LETTER

doi:10.1038/nature11837

The genomic signature of dog domestication reveals adaptation to a starch-rich diet

Erik Axelsson[1], Abhirami Ratnakumar[1], Maja-Louise Arendt[1], Khurram Maqbool[1], Matthew T. Webster[1], Michele Perloski[2], Olof Liberg[3], Jon M. Arnemo[4,5], Åke Hedhammar[6] & Kerstin Lindblad-Toh[1,2]

The domestication of dogs was an important episode in the development of human civilization. The precise timing and location of this event is debated[1–5] and little is known about the genetic changes that accompanied the transformation of ancient wolves into domestic dogs. Here we conduct whole-genome resequencing of dogs and wolves to identify 3.8 million genetic variants used to identify 36 genomic regions that probably represent targets for selection during dog domestication. Nineteen of these regions contain genes important in brain function, eight of which belong to nervous system development pathways and potentially underlie behavioural changes central to dog domestication[6]. Ten genes with key roles in starch digestion and fat metabolism also show signals of selection. We identify candidate mutations in key genes and provide functional support for an increased starch digestion in dogs relative to wolves. Our results indicate that novel adaptations allowing the early ancestors of modern dogs to thrive on a diet rich in starch, relative to the carnivorous diet of wolves, constituted a crucial step in the early domestication of dogs.

Domestic animals are crucial to modern human society, and it is likely that the first animal to be domesticated was the dog. Claims of early, fossilised dog remains include a 33,000-year-old doglike canid from the Altai Mountains in Siberia[1], whereas fossils dating from 12,000–11,000 years BP found buried together with humans in Israel[2] could represent the earliest verified dog remains. Patterns of genomic variation indicate that dog domestication started at least 10,000 years BP[3,4] in southern East Asia[4] or the Middle East[5]. Dog domestication may however have been more complex, involving multiple source populations and/or backcrossing with wolves.

It is unclear why and how dogs were domesticated. Humans may have captured wolf pups for use in guarding or hunting, resulting in selection for traits of importance for these new roles. Alternatively, as humans changed from a nomadic to sedentary lifestyle during the dawn of the agricultural revolution, wolves may themselves have been attracted to dumps near early human settlements to scavenge[6]. Natural selection for traits allowing for efficient use of this new resource may have led to the evolution of a variety of scavenger wolves that constituted the ancestors of modern dogs. Regardless of how dog domestication started, several characteristics separating modern dogs from wolves, including reduced aggressiveness and altered social cognition capabilities[7], suggest that behavioural changes were early targets of this process[6]. Dogs also differ morphologically from wolves, showing reduced skull, teeth and brain sizes[6]. Artificial selection for tameness in silver foxes indicates that selection on genetic variation in developmental genes may underlie both behavioural and morphological changes, potentially representing an important mechanism throughout animal domestication[7,8].

At present, only a handful of genes separating wild from domestic forms have been identified in any domestic animals, including coat colour variants in *MC1R* in pig[9] and a mutation in *TSHR* likely to affect seasonal reproduction in chicken[10], but to our knowledge in dogs no genome-wide sequence-based searches have been performed until now. To identify genomic regions under selection during dog domestication we performed pooled whole-genome resequencing of dogs and wolves followed by functional characterization of candidate genes.

Uniquely placed sequence reads from pooled DNA representing 12 wolves of worldwide distribution and 60 dogs from 14 diverse breeds (Supplementary Table 1) covered 91.6% and 94.6%, respectively, of the 2,385 megabases (Mb) of autosomal sequence in the CanFam 2.0 genome assembly[11]. The aligned coverage depth was 29.8× for all dog pools combined and 6.2× for the single wolf pool (Supplementary Table 1 and Supplementary Fig. 1). We identified 3,786,655 putative single nucleotide polymorphisms (SNPs) in the combined dog and wolf data, 1,770,909 (46.8%) of which were only segregating in the dog pools, whereas 140,818 (3.7%) were private to wolves (Supplementary Table 2). Similarly we detected 506,148 short indels and 26,619 copy-number variations (CNVs) (Supplementary Files 1 and 2). We were able to experimentally validate 113 out of 114 tested SNPs (Supplementary Table 3 and Supplementary Discussion, section 1).

To detect signals of strong recent selection we searched the dog genome for regions with reduced pooled heterozygosity (H_P)[10] and/or increased genetic distance to wolf (F_{ST}). As evident from the skewed distribution of heterozygosity scores in dog relative to wolf (Fig. 1a and Supplementary Fig. 2), a major challenge to this approach is to separate true signals of selection from those caused by random fixation of large genomic regions during the formation of dog breeds[11]. We alleviate this problem by combining sequence data from all dog pools before selection analyses and require that detected signals span at least 200 kilobases (kb; Methods and Supplementary Discussion, sections 2 and 3). Given the complex and partly unknown demographic history of dogs, it is furthermore difficult to assign strict thresholds that distinguish selection and drift. We propose that the best way to validate regions detected here is to study genetic data from additional individuals and provide evidence for functional change associated with putatively selected regions. Eventually, indications that similar pathways changed during independent domestication events may provide conclusive evidence for selection. Here we *Z*-transform the autosomal H_P ($Z(H_P)$) and F_{ST} ($Z(F_{ST})$) distributions (see Supplementary Discussion, section 4 for an analysis of the X chromosome) and focus our description of putatively selected regions to those that fall at least five standard deviations away from the mean ($Z(H_P) < -5$ and $Z(F_{ST}) > 5$), as these represent the extreme ends of the distributions. By applying these thresholds we identified 14 regions in the dog genome with extremely low levels of heterozygosity (average length = 400 kb, average $H_P = 0.036$ (range 0.015–0.056), average autosomal $H_P = 0.331$) (Fig. 1c and Supplementary Table 4) and 35 regions with strongly elevated F_{ST} values (average length = 340 kb, average

