

## Diversifying selection between pure-breed and free-breeding dogs inferred from genome-wide SNP analysis

Małgorzata Pilot,<sup>\*,†</sup> Tadeusz Malewski,<sup>†</sup> Andre E. Moura,<sup>\*</sup> Tomasz Grzybowski,<sup>‡</sup> Kamil Oleński,<sup>§</sup> Stanisław Kamiński,<sup>§</sup> Fernanda Ruiz Fadel,<sup>\*</sup> Abdulaziz N. Alagaili,<sup>\*\*</sup> Osama B. Mohammed,<sup>\*\*</sup> and Wiesław Bogdanowicz<sup>†,1</sup>

<sup>\*</sup>School of Life Sciences, University of Lincoln, Green Lane, LN6 7DL Lincoln, UK.

<sup>†</sup>Museum and Institute of Zoology, Polish Academy of Sciences, Wilcza 64, 00-679 Warszawa, Poland.

<sup>‡</sup>Division of Molecular and Forensic Genetics, Department of Forensic Medicine, Ludwik Rydygier Collegium Medicum, Nicolaus Copernicus University, Skłodowskiej-Curie 9, 85-094 Bydgoszcz, Poland.

<sup>§</sup>Department of Animal Genetics, University of Warmia and Mazury, Oczapowskiego 5, 10-711 Olsztyn, Poland.

<sup>\*\*</sup>KSU Mammals Research Chair, Department of Zoology, College of Science, King Saud University, P.O. Box 2455, Riyadh 11451, Saudi Arabia.

<sup>1</sup>Corresponding author: Wiesław Bogdanowicz, Museum and Institute of Zoology, Polish Academy of Sciences, Wilcza 64, 00-679 Warszawa, Poland. e-mail: [wieslawb@miiz.waw.pl](mailto:wieslawb@miiz.waw.pl)

## **Abstract**

Domesticated species are often composed of distinct populations differing in the character and strength of artificial and natural selection pressures, providing a valuable model to study adaptation. In contrast to pure-breed dogs that constitute artificially maintained inbred lines, free-ranging dogs are typically free-breeding, i.e. unrestrained in mate choice. Many traits in free-breeding dogs (FBDs) may be under similar natural and sexual selection conditions to wild canids, while relaxation of sexual selection is expected in pure-breed dogs. We used a Bayesian approach with strict false-positive control criteria to identify  $F_{ST}$ -outlier SNPs between FBDs and either European or East Asian breeds, based on 167,989 autosomal SNPs. By identifying outlier SNPs located within coding genes, we found four candidate genes under diversifying selection shared by these two comparisons. Three of them are associated with the Hedgehog (HH) signalling pathway regulating vertebrate morphogenesis. A comparison between FBDs and East Asian breeds also revealed diversifying selection on BBS6 gene, which was earlier shown to cause snout shortening and dental crowding via disrupted HH signalling. Our results suggest that relaxation of natural and sexual selection in pure-breed dogs as opposed to FBDs could have led to mild changes in regulation of the HH signalling pathway. HH inhibits adhesion and migration of neural crest cells from neural tube, and minor deficits of these cells during embryonic development have been proposed as the underlying cause of “domestication syndrome”. This suggests that the process of breed formation involved the same genetic and developmental pathways as the process of domestication.

**Keywords:** Artificial selection, *Canis lupus familiaris*, Diversifying selection, Domestication syndrome, Hedgehog signalling pathway

**Running title:** Selection in pure-breed and non-breed dogs

## Introduction

Large phenotypic differentiation between domesticated species and their wild conspecifics provides a unique opportunity to study the process of selection at an intra-specific level.

Numerous studies were dedicated to the search of “domestication genes”, i.e. genes that have been under diversifying selection at the initial stages of the domestication process (e.g. Rubin *et al.* 2010, 2012, Axelsson *et al.* 2013, Li *et al.* 2013a,b, Olsen and Wendel 2013, Carneiro *et al.* 2014, Montague *et al.* 2014, Ramirez *et al.* 2014, Freedman *et al.* 2016, Wang *et al.* 2016). In such studies, a domesticated species is typically assumed to constitute a relatively uniform genetic population as compared to its wild ancestor, which may be justified by an assumption that all representatives of the domesticated species should share a common genetic signature of domestication.

