic Variation between Dog Breeds using Selection Mapping. *Plos Genetics* **7**, doi:Artn E1002316
Doi 10.1371/Journal.Pgen.1002316 (2011).

42 vonHoldt, B. M. *et al.* Genome-wide SNP and haplotype analyses reveal a rich history underlying dog domestication. *Nature* **464**, 898-U109, doi:Doi 10.1038/Nature08837 (2010).

43 Hochegger, H. *et al.* New B-type cyclin synthesis is required between melosis I and II during Xenopus oocyte maturation. *Development* **128**, 3795-3807 (2001).

44 Xie, C. & Tammi, M. T. CNV-seq, a new method to detect copy number variation using high-throughput sequencing. *Bmc Bioinformatics* **10**, doi:Artn 80
Doi 10.1186/1471-2105-10-80 (2009).

45 Abyzov, A., Urban, A. E., Snyder, M. & Gerstein, M. CNVnator: An approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. *Genome Res* **21**, 974-984, doi:Doi 10.1101/Gr.114876.110 (2011).

46 Berglund, J. *et al.* Novel origins of copy number variation in the dog genome. *Genome Biol* **13**, R73, doi:gb-2012-13-8-r73 [pii]
10.1186/gb-2012-13-8-r73 (2012).

47 Nielsen, R. *et al.* A scan for positively selected genes in the genomes of humans and chimpanzees. *PLoS Biol* **3**, e170, doi:04-PLBI-RA-0748R3 [pii]
10.1371/journal.pbio.0030170 (2005).

48 Lindblad-Toh, K. *et al.* Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature* **438**, 803-819, doi:Doi 10.1038/Nature04338 (2005).

49 Lee, K. K., Ohyama, T., Yajima, N., Tsubuki, S. & Yonehara, S. MST, a physiological caspase substrate, highly sensitizes apoptosis both upstream and downstream of caspase activation. *J Biol Chem* **276**, 19276-19285 (2001).

50 Tate, S. *et al.* Evidence of the Existence of Adenylyl Cyclase 10 (ADCY10) Ortholog Proteins in the Heads and Connecting Pieces of Boar Spermatozoa. *J Reprod Develop* **56**, 271-278 (2010).

51 Visser, L. *et al.* A comprehensive gene mutation screen in men with asthenozoospermia. *Fertil Steril* **95**, 1020-1024 e1021-1029, doi:S0015-0282(10)02974-2 [pii]
10.1016/j.fertnstert.2010.11.067 (2011).

52 Olszewski, P. K. *et al.* Neurobeachin, a Regulator of Synaptic Protein Targeting, Is Associated with Body Fat Mass and Feeding Behavior in Mice and Body-Mass Index in Humans. *Plos Genetics* **8**, doi:Artn E1002568
Doi 10.1371/Journal.Pgen.1002568 (2012).

53 Karlsson, E. K. *et al.* Efficient mapping of mendelian traits in dogs through genome-wide association. *Nat Genet* **39**, 1321-1328, doi:Doi 10.1038/Ng.2007.10 (2007).

Supplemental Figures and Legends

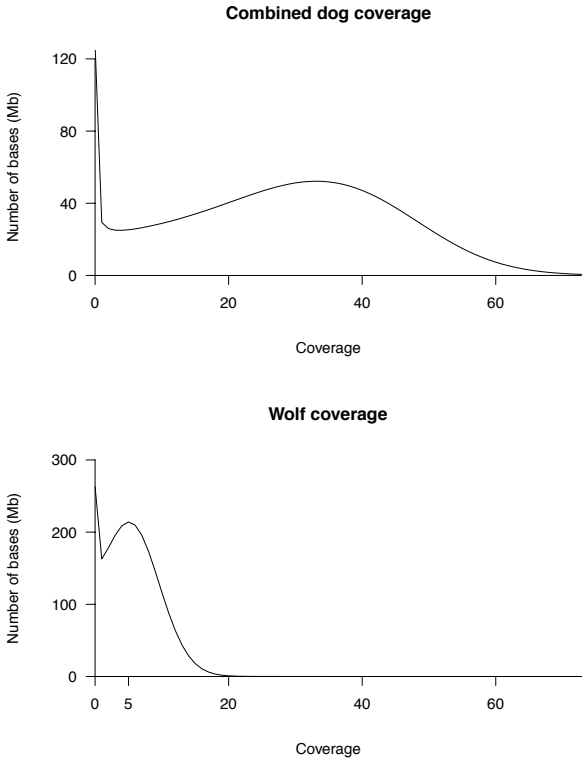

Figure S1. Distribution of sequence coverage for the five dog pools combined and a single wolf pool.

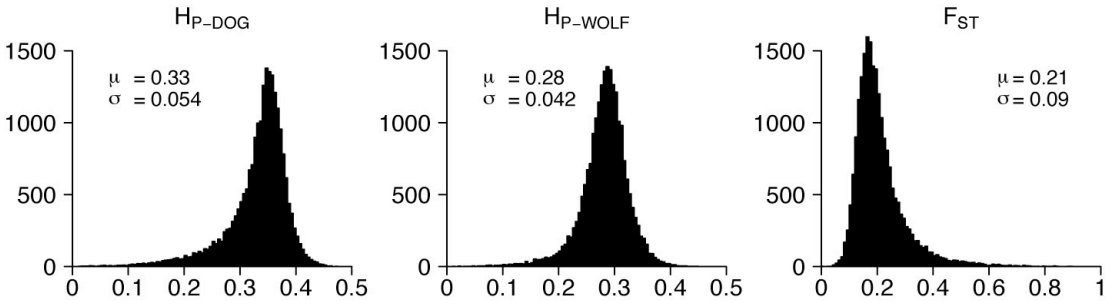

Figure S2. Distribution of heterozygosity and fixation index. Distribution of average pooled heterozygosity in dog ($H_{P\text{-}DOG}$) and wolf ($H_{P\text{-}WOLF}$) respectively, as well as average fixation index (F_{ST}), for autosomal 200 Kb windows (σ, standard deviation; μ, average).

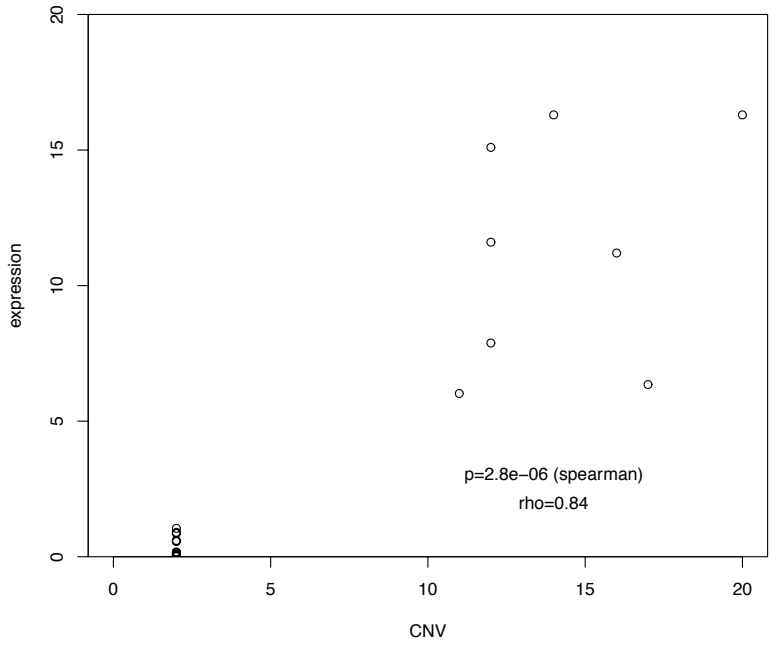

Figure S3. *AMY2B* expression measured in pancreas is significantly correlated with copy number in 12 wolves and 8 dogs (rho=0.84, p<0.0001, Spearman).

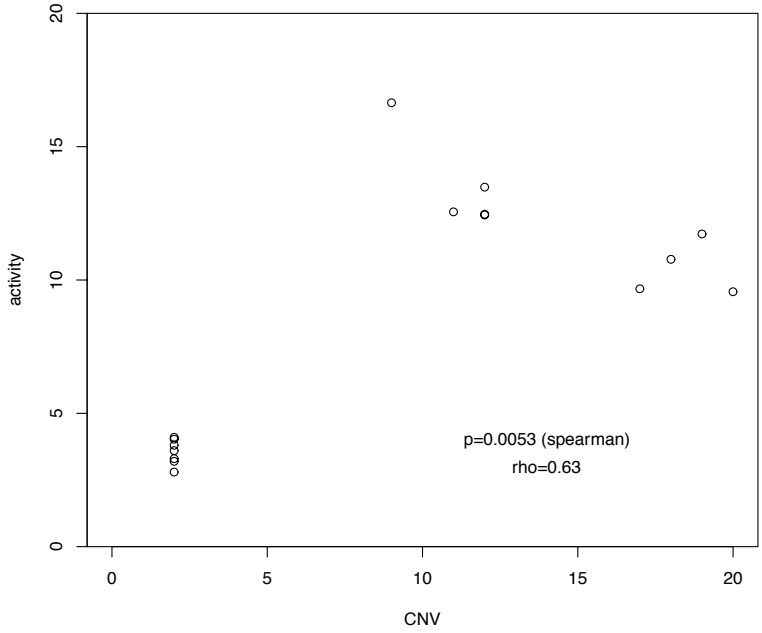

Figure S4. Serum amylase activity is significantly correlated with *AMY2B* copy number in 7 wolves and 11 dogs (rho=0.63, p=0.0053, Spearman).

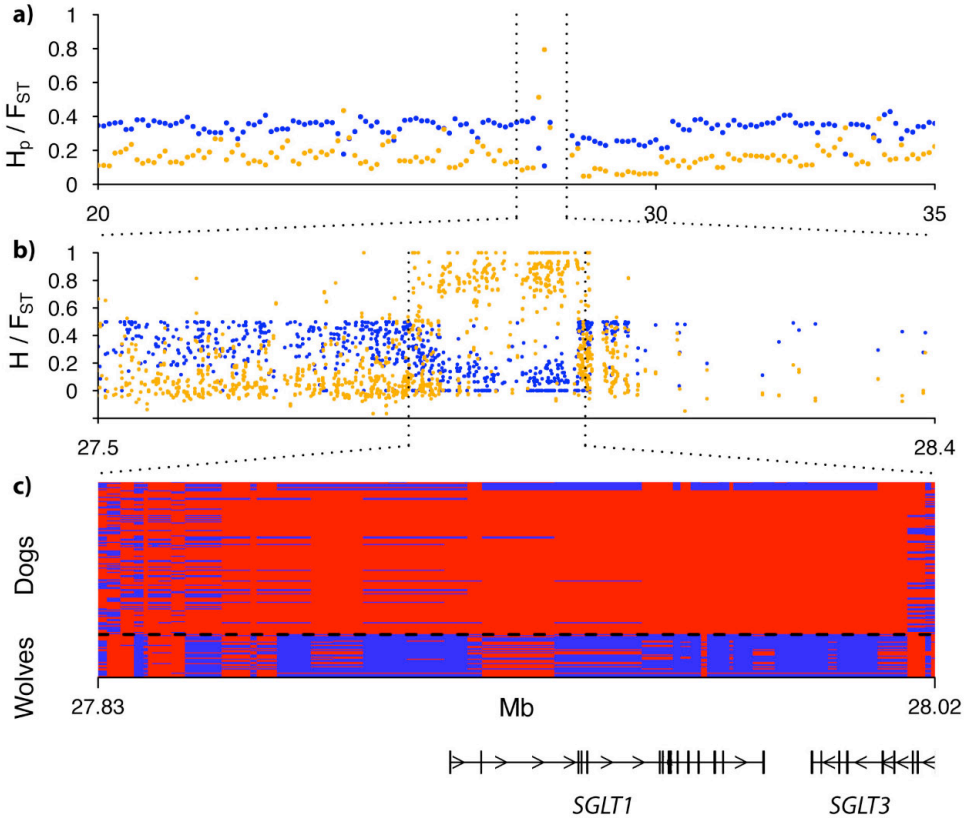

Figure S5. Selection affected main glucose transporter *SGLT1*. a) Pooled heterozygosity, H_P (blue dots), and average fixation index, F_{ST} (orange dots) plotted for 200 Kb windows across a region spanning 20-35 Mb on chromosome 26. Dashed vertical lines indicate the location of the selected region harbouring *SGLT1*. **b)** Magnification of the region affected by selection showing heterozygosity, H (blue dots) and fixation index, F_{ST} (orange dots) estimated for single SNPs. Dashed horizontal lines delineate genotyped region shown in panel c. **c)** Haplotypes inferred from genotyping of 48 SNPs in 71 dogs and 19 wolves, shows the location of a 50.5 Kb region, spanning approximately 27.96-28.01 Mb, that is nearly fixed in all dogs. Red colour represents the major dog allele, while blue is minor dog allele. Genes residing in the genotyped region are shown below panel c.

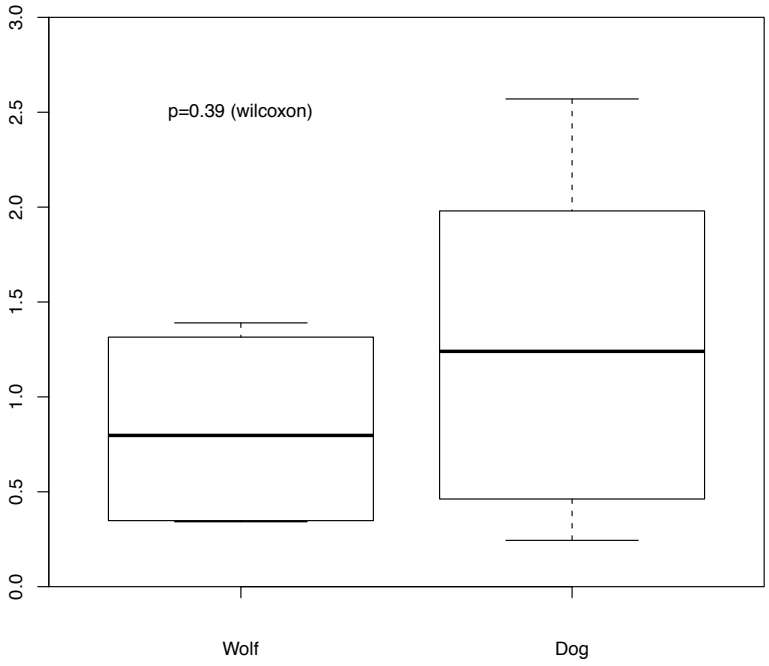

Figure S6. *SGLT1* is not differentially significantly expressed in dog and wolf pancreas (n_{WOLF}=4, n_{DOG}=9, P=0.39, Wilcoxon).

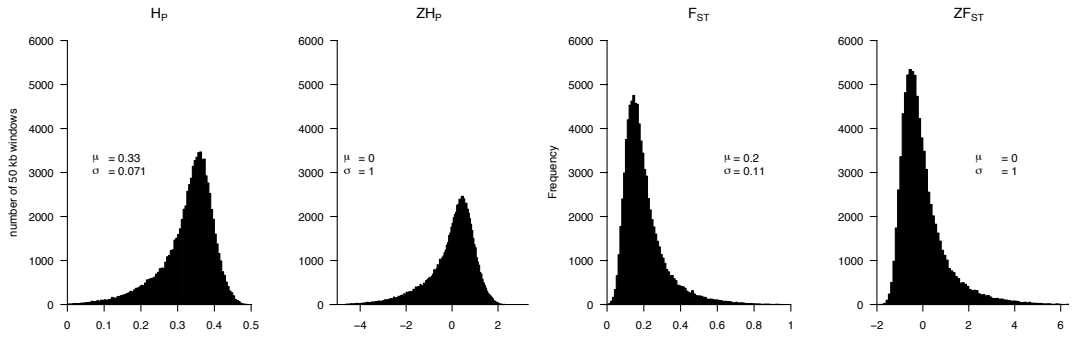

Figure S7. Distribution of pooled heterozygosity, H_P and average fixation index, F_{ST}, and corresponding Z-tranformations, $Z(H_P)$ and $Z(F_{ST})$, estimated in 50 Kb windows across all dog autosomes. The standard deviation (σ) of the autosomal average (μ) is indicated for each distribution.

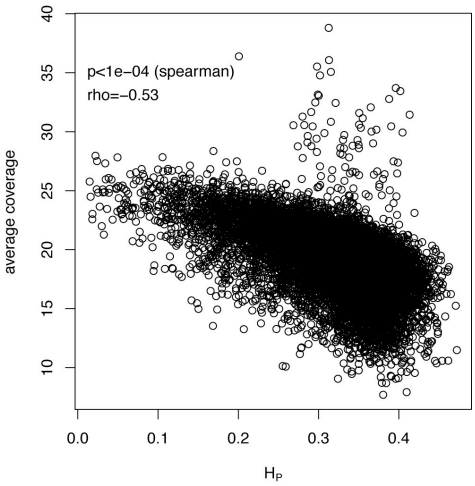

Figure S8. The average H_P of all autosomal 200 Kb windows plotted against the corresponding average read coverage. H_P is negatively correlated with sequence coverage (rho=-0.53, p<0.0001, (Spearman)).