[1]Science for Life Laboratory, Department of Medical Biochemistry and Microbiology, Uppsala University, 75237 Uppsala, Sweden. [2]Broad Institute of Massachusetts Institute of Technology and Harvard, Cambridge, Massachusetts 02139, USA. [3]Grimsö Wildlife Research Station, Department of Ecology, Swedish University of Agricultural Sciences, 73091 Riddarhyttan, Sweden. [4]Department of Forestry and Wildlife Management, Faculty of Applied Ecology and Agricultural Sciences, Hedmark University College, Campus Evenstad, NO-2418 Elverum, Norway. [5]Department of Wildlife, Fish and Environmental Studies, Faculty of Forest Sciences, Swedish University of Agricultural Sciences, 901 83 Umeå, Sweden. [6]Science for Life Laboratory, Department of Clinical Sciences, Swedish University of Agricultural Sciences, 75651 Uppsala, Sweden.

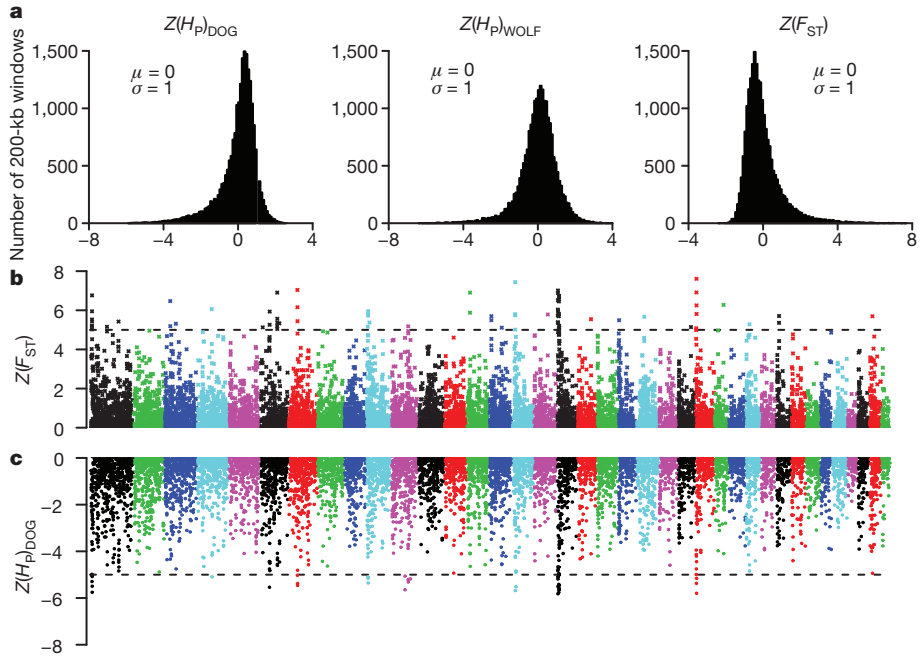

Figure 1 | Selection analyses identified 36 candidate domestication regions. **a**, Distribution of *Z*-transformed average pooled heterozygosity in dog ($Z(H_P)_{DOG}$) and wolf ($Z(H_P)_{WOLF}$) respectively, as well as average fixation index ($Z(F_{ST})$), for autosomal 200 kb windows (σ, standard deviation; μ, average). **b**, 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. **c**, 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.

$F_{ST} = 0.734$ (range 0.654–0.903), average autosomal $F_{ST} = 0.223$) (Fig. 1b and Supplementary Table 5). All F_{ST} regions are characterized by low levels of heterozygosity in either dog or wolf (although all do not pass the $Z(H_P) < -5$ threshold), indicating that the two statistics detect the same events (Methods and Supplementary Discussion, sections 2 and 3). In total, 36 unique autosomal candidate domestication regions (CDRs) containing 122 genes were identified by the two approaches combined (Supplementary Table 6 and Fig. 1b, c). None of these regions overlaps those of a previous genotype-based study[5] (Supplementary Discussion, section 3), stressing the importance of identifying domestication regions directly by sequencing or by comprehensively ascertaining SNPs in wild ancestors before genotyping.

We searched for significantly overrepresented gene ontology terms among genes in autosomal CDRs and identified 25 categories, representing several groups of interrelated terms (Table 1 and Supplementary Table 7), none of which was indicated in a separate analysis of selection in wolf (Supplementary Discussion, section 8). The most conspicuous cluster (11 terms) relates to the term 'nervous system development'. The eight genes belonging to this category (Supplementary Tables 7 and 8) include *MBP*, *VWC2*, *SMO*, *TLX3*, *CYFIP1* and *SH3GL2*, of which several affect developmental signalling and synaptic strength and plasticity[12–16]. We surveyed published literature and identified 11 additional CDR genes with central nervous system function (Supplementary Table 9), adding to a total of 19 CDRs that contain brain genes. These findings support the hypothesis that selection for altered behaviour was important during dog domestication and that mutations affecting developmental genes may underlie these changes[7].

The gene ontology analysis also pinpoints two genes involved in the binding of sperm and egg: *ZPBP* encodes the zona pellucida binding protein that mediates binding of sperm to the zona pellucida glycoprotein layer (ZP) of the egg, and *ZP2* codes for one of the proteins that make up ZP itself. In addition, a CDR on chromosome 6 encompasses *PDILT* that also affects binding of sperm to ZP[17], altogether indicating that sperm competition may have been an important evolutionary force during dog domestication[18].