However, domesticated species are often composed of distinct populations, differing in the character and strength of artificial and natural selection pressures. Most domesticates consist of multiple pure-breed forms that constitute close populations artificially selected for a particular set of traits, but some species also include feral or semi-feral populations that are unrestrained in mate choice. The comparison between populations of a domestic species that considerably differ in artificial, natural and sexual selection pressures provides a valuable model to study adaptation.

The domestic dog (*Canis lupus familiaris*) constitutes a particularly useful model for this purpose. Although the time of dog domestication is still under debate (e.g. Freedman *et al.* 2014, Duleba *et al.* 2015, Skoglund *et al.* 2015, Wang *et al.* 2016), it is generally agreed that the dog was the first species to be domesticated (e.g. see Larson and Fuller 2014). This implies that dogs have been separated from their wild ancestor for considerably longer time than any other species, and therefore distinct populations experiencing different levels of artificial selection could have existed for many generations. Dog breeds exhibit very high

level of morphological differentiation, exceeding that seen among all wild canids (Drake and Klingenberg 2010). Furthermore, dogs have numerous free-ranging and free-breeding populations, which represent a considerably broader range of genetic diversity as compared with pure-breed dogs (Pilot *et al.* 2015, Shannon *et al.* 2015). Although in some regions like the Neotropics, South Pacific and parts of Africa, native dog populations have been mostly replaced by dogs of European origin (Shannon *et al.* 2015), in Europe and continental Asia the majority of free-breeding dogs (FBDs; a term introduced in Boyko and Boyko 2014) constitute distinct genetic units rather than being an admixture of breeds (Pilot *et al.* 2015, Shannon *et al.* 2015). This implies that FBD populations in mainland Eurasia were free from artificial breeding constraints throughout multiple generations, while experiencing sexual selection pressures (resulting from free mate choice) similar to populations of wild canids.

Free-ranging dogs represent a broad spectrum of ecological conditions, from truly wild populations such as the Australian dingo to dogs that are mostly unrestrained in their ranging behaviour, but rely on humans for subsistence (Gompper *et al.* 2014). FBD populations are therefore expected to experience natural selection on traits that are important to survival outside the domestic environment (e.g. traits related to pathogen resistance, hunting skills or to social interactions with conspecifics), although the strength of this selection may vary between populations. Natural and sexual selection likely accounts for considerably smaller morphological variation between FBD populations as compared to morphological variation between dog breeds (Coppinger and Coppinger 2016).

Ongoing gene flow from pure-breed populations into FBDs may counteract natural selection in FBDs to some extent, but it does not necessarily prevent evolutionary diversification between these two groups (e.g. see Smadja and Butlin 2011). Indeed, previous studies have found evidence of diversifying selection at immune system genes (Li *et al.* 2013b) and segregating olfactory receptor pseudogenes (Chen *et al.* 2012). The capability for

fast adaptation in response to environmental pressures has also been demonstrated in native dogs of the Tibetan Plateau (Tibetan Mastiffs and Diqing indigenous dogs), which showed signals of positive selection at several genes involved in response to high-altitude hypoxia (Gou *et al.* 2014, Li *et al.* 2014). Marsden *et al.* (2016) have shown that pure-breed dogs have higher levels of deleterious genetic variation genome-wide than grey wolves, with FBDs displaying intermediate values. This implies relaxation of natural selection in dogs as compared to their wild ancestors, but also stronger natural selection in FBDs as compared with pure-breed dogs.

The majority of breeds registered by kennel clubs have European origin and show close genetic similarity, as reflected in poorly-resolved phylogenies based on microsatellite loci (Parker *et al.* 2004) and genome-wide SNPs (vonHoldt *et al.* 2010, Larson *et al.* 2012, Pilot *et al.* 2015). However, several breeds of non-European origin branch from basal nodes in the pure-breed dog phylogeny, suggesting their distinct origin (Parker *et al.* 2004, vonHoldt *et al.* 2010, Larson *et al.* 2012, Pilot *et al.* 2015). This group includes East Asian and Arctic spitz-type breeds, which were shown to have a common origin in East Asia (Brown *et al.* 2013, 2015, van Asch *et al.* 2013). There is a strong evidence for the genetic distinctiveness of East Asian and Arctic breeds from modern European breeds, including the phylogeny based on 186 canid whole-genome sequences (Decker *et al.* 2015). The East Asian and Arctic breeds show close genetic similarity to East Asian FBDs, while modern European breeds show closest similarity to European FBDs (Decker *et al.* 2015