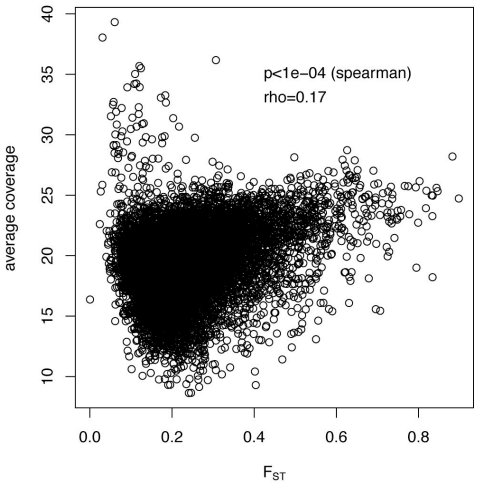

Figure S9. The average F_{ST} of all autosomal 200 Kb windows plotted against the corresponding average read coverage. F_{ST} is positively correlated with sequence coverage (rho=0.17, p<0.0001, (Spearman)).

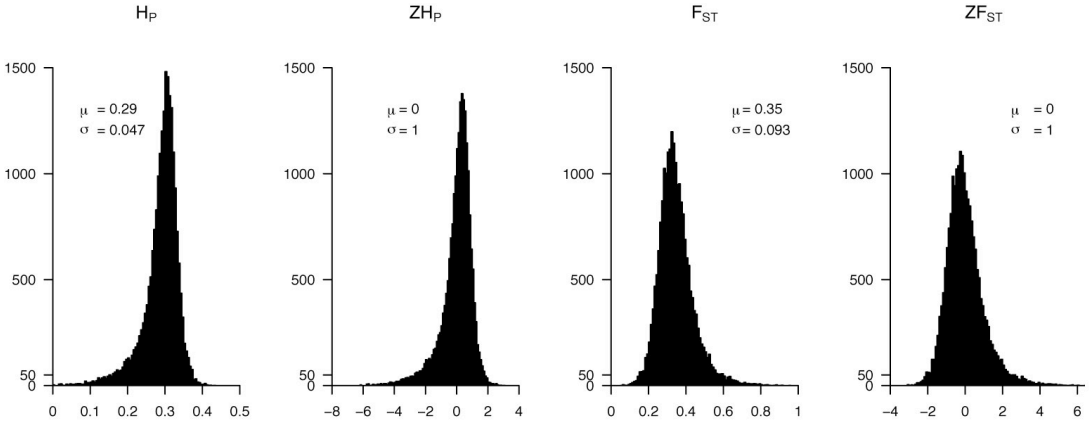

Figure S10. Distribution of average pooled heterozygosity, H_P and average fixation index, F_{ST}, and corresponding Z-tranformations, $Z(H_P)$ and $Z(F_{ST})$ estimated in 200 Kb windows across all dog autosomes. The analyses are based on the original wolf data and a sub sampled dog data set that matches the wolf data in terms of sequence coverage.

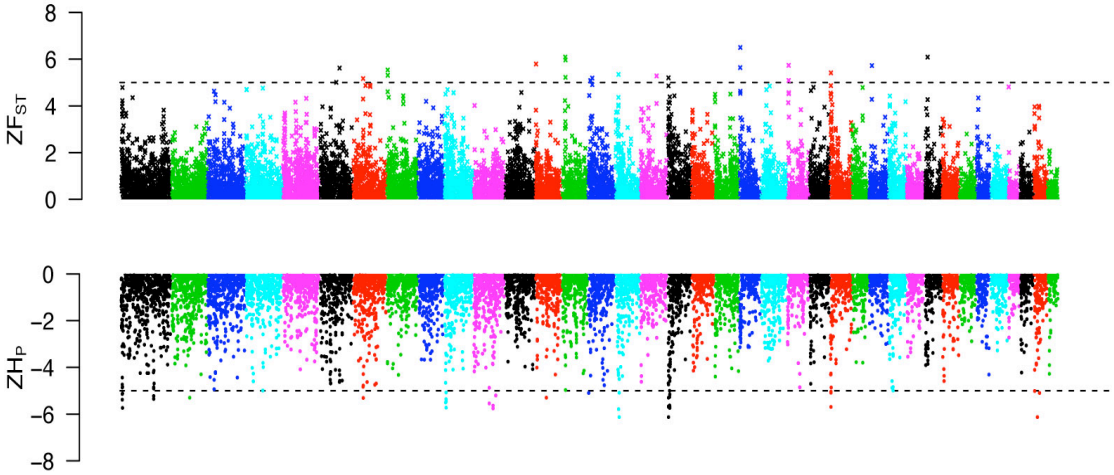

Figure S11. The positive end of the $Z(F_{ST})$ distribution plotted along dog autosomes 1-38 (chromosomes are separated by colour). A dashed horizontal line indicates the cut-off (Z>5) used for extracting outliers. The negative end of the $Z(H_P)$ distribution plotted along dog autosomes 1-38. A dashed horizontal line indicates the cut-off (Z<-5) used for extracting outliers. The analyses are based on the original wolf data and a sub sampled dog data set that matches the wolf data in terms of sequence coverage.

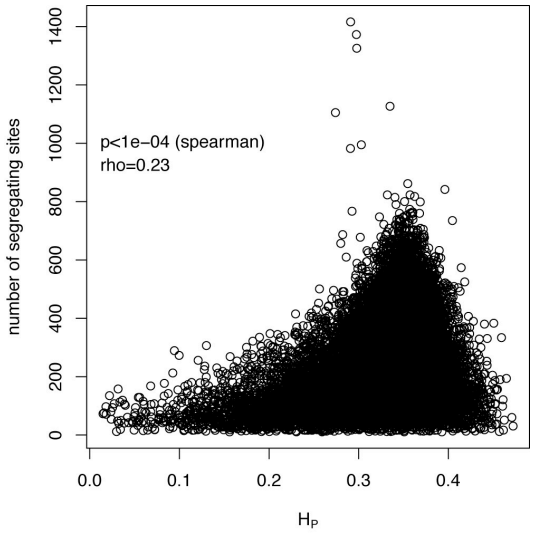

Figure S12. The average H_P of all autosomal 200 Kb windows plotted against the corresponding number of segregating sites per window. H_P is positively correlated with the number of segregating sites (rho=0.23, p<0.0001, (Spearman)).

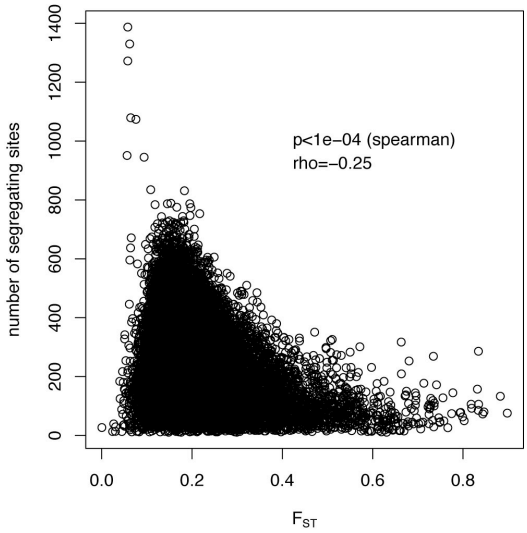

Figure S13. The average F_{ST} of all autosomal 200 Kb windows plotted against the corresponding number of segregating sites per window. F_{ST} is negatively correlated with the number of segregating sites (rho=-0.25, p<0.0001, (Spearman)).

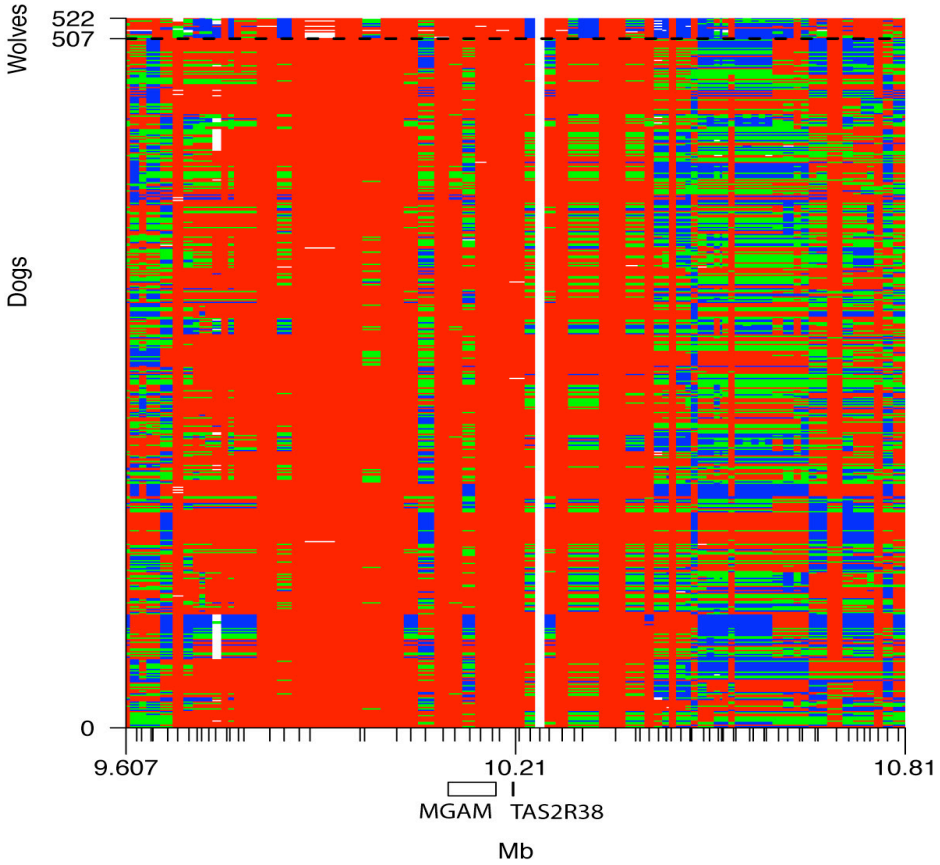

Figure S14. Genetic variation in a region spanning 9.6 - 10.8 Mb on chr. 16 that encompass the MGAM CDR and the nearby T-cell receptor cluster CDR (located at approx. 9.8 – 10 Mb). 507 dogs and 15 wolves were genotyped using the Illumina 170K Canine HD array. A dashed horizontal line separates dogs from wolves. Short tick marks represents individual SNPs. Long tick marks indicates position in Mb. Red colour: homozygous A-allele; green colour: heterozygous; blue colour: homozygous a-allele, white colour: missing genotype call. Four SNPs spanning the *MGAM* and *TAS2R38* genes are completely fixed in all dogs.

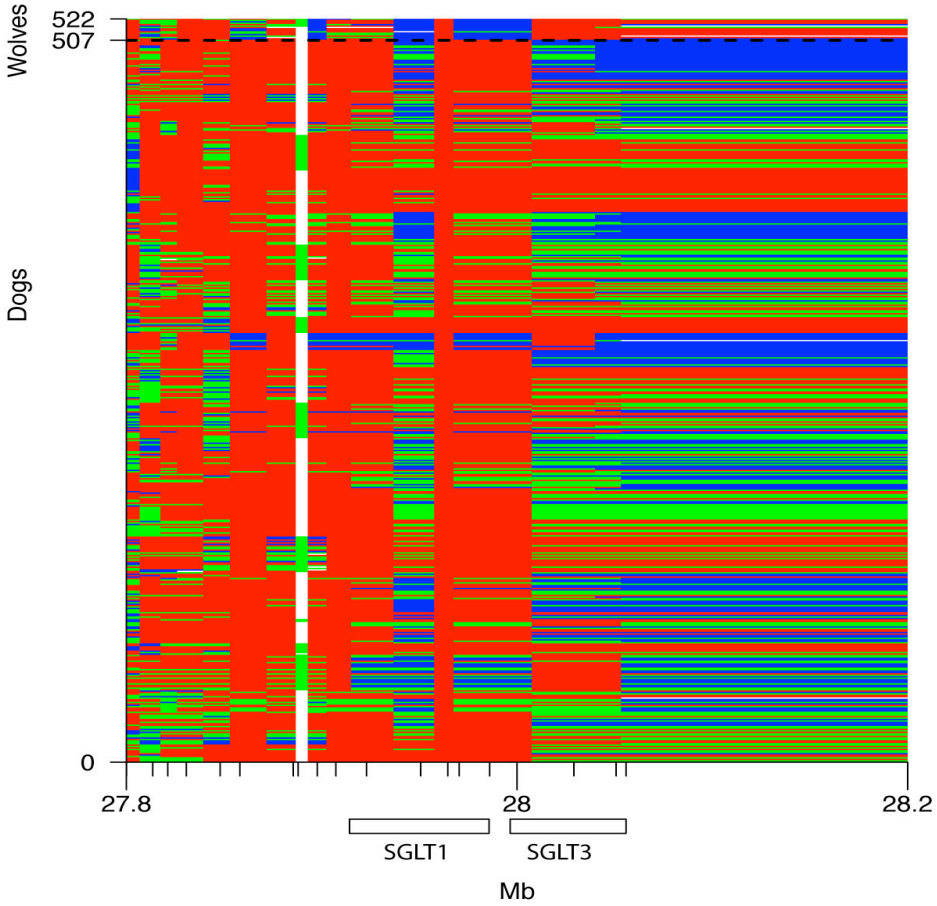

Figure S15. Genetic variation in a region spanning 27.8 – 28.2 Mb on chr. 26 that encompass the SGLT1 CDR. 507 dogs and 15 wolves were genotyped using the Illumina 170K Canine HD array. A dashed horizontal line separates dogs from wolves. Short tick marks represents individual SNPs. Long tick marks indicates position in Mb. Red colour: homozygous A-allele; green colour: heterozygous; blue colour: homozygous a-allele, white colour: missing genotype call. A single SNP in the *SGLT1* gene is completely fixed in all dogs.

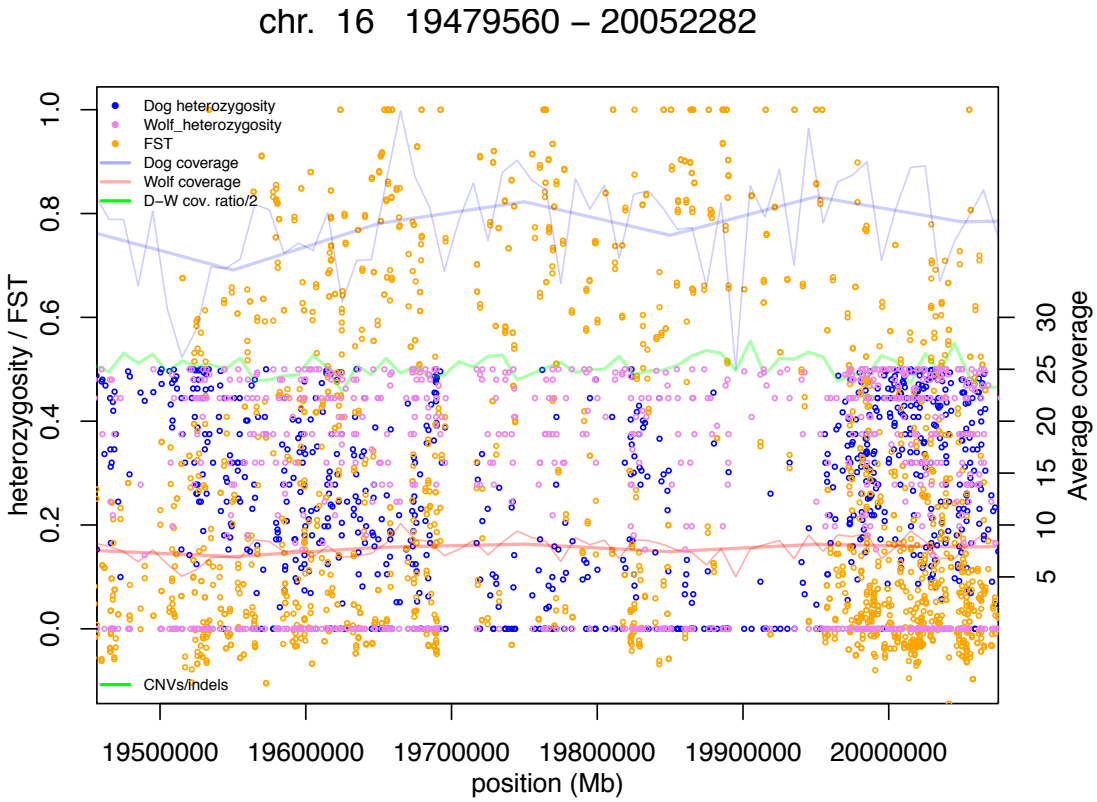

Figure S16. Example of a VonHoldt region with moderate evidence of selection in the pooled resequencing data (maximum $Z(F_{ST})$ = 3.664, minimum $Z(H_P)$ = -4.04).