Overrepresented terms 'starch metabolic process', 'digestion' and 'fatty acid metabolism' include genes involved in starch digestion (*MGAM*) and glucose uptake (*SGLT1*), as well as a candidate gene for

Table 1 | Enriched gene ontology terms among CDR genes

Gene ontology term	P_{FDR} value	Gene count
Regulation of neuron differentiation	0.005	3 (26)
Multicellular organismal process	0.005	21 (3,822)
Digestion	0.008	4 (95)
Neuron differentiation	0.010	5 (210)
Regulation of molecular function	0.011	8 (671)
Central nervous system development	0.013	5 (235)
Regulation of developmental process	0.013	5 (236)
Generation of neurons	0.013	5 (242)
Nervous system development	0.013	8 (716)
Binding of sperm to zona pellucida	0.015	2 (12)
Sperm–egg recognition	0.015	2 (12)
Neurogenesis	0.015	5 (262)
Cell–cell recognition	0.019	2 (14)
Regulation of catalytic activity	0.020	7 (605)
Regulation of hydrolase activity	0.026	5 (307)
Fatty acid metabolic process	0.031	4 (191)
System development	0.034	11 (1,605)
Regulation of GTPase activity	0.039	4 (211)
Anatomical structure development	0.039	12 (2,005)
Intramembranous ossification	0.039	1 (1)
Quinolinate metabolic process	0.039	1 (1)
Starch metabolic process	0.039	1 (1)
Starch catabolic process	0.039	1 (1)
Glucocorticoid catabolic process	0.039	1 (1)
Cell development	0.039	9 (1,242)

Enriched terms are colour-coded to reflect relatedness in the ontology or functional proximity. Blue, nervous system development; green, sperm–egg recognition; grey, regulation of molecular function; orange, digestion. For each term, gene count shows number of genes in CDRs relative to total number of annotated genes (in parentheses).

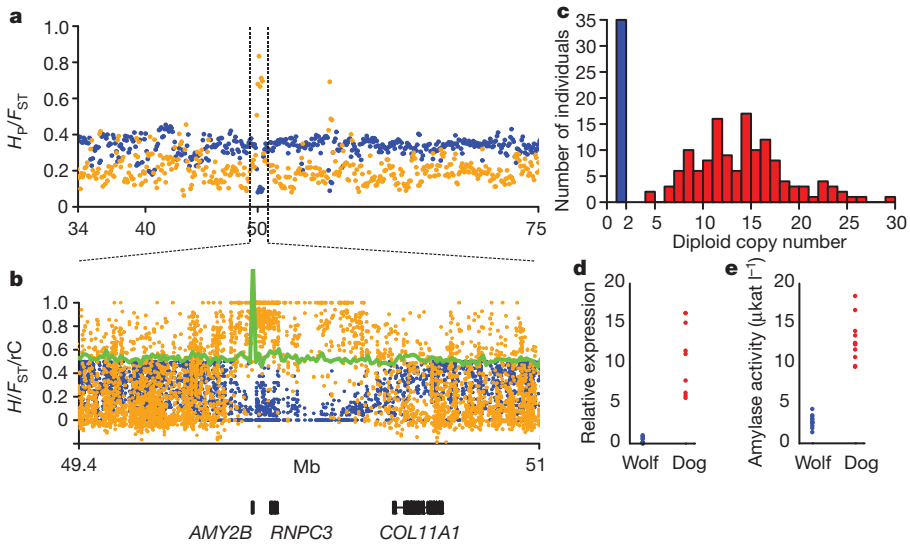

Figure 2 | Selection for increased amylase activity. **a**, Pooled heterozygosity, H_P (blue), and average fixation index, F_{ST} (orange), plotted for 200-kb windows across a chromosome 6 region harbouring *AMY2B*. **b**, Heterozygosity, H (blue), and fixation index, F_{ST} (orange), for single SNPs in the selected region. Dog relative to wolf coverage, rC (green line), indicates increase in *AMY2B* copy number in dog. Genes in the region are shown below panel **b**. **c**, Histogram showing the distribution of diploid amylase copy number in wolf ($n = 35$) (blue) and dog ($n = 136$) (red). **d**, Amylase messenger RNA expression levels in pancreas of wolf ($n = 12$) and dog ($n = 9$). **e**, Amylase activity in serum from wolf ($n = 13$) and dog ($n = 12$).

insulin resistance (*ACSM2A*) that initiates the fatty acid metabolism[19]. A total of 6 CDRs harbour 10 genes with functions related to starch and fat metabolism (Supplementary Table 10). We propose that genetic variants within these genes may have been selected to aid adaptation from a mainly carnivorous diet to a more starch rich diet during dog domestication.

The breakdown of starch in dogs proceeds in three stages: (1) starch is first cleaved to maltose and other oligosaccharides by alpha-amylase in the intestine; (2) the oligosaccharides are subsequently hydrolysed by maltase-glucoamylase[20], sucrase and isomaltase to form glucose; and (3) finally, glucose is transported across the plasma membrane by brush border protein SGLT1[21]. Here we present evidence for selection on all three stages of starch digestion during dog domestication.

Whereas humans have acquired amylase activity in the saliva[22] via an ancient duplication of the pancreatic amylase gene, dogs only express amylase in the pancreas[23]. In dogs the *AMY2B* gene, encoding the alpha-2B-amylase, resides in a 600-kb CDR on chromosome 6 with $Z(H_P)$ and $Z(F_{ST})$ scores of -4.60 and 7.16, respectively (Figs 1 and 2a). Interestingly, an 8-kb sequence spanning the *AMY2B* locus showed a several-fold increase in aligned read depth in dog relative to wolf (Fig. 2b), suggestive of a copy number change. Formal comparisons of regional and local pool coverage, and wolf and dog coverage (Methods), respectively, also suggest a substantial increase in copy numbers in all dog pools compared to wolf at this locus (Supplementary Discussion, section 5).

