Supplementary information, Data S15 Selected genes and gene ontology (GO) analysis

We did two types of analysis targeting artificial selection occurring at different stages of dog domestication. First, we used 11 individuals sampled from southern East Asia and applied a site frequency-based approach (SweepFinder) to locate positively selected regions[24]. Since there is very little selection for specific breed phenotypes (e.g., coat color, etc.) in this group, the selective signal along the southern East Asian lineage should represent artificial selection in the early stage of domestication (Supplementary information, Table S9). In this set, we see three strongly enriched signals of positive selection, namely, genes related to metabolism and motility, genes related to neurological process and perception, and genes related to sexual reproduction.

Since our 46 dog samples represent a diverse collection of populations, we also combined all dogs and used a method that is more tolerant/robust to population differentiation within dogs. We extracted positively selected genes in all dogs by looking for regions with a) high Fst between wolves and dogs (we used a top 2% cutoff) and b) a strong reduction in genetic diversity in dogs (we used $\theta\pi(\text{wolves})/\theta\pi(\text{dogs})$ and used a top 2% as the cutoff). By combining these two measures, we extracted a second list of positively selected genes in dogs. Genes that show strong differentiation between dogs and wolves include those due to selection occurring at the first stage, but also partially reflect selection during breed formation. As we can see from Supplementary information, Table S10, in addition to the three categories above, we see an extra class of genes, namely, genes related to controlling the developmental process. Since domesticated animals often show a diverse array of developmental changes (e.g. neoteny), this group of genes might be responsible for the changes in the developmental programs. In addition, we observed a further enhancement in the category of metabolism, especially genes related to lipid/cholesterol metabolism showing strong selection. These genes might reflect additional selective forces acting on dietary changes in the course of dog evolution.