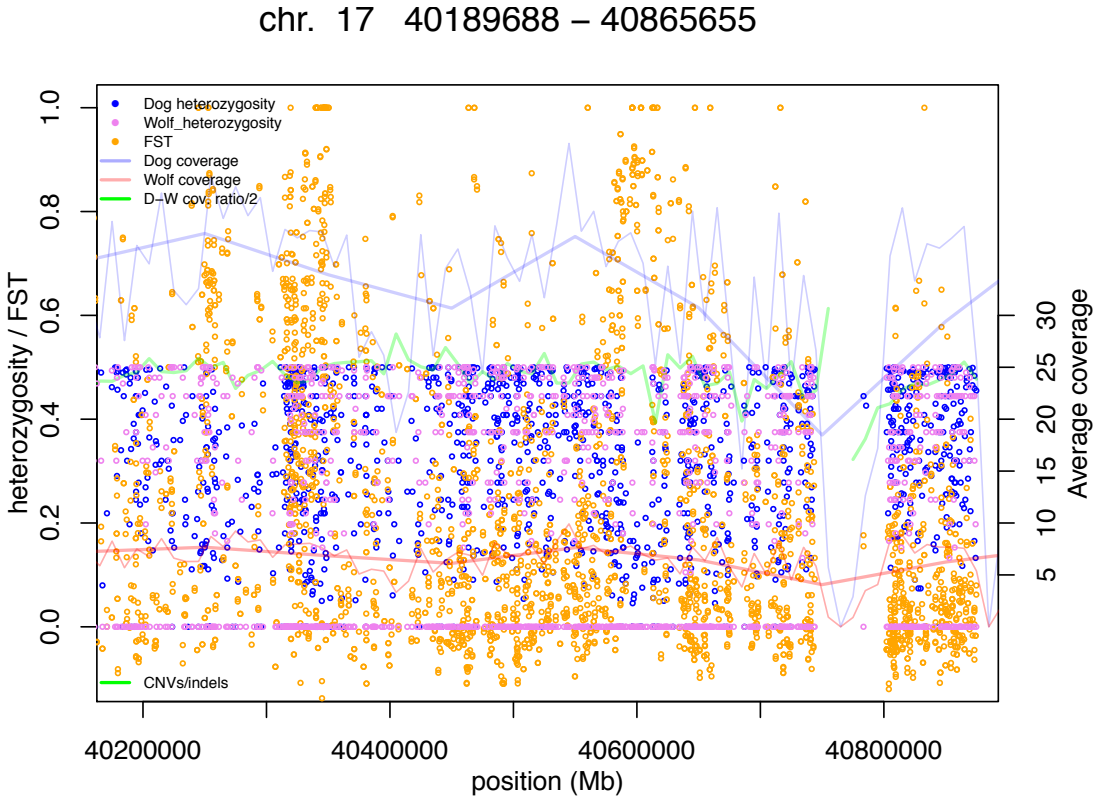

Figure S17. Example showing how markers from two narrow regions, with skewed allele frequencies, may have been mistaken to represent a single region spanning ~280 Kb in the analysis by VonHoldt et al. (maximum $Z(F_{ST})$ = 2.67 , minimum $Z(H_P)$ = -1.47).

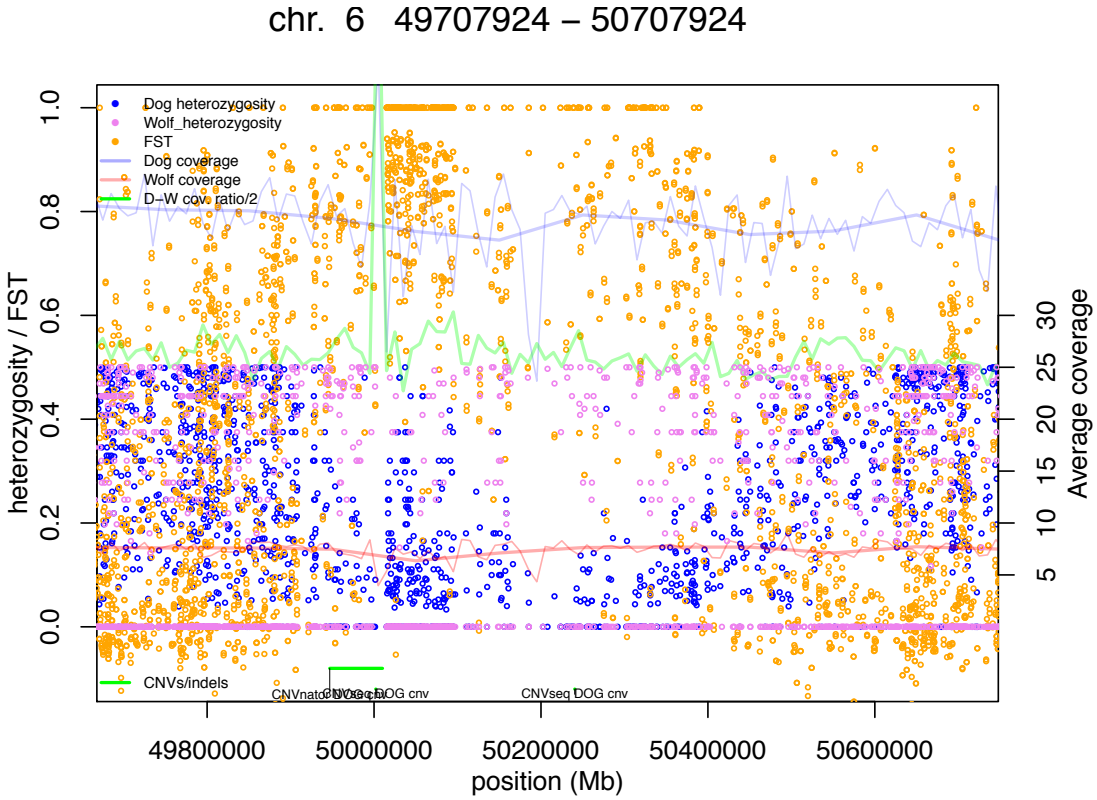

Figure S18. Example of a significant F_{ST} region ($Z(F_{ST})$>5) that did not pass the $Z(H_P)$ threshold ($Z(H_P)$<-4.60).

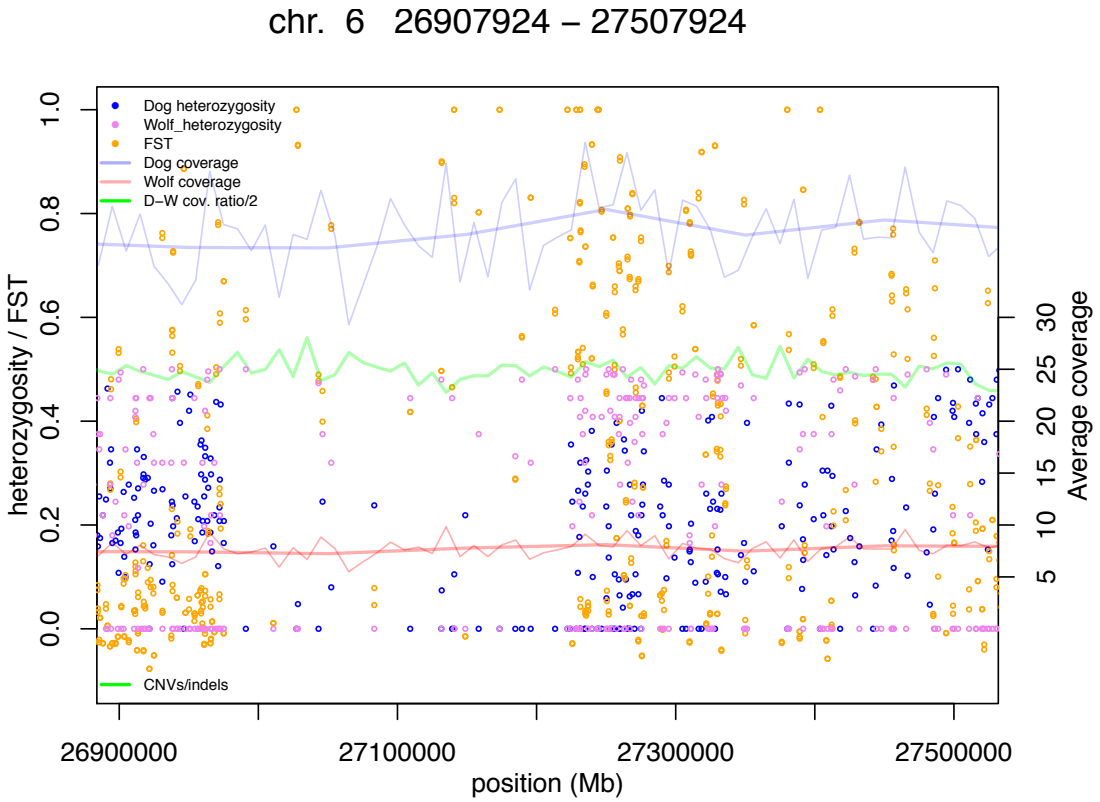

Figure S19. Example of a significant H_P region ($Z(H_P)$<-5) that did not pass the $Z(F_{ST})$ threshold ($Z(F_{ST})$=4.53). In this particular region the most extreme 200 Kb window in terms of H_P only harboured 12 segregating sites.

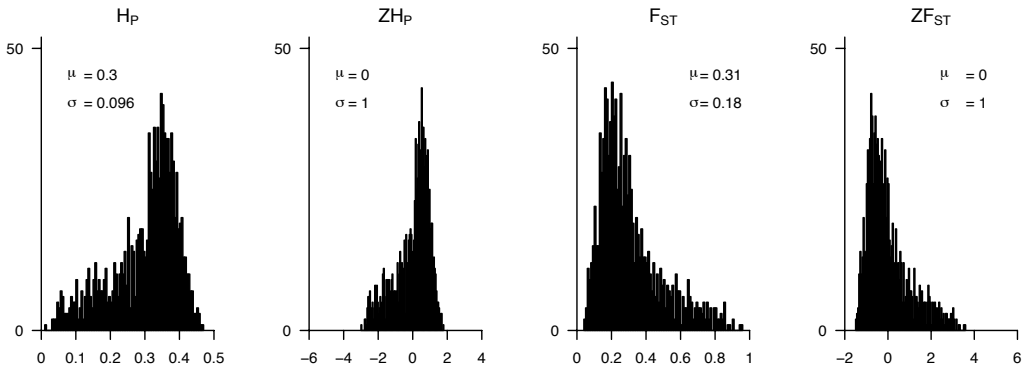

Figure S20. Distribution of pooled heterozygosity, H_P and average fixation index, F_{ST}, and corresponding Z-tranformations, $Z(H_P)$ and $Z(F_{ST})$, estimated in 200 Kb windows across chromosome X. The standard deviation (σ) of the autosomal average (μ) is indicated for each distribution.

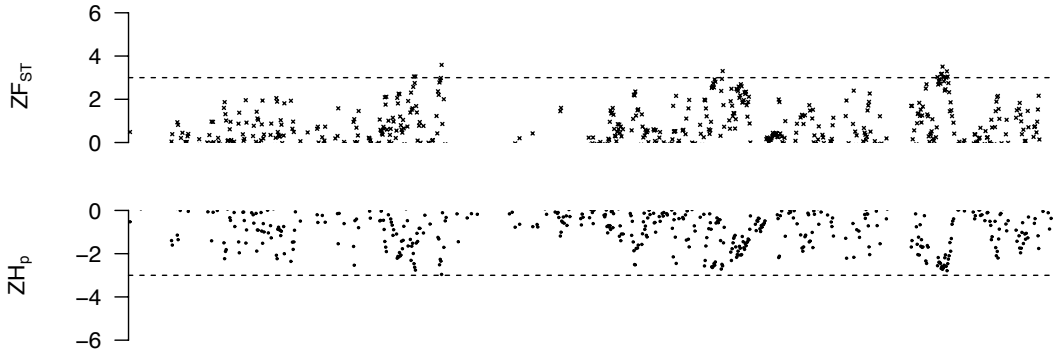

Fig S21. Z-transformed average fixation index (only positive values shown), $Z(F_{ST})$ and pooled heterozygosity (only negative values shown), $Z(H_P)$, in 200 Kb windows across chromosome X. Dashed horizontal lines show $Z(F_{ST})>3$ and $Z(H_P)<-3$, respectively.

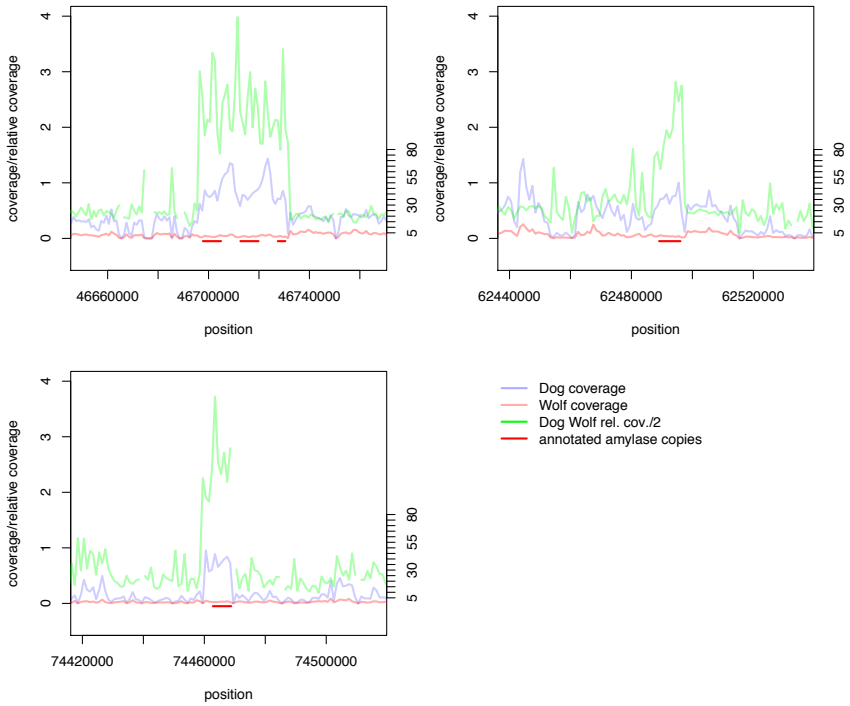

Figure S22. Sequence coverage averaged across 1 Kb windows in dog (blue line) and wolf (red line) at five annotated amylase gene copies residing on chromosome unknown. The relative dog to wolf coverage (green line) shows a 4 to 8 fold increase (half the ratio is plotted) across the amylase copies, indicative of a significant increase in copy number in dog.

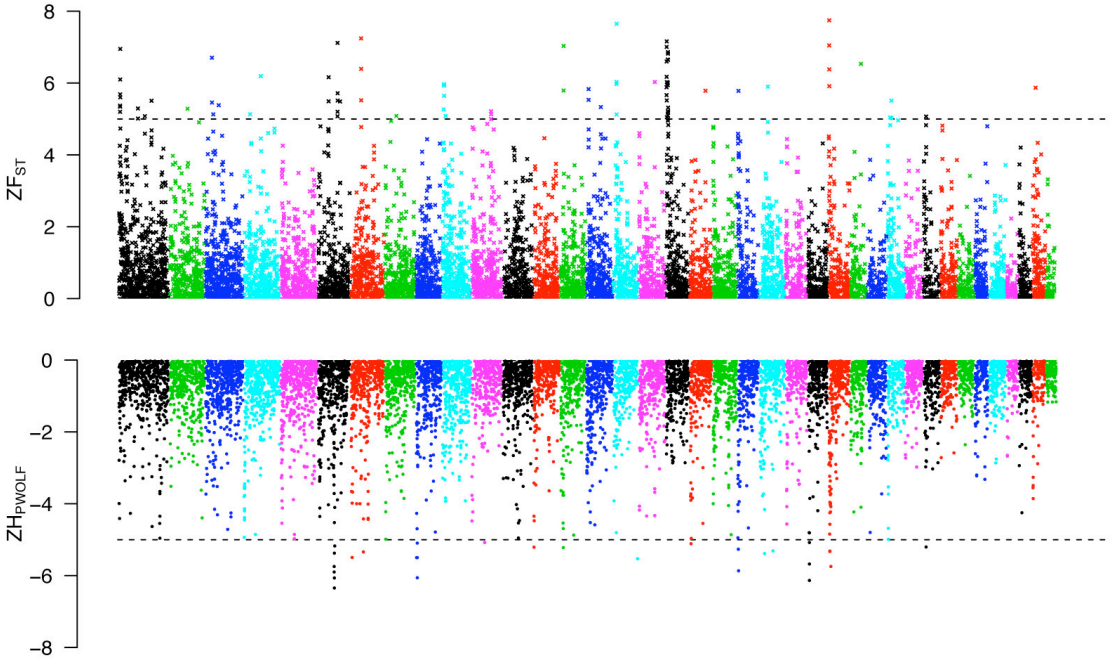

Figure S23. The positive end of the $Z(F_{ST})$ distribution plotted along autosomes 1-38 (chromosomes are separated by colour). A dashed horizontal line indicates the cut-off (Z>5) used for extracting outliers. The negative end of the $Z(H_P)$ distribution in wolf plotted along autosomes 1-38. A dashed horizontal line indicates the cut-off (Z<-5) used for extracting outliers.