We confirmed this CNV by quantifying *AMY2B* copy numbers in 136 dogs and 35 wolves (Supplementary Table 11) using real-time quantitative PCR (qPCR). Whereas all wolves tested carried only 2 copies ($2N = 2$), diploid copy numbers in dog ranged from 4 to 30 ($P < 0.001$, Wilcoxon) (Fig. 2c), corresponding to a remarkable 7.4-fold average increase in dog *AMY2B* copy numbers. To assess whether this change correspond to a difference in amylase activity, we first compared *AMY2B* gene expression in pancreas from dog ($n = 9$) and wolf ($n = 12$) and noted a 28-fold higher average expression in dog ($P < 0.001$, Wilcoxon, Fig. 2d). We then quantified amylase activity in frozen serum (Fig. 2e) and found a 4.7-fold higher activity in dog (9.6–18.4 μkat l^{-1} ($n = 12$)) relative to wolf (1.4–4.3 μkat l^{-1} ($n = 13$)) ($P < 0.001$, Wilcoxon). Similar results were obtained in comparisons of a limited number of fresh samples (Supplementary Tables 12 and 13). The change in *AMY2B* gene copy number together with a correlated increase in both expression level (rho = 0.84, $P < 0.0001$, Spearman) (Supplementary Fig. 3) and enzyme activity (rho = 0.63, $P < 0.01$, Spearman) (Supplementary Fig. 4) indicates that duplications of the alpha-amylase locus conferred a selective advantage to early dogs by causing an increase in amylase activity.

Maltase-glucoamylase is responsible for the second step in the breakdown of starch, catalysing the hydrolysis of maltose to glucose[20]. No copy number changes were observed in the *MGAM* locus so we decided to study haplotype diversity across the region to facilitate the identification of causal variants. We genotyped 47 randomly selected SNPs in 71 dogs representing 38 diverse breeds and 19 wolves of worldwide distribution (referred to as 'the reference panel', Supplementary Table 14). Sixty-eight of the seventy-one dogs tested carried at least one copy of a 124-kb long haplotype spanning the entire *MGAM* and a small neighbouring locus encoding the bitter taste mediating taste receptor 2 member 38 (TASR38) (Fig. 3a–c). Whereas none of the wolves carried the selected haplotype, 55 dogs were homozygous for it, 13 were heterozygous and only three dogs lacked it (2 West Highland White Terriers and 1 Chinese Crested Dog). This high degree of haplotype differentiation between dog and wolf (average F_{ST} for genotyped SNPs = 0.75) indicates that this haplotype may harbour genetic variation of selective advantage to dogs (Supplementary Discussion, sections 3 and 6).

We identified several candidate mutations within *MGAM* that may have been targeted by selection in this region (Supplementary Table 15). First a conservative amino acid substitution located in the duplicated trefoil domain of MGAM (residue 1001) is nearly fixed for isoleucine in wolf and for valine in dogs. Eleven out of fourteen mammals have valine at this position, whereas the omnivorous rat, and the insectivorous hedgehog and short-tailed opossum, carry isoleucine like the wolf (Supplementary Table 16). Second, another conservative substitution, methionine to valine, located in the beta sheet of the maltase enzyme (residue 797), is segregating in wolf but fixed for methionine in dog. The insectivorous hedgehog and common shrew are the only mammals without methionine at this evolutionarily conserved position (Supplementary Table 17) and *in silico* modelling using the SDM-server indicates that a change from methionine to valine at this residue is destabilizing[24]. Third, a fixed two-base-pair deletion in dog disrupts the stop codon, thereby extending the carboxy-terminal end of dog

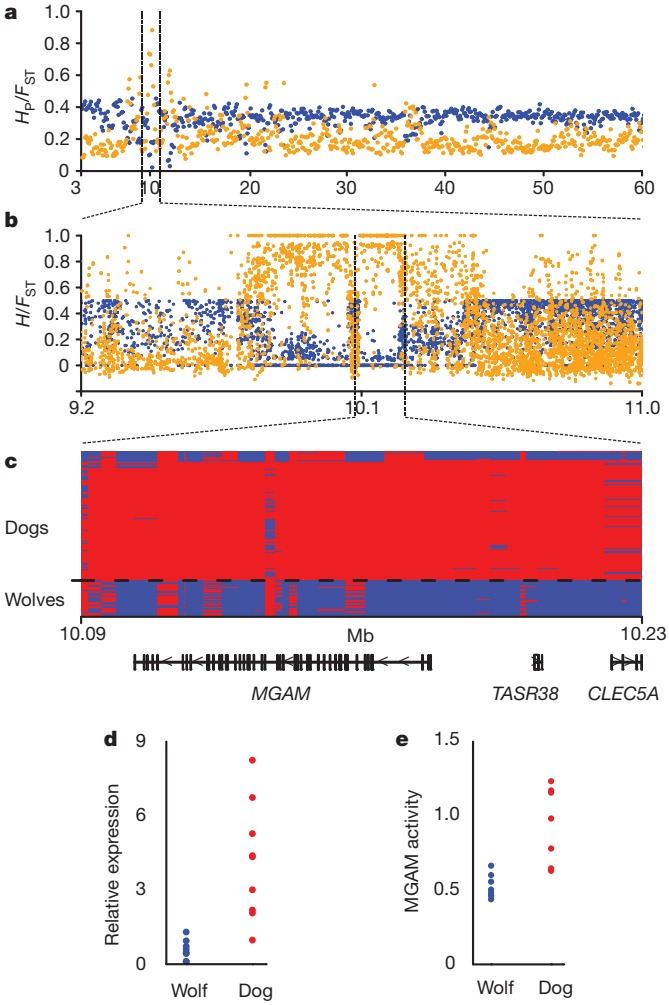

Figure 3 | Selection is associated with increased maltase activity. **a**, Pooled heterozygosity, H_P (blue), and average fixation index, F_{ST} (orange), plotted for 200-kb windows across a chromosome 16 region harbouring *MGAM*. **b**, Heterozygosity, H (blue), and fixation index, F_{ST} (orange), for single SNPs in the selected region. **c**, Haplotypes inferred from genotyping of 47 SNPs across the *MGAM* locus in 71 dogs and 19 wolves (red and blue colour are major and minor dog allele, respectively). Genes in the genotyped region are shown below panel **c**. **d**, *MGAM* mRNA expression levels in pancreas of wolf ($n = 8$) and dog ($n = 9$). **e**, MGAM activity in serum from wolf ($n = 8$) and dog ($n = 7$).

MGAM by two amino acids: asparagine and phenylalanine. In 32 mammals studied only herbivores (rabbit, pika, alpaca and cow) and omnivores (mouse lemur and rat) share an extension like that seen in dog (Supplementary Table 18). A fourth candidate mutation in intron 37 affects a predicted binding site for the glucose metabolism regulator NR4A2 protein[25] by shifting the wolf sequence away from the canonical NR4A2-binding motif. Three out of four mammals with the wolf allele at this site rely heavily on insects or fish for their nutritional requirements (Supplementary Table 19).

To decipher whether the candidate mutations act primarily on expression or protein activity we examined *MGAM* expression in pancreas and the resulting enzymatic activity in serum. Dogs showed a ~12-fold higher expression ($P < 0.001$, Wilcoxon, $n_{DOG} = 9, n_{WOLF} = 8$) (Fig. 3d) and a ~twofold increase in maltose to glucose turnover compared to wolves (average glucose produced in dogs: 0.94 $\Delta A_{570\,nm}$ (0.64–1.23, $n = 7$) and wolves: 0.52 $\Delta A_{570\,nm}$ (0.44–0.66, $n = 8$), $P = 0.0012$, Wilcoxon) (Fig. 3e). Although we cannot rule out that diet-induced plasticity contributed to this difference[26], our results indicate that the mutation affecting a NR4A2-binding site or another unknown variant probably affect the expression of *MGAM*. Selection may thus clearly have led to increased *MGAM* expression, but we cannot rule out that the strong selection affecting this locus may have favoured the accumulation of protein-coding changes on the same haplotype. Similar scenarios have been seen for white coat colour in dogs and pigs, where repeated selection for additional mutations has resulted in an allelic series of white spotting at the *MITF* and *KIT* loci, respectively[27].

Once starch has been digested to glucose it is absorbed through the luminal plasma membrane of the small intestine by the sodium/glucose cotransporter 1 (SGLT1)[21]. To benefit from an increased capacity to digest starch, dogs would therefore be expected to show a parallel increase in glucose uptake. A CDR on chromosome 26 (Supplementary Fig. 5a, b) encompasses *SGLT1* and a gene (*SGLT3*) encoding the glucose-sensing sodium/glucose cotransporter 3 protein[28]. To characterize the haplotype diversity we genotyped 48 randomly chosen SNPs across this CDR in the reference panel and identified a 50.5-kb region, spanning the 3′ section of *SGLT1* as well as the 3′ end of *SGLT3*, that is highly divergent between dog and wolf (Supplementary Fig. 5c). In this region all dogs tested were carriers of a particular haplotype, for which 63 were homozygous and eight heterozygous. This contrasts to 19 wolves where a single individual carried one copy of the haplotype. Based on the high haplotype differentiation (average F_{ST} for 18 SNPs in 50.5-kb haplotype = 0.81) it is likely that *SGLT1* and its 3′ region represents an additional dog domestication locus.

The 50.5-kb region includes a conservative isoleucine to valine substitution in SGLT1 (residue 244) that affects a loop facing the extracellular side of the luminal membrane (Supplementary Table 15). Heterologous expression analysis[29] shows that glycosylation at a nearby site (residue 248) affects glucose transport, indicating that it is possible that dogs acquired improved glucose uptake as a result of the observed substitution. In addition, we see only non-significant differences in *SGLT1* expression in pancreas of dog ($n = 9$) and wolf ($n = 4$) ($P = 0.39$, Wilcoxon) (Supplementary Fig. 6), indicating that selection primarily targeted a structural rather than regulatory mutation in *SGLT1*.

In conclusion, we have presented evidence that dog domestication was accompanied by selection at three genes with key roles in starch digestion: *AMY2B*, *MGAM* and *SGLT1*. Our results show that adaptations that allowed the early ancestors of modern dogs to thrive on a diet rich in starch, relative to the carnivorous diet of wolves, constituted a crucial step in early dog domestication. This may suggest that a change of ecological niche could have been the driving force behind the domestication process, and that scavenging in waste dumps near the increasingly common human settlements during the dawn of the agricultural revolution may have constituted this new niche[6]. In light of previous results describing the timing and location of dog domestication, our findings may suggest that the development of agriculture catalysed the domestication of dogs.