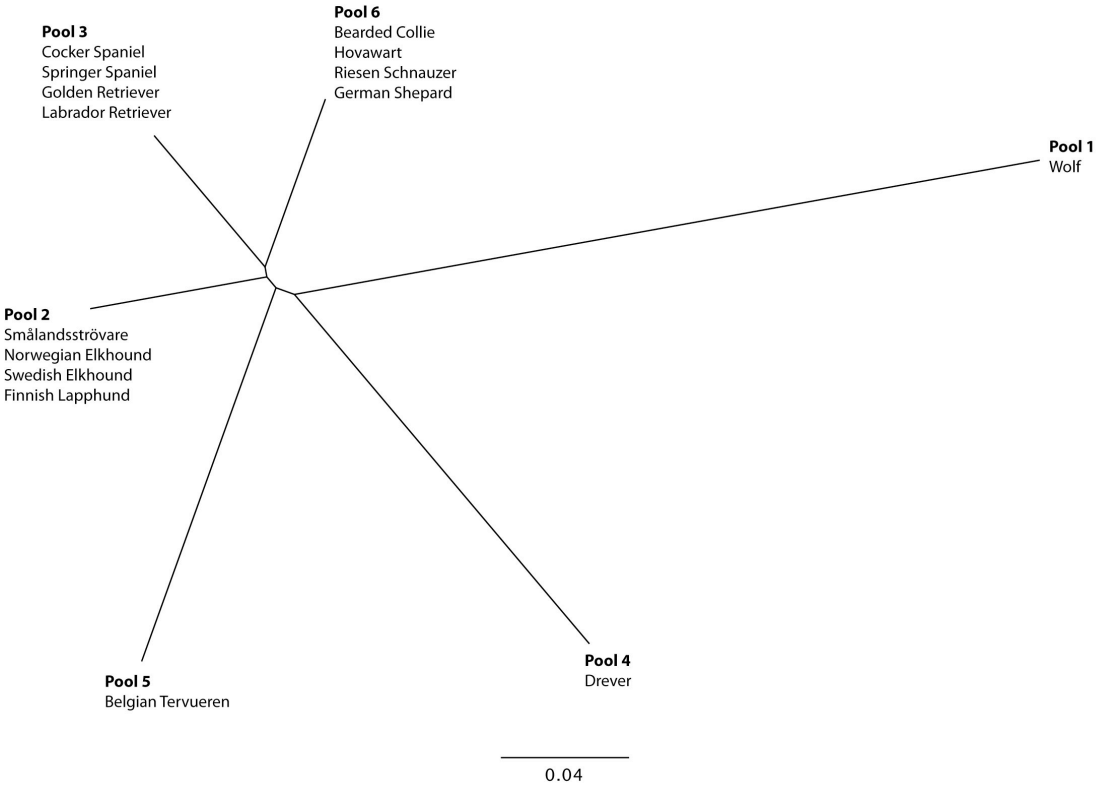

Figure S24. Phylogenetic tree depicting the relationship of the 5 dog pools (pool2-6) and the single wolf pool (pool1). Pools 2, 3 and 6 are a mixture of breeds, and these have shorter branches showing more shared variation than the single breed pools.

Supplemental Tables

Table S1. Pool information for resequencing data from dogs and wolves. A single wolf pool represents wolf diversity across Eurasia and North America, three dog pools represent four separate breeds respectively and the remaining two dog pools contain DNA from representatives of a single breed. The number of individuals from each breed (n) and the average sequence and assembly coverage per pool is indicated.

	WOLVES	**n**	**sequence cov.**	**assembly cov. (%)**
Pool 1	Sweden	1		
	Spain	3		
	Russia	3		
	Belarus	2		
	Bulgaria	1		
	USA	1		
	Canada	1		
	Total	**12**	**6.2x**	**89**
	DOGS			
Pool 2	Smålandsstövare	3		
	Norwegian Elkhound	3		
	Swedish Elkhound	3		
	Finnish Lapphund	3	5.9x	89
Pool 3	Cocker Spaniel	3		
	Springer Spaniel	3		
	Golden Retriever	3		
	Labrador Retriever	3	5.3x	87
Pool 4	Drever	12	6.8x	90
Pool 5	Belgian Tervueren	12	6.8x	90
Pool 6	Bearded Collie	3		
	Hovawart	3		
	Giant Schnauzer	3		
	German Shephard	3	6.7x	90
	Total	**60**	**29.8x**	**94**

Table S2. SNP summary statistics. Table shows the number of SNPs called in all dog and wolf data combined (SNP count); the number of SNPs covered by at least one sequencing read in wolf (Wolf coverage) and dog (Dog coverage); number of SNPs segregating only in dog (Private to dog) and wolf (Private to wolf); number of SNPs fixed in dogs and at the same time either fixed or segregating for the alternate allele in wolves (Fixed). "Fixed" refers to SNPs with allele frequency > 0.95.

Chr.	SNP count	Wolf coverage	Dog coverage	Private to dog	Private to wolf	Fixed between wolf dog
1	185673	180288	185070	86448	8341	1575
2	120703	116954	120287	56306	4653	807
3	157222	152698	156727	74600	5995	1136
4	149078	144899	148609	70776	5572	1021
5	146796	142581	146365	69744	5078	921
6	123932	120182	123535	59461	4584	961
7	128151	124714	127747	59214	4555	772
8	119933	116405	119461	55274	4252	754
9	80984	78157	80760	39005	3064	642
10	102917	99712	102583	49653	4483	933
11	113233	110204	112851	51840	4797	811
12	126438	122677	126056	59529	4799	879
13	112292	109241	111944	51293	3839	706
14	102850	100098	102496	47457	3592	645
15	99480	96653	99163	47229	4080	760
16	103571	100413	103208	47653	3798	804
17	110635	107581	110268	51414	3764	669
18	88670	85779	88404	42447	4682	1128
19	102612	99717	102284	46984	3422	566
20	82422	80021	82170	37474	2831	507
21	98244	95538	97900	45529	2883	472
22	102249	99438	101876	47366	3634	683
23	93155	90729	92854	41414	2903	469
24	77485	75048	77230	36819	2725	503
25	87717	85157	87460	42343	3428	745
26	70446	68159	70184	32953	2111	407
27	81738	79398	81454	37809	2195	359
28	69287	67126	69089	32497	2515	473
29	81532	79372	81267	37236	2306	363
30	64607	62583	64415	29444	1713	285
31	76801	74411	76514	35789	2724	539
32	77181	74947	76904	34530	2298	349
33	59705	58021	59517	27568	2128	366
34	83556	81072	83308	38564	2793	472
35	59015	57419	58807	26476	1416	223
36	55354	53823	55170	25170	1958	359
37	54015	52460	53838	24531	2106	393
38	53605	51836	53406	24578	1488	265
X	83371	80095	83182	46492	7313	3025
total	**3786655**	**3675606 (97.1%)**	**3774363 (99.7%)**	**1770909 (46.8%)**	**140818 (3.7%)**	**27747 (0.7%)**

Table S3. 99.1% of SNP validated using iPLEX technology.

Assays	Number of SNPs
Total number of SNP assays	124
Failed SNPs assay	0
Excluded SNPs due to low call rates	10
Erroneous base call in genome assembly	0
Total number of successful SNP assays	114
Erroneous base call in SOLiD data	1
Verified SNPs	113

Table S4. Ranking of 200 Kb windows in the dog genome based on a significantly reduced average pooled heterozygosity, H_P, sorted by Z-sore. CDR indicates which candidate domestication region the window is part of. Z-score refers to the value of the window after Z-transformation of the H_P distribution. Ensemble ID and gene name or gene description is shown for genes residing in these windows.

CDR	Position (Chrom:Mb)	H_P	Z-score	Ensemble ID	Gene
12	18: 3.9-4.1	0.015	-5.82		
14	25: 4.1-4.3	0.016	-5.79		
2	1: 6.0-6.2	0.018	-5.76	ENSCAFG00000000017	*ZNF236*
13	18: 6.3-6.5	0.018	-5.75		
10	16: 10.1-10.3	0.022	-5.68	ENSCAFG00000003841	*MGAM*
10	16: 10.1-10.3	0.022	-5.68	ENSCAFG00000003856	*Q2ABD2_CANFA*
10	16: 10.1-10.3	0.022	-5.68	ENSCAFG00000003864	*CLEC5A*
10	16: 10.1-10.3	0.022	-5.68	ENSCAFG00000003872	
10	16: 10.1-10.3	0.022	-5.68	ENSCAFG00000003876	*XM_539875.1*
10	16: 10.1-10.3	0.022	-5.68	ENSCAFG00000003879	
13	18: 6.5-6.7	0.024	-5.65		
7	11: 40.9-41.1	0.024	-5.65		
13	18: 6.9-7.1	0.025	-5.62		
13	18: 7.0-7.2	0.027	-5.59		
4	6: 27.1-27.3	0.03	-5.54	ENSCAFG00000017807	*CRYM*
4	6: 27.1-27.3	0.03	-5.54	ENSCAFG00000017810	*ANKS4B*
4	6: 27.1-27.3	0.03	-5.54	ENSCAFG00000017814	*ZP2_CANFA*
4	6: 27.1-27.3	0.03	-5.54	ENSCAFG00000017819	*TMEM159*
4	6: 27.1-27.3	0.03	-5.54	ENSCAFG00000017844	*DNAH3*
4	6: 27.1-27.3	0.03	-5.54	ENSCAFG00000023577	*XM_547099.2*
11	16: 11.9-12.1	0.031	-5.52	ENSCAFG00000004011	*TBXAS1*
11	16: 11.9-12.1	0.031	-5.52	ENSCAFG00000004038	*HIPK2*
12	18: 3.8-4.0	0.032	-5.51		
2	1: 6.1-6.3	0.032	-5.49		
13	18: 6.4-6.6	0.034	-5.47		
5	7: 27.9-28.1	0.034	-5.46	ENSCAFG00000025140	*RAB GTPase activating protein 1-like*
5	7: 27.9-28.1	0.034	-5.46	ENSCAFG00000025165	*RAB GTPase activating*

					protein 1-like
13	18: 7.1-7.3	0.035	-5.45		
13	18: 6.2-6.4	0.037	-5.42		
5	7: 27.8-28.0	0.039	-5.38	ENSCAFG00000025165	*RAB GTPase activating protein 1-like*
14	25: 4.0-4.2	0.039	-5.38	ENSCAFG00000005968	
6	10: 6.8-7.0	0.04	-5.36		
8	11: 50.4-50.6	0.042	-5.32		
1	1: 5.5-5.7	0.043	-5.29		
8	11: 50.3-50.5	0.048	-5.21		
8	11: 50.5-50.7	0.048	-5.21	ENSCAFG00000001734	
12	18: 3.7-3.9	0.049	-5.18		
9	11: 56.9-57.1	0.049	-5.18	ENSCAFG00000002386	*FRMPD1*
9	11: 56.9-57.1	0.049	-5.18	ENSCAFG00000002389	*RG9MTD3*
14	25: 4.2-4.4	0.05	-5.16		
6	10: 6.6-6.8	0.051	-5.15		
6	10: 6.7-6.9	0.053	-5.12		
6	10: 6.9-7.1	0.054	-5.1	ENSCAFG00000000319	*HSPD1*
12	18: 3.6-3.8	0.054	-5.09		
3	4: 44.0-44.2	0.054	-5.09	ENSCAFG00000016919	*TLX3*
3	4: 44.0-44.2	0.054	-5.09	ENSCAFG00000024671	*XM_536433.2*
12	18: 4.1-4.3	0.055	-5.08	ENSCAFG00000003354	*VWC2*
2	1: 5.9-6.1	0.055	-5.07	ENSCAFG00000000016	*MBP*
2	1: 5.9-6.1	0.055	-5.07	ENSCAFG00000000017	*ZNF236*
7	11: 40.8-41.0	0.056	-5.06		
2	1: 5.8-6.0	0.057	-5.04	ENSCAFG00000000016	*MBP*
2	1: 5.8-6.0	0.057	-5.04	ENSCAFG00000000017	*ZNF236*
5	7: 27.6-27.8	0.058	-5.02	ENSCAFG00000014399	*XM_856082.1*
13	18: 6.7-6.9	0.058	-5.02		

Table S5. Ranking of 200 Kb windows in the dog genome based on a significantly increased fixation index between dog and wolf, F_{ST}, sorted by Z-sore. CDR indicates which candidate domestication region the window is part of. Z-score refers to the value of the window after Z-transformation of the F_{ST} distribution. Ensemble ID and gene name or gene description is shown for genes residing in these windows.

CDR	Position	F_{ST}	Z-score	Ensemble ID	Gene
30	25: 4.1-4.3	0.9	7.74		
23	16: 10.1-10.3	0.89	7.65	ENSCAFG00000003841	*MGAM*
23	16: 10.1-10.3	0.89	7.65	ENSCAFG00000003856	*Q2ABD2_CANFA*
23	16: 10.1-10.3	0.89	7.65	ENSCAFG00000003864	*CLEC5A*
23	16: 10.1-10.3	0.89	7.65	ENSCAFG00000003872	
23	16: 10.1-10.3	0.89	7.65	ENSCAFG00000003876	*XM_539875.1*
23	16: 10.1-10.3	0.89	7.65	ENSCAFG00000003879	
14	7: 27.6-27.8	0.86	7.25	ENSCAFG00000014399	*XM_856082.1*
25	18: 3.9-4.1	0.85	7.17		
12	6: 50.0-50.2	0.85	7.16	ENSCAFG00000019972	*RNPC3*
20	14: 10.3-10.5	0.84	7.06	ENSCAFG00000001524	*AHCYL2*

20	14: 10.3-10.5	0.84	7.06	ENSCAFG00000001528	*XM_856148.1*
20	14: 10.3-10.5	0.84	7.06	ENSCAFG00000001531	*SMO*
30	25: 4.0-4.2	0.84	7.03	ENSCAFG00000005968	
25	18: 3.8-4.0	0.84	7		
1	1: 6.0-6.2	0.83	6.96	ENSCAFG00000000017	*ZNF236*
26	18: 6.9-7.1	0.82	6.87		
26	18: 7.0-7.2	0.82	6.81		
6	3: 18.2-18.4	0.81	6.71	ENSCAFG00000008064	*FAM172A*
26	18: 6.5-6.7	0.8	6.66		
26	18: 7.1-7.3	0.8	6.64		
25	18: 3.7-3.9	0.8	6.6		
31	26: 27.9-28.1	0.79	6.48	ENSCAFG00000013439	*SGLT1*
31	26: 27.9-28.1	0.79	6.48	ENSCAFG00000013498	*SGLT3*
31	26: 27.9-28.1	0.79	6.48	ENSCAFG00000023285	*LOC612066*
14	7: 27.7-27.9	0.78	6.41	ENSCAFG00000025165	*RAB GTPase activating protein 1-like*
30	25: 4.2-4.4	0.77	6.24		
25	18: 3.6-3.8	0.76	6.18		
10	4: 44.0-44.2	0.76	6.17	ENSCAFG00000016919	*TLX3*
10	4: 44.0-44.2	0.76	6.17	ENSCAFG00000024671	*XM_536433.2*
1	1: 5.9-6.1	0.76	6.12	ENSCAFG00000000016	*MBP*
1	1: 5.9-6.1	0.76	6.12	ENSCAFG00000000017	*ZNF236*
23	16: 9.8-10.0	0.75	6.07	ENSCAFG00000003811	
23	16: 9.8-10.0	0.75	6.07	ENSCAFG00000003812	*T cell receptor beta variable 9*
23	16: 9.8-10.0	0.75	6.07	ENSCAFG00000003814	
23	16: 9.8-10.0	0.75	6.07	ENSCAFG00000003815	
23	16: 9.8-10.0	0.75	6.07	ENSCAFG00000003817	
23	16: 9.8-10.0	0.75	6.07	ENSCAFG00000003818	*TRY1_CANFA*
23	16: 9.8-10.0	0.75	6.07	ENSCAFG00000003820	*XM_846299.1*
23	16: 9.8-10.0	0.75	6.07	ENSCAFG00000003823	*PRSS58*
23	16: 9.8-10.0	0.75	6.07	ENSCAFG00000003827	*XM_539871.2*
23	16: 9.8-10.0	0.75	6.07	ENSCAFG00000014478	*T cell receptor beta constant 2*
23	16: 9.8-10.0	0.75	6.07	ENSCAFG00000024325	
23	16: 9.8-10.0	0.75	6.07	ENSCAFG00000024754	
23	16: 9.8-10.0	0.75	6.07	ENSCAFG00000024805	
23	16: 9.8-10.0	0.75	6.07	ENSCAFG00000024808	
23	16: 9.8-10.0	0.75	6.07	ENSCAFG00000024810	*T cell receptor beta variable 2*
23	16: 9.8-10.0	0.75	6.07	ENSCAFG00000024819	*T cell receptor beta variable 19*
11	6: 28.1-28.3	0.75	6.04	ENSCAFG00000018025	*GPR139*
23	16: 9.9-10.1	0.75	6.03	ENSCAFG00000003814	
23	16: 9.9-10.1	0.75	6.03	ENSCAFG00000003815	
23	16: 9.9-10.1	0.75	6.03	ENSCAFG00000003817	
23	16: 9.9-10.1	0.75	6.03	ENSCAFG00000003818	*TRY1_CANFA*
23	16: 9.9-10.1	0.75	6.03	ENSCAFG00000003820	*XM_846299.1*
23	16: 9.9-10.1	0.75	6.03	ENSCAFG00000003823	*PRSS58*
23	16: 9.9-10.1	0.75	6.03	ENSCAFG00000003827	*XM_539871.2*
23	16: 9.9-10.1	0.75	6.03	ENSCAFG00000003841	*MGAM*
23	16: 9.9-10.1	0.75	6.03	ENSCAFG00000024325	
23	16: 9.9-10.1	0.75	6.03	ENSCAFG00000024754	