The results presented here demonstrate a striking case of parallel evolution whereby the benefits of coping with an increasingly starch-rich diet during the agricultural revolution caused similar adaptive responses in dog and human[30]. This emphasizes how insights from dog domestication may benefit our understanding of human recent evolution and disease. Finally, by understanding the genetic basis of adaptive traits in dogs we have come closer to unlocking the potential in dog and wolf comparisons to decipher the genetics of behaviour.

METHODS SUMMARY

Sequencing. We pooled genomic DNA from 12 individuals before mate-pair library construction and sequencing on the AB SOLiD system, version 3, according to standard manufacturer protocols. Sequencing reads were aligned to the CanFam 2.0 reference sequence using the Bioscope 1.1 software.

Selection analyses. We identified variable sites in data combined from all pools and required a minimum of three reads supporting an alternative allele to call a SNP. We used allele counts at variable sites to identify signals of selection in 200-kb windows using two approaches: for each window we calculated (1) the average pooled heterozygosity, H_P (ref. 10), and (2) the average fixation index, F_{ST}, between dog and wolf. Putatively selected regions were located by extracting

windows from the extreme tails of the Z-transformed H_P and F_{ST} distributions by applying a threshold of 5 standard deviations.

Functional assays. We used multiplex TaqMan assays and SYBR Green real-time PCR to quantify CNVs and gene expression, respectively. Serum amylase activity was analysed using an Architect e400 instrument and serum maltase activity was quantified based on the amount of maltose to glucose turnover.

Full Methods and any associated references are available in the online version of the paper.

Received 1 July; accepted 11 December 2012.
Published online 23 January 2013.

1. Ovodov, N. D. *et al.* A 33,000-year-old incipient dog from the Altai mountains of Siberia: evidence of the earliest domestication disrupted by the last glacial maximum. *PLoS ONE* **6,** e22821 (2011).
2. Davis, S. J. M. & Valla, F. R. Evidence for domestication of the dog 12,000 years ago in the Natufian of Israel. *Nature* **276,** 608–610 (1978).
3. Skoglund, P., Gotherstrom, A. & Jakobsson, M. Estimation of population divergence times from non-overlapping genomic sequences: examples from dogs and wolves. *Mol. Biol. Evol.* **28,** 1505–1517 (2011).
4. Pang, J. F. *et al.* mtDNA data indicate a single origin for dogs south of Yangtze River, less than 16,300 years ago, from numerous wolves. *Mol. Biol. Evol.* **26,** 2849–2864 (2009).
5. vonHoldt, B. M. *et al.* Genome-wide SNP and haplotype analyses reveal a rich history underlying dog domestication. *Nature* **464,** 898–902 (2010).
6. Coppinger, R. & Coppinger, L. *Dogs: a Startling New Understanding of Canine Origin, Behaviour and Evolution* (Scribner, 2001).
7. Hare, B., Wobber, V. & Wrangham, R. The self-domestication hypothesis: evolution of bonobo psychology is due to selection against aggression. *Anim. Behav.* **83,** 573–585 (2012).
8. Belyaev, D. K. Destabilizing selection as a factor in domestication. *J. Hered.* **70,** 301–308 (1979).
9. Fang, M., Larson, G., Ribeiro, H. S., Li, N. & Andersson, L. Contrasting mode of evolution at a coat color locus in wild and domestic pigs. *PLoS Genet.* **5,** e1000341 (2009).
10. Rubin, C. J. *et al.* Whole-genome resequencing reveals loci under selection during chicken domestication. *Nature* **464,** 587–591 (2010).
11. Lindblad-Toh, K. *et al.* Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature* **438,** 803–819 (2005).
12. Koike, N. *et al.* Brorin, a novel secreted bone morphogenetic protein antagonist, promotes neurogenesis in mouse neural precursor cells. *J. Biol. Chem.* **282,** 15843–15850 (2007).
13. Cheng, L. *et al. Tlx3* and *Tlx1* are post-mitotic selector genes determining glutamatergic over GABAergic cell fates. *Nature Neurosci.* **7,** 510–517 (2004).
14. Napoli, I. *et al.* The fragile X syndrome protein represses activity-dependent translation through CYFIP1, a new 4E-BP. *Cell* **134,** 1042–1054 (2008).
15. Weston, M. C., Nehring, R. B., Wojcik, S. M. & Rosenmund, C. Interplay between VGLUT isoforms and endophilin A1 regulates neurotransmitter release and short-term plasticity. *Neuron* **69,** 1147–1159 (2011).
16. Varga, Z. M. *et al.* Zebrafish smoothened functions in ventral neural tube specification and axon tract formation. *Development* **128,** 3497–3509 (2001).
17. Tokuhiro, K., Ikawa, M., Benham, A. M. & Okabe, M. Protein disulfide isomerase homolog PDILT is required for quality control of sperm membrane protein ADAM3 and male fertility. *Proc. Natl Acad. Sci. USA* **109,** 3850–3855 (2012).
18. Gardner, A. J. & Evans, J. P. Mammalian membrane block to polyspermy: new insights into how mammalian eggs prevent fertilisation by multiple sperm. *Reprod. Fertil. Dev.* **18,** 53–61 (2006).
19. Boomgaarden, I., Vock, C., Klapper, M. & Doring, F. Comparative analyses of disease risk genes belonging to the acyl-CoA synthetase medium-chain (ACSM) family in human liver and cell lines. *Biochem. Genet.* **47,** 739–748 (2009).
20. Nichols, B. L. *et al.* The maltase-glucoamylase gene: common ancestry to sucrase-isomaltase with complementary starch digestion activities. *Proc. Natl Acad. Sci. USA* **100,** 1432–1437 (2003).
21. Wright, E. M., Loo, D. D. F. & Hirayama, B. A. Biology of human sodium glucose transporters. *Physiol. Rev.* **91,** 733–794 (2011).
22. Meisler, M. H. & Ting, C. N. The remarkable evolutionary history of the human amylase genes. *Crit. Rev. Oral Biol. Med.* **4,** 503–509 (1993).
23. Simpson, J. W., Doxey, D. L. & Brown, R. Serum isoamylase values in normal dogs and dogs with exocrine pancreatic insufficiency. *Vet. Res. Commun.* **8,** 303–308 (1984).
24. Worth, C. L., Preissner, R. & Blundell, T. L. SDM-a server for predicting effects of mutations on protein stability and malfunction. *Nucleic Acids Res.* **39,** W215–W222 (2011).
25. Pei, L. *et al.* NR4A orphan nuclear receptors are transcriptional regulators of hepatic glucose metabolism. *Nature Med.* **12,** 1048–1055 (2006).
26. Mochizuki, K., Honma, K., Shimada, M. & Goda, T. The regulation of jejunal induction of the maltase-glucoamylase gene by a high-starch/low-fat diet in mice. *Mol. Nutr. Food Res.* **54,** 1445–1451 (2010).
27. Andersson, L. Studying phenotypic evolution in domestic animals: a walk in the footsteps of Charles Darwin. *Cold Spring Harb. Symp. Quant. Biol.* **74,** 319–325 (2009).
28. Diez-Sampedro, A. *et al.* A glucose sensor hiding in a family of transporters. *Proc. Natl Acad. Sci. USA* **100,** 11753–11758 (2003).
29. Hediger, M. A., Mendlein, J., Lee, H. S. & Wright, E. M. Biosynthesis of the cloned intestinal Na^+ glucose cotransporter. *Biochim. Biophys. Acta* **1064,** 360–364 (1991).
30. Perry, G. H. *et al.* Diet and the evolution of human amylase gene copy number variation. *Nature Genet.* **39,** 1256–1260 (2007).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank Järvzoo, Nordens ark and the Canine Biobank at Uppsala University and the Swedish University of Agricultural Sciences for providing samples, Uppsala Genomics Platform at SciLifeLab Uppsala for generating the resequencing data, the UPPNEX platform for assisting with computational infrastructure for data analysis and the Broad Institute Genomics Platform for validation genotyping. The project was funded by the SSF, the Swedish Research Council, the Swedish Research Council Formas, Uppsala University and a EURYI to K.L.-T. funded by the ESF supporting also E.A.; K.M. was funded by the Higher Education Commission, Pakistan.