23	16: 9.9-10.1	0.75	6.03	ENSCAFG00000024805	
23	16: 9.9-10.1	0.75	6.03	ENSCAFG00000025371	
23	16: 9.9-10.1	0.75	6.03	ENSCAFG00000025387	*XM_539872.2*
25	18: 4.1-4.3	0.75	6.03	ENSCAFG00000003354	*VWC2*
26	18: 6.4-6.6	0.75	6.02		
24	17: 41.7-41.9	0.74	5.99	ENSCAFG00000007479	*RNF103-CHMP3*
24	17: 41.7-41.9	0.74	5.99	ENSCAFG00000007522	*KDM3A*
24	17: 41.7-41.9	0.74	5.99	ENSCAFG00000007549	*REEP1*
26	18: 6.2-6.4	0.74	5.99		
17	10: 6.7-6.9	0.74	5.96		
26	18: 6.3-6.5	0.74	5.94		
29	22: 22.9-23.1	0.73	5.88		
25	18: 3.4-3.6	0.73	5.87		
30	25: 4.3-4.5	0.73	5.85	ENSCAFG00000005979	*COG6*
17	10: 6.6-6.8	0.73	5.84		
20	14: 10.2-10.4	0.73	5.82	ENSCAFG00000001515	*FAM40B*
20	14: 10.2-10.4	0.73	5.82	ENSCAFG00000001524	*AHCYL2*
21	15: 8.2-8.4	0.73	5.81		
27	19: 40.7-40.9	0.72	5.77	ENSCAFG00000005020	*ACMSD*
27	19: 40.7-40.9	0.72	5.77	ENSCAFG00000005031	*CCNT2*
27	19: 40.7-40.9	0.72	5.77	ENSCAFG00000005038	*YSK4*
27	19: 40.7-40.9	0.72	5.77	ENSCAFG00000005042	*RAB3GAP1*
28	21: 4.3-4.5	0.72	5.73	ENSCAFG00000025022	
12	6: 50.2-50.4	0.72	5.73		
1	1: 5.8-6.0	0.72	5.71	ENSCAFG00000000016	*MBP*
1	1: 5.8-6.0	0.72	5.71	ENSCAFG00000000017	*ZNF236*
25	18: 3.5-3.7	0.71	5.67		
35	37: 9.9-10.1	0.71	5.67	ENSCAFG00000010826	*SF3B1*
35	37: 9.9-10.1	0.71	5.67	ENSCAFG00000010837	*COQ10B*
35	37: 9.9-10.1	0.71	5.67	ENSCAFG00000010865	
35	37: 9.9-10.1	0.71	5.67	ENSCAFG00000010890	*HSPE1*
35	37: 9.9-10.1	0.71	5.67	ENSCAFG00000010899	*MOB4*
35	37: 9.9-10.1	0.71	5.67	ENSCAFG00000010934	*RFTN2*
1	1: 6.1-6.3	0.71	5.62		
17	10: 6.8-7.0	0.71	5.61		
21	15: 8.1-8.3	0.71	5.58		
4	1: 83.0-83.2	0.7	5.55		
14	7: 27.8-28.0	0.7	5.53	ENSCAFG00000025165	*RAB GTPase activating protein 1-like*
12	6: 50.3-50.5	0.7	5.53	ENSCAFG00000019985	*COL11A1*
33	28: 12.3-12.5	0.7	5.52	ENSCAFG00000008433	*ENTPD1*
33	28: 12.3-12.5	0.7	5.52	ENSCAFG00000008439	
33	28: 12.3-12.5	0.7	5.52	ENSCAFG00000008441	
33	28: 12.3-12.5	0.7	5.52	ENSCAFG00000008444	
33	28: 12.3-12.5	0.7	5.52	ENSCAFG00000008446	
33	28: 12.3-12.5	0.7	5.52	ENSCAFG00000008454	*CCNJ*
33	28: 12.3-12.5	0.7	5.52	ENSCAFG00000025587	
25	18: 4.0-4.2	0.7	5.52		
13	6: 56.3-56.5	0.7	5.5		
6	3: 18.3-18.5	0.7	5.48	ENSCAFG00000008064	*FAM172A*

11	6: 28.0-28.2	0.7	5.46	ENSCAFG00000018015	*PDILT*
11	6: 28.0-28.2	0.7	5.46	ENSCAFG00000018021	*UROM_CANFA*
11	6: 28.0-28.2	0.7	5.46	ENSCAFG00000018023	*GP2_CANFA*
12	6: 50.1-50.3	0.69	5.39		
1	1: 5.5-5.7	0.69	5.37		
8	3: 34.9-35.1	0.69	5.36	ENSCAFG00000009720	*CYFIP1*
8	3: 34.9-35.1	0.69	5.36	ENSCAFG00000009743	*NIPA2*
22	15: 38.2-38.4	0.69	5.34	ENSCAFG00000006331	*VEZT*
22	15: 38.2-38.4	0.69	5.34	ENSCAFG00000006353	*METAP2*
26	18: 7.4-7.6	0.68	5.34		
1	1: 5.7-5.9	0.68	5.32	ENSCAFG00000000015	*GALR1*
2	1: 49.6-49.8	0.68	5.29	ENSCAFG00000000608	*ARID1B*
5	2: 46.5-46.7	0.68	5.28	ENSCAFG00000006911	
19	11: 50.3-50.5	0.68	5.23		
26	18: 7.5-7.7	0.67	5.22		
16	10: 5.7-5.9	0.67	5.21	ENSCAFG00000000311	*LRIG3*
26	18: 6.7-6.9	0.67	5.19		
1	1: 5.6-5.8	0.67	5.18	ENSCAFG00000000015	*GALR1*
19	11: 50.5-50.7	0.67	5.16	ENSCAFG00000001734	
7	3: 21.5-21.7	0.67	5.13		
15	8: 30.7-30.9	0.67	5.13		
23	16: 10.0-10.2	0.67	5.12	ENSCAFG00000003841	*MGAM*
23	16: 10.0-10.2	0.67	5.12	ENSCAFG00000003856	*Q2ABD2_CANFA*
23	16: 10.0-10.2	0.67	5.12	ENSCAFG00000025371	
23	16: 10.0-10.2	0.67	5.12	ENSCAFG00000025387	*XM_539872.2*
9	4: 17.7-17.9	0.66	5.11		
12	6: 49.9-50.1	0.66	5.1	ENSCAFG00000019972	*RNPC3*
34	30: 9.1-9.3	0.66	5.09		
26	18: 6.8-7.0	0.66	5.08		
18	10: 10.5-10.7	0.66	5.07	ENSCAFG00000000364	*WIF1*
25	18: 4.2-4.4	0.66	5.04	ENSCAFG00000003354	*VWC2*
25	18: 4.2-4.4	0.66	5.04	ENSCAFG00000003356	*ZPBP*
32	28: 9.4-9.6	0.66	5.04	ENSCAFG00000007355	*BTAF1*
32	28: 9.4-9.6	0.66	5.04	ENSCAFG00000007423	*CPEB3*
26	18: 7.3-7.5	0.66	5.03		
3	1: 66.6-66.8	0.66	5.03	ENSCAFG00000001026	*NKAIN2*
19	11: 50.4-50.6	0.66	5.03		
33	28: 12.2-12.4	0.66	5.02	ENSCAFG00000008433	*ENTPD1*
33	28: 12.2-12.4	0.66	5.02	ENSCAFG00000008439	
33	28: 12.2-12.4	0.66	5.02	ENSCAFG00000008441	
2	1: 49.7-49.9	0.65	5	ENSCAFG00000000608	*ARID1B*
2	1: 49.7-49.9	0.65	5	ENSCAFG00000000616	*TMEM242*

Table S6. Summary of 36 autosomal candidate domestication regions (CDRs). Start and end of regions with significantly reduced average pooled heterozygosity, H_P, and/or increased fixation index, F_{ST}, shows the position of 36 CDRs. Individual CDRs are separated by thick horizontal lines. In two cases the F_{ST} analysis identifies a single coherent region, while the H_P analysis indicates two separate regions. For these regions a dashed horizontal line separates the individual H_P regions. Ensemble ids, gene descriptions and gene names of genes residing in CDRs are shown. The * indicates genes that reside in regions 100 Kb up- or downstream of the CDR.

CDR	chr	H_P region	start	end	F_{ST} region	start	end	ensemble gene id	gene description	gene name
1	1	1	5517430	5717430	1	5517430	6317430	ENSCAFG00000000015	galanin receptor 1	*GALR1*
	1	2	5817430	6317430		5517430	6317430	ENSCAFG00000000016	myelin basic protein	*MBP*
								ENSCAFG00000000017	zinc finger protein 236	*ZNF236*
							*	ENSCAFG00000000019	zinc finger protein 516	*ZNF516*
2	1				2	49617430	49917430	ENSCAFG00000000608	AT rich interactive domain 1B (SWI1-like)	*ARID1B*
								ENSCAFG00000000616	UPF0463 transmembrane protein C6orf35	*TMEM242*
							*	ENSCAFG00000023184		
3	1				3	66617430	66817430	ENSCAFG00000001026	Na+/K+ transporting ATPase interacting 2	*NKAIN2*
4	1				4	83017430	83217430			
5	2				5	46500196	46700196	ENSCAFG00000006911		
6	3				6	18207515	18507515	ENSCAFG00000008064	microRNA 2277	*FAM172A*
7	3				7	21507515	21707515			
							*	ENSCAFG00000008286	centrin, EF-hand protein, 3	*CETN3*
8	3				8	34907515	35107515	ENSCAFG00000009720	cytoplasmic FMR1 interacting protein 1	*CYFIP1*
								ENSCAFG00000009743	non imprinted in Prader-Willi/Angelman syndrome 2	*NIPA2*
							*	ENSCAFG00000009579	tubulin, gamma complex associated protein 5	*TUBGCP5*
							*	ENSCAFG00000009863		*HERC2*
9	4				9	17700233	17900233			
							*	ENSCAFG00000013045		*XM_546121.2*
10	4	3	44000233	44200233	10	44000233	44200233	ENSCAFG00000016919	T-cell leukemia homeobox 3	*TLX3*
								ENSCAFG00000024671	RAN binding protein 17	*XM_536433.2*
							*	ENSCAFG00000016903	fibroblast growth factor 18	*FGF18*
							*	ENSCAFG00000016912	similar to nucleophosmin 1	*XM_861675.1*
11	6	4	27107924	27307924				ENSCAFG00000017807	crystallin, mu	*CRYM*
								ENSCAFG00000017810	ankyrin repeat and sterile alpha motif domain containing 4B	*ANKS4B*
								ENSCAFG00000017814	Zona pellucida sperm-binding protein 2 Precursor	*ZP2_CANFA*
								ENSCAFG00000017819	transmembrane protein 159	*TMEM159*
								ENSCAFG00000017844	dynein, axonemal, heavy chain 3	*DNAH3*
								ENSCAFG00000023577	similar to ATP-binding cassette transporter sub-family A member 15	*XM_547099.2*
							*	ENSCAFG00000024044		
12	6				11	28007924	28307924	ENSCAFG00000018015	protein disulfide isomerase-like, testis expressed	*PDILT*
								ENSCAFG00000018021	Uromodulin Precursor	*UROM_CANFA*
								ENSCAFG00000018023	Pancreatic secretory granule membrane major glycoprotein	*GP2_CANFA*

										GP2 Precursor
									ENSCAFG00000018025	G protein-coupled receptor 139
								*	ENSCAFG00000023658	acyl-CoA synthetase medium-chain family member 5
								*	ENSCAFG00000024109	acyl-CoA synthetase medium-chain family member 2A
13	6				12	49907924	50507924	ENSCAFG00000019972		RNA-binding region (RNP1, RRM) containing 3
								ENSCAFG00000019985		collagen, type XI, alpha 1
14	6				13	56307924	56507924			
15	7	5	27600316	28100316	14	27600316	28000316	ENSCAFG00000014399		RAB GTPase activating protein 1-like
								ENSCAFG00000025140		RAB GTPase activating protein 1-like
								ENSCAFG00000025165		RAB GTPase activating protein 1-like
16	8				15	30702583	30902583			
								*	ENSCAFG00000014633	FERM domain containing 6
17	10				16	5701010	5901010	ENSCAFG00000000311		leucine-rich repeats and immunoglobulin-like domains 3
18	10	6	6601010	7101010	17	6601010	7001010	ENSCAFG00000000319		similar to heat shock protein 1 (chaperonin)
19	11	7	40800086	41100086						
								*	ENSCAFG00000001567	SH3-domain GRB2-like 2
20	11	8	50300086	50700086	19	50300086	50700086	ENSCAFG00000001734		
21	11	9	56900086	57100086				ENSCAFG00000002386		FERM and PDZ domain containing 1
								ENSCAFG00000002389		RNA (guanine-9-) methyltransferase domain containing 3
								*	ENSCAFG00000002377	polymerase (RNA) I polypeptide E, 53kDa
								*	ENSCAFG00000002382	F-box protein 10
22	14				20	10200337	10500337	ENSCAFG00000001515		family with sequence similarity 40, member B
								ENSCAFG00000001524		adenosylhomocysteinase-like 2
								ENSCAFG00000001528		ribosomal protein L12
								ENSCAFG00000001531		smoothened homolog (Drosophila)
								*	ENSCAFG00000001507	nuclear respiratory factor 1
23	15				21	8103479	8403479			
								*	ENSCAFG00000003333	glutamate receptor, ionotropic, kainate 3
24	15				22	38203479	38403479	ENSCAFG00000006331		vezatin, adherens junctions transmembrane protein
								ENSCAFG00000006353		methionyl aminopeptidase 2
								*	ENSCAFG00000006273	FYVE, RhoGEF and PH domain containing 6
25	16	10	10107391	10307391	23	9807391	10307391	ENSCAFG00000003841		maltase-glucoamylase (alpha-glucosidase)
								ENSCAFG00000003856		Taste receptor type 2
								ENSCAFG00000003864		C-type lectin domain family 5, member A
								ENSCAFG00000003872		
								ENSCAFG00000003876		cOR9A7 olfactory receptor family 9 subfamily A-like
								ENSCAFG00000003879		
					23	9807391	10307391	ENSCAFG00000003811		
								ENSCAFG00000003812		T cell receptor beta variable 9
								ENSCAFG00000003814		
								ENSCAFG00000003815		
								ENSCAFG00000003817		

									Cationic trypsin Precursor (EC 3.4.21.4)	*TRY1_CANFA*
								ENSCAFG00000003820	similar to trypsin X5	*XM_846299.1*
								ENSCAFG00000003823	similar to trypsin X3	*PRSS58*
								ENSCAFG00000003827	similar to monooxygenase, DBH-like 2	*XM_539871.2*
								ENSCAFG00000014478	T cell receptor beta constant 2	
								ENSCAFG00000024325		
								ENSCAFG00000024754		*XM_539869.2*
								ENSCAFG00000024805		
								ENSCAFG00000024808		
								ENSCAFG00000024810	T cell receptor beta variable 2	
								ENSCAFG00000024819	T cell receptor beta variable 19	
								ENSCAFG00000025371		
								ENSCAFG00000025387	similar to Maltase-glucoamylase, intestinal	*XM_539872.2*
							*	ENSCAFG00000024822	T cell receptor beta variable 28	*TRBV25OR9-2*
							*	ENSCAFG00000014481	Anionic trypsin Precursor	*TRY2_CANFA*
							*	ENSCAFG00000014471	T cell receptor beta variable 29-1	
26	16	11	11907391	12107391				ENSCAFG00000004011	thromboxane A synthase 1 (platelet)	*TBXAS1*
								ENSCAFG00000004038	homeodomain interacting protein kinase 2	*HIPK2*
							*	ENSCAFG00000003997	poly (ADP-ribose) polymerase family, member 12	*PARP12*
27	17				24	41722203	41922203	ENSCAFG00000007479	vacuolar protein sorting 24 homolog (S. cerevisiae)	*RNF103-CHMP3*
								ENSCAFG00000007522	lysine (K)-specific demethylase 3A	*KDM3A*
								ENSCAFG00000007549	receptor accessory protein 1	*REEP1*
							*	ENSCAFG00000007472	ring finger protein 103	*RNF103*
28	18	12	3604681	4304681	25	3404681	4404681	ENSCAFG00000003354	von Willebrand factor C domain containing 2	*VWC2*
								ENSCAFG00000003356	zona pellucida binding protein	*ZPBP*
29	18	13	6204681	7304681	26	6204681	7704681			
30	19				27	40704062	40904062	ENSCAFG00000005020	aminocarboxymuconate semialdehyde decarboxylase	*ACMSD*
								ENSCAFG00000005031	cyclin T2	*CCNT2*
								ENSCAFG00000005038	YSK4 Sps1/Ste20-related kinase homolog (S. cerevisiae)	*YSK4*
								ENSCAFG00000005042	RAB3 GTPase activating protein subunit 1 (catalytic)	*RAB3GAP1*
							*	ENSCAFG00000003662	TBC1 domain family, member 9	
31	22				29	22941181	23141181			
32	25	14	4000488	4400488	30	4000488	4500488	ENSCAFG00000005968		
								ENSCAFG00000005979	component of oligomeric golgi complex 6	*COG6*
33	26				31	27900108	28100108	ENSCAFG00000013439	solute carrier family 5 (sodium/glucose cotransporter), member 1	*NP_001007142.1*
	26							ENSCAFG00000013498	solute carrier family 5 (low affinity glucose cotransporter), member 4	*SLC5A4*
	26							ENSCAFG00000023285	similar to Ig lambda chain V region 4A precursor	*LOC612066*
							*	ENSCAFG00000013330	tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, eta polypeptide	*YWHAH*
34	28				32	9400594	9600594	ENSCAFG00000007355	BTAF1 RNA polymerase II, B-TFIID transcription	*BTAF1*

						factor-associated, 170kDa	
					ENSCAFG00000007423	cytoplasmic polyadenylation element binding protein 3	*CPEB3*
				*	ENSCAFG00000007302	tankyrase, TRF1-interacting ankyrin-related ADP-ribose polymerase 2	*TNKS2*
				*	ENSCAFG00000007314	fibroblast growth factor binding protein 3	*FGFBP3*
35	28	33	12200594	12500594	ENSCAFG00000008433	ectonucleoside triphosphate diphosphohydrolase 1	*ENTPD1*
					ENSCAFG00000008439		
					ENSCAFG00000008441		
					ENSCAFG00000008444		
					ENSCAFG00000008446		
					ENSCAFG00000008454	cyclin J	*CCNJ*
					ENSCAFG00000025587		
				*	ENSCAFG00000008339	aldehyde dehydrogenase 18 family, member A1	*ALDH18A1*
				*	ENSCAFG00000008406	tectonic family member 3	*TCTN3*
36	37	35	9915022	10115022	ENSCAFG00000010826	splicing factor 3b, subunit 1, 155kDa	*SF3B1*
					ENSCAFG00000010837	coenzyme Q10 homolog B	*COQ10B*
					ENSCAFG00000010865		
					ENSCAFG00000010890	heat shock 10kDa protein 1 (chaperonin 10)	*HSPE1*
					ENSCAFG00000010899	MOB1, Mps One Binder kinase activator-like 3	*MOB4*
					ENSCAFG00000010934	raftlin family member 2	*RFTN2*

Table S7. Significantly overrepresented Gene Ontology terms (GO-terms) among genes residing in CDRs.

Ensembl id	Genes	Group count	Total count	P_{FDR}-value	GO term
ENSCAFG00000013330	*YWHAH*				
ENSCAFG00000003354	*VWC2*				
ENSCAFG00000016919	*TLX3*	3	26	0.005	regulation of neuron differentiation
ENSCAFG00000013439	*SLC5A1*				
ENSCAFG00000008433	*ENTPD1*				
ENSCAFG00000014481	*PRSS3*				
ENSCAFG00000010826	*SF3B1*				
ENSCAFG00000016903	*FGF18*				
ENSCAFG00000000015	*GALR1*				
ENSCAFG00000014481	*PRSS1*				
ENSCAFG00000000016	*MBP*				
ENSCAFG00000009720	*CYFIP1*				
ENSCAFG00000013330	*YWHAH*				
ENSCAFG00000003354	*VWC2*				
ENSCAFG00000016919	*TLX3*				
ENSCAFG00000019985	*COL11A1*				
ENSCAFG00000007472	*RNF103*				
ENSCAFG00000003872	*OR9A4*				

ENSCAFG00000001531	*SMO*				
ENSCAFG00000004011	*TBXAS1*				
ENSCAFG00000003856	*TAS2R38*				
ENSCAFG00000005968	*FABP5*				
ENSCAFG00000001567	*SH3GL2*				
ENSCAFG00000017807	*CRYM*	21	3822	0.005	multicellular organismal process
ENSCAFG00000014481	*PRSS1*				
ENSCAFG00000013439	*SLC5A1*				
ENSCAFG00000014481	*PRSS3*				
ENSCAFG00000000015	*GALR1*	4	95	0.008	digestion
ENSCAFG00000009720	*CYFIP1*				
ENSCAFG00000013330	*YWHAH*				
ENSCAFG00000003354	*VWC2*				
ENSCAFG00000016919	*TLX3*				
ENSCAFG00000001531	*SMO*	5	210	0.010	neuron differentiation
ENSCAFG00000003662	*TBC1D9*				
ENSCAFG00000010890	*HSPE1*				
ENSCAFG00000001531	*SMO*				
ENSCAFG00000000015	*GALR1*				
ENSCAFG00000025165	*RABGAP1L*				
ENSCAFG00000005042	*RAB3GAP1*				
ENSCAFG00000006273	*FGD6*				
ENSCAFG00000005031	*CCNT2*	8	671	0.011	regulation of a molecular function
ENSCAFG00000007472	*RNF103*				
ENSCAFG00000000016	*MBP*				
ENSCAFG00000016919	*TLX3*				
ENSCAFG00000001567	*SH3GL2*				
ENSCAFG00000001531	*SMO*	5	235	0.013	central nervous system development
ENSCAFG00000009720	*CYFIP1*				
ENSCAFG00000013330	*YWHAH*				
ENSCAFG00000003354	*VWC2*				
ENSCAFG00000016919	*TLX3*				
ENSCAFG00000006273	*FGD6*	5	236	0.013	regulation of developmental process
ENSCAFG00000009720	*CYFIP1*				
ENSCAFG00000013330	*YWHAH*				
ENSCAFG00000003354	*VWC2*				
ENSCAFG00000016919	*TLX3*				
ENSCAFG00000001531	*SMO*	5	242	0.013	generation of neurons
ENSCAFG00000000016	*MBP*				
ENSCAFG00000009720	*CYFIP1*				
ENSCAFG00000007472	*RNF103*				
ENSCAFG00000003354	*VWC2*				
ENSCAFG00000013330	*YWHAH*				
ENSCAFG00000016919	*TLX3*				
ENSCAFG00000001567	*SH3GL2*				
ENSCAFG00000001531	*SMO*	8	716	0.013	nervous system development
ENSCAFG00000003356	*ZPBP*				
ENSCAFG00000017814	*ZP2*	2	12	0.015	binding of sperm to zona pellucida
ENSCAFG00000003356	*ZPBP*				

ENSCAFG00000017814	*ZP2*	2	12	0.015	sperm-egg recognition
ENSCAFG00000009720	*CYFIP1*				
ENSCAFG00000013330	*YWHAH*				
ENSCAFG00000003354	*VWC2*				
ENSCAFG00000016919	*TLX3*				
ENSCAFG00000001531	*SMO*	5	262	0.015	neurogenesis
ENSCAFG00000003356	*ZPBP*				
ENSCAFG00000017814	*ZP2*	2	14	0.019	cell-cell recognition
ENSCAFG00000025165	*RABGAP1L*				
ENSCAFG00000003662	*TBC1D9*				
ENSCAFG00000010890	*HSPE1*				
ENSCAFG00000005042	*RAB3GAP1*				
ENSCAFG00000006273	*FGD6*				
ENSCAFG00000005031	*CCNT2*				
ENSCAFG00000000015	*GALR1*	7	605	0.020	regulation of catalytic activity
ENSCAFG00000025165	*RABGAP1L*				
ENSCAFG00000003662	*TBC1D9*				
ENSCAFG00000010890	*HSPE1*				
ENSCAFG00000005042	*RAB3GAP1*				
ENSCAFG00000006273	*FGD6*	5	307	0.026	regulation of hydrolase activity
ENSCAFG00000024109	*ACSM2A*				
ENSCAFG00000023658	*ACSM5*				
ENSCAFG00000024109	*ACSM2B*				
ENSCAFG00000004011	*TBXAS1*	4	191	0.031	fatty acid metabolic process
ENSCAFG00000007472	*RNF103*				
ENSCAFG00000001531	*SMO*				
ENSCAFG00000016903	*FGF18*				
ENSCAFG00000005968	*FABP5*				
ENSCAFG00000000016	*MBP*				
ENSCAFG00000009720	*CYFIP1*				
ENSCAFG00000013330	*YWHAH*				
ENSCAFG00000003354	*VWC2*				
ENSCAFG00000016919	*TLX3*				
ENSCAFG00000001567	*SH3GL2*				
ENSCAFG00000019985	*COL11A1*	11	1605	0.034	system development
ENSCAFG00000025165	*RABGAP1L*				
ENSCAFG00000003662	*TBC1D9*				
ENSCAFG00000005042	*RAB3GAP1*				
ENSCAFG00000006273	*FGD6*	4	211	0.039	regulation of GTPase activity
ENSCAFG00000007472	*RNF103*				
ENSCAFG00000001531	*SMO*				
ENSCAFG00000016903	*FGF18*				
ENSCAFG00000005968	*FABP5*				
ENSCAFG00000000016	*MBP*				
ENSCAFG00000009720	*CYFIP1*				
ENSCAFG00000003354	*VWC2*				
ENSCAFG00000013330	*YWHAH*				
ENSCAFG00000006273	*FGD6*				
ENSCAFG00000016919	*TLX3*				
ENSCAFG00000001567	*SH3GL2*				
ENSCAFG00000019985	*COL11A1*	12	2005	0.039	anatomical structure development
ENSCAFG00000016903	*FGF18*	1	1	0.039	intramembranous

					ossification
ENSCAFG00000005020	*ACMSD*	1	1	0.039	quinolinate metabolic process
ENSCAFG00000003841	*MGAM*	1	1	0.039	starch metabolic process
ENSCAFG00000003841	*MGAM*	1	1	0.039	starch catabolic process
ENSCAFG00000013330	*YWHAH*	1	1	0.039	glucocorticoid catabolic process
ENSCAFG00000010890	*HSPE1*				
ENSCAFG00000004038	*HIPK2*				
ENSCAFG00000001531	*SMO*				
ENSCAFG00000016903	*FGF18*				
ENSCAFG00000009720	*CYFIP1*				
ENSCAFG00000008406	*TCTN3*				
ENSCAFG00000013330	*YWHAH*				
ENSCAFG00000003354	*VWC2*				
ENSCAFG00000016919	*TLX3*	9	1242	0.039	cell development

Table S8. Nervous system development genes classified by GO-analysis.

Position (Chrom:Mb)	Gene name	Function
1: 5.8-6.3	MBP	Insulates axons
3: 34.9-35.1	CYFIP1	Regulates synaptic plasticity and brain development
4: 44-44.2	TLX3	Determine excitatory over inhibitory cell fates in dorsal spinal cord
11: 40.8-41.1	SH3GL2	Regulates Neurotransmitter Release and Short-Term Plasticity
14: 10.2-10.5	SMO	Mediates signals from hedgehog protein
17: 41.7-41.9	RNF103	Nervous system development
18: 3.6-4.3	VWC2	Promotes neurogenesis
26: 27.9-28.1	YWHAH	Nervous system development

Table S9. Additional CDR genes with a function in the central nervous system.

Position (Chrom:Mb)	Gene name	Function
1: 5,5-5,7	GALR1	Modulation of action potentials
1: 49,6-49,9	ARID1B	Neural development
1: 66,6-66,8	NKAIN2	Neural development
6: 27,1-27,3	CRYM	Binds thyroid hormone for possible regulatory or developmental roles.
6: 28-28,3	GPR139	Specifically expressed in brain
8: 30,7-30,9	FRMD6	Regulating cell contact inhibition, organ size control etc.
15: 8,1-8,4	GRIK3	Excitatory neurotransmitter receptor
15: 38,2-38,4	VEZT	Regulates the dendritic formation of hippocampal neurons
16: 11,9-12,1	HIPK2	Regulates Postnatal Development of Enteric Dopaminergic Neurons and Glia
19: 40,7-40,9	ACMSD	Synaptic plasticity and neurodegeneration
28: 12,2-12,5	TCTN3	Homologous to a brain photoreceptor mediating photoperiodic response in birds

Table S10. Genes in CDRs with function related to starch digestion, glucose uptake and storage.

Position (Chrom:Mb)	Gene	Function
6: 49.9-50.5	AMY2B	Starch digestion
16: 10.1-10.3	MGAM	Starch digestion
26: 27.9-28.1	SGLT1	Glucose uptake
6: 28-28.3	GP2	Pancreatic digestive enzyme storage
26: 27.9-28.1	SGLT3	Sugar sensor
16: 10.1-10.3	TAS2R38	Perception of bitterness
6: 28-28.3	ACSM5	Fatty acid metabolism
6: 28-28.3	ACSM2B	Fatty acid metabolism
15: 38.2-38.4	METAP2	Fat metabolism and appetite
25: 4-4.4	FABP5	Fatty acid metabolism

Table S11. Amylase copy numbers in 136 dogs and 35 wolves quantified using real time PCR. Mean estimates from three replicates are reported.

DOGS	Copy number	WOLVES	Copy number
Swedish Lapphound 1	8.1	22799, Belarus	1.6
Swedish Lapphound 2	15.1	22801, Russia	2.0
American Staffordshire Terrier 1	11.5	22802, Russia	1.7
American Staffordshire Terrier 2	18.2	22803, Russia	1.7
Beagle 1	8.6	22804, Bulgaria	1.8
Bearded Collie 1	23.0	22807, Spain	1.7
Bearded Collie 2	10.2	22808, Spain	1.8
Bearded Collie 3	12.5	22809, Spain	1.6
Beddlington Terrier 1	15.3	22810, Spain	1.8
Belgian Shepard 1	9.7	22800, Belarus	1.7
Belgian Shepard 2	7.3	10w, Canada	2.4
Belgian Tervern 1	12.0	11w, USA	2.1
Bichon Frise 1	16.2	Scandinavian Wolf 1	2.2
Border Collie 1	6.9	Scandinavian Wolf 17	2.3
Border Collie 2	7.7	Scandinavian Wolf 19	1.5
Border Collie10	9.4	Scandinavian Wolf 2	2.2
Border Collie11	21.3	Scandinavian Wolf 20	2.2
Border Collie12	17.0	Scandinavian Wolf 21	2.2
Border Collie13	17.6	Scandinavian Wolf 22	1.6
Border Collie14	17.6	Scandinavian Wolf 23	1.6
Border Collie15	14.8	Scandinavian Wolf 24	2.2
Border Collie16	13.3	Scandinavian wolf 25	2.2
Border Collie3	16.5	Scandinavian wolf 26	2.4
Border Collie4	15.1	Scandinavian Wolf 27	2.0
Border Collie5	15.4	Scandinavian Wolf 28	2.2
Border Collie6	13.2	Scandinavian Wolf 29	2.0
Border Collie7	21.1	Scandinavian Wolf 3	2.0
Border Collie9	14.7	Scandinavian Wolf 30	2.2
Border Terrier 1	11.7	Scandinavian Wolf 31	2.0
Border Terrier 2	13.1	Scandinavian Wolf 4	2.0
Boxer 1	14.8	Scandinavian Wolf 5	2.0
Boxer 1	10.3	Scandinavian Wolf 5	2.4
Boxer 2	15.9	Scandinavian Wolf 6	1.8
Boxer 2	10.8	Scandinavian Wolf 7	2.2
Boxer 3	15.2	Scandinavian Wolf 18	2.1
Bullterrier 1	12.4	**Average**	**2.0**
Bullterrier 2	10.8		
Chinese Crested 1	11.1		
Chinese Crested 2	9.6		
Cocker Spaniel 1	12.4		
Cocker Spaniel 3	10.2		
Collie 1	22.8		
Dalmation 1	9.8		
English Springer Spaniel 1	11.0		
English Springer Spaniel 2	19.9		
English Springer Spaniel 3	23.0		
English Springer Spaniel 4	14.2		
Eurasier 1	15.3		
Field Spaniel 1	15.7		
German Shepard 1	14.8		

German Shepard 2	20.0
German Shepard 4	27.3
German Shepard 5	16.9
Golden Retriever 1	17.3
Golden Retriever 2	26.1
Golden retriever 3	21.2
Great Dane 1	17.0
Great Dane 2	12.3
Great Dane 3	12.2
Howawart 1	11.1
Howawart 2	15.6
Howawart 3	15.3
Irish Wolfhound 1	13.3
Irish Wolfhound 2	12.0
Irish Wolfhound 3	11.6
Karelian Beardog 1	30.3
Karelian Beardog 2	16.7
Labrador 1	16.2
Labrador 2	12.4
Labrador 3	13.9
Labrador 4	15.2
Labrador 5	12.7
Labrador 6	16.7
Labrador 7	16.7
Leonberger 1	19.0
Leonberger 2	21.9
Leonberger 3	25.0
Löwchen 1	17.6
Miniature Schnauzer 1	17.4
Miniature Schnauzer 2	24.4
Newfoundland 1	13.3
Newfoundland 2	18.6
Nova Scotia Duck Trolling Retriever 1	12.5
Nova Scotia Duck trolling retriever 2	11.7
Nova Scotia Duck trolling retriever 3	9.4
Nova Scotia Duck trolling retriever 4	11.7
Papillion 1	17.4
Papillion 2	15.3
Polish Lowland Sheepdog 1	15.1
Polish Lowland Sheepdog 2	11.2
Polish Lowland Sheepdog 3	15.8
Poodle 1	8.0
Poodle 2	15.6
Poodle 3	8.7
Poodle 4	22.9
Poodle 5	9.5
Portuguese Waterdog 1	12.9
Pug 2	5.1
Pug1	8.0
Pumi 1	9.4
Pumi 2	24.0
Giant Schnauzer 1	16.2
Giant Schnauzer 2	8.7
Rottweiler 1	11.4

Rottweiler 2	23.9
Rottweiler 3	20.3
Samoyed 1	8.6
Samoyed 2	4.5
Shar-Pei 1	7.5
Shar Pei 2	9.5
Sheltie 1	12.6
Smaland Hound 1	18.5
Smaland Hound 2	12.2
Smaland Hound 3	17.9
Swedish Elkhound 1	16.9
Swedish Elkhound 10	14.1
Swedish Elkhound 17	17.1
Swedish Elkhound 18	13.8
Swedish Elkhound 3	17.5
Swedish Elkhound11	18.7
Swedish Elkhound12	16.4
Swedish Elkhound13	15.0
Swedish Elkhound14	15.2
Swedish Elkhound15	14.9
Swedish Elkhound16	12.5
Swedish Elkhound4	15.7
Swedish Elkhound5	14.0
Swedish Elkhound7	9.3
Swedish Elkhound8	18.4
Swedish Elkhound9	14.5
Swedish Vallhund 1	19.5
Swedish Vallhund 2	12.2
Tibetan Spaniel 1	7.0
West Highland White Terrier 1	12.2
West Highland White Terrier 2	8.2
West Siberian Laika	25.4
Average	**14.7**

Table S12. Amylase activity in fresh serum from 6 dogs and 6 wolves.