Author Contributions K.L.-T. and Å.H. designed the study. K.L.-T. and E.A. oversaw the study. M.-L.A. coordinated and performed the majority of the sample collecting and O.L. and J.M.A. provided samples of critical importance. E.A. performed the SNP detection and selection analyses; A.R. identified candidate causative mutations and analysed haplotypes in CDRs; K.M. detected CNVs bioinformatically; M.T.W. performed phylogenetic analysis and analysed the Canine HD-array data; A.R. performed the maltase activity assay; M.-L.A. validated CNVs and quantified mRNA expression of candidate genes; M.P. performed validation SNP genotyping; E.A., A.R., M.-L.A. and K.L-T. interpreted the data; E.A. and K.L.-T. wrote the paper with input from the other authors.

Author Information Sequence reads are available under the accession number SRA061854 (NCBI Sequence Read Archive). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to E.A. (Erik.Axelsson@imbim.uu.se) and K.L-T. (kersli@broadinstitute.org).

METHODS

DNA extraction. DNA was extracted from tissue using Qiagen tissue DNA extraction kits or from EDTA blood using either manual salt precipitation or the QIASymphony DNA Midi kit (Qiagen) on the QIASymphony robot (Qiagen).

Sequencing. We pooled DNA from 12 individuals per pool before mate-pair library construction and sequencing on the AB SOLiD system, version 3, according to standard manufacturer protocols (Applied Biosystems). Sequencing reads were aligned to the CanFam 2.0 reference sequence using the Bioscope 1.1 software. We removed duplicated (http://picard.sourceforge.net) and poorly mapped reads (mapping quality <20 in Samtools)[31] and retained only uniquely mapped reads for further analyses.

SNP detection. We searched for variable sites in data combined from all pools (including wolf) to increase sensitivity to rare alleles. We required a minimum of three reads supporting an alternative allele to call a SNP, and applied a further filtering step implemented in samtools.pl varFilter (settings: –Q25 –q10 –d3 –D120 –G25 –w10 –N2 –l30) to ensure a high call accuracy that is largely unaffected by, for example, paralogous sequence variants. We called genotypes for all SNPs in all dog pools and the single wolf pool by counting sequencing reads supporting the reference and variant allele, given a minimum base quality of 20, to estimate allele frequencies in the dog and wolf populations. A random selection representing 25% of the sequencing reads from pools 4 and 5 were included in this process to achieve unbiased allele frequency estimates.

Selection analyses. Allele counts and allele frequencies at all identified variable sites were used to search the dog genome for regions that may have been affected by selection during the early phase of dog domestication using two complementary approaches. First we calculated the average pooled heterozygosity (H_P) in 200-kb windows sliding 100 kb at a time, for all five dog pools combined, and in the single wolf pool separately, following the methodology described in ref. 10. Briefly, this method sums all minor and major allele counts, respectively, at all variable sites within a window, and estimates the heterozygosity based on the combined allele counts for the entire window. The advantage of this method over calculating a simple arithmetic mean of all single-site heterozygosity estimates is that it accounts for variable sequence coverage across the window. To avoid spurious selection signals we discarded 49 out of 21,927 windows containing fewer than 10 informative sites from both this and the subsequent F_{ST} analysis. We Z-transformed the resultant distribution of H_P scores and extracted putatively selected windows in the extreme tail of the distribution by applying a $Z(H_P) < -5$ cut-off.