Wolves	**Amylase activity (ukat/L)**
Järvsö zoo 1	3.6
Järvsö zoo 2	3.3
Järvsö zoo 3	2.8
Järvsö zoo 4	4.1
Järvsö zoo 5	3.8
Järvsö zoo 6	3.2
Average	**3.5**
Dogs	
Bernese Mountain dog	17
Irish Setter	10.6
Mixed breed	15.1
Bichon Frisé	15.1
Boxer	16
English Springer Spaniel	9.3
Average	**13.85**

Table S13. Amylase activity from serum of 6 wolves. Note that these samples were assayed using a different instrument (VetScan® (Abaxis Inc., USA)) compared to the other measurements. Range of dog reference values is noted for comparison.

Wolves	Amylase activity (IU/L)	Dog reference activity (IU/L)
Nordens ark zoo	176	
Nordens ark zoo	197	
Nordens ark zoo	193	
Nordens ark zoo	167	
Nordens ark zoo	54	
Nordens ark zoo	198	
Nordens ark zoo	171	
Nordens ark zoo	223	
Average	**172.4**	**200-1200**

Table S14. Canine reference panel. Panel of 74 dogs representing 38 diverse breeds and 21 wolves of wide geographical origin used for the iPLEX genotyping assay.

	Dog breed	Count
	Polish Lowland	2
	American Staffordshire terrier	2
	Karelsk Björnhund	2
	Pumi	2
	Golden Retriever	2
	Bearded Collie	2
	Border Collie	2
	Smålandsstövere	2
	English Springer Spaniel	2
	German Shepard	2
	Samoyed	2
	Pug	2
	Miniature Schnauzers	2
	Swedishk Lapphound	2
	Poodle	2
	Elkhound	2
	Hovawart	4
	Shar-Pei	1
	Toller	1
	West Highland White Terrier	2
	Papillion	2
	Bichon Frisé	2
	Västgötaspets	2
	Cocker Spaniel	2
	Irish Wolfhound	2
	Boxer	2
	Belgian Tervuren	2
	Rottweiler	2
	Gran Dane	2
	Bull Terrier	2
	Chinese Crested	2
	Boston Terrier	2
	Löwchen	1
	Labrador	3
	Leonberger	2
	Sheltie	1
	Giant Schnauzer	2
	Dalmation	1
Total	**38**	**74**

Wolf origin	Count
Sweden	6
Spain	4
Russia	3
Belarus	2

	Bulgaria	1
	USA	1
	Canada	1
	Copenhagen Zoo	1
	Nordens Ark Zoo	2
Total		**21**

Table S15. Candidate causative mutations in MGAM and SGLT1. Alleles, allele frequencies (freq.) and number of chromosomes (sequencing reads) sampled (cov.) for five candidate causative mutations residing in *MGAM* and *SGLT1* are shown for wolf and dog. Results are from the resequencing data and genotyping assay (when available).

		RESEQUENCING DATA						GENOTYPING DATA					
		WOLF			DOG			WOLF			DOG		
chr.	position	allele	freq.	cov.	allele	freq.	cov.	allele	freq.	cov.	allele	freq.	cov.
16	10103702	CA	1	6	-	1	50	-	-	-	-	-	-
16	10117660	A	0.57	7	G	1	53	-	-	-	-	-	-
16	10135196	T	1	11	C	1	54	T	1	40	C	0.92	142
16	10143343	T	0.67	9	T	1	27	T	0.68	40	T	1	142
26	27964111	A	0.83	6	G	0.98	41	A	0.98	40	G	0.94	142

Table S16. MGAM non synonymous candidate causative mutation (chr16: 10,135,196). Wolf, the omnivorous rat and the insectivorous hedgehog and short tailed opossum lack valine at MGAM residue 1001.

```
Dog          ASSSPGVPFCYFVNDLYSVSDVQYDSHGATATISLKSSVYASALPSVPVTSL
Wolf         ASSSPGVPFCYFVNDLYSVSDIQYDSHGATATISLKSSVYASALPSVPVTSL
Human        ASNSSGVPFCYFVNDLYSVSDVQYNSHGATADISLKSSVYANAFPSTPVNPL
Chimpanzee   ASNSSGVPFCYFVNDLYSVSDVQYNSHGATADISLKSSIYANAFPSTPVNPL
Orangutan    ASNSSGVPFCYFVNDLYSVSDVQYNSHGATADISLKSSVYASAFPSTPVNPL
Macaque      ASNSSGVPFCYFVNDLYSVSNVQYSSHGATADISLKSSVYANAFPSTPVNPL
Baboon       ASNSSGVPFCYFVNDLYSVSNV----------------------------
Mouse        ESNTIGVPTCYFAHELYSVSNVQYDSHGATADISLKASTYSNAFPSTPVNKL
Rat          VSNTPGVPHCYFANELYSVSNIQYNSHGATADIFLKASTYSNAFPSTPVNQL
Guinea pig   ESSTTGVPFCYFVTDLYSVSNVQYDSQGASADISLKSSSYANAFPSTPVSPL
Rabbit       ESASPGVPFCYFVNDLYSVSNVQYNSDGATADISLKSSVEANAFPSTPVNPL
Horse        ESSSPGVPFCYFVSDLYSVSDVQYDTHGATAVISLNSSPYAYALPSIPVNSL
Hedgehog     VSTIDRVPHCYFVKDLYSVSDIQYNSNGASAVISLSSSLYANAFPSTPVNPL
Elephant     ESSISGVPFCYFVSDLYSVDVVQYKADGATADISLKTGVYADAFPSTPVTSL
Opossum      LSNSPGVPNCYVINHLYSVSSIQYNPTGITADIFLNSPVRASAGLSTPVNPL
```

Table S17. MGAM non synonymous candidate causative mutation (chr16: 10,143,343). Wolf and the insectivorous shrew and hedgehog lack methionine at MGAM residue 797.

```
Dog          GARARWRKQRVEMGLPADKIGLHLRGGHIFPTQQPATTTVAS
Wolf         GARARWRKQRVEVGLPADKIGLHLRGGHIFPTQQPATTTVAS
Human        GSQVRWRKQKVEMELPGDKIGLHLRGGYIFPTQQPNTTTLAS
Chimpanzee   GSQVRWRKQKVEMELPGDKIGLHLRGGYIFPTQQPNTTTLAS
Orangutan    GSQVRWRKQKVEMELPGDKIGLHLRGGYIFPTQQPNTTTLAS
Macaque      GNQVRWRKQKVEMELPGDKIGLHLRGGYIFPTQQPNTTTLAS
Baboon       GNQVRWRKQKVEMELPGDKIGLHLRGGYIFPTQQPNTTTLAS
Mouse        GEELGWRKQSIEMQLPGDKIGLHLRGGYIFPTQQPATTTEAS
Rat          GEQLAWRKQSVEMELPEDKIGLHLRGGYIFPTQQPATTTEAS
Guinea pig   GGQLGWRKQNIEMELPGDKIGLHLRGGYIFPIQQPSTTTVAS
Cow          -----WRKQFVEMLLPGDRIGLHLRGGYIFPIQQPNTTTETS
Horse        GGRVRWRKQQVEMDLPGDKIGLHLRGGYIFPTQQPATTTVAS
Cat          GARTRWRKQRVEMELPGDKIGLHLRGGHVFPTQQPATTTVVS
Bat          GSQLRWRKQKVEMQLPGDKIGLHLRGGYIFPTQQPATTTVA-
Hedgehog     GAKMNWRGNKVELQLPKDKIGLHFRGGYIFPIQEPAMTTVAS
Shrew        GAQLNWRGNKD--MLPKDKIGLHLRGGYIFPTQQPATTTVAS
Elephant     GARIRWRKQQVEMELPGDKIGLHLRGGYIFPTQEPSTTTEAS
Sloth        GGQIRWRKQKVEMLLPGDKIGLHLRGGYIFPTQQPATTTVL-
Opossum      GGQIPERKQQVEMLFSPEQIGLHLRGGYIFPIQQPAITTVAS
```

Table S18. *MGAM* candidate causative C-terminal deletion disrupts a stop codon and thereby extends dog MGAM by two amino acids (chr16: 10,103,702) relative to in wolf. Blue capital letters refer to coding sequence. Black, small letters denote 3' UTR. Omnivorous mouse lemur and rat, as well as herbivorous rabbit, pika, alpaca and cow share a similar extension of MGAM as seen in dog.

```
Dog                GAGA-AGCACTC--TGAATTTTTAGagc
Wolf               GAGA-AGCACTCCATGAattttagagc
Human              GATA-AGCACTCTGTGAattttacagc
Chimp              GATA-ATCACTCTGTGAattttacagc
Gorilla            GATA-AGCACTCTGTGAattttacagc
Orangutan          GATA-AGCACTCTGTGAattttacagc
Rhesus             GAGA-AGCACTCCGTGAattttagagc
Baboon             GAGA-AGCACTCCGTGAattttagagc
Marmoset           GAGA-AGCACTCCTTGAattttcagagc
Tarsier            GAAA-AGCACTCTGTGAattttaaaac
Mouse lemur        GAGA-AGCACTC--TGAATTTTTAGagc
Bushbaby           GA-A-TACTCTG--TGAactttagagc
Tree shrew         GGGA-AGCAGCC--TGAacttctgggac
Rat                GAGCTAGCTCTTCATAATGTTTTAAagc
Kangaroo rat       GAAT-ATCTCTCCATAAaactttagtgt
Mouse              G----AGCTCTTCATAAt-ttttaaagc
Guinea pig         GAAA-AGCTCTCTGTAAattttaggac
Squirrel           GAAA-AGCTCTCCATAAattttagagc
Rabbit             GAGA-AGCACA--ACGAGCTCTCAGCGC
Pika               GAGA-AGCACT--ACGGATTCTTAGtac
Alpaca             GAGA-AGCCCTC--TGAGTTCCTAGagc
Dolphin            G-AG-AGCACTC--TGAattttagagc
Cow                A-GA-AGCACGC--GAA--TTTTAGagc
Microbat           GAAA-AGCACTCTGTGAattttaagagc
Horse              GAGC-AGCACTCTGTGAattttagagc
Megabat            GAGA-AGCACTCTGTGAattttaagaac
Shrew              GTAA-AACACTCTACAGacttttaaagc
Hedgehog           GACA-AACAACCCATGAactgctagagc
Elephant           GGCA-AGCACTCTGTGAatactcagagc
Rock hyrax         GACA-AGCACTCCATGAattccaagagc
```

```
Tenrec          GACA-GGCACTCTGTGAacgtaaagagc
Sloth           GAGC-AGCACTCTGTGAatcttcagagt
Wallaby         GCAG-AGTTCCTCATGAccatttggagc
Opossum         GTGA-AGTTTCCAATGGcccttgaagc
```

Table S19. MGAM candidate causative mutation affecting NR4A2 transcription factor binding site (chr16: 10,117,660). Wolf share a T with 4 other mammals at this site: the mainly insectivorous tarsier and marmoset, the fish eating dolphin and the primarily seed eating kangaroo rat. Dog and 21 other mammals share a C, as in the canonical binding motif of the NR4A2 protein.

Canonical NR4A2 binding motif	**TGACCTT**
Dog	TGACATC
Wolf	TGATATC
Human	TGACCTT
Common marmoset	TGTTCTT
Macaque	-----TT
Orangutan	TGACCTT
Gorilla	TGACCTT
Chimapnzee	TGACCTT
Grey mouse lemur	T----CC
Philippine tarsier	TTATTTT
Rat	TGACC--
Kangaroo Rat	TGATTTC
Guinea pig	TGACATC
Rabbit	TGACTTC
Pika	TGACTTC
Horse	TGACCTC
Micro bat	TGACCTC
Large flying fox	CGACCCC
Cow	CGACCTC
Dolphin	TTATCTC
Lama	-GACCTC
Common shrew	TGACCTT
Thirteen lined ground squirrel	TTACCTC
Rock hyrax	TGCCCTT
Hoffmann's two-toed sloth	TGACCAT
Platypus	-GGCCAG
Mouse	TGACCT-
Elephant	TGCCCTT
Hedgehog	TCACTTC

Table S20. TaqMan realtime PCR primers and probe for the *AMYB2* CNV quantification.

Primers

AMYLASE-5'	CCAAACCTGGACGGACATCT
AMYLASE-3'	TATCGTTCGCATTCAAGAGCAA

Probe

Amylase: 6FAM - TTT GAG TGG CGC TGG G – MGBNFQ

Table S21. Real time PCR primers for mRNA expression analyses.

Primers:	
MGAM-5'	GGTTGCTTTGGATGATGAGG
MGAM-3'	AATGGAAACACTGCCCACTC
Amyl-5'	CTGGTGGGATAATGGTAGCAA
Amyl-3'	GAAAAATGAGCATTCCCATCC
SGLT1-5'	TGCCAGTAACATTGGGAGTG
SGLT1-3'	GGTAGATCTGGATTCGCTTGC

Table S22. Comparison of patterns of genetic variation in CDRs with the genomic average using genotyping data from the Illumina 170K Canine HD-array. The SNP density (SNPs/Kb), the minor allele frequency in dogs (MAF dog), the minor allele frequency in wolf (MAF wolf) and the fixation index between dog and wolf (F_{ST}) is stated for 36 CDRs and for the whole genome (genome).

CDR id	Chr.	Start	End	Length	SNPs	SNPs/ Kb	MAF dog	MAF wolf	F_{ST}
genome	**genome**	**genome**	**genome**	**1897311618**	**165542**	**0.09**	**0.24**	**0.29**	**0.13**
1	1	5517430	6317430	800001	46	0.06	0.08	0.04	0.21
2	1	49617430	49917430	300001	19	0.06	0.16	0.18	0.23
3	1	66617430	66817430	200001	11	0.05	0.08	0.28	0.30
4	1	83017430	83217430	200001	10	0.05	0.07	0.16	0.24
5	2	46500196	46700196	200001	8	0.04	0.01	0.02	0.04
6	3	18207515	18507515	300001	26	0.09	0.09	0.57	0.49
7	3	21507515	21707515	200001	9	0.04	0.04	0.10	0.12
8	3	34907515	35107515	200001	7	0.03	0.07	0.02	0.09
9	4	17700233	17900233	200001	12	0.06	0.11	0.17	0.12
10	4	44000233	44200233	200001	12	0.06	0.06	0.14	0.35
11	6	27107924	27307924	200001	11	0.05	0.04	0.14	0.32
12	6	28007924	28307924	300001	10	0.03	0.03	0.17	0.58
13	6	49907924	50507924	600001	36	0.06	0.09	0.44	0.50
14	6	56307924	56507924	200001	9	0.04	0.02	0.31	0.77
15	7	27600316	28100316	500001	14	0.03	0.01	0.02	0.08
16	8	30702583	30902583	200001	8	0.04	0.12	0.29	0.29
17	10	5701010	5901010	200001	8	0.04	0.03	0.05	0.13
18	10	6601010	7101010	500001	23	0.05	0.11	0.12	0.26
19	11	40800086	41100086	300001	16	0.05	0.01	0.00	0.03
20	11	50300086	50700086	400001	15	0.04	0.03	0.11	0.25
21	11	56900086	57100086	200001	5	0.02	0.02	0.21	0.32
22	14	10200337	10500337	300001	22	0.07	0.10	0.46	0.57
23	15	8103479	8403479	300001	13	0.04	0.06	0.12	0.34
24	15	38203479	38403479	200001	8	0.04	0.06	0.00	0.14
25	16	9807391	10307391	500001	19	0.04	0.07	0.18	0.37
26	16	11907391	12107391	200001	8	0.04	0.06	0.03	0.05

27	17	41722203	41922203	200001	8	0.04	0.13	0.20	0.12
28	18	3404681	4404681	1000001	47	0.05	0.09	0.15	0.38
29	18	6204681	7704681	1500001	75	0.05	0.05	0.14	0.45
30	19	40704062	40904062	200001	12	0.06	0.23	0.25	0.18
31	22	22941181	23141181	200001	14	0.07	0.12	0.42	0.46
32	25	4000488	4500488	500001	26	0.05	0.07	0.22	0.57
33	26	27900108	28100108	200001	9	0.04	0.25	0.40	0.35
34	28	9400594	9600594	200001	0	0.00	NA	NA	0.00
35	28	12200594	12500594	300001	0	0.00	NA	NA	0.00
36	37	9915022	10115022	200001	0	0.00	NA	NA	0.00
CDR average						**0.04**	**0.08**	**0.19**	**0.27**

Table S23. Candidate domestication regions on chromosome X. Start and end of regions with significantly increased ($Z(F_{ST})>3$) fixation index, F_{ST}, shows the position of 6 CDRs. Individual CDRs are separated by horizontal lines. Ensemble IDs, gene descriptions and gene names of genes residing in CDRs are shown.

Chr	F_{ST} region	start	end	F_{ST}	ZF_{ST}	Ensemble ID	Gene description	Gene
X	1	39105187	39505187	0.86	3.1	ENSCAFG00000014691		*HNRNA1*
X	2	42905187	43105187	0.96	3.6	ENSCAFG00000015996	G2/mitotic-specific cyclin-B3	*CCNB3*
X	2					ENSCAFG00000016002		*RBM3*
X	2					ENSCAFG00000016018		*DGKK*
X	3	81505187	81705187	0.91	3.3			
X	4	111005187	111205187	0.85	3.0			
X	5	111505187	112005187	0.88	3.2	ENSCAFG00000018988		*MAGEC3*
X	5					ENSCAFG00000018995	fibroblast growth factor 13	*FGF13*
X	6	112205187	112605187	0.85	3.0	ENSCAFG00000018998	Coagulation factor IX Precursor	*FA9 CANFA*

Table S24. Ranking of 200 Kb windows in the wolf genome based on a significantly reduced average pooled heterozygosity, H_P, sorted by Z-sore. Wolf CSR indicates which candidate selection region the window is part of. Z-score refers to the value of the window after Z-transformation of the H_P distribution. Ensemble ID and gene name or gene description is shown for genes residing in these windows.

wolf CSR	Position (Chrom:Mb)	H_P	Z-score	Ensemble ID	Gene
1	6: 42.1-42.3	0.013	-6.34	ENSCAFG00000014270	*CRAMP1L*
1	6: 42.1-42.3	0.013	-6.34	ENSCAFG00000019486	*C16orf73*
1	6: 42.1-42.3	0.013	-6.34	ENSCAFG00000019488	*FAHD1*
1	6: 42.1-42.3	0.013	-6.34	ENSCAFG00000019490	*HAGH*
1	6: 42.1-42.3	0.013	-6.34	ENSCAFG00000019493	*IGFALS*
1	6: 42.1-42.3	0.013	-6.34	ENSCAFG00000019494	*NUBP2*
1	6: 42.1-42.3	0.013	-6.34	ENSCAFG00000019495	*SPSB3*
1	6: 42.1-42.3	0.013	-6.34	ENSCAFG00000019496	*MRPS34*
1	6: 42.1-42.3	0.013	-6.34	ENSCAFG00000019497	*NME3*
1	6: 42.1-42.3	0.013	-6.34	ENSCAFG00000019513	*MAPK8IP3*
1	6: 42.1-42.3	0.013	-6.34	ENSCAFG00000019540	*HN1L*
1	6: 42.1-42.3	0.013	-6.34	ENSCAFG00000019542	*XM_547195.2*
1	6: 42.1-42.3	0.013	-6.34	ENSCAFG00000023861	*EME2*
15	24: 5.9-6.1	0.022	-6.13	ENSCAFG00000005251	*RALGAPA2*
15	24: 5.9-6.1	0.022	-6.13	ENSCAFG00000005264	*INSM1*
15	24: 5.9-6.1	0.022	-6.13	ENSCAFG00000005270	*C20orf26*
1	6: 42-42.2	0.025	-6.06	ENSCAFG00000019472	*XM_859963.1*
1	6: 42-42.2	0.025	-6.06	ENSCAFG00000019476	*NDUFB10*
1	6: 42-42.2	0.025	-6.06	ENSCAFG00000019480	*RPL3L*
1	6: 42-42.2	0.025	-6.06	ENSCAFG00000019484	
1	6: 42-42.2	0.025	-6.06	ENSCAFG00000019486	*C16orf73*
1	6: 42-42.2	0.025	-6.06	ENSCAFG00000019488	*FAHD1*
1	6: 42-42.2	0.025	-6.06	ENSCAFG00000019490	*HAGH*
1	6: 42-42.2	0.025	-6.06	ENSCAFG00000019493	*IGFALS*
1	6: 42-42.2	0.025	-6.06	ENSCAFG00000019494	*NUBP2*
1	6: 42-42.2	0.025	-6.06	ENSCAFG00000019495	*SPSB3*
1	6: 42-42.2	0.025	-6.06	ENSCAFG00000019496	*MRPS34*
1	6: 42-42.2	0.025	-6.06	ENSCAFG00000019497	*NME3*
1	6: 42-42.2	0.025	-6.06	ENSCAFG00000019513	*MAPK8IP3*
1	6: 42-42.2	0.025	-6.06	ENSCAFG00000023861	*EME2*
1	6: 42-42.2	0.025	-6.06	ENSCAFG00000024261	*HS3ST6*
1	6: 41.8-42	0.032	-5.9	ENSCAFG00000019392	*RAB26*
1	6: 41.8-42	0.032	-5.9	ENSCAFG00000019397	*RNPS1*
1	6: 41.8-42	0.032	-5.9	ENSCAFG00000019412	*CASKIN1*
1	6: 41.8-42	0.032	-5.9	ENSCAFG00000019418	*TRAF7*
1	6: 41.8-42	0.032	-5.9	ENSCAFG00000019424	*Q8MJF3_CANFA*
1	6: 41.8-42	0.032	-5.9	ENSCAFG00000019430	*TBL3*
1	6: 41.8-42	0.032	-5.9	ENSCAFG00000019438	*Q9XSY8_CANFA*
1	6: 41.8-42	0.032	-5.9	ENSCAFG00000019439	*NTHL1*
1	6: 41.8-42	0.032	-5.9	ENSCAFG00000019449	*SLC9A3R2*
1	6: 41.8-42	0.032	-5.9	ENSCAFG00000019456	*ZNF598*
1	6: 41.8-42	0.032	-5.9	ENSCAFG00000019459	*SYNGR3*

1	6: 41.8-42	0.032	-5.9	ENSCAFG00000019461	*GFER*
1	6: 41.8-42	0.032	-5.9	ENSCAFG00000019462	*NOXO1*
1	6: 41.8-42	0.032	-5.9	ENSCAFG00000019463	*RNF151*
12	21: 4.3-4.5	0.033	-5.86	ENSCAFG00000025022	
17	25:8.3-8.5	0.038	-5.74	ENSCAFG00000006255	*NBEA*
1	6: 41.9-42.1	0.038	-5.74	ENSCAFG00000019397	*RNPS1*
1	6: 41.9-42.1	0.038	-5.74	ENSCAFG00000019430	*TBL3*
1	6: 41.9-42.1	0.038	-5.74	ENSCAFG00000019438	*Q9XSY8_CANFA*
1	6: 41.9-42.1	0.038	-5.74	ENSCAFG00000019439	*NTHL1*
1	6: 41.9-42.1	0.038	-5.74	ENSCAFG00000019449	*SLC9A3R2*
1	6: 41.9-42.1	0.038	-5.74	ENSCAFG00000019456	*ZNF598*
1	6: 41.9-42.1	0.038	-5.74	ENSCAFG00000019459	*SYNGR3*
1	6: 41.9-42.1	0.038	-5.74	ENSCAFG00000019461	*GFER*
1	6: 41.9-42.1	0.038	-5.74	ENSCAFG00000019462	*NOXO1*
1	6: 41.9-42.1	0.038	-5.74	ENSCAFG00000019463	*RNF151*
1	6: 41.9-42.1	0.038	-5.74	ENSCAFG00000019472	*XM_859963.1*
1	6: 41.9-42.1	0.038	-5.74	ENSCAFG00000019476	*NDUFB10*
1	6: 41.9-42.1	0.038	-5.74	ENSCAFG00000019480	*RPL3L*
1	6: 41.9-42.1	0.038	-5.74	ENSCAFG00000019484	
1	6: 41.9-42.1	0.038	-5.74	ENSCAFG00000019486	*C16orf73*
1	6: 41.9-42.1	0.038	-5.74	ENSCAFG00000024261	*HS3ST6*
15	24: 5.6-5.8	0.041	-5.68	ENSCAFG00000005251	*RALGAPA2*
6	9: 5.3-5.5	0.049	-5.5	ENSCAFG00000005528	*TIMP2*
6	9: 5.3-5.5	0.049	-5.5	ENSCAFG00000005537	*LGALS3BP*
6	9: 5.3-5.5	0.049	-5.5	ENSCAFG00000005540	*CANT1*
6	9: 5.3-5.5	0.049	-5.5	ENSCAFG00000005542	*C1QTNF1*
6	9: 5.3-5.5	0.049	-5.5	ENSCAFG00000005545	*ENGASE*
6	9: 5.3-5.5	0.049	-5.5	ENSCAFG00000005547	
5	9: 3.7-3.9	0.049	-5.5	ENSCAFG00000005697	*CHMP6*
5	9: 3.7-3.9	0.049	-5.5	ENSCAFG00000005700	*Q7YS42_CANFA*
5	9: 3.7-3.9	0.049	-5.5	ENSCAFG00000005705	*AATK*
5	9: 3.7-3.9	0.049	-5.5	ENSCAFG00000005711	*AZI1*
5	9: 3.7-3.9	0.049	-5.5	ENSCAFG00000005711	*AZI1*
5	9: 3.7-3.9	0.049	-5.5	ENSCAFG00000005719	*C17orf56*
5	9: 3.7-3.9	0.049	-5.5	ENSCAFG00000005719	*C17orf56*
5	9: 3.7-3.9	0.049	-5.5	ENSCAFG00000005725	*SLC38A10*
5	9: 3.7-3.9	0.049	-5.5	ENSCAFG00000005725	*SLC38A10*
3	7: 5.5-5.7	0.049	-5.49	ENSCAFG00000011120	*DDX59*
3	7: 5.5-5.7	0.049	-5.49	ENSCAFG00000011140	*KIF14*
3	7: 5.5-5.7	0.049	-5.49	ENSCAFG00000011151	*ZNF281*
13	22: 14.9-15.1	0.054	-5.38		
1	6: 42.2-42.4	0.054	-5.37	ENSCAFG00000014270	*CRAMP1L*
1	6: 42.2-42.4	0.054	-5.37	ENSCAFG00000019513	*MAPK8IP3*
1	6: 42.2-42.4	0.054	-5.37	ENSCAFG00000019540	*HN1L*
1	6: 42.2-42.4	0.054	-5.37	ENSCAFG00000019542	*XM_547195.2*
1	6: 42.2-42.4	0.054	-5.37	ENSCAFG00000019547	*TMEM204*
1	6: 42.2-42.4	0.054	-5.37	ENSCAFG00000019549	*TELO2*
1	6: 42.2-42.4	0.054	-5.37	ENSCAFG00000019558	*PTX4*
1	6: 42.2-42.4	0.054	-5.37	ENSCAFG00000019565	*CLCN7*
1	6: 42.2-42.4	0.054	-5.37	ENSCAFG00000019581	
4	7: 33.5-33.7	0.055	-5.34	ENSCAFG00000015343	*DCAF6*
4	7: 33.5-33.7	0.055	-5.34	ENSCAFG00000015358	*BRP44*
4	7: 33.5-33.7	0.055	-5.34	ENSCAFG00000015387	*ADCY10*
4	7: 33.5-33.7	0.055	-5.34	ENSCAFG00000015405	*MPZL1*

14	22: 35.5-35.7	0.057	-5.31	ENSCAFG00000005242	*RBM26*
14	22: 35.5-35.7	0.057	-5.31	ENSCAFG00000005256	*NDFIP2*
12	21: 4.2-4.4	0.059	-5.27	ENSCAFG00000004018	
12	21: 4.2-4.4	0.059	-5.27	ENSCAFG00000025022	
9	14: 9.3-9.5	0.061	-5.22	ENSCAFG00000001426	*COPG2*
9	14: 9.3-9.5	0.061	-5.22	ENSCAFG00000001450	*MEST*
9	14: 9.3-9.5	0.061	-5.22	ENSCAFG00000001459	*TSGA14*
8	13: 3.5-3.7	0.061	-5.21	ENSCAFG00000000484	
8	13: 3.5-3.7	0.061	-5.21	ENSCAFG00000000488	*KCNS2*
18	30: 9.1-9.3	0.061	-5.21		
2	6: 43.1-43.3	0.062	-5.18	ENSCAFG00000015363	*ARHGDIG*
2	6: 43.1-43.3	0.062	-5.18	ENSCAFG00000015429	*PDIA2*
2	6: 43.1-43.3	0.062	-5.18	ENSCAFG00000019675	*DECR2*
2	6: 43.1-43.3	0.062	-5.18	ENSCAFG00000019677	
2	6: 43.1-43.3	0.062	-5.18	ENSCAFG00000019680	*TMEM8A*
2	6: 43.1-43.3	0.062	-5.18	ENSCAFG00000019682	*MRPL28*
2	6: 43.1-43.3	0.062	-5.18	ENSCAFG00000019684	*AXIN1*
2	6: 43.1-43.3	0.062	-5.18	ENSCAFG00000019687	*ITFG3*
2	6: 43.1-43.3	0.062	-5.18	ENSCAFG00000023874	*RGS11*
6	9: 5-5.2	0.066	-5.1		
15	24: 5.7-5.9	0.067	-5.08	ENSCAFG00000005251	*RALGAPA2*
7	11: 34.7-34.9	0.067	-5.08		

Table S25. Enriched GO-categories among genes residing in wolf candidate selected regions (wolf CSRs).

Ensembl id	Genes	Group count	Total count	P-value	GO term
ENSCAFG00000019430	*TBL3*				
ENSCAFG00000003662	*TBC1D9*				
ENSCAFG00000005540	*CANT1*				
ENSCAFG00000000484	*STK3*				
ENSCAFG00000019495	*SPSB3*				
ENSCAFG00000019513	*MAPK8IP3*				
ENSCAFG00000015429	*PDIA2*				
ENSCAFG00000015363	*ARHGDIG*				
ENSCAFG00000005256	*NDFIP2*				
ENSCAFG00000019418	*TRAF7*				
ENSCAFG00000019438	*TSC2*				
ENSCAFG00000019392	*RAB26*				
ENSCAFG00000005528	*TIMP2*				
ENSCAFG00000023874	*RGS11*	14	1965	0.0184	intracellular signaling cascade
ENSCAFG00000019418	*TRAF7*				
ENSCAFG00000019392	*RAB26*	2	8	0.0184	exocrine system development
ENSCAFG00000005256	*NDFIP2*				
ENSCAFG00000019418	*TRAF7*				
ENSCAFG00000005528	*TIMP2*				
ENSCAFG00000005540	*CANT1*				
ENSCAFG00000000484	*STK3*				
ENSCAFG00000019513	*MAPK8IP3*	6	376	0.0188	protein kinase cascade
ENSCAFG00000005256	*NDFIP2*				
ENSCAFG00000003662	*TBC1D9*				
ENSCAFG00000019418	*TRAF7*				
ENSCAFG00000019438	*TSC2*				
ENSCAFG00000005528	*TIMP2*				
ENSCAFG00000005540	*CANT1*				
ENSCAFG00000023874	*RGS11*				
ENSCAFG00000019513	*MAPK8IP3*	8	800	0.033	regulation of signal transduction
ENSCAFG00000019418	*TRAF7*				
ENSCAFG00000005528	*TIMP2*	2	15	0.0341	regulation of MAPKKK cascade
ENSCAFG00000019418	*TRAF7*				
ENSCAFG00000019392	*RAB26*	2	15	0.0341	regulation of exocytosis