Second we calculated F_{ST} values between dog and wolf for individual SNPs using a method that adjusts for sample size differences[32]. We averaged F_{ST} values across 200-kb windows, sliding 100 kb at a time and Z-transformed the resultant distribution. Putative selection targets were extracted from the extreme tail of the distribution by applying a $Z(F_{ST}) > 5$ cut-off, and attributed to selection in dog if the corresponding $Z(H_P)_{DOG} < Z(H_P)_{WOLF}$, to selection in wolf if $Z(H_P)_{WOLF} < Z(H_P)_{DOG}$ (three regions), and to selection in both taxa if $Z(H_P)_{WOLF} < -4$ and $Z(H_P)_{WOLF} < -4$.

Gene ontology analysis. We used the Ensembl gene annotations to identify genes residing within regions extending 100 kb up- and downstream of CDRs to include potential effects of regulatory changes on loci at some distance, and to reduce the risk of excluding the outermost portions of the selected haplotypes by using sliding windows of fixed size. We tested for enrichment of gene ontology terms (GOa-human) assigned to the subset of these CDR genes for which human orthology could be established (79 out of 122) using the GOstat program[33].

Genotyping validation. We designed an iPLEX assay targeting 124 SNPs located in CDRs showing a high degree of homozygosity or population differentiation. A total of 71 dogs, representing 38 different breeds, and 19 wolves (Supplementary Table 14) were genotyped using standard protocols provided by the manufacturer (Sequenome). Haplotypes were phased using fastPHASE[34].

qPCR CNV detection. We quantified DNA copy number variation using Multiplex TaqMan assays containing primers and probes (Supplementary Table 20) matching both the target and reference sequence (housekeeping gene *C7orf28b*) according to the manufacturer's protocol. All reactions were run in triplicate and data was analysed using the CopyCaller software (Applied Biosystems). Copy numbers for each target were normalized to the same wolf to account for inter-plate variability.

qPCR expression analyses. Pancreatic tissue samples from dogs and wolves where collected post mortem, stored in RNAlater at 4 °C for 24 h and subsequently freeze-stored at −80 °C. We used TRIzol to isolate RNA from these samples, followed by complementary DNA synthesis using the Advantage RT for PCR kit according to the manufacturers' protocols (Life Technologies and Clontech, respectively). We designed exonic primers (Supplementary Table 21) and quantified the amount of cDNA using SYBR Green real-time PCR (Applied biosciences) on a 7900HT Fast real time PCR system (Applied Biosystems) and analysed the data using the qbasePLUS (Biogazelle) software according to the $\Delta\Delta C_T$ method. All reactions where run in triplicate and normalized by comparisons to housekeeping genes *RPL32* and *RPL13A*.

Amylase activity. Peripheral EDTA and serum blood samples where collected from dogs and both captive and free-ranging wolves. Serum amylase activity was analysed at the Clinical Pathology service (Swedish Agricultural University) using an Architect e400 instrument (Abott Laboratories), except for 8 serum samples (Supplementary Table 13) which were run on a VetScan instrument (Abaxis).

Maltase activity. Maltase activity was assayed according to the principle outlined in ref. 35, whereby a known amount of maltose substrate is added to serum and the resultant glucose produced is measured as the change in absorbance after five minutes ($\Delta A_{570\,nm}$). We used reagents from the ab83388 Maltose assay kit (Abcam) and serum sampled as described above. For each individual, glucose residuals were measured in duplicate and maltase assays were performed in triplicate.

Indel calling. We used Bioscope 1.1 to call small insertions and deletions in each pool separately. We then combined the results of all pools and extracted a set of high confident indels by requiring that indels were supported by at least three sequencing reads.

CNV detection. Four methods were used to detect structural variation in the dog genome. We searched for deviations in insert size using the large indel tool implemented in Bioscope1.1. We compared the coverage depth between the pooled samples using CNVseq[36] and the Fixed deletions method[10] and finally identified regions in which the coverage depth deviated from the pool average using CNVnator[37]. Methods relying on comparisons of sequence coverage between pools always used the wolf as reference pool.

Ethics. All animals contributing tissue samples to this study died for other reasons than participating in this study. All dog samples were taken with the owners consent. The sampling conformed to the decision of the Swedish Animal Ethical Committee (no. C62/10) and the Swedish Animal Welfare Agency (no.31-1711/10).

31. Li, H. *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics* **25,** 2078–2079 (2009).
32. Weir, B. S. & Cockerham, C. C. Estimating *F*-statistics for the analysis of population-structure. *Evolution* **38,** 1358–1370 (1984).
33. Beissbarth, T. & Speed, T. P. GOstat: find statistically overrepresented Gene Ontologies within a group of genes. *Bioinformatics* **20,** 1464–1465 (2004).
34. Scheet, P. & Stephens, M. A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. *Am. J. Hum. Genet.* **78,** 629–644 (2006).
35. Dahlqvist, A. Method for assay of intestinal disaccharidases. *Anal. Biochem.* **7,** 18–25 (1964).
36. Xie, C. & Tammi, M. T. CNV-seq, a new method to detect copy number variation using high-throughput sequencing. *BMC Bioinformatics* **10,** 80 (2009).
37. 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 (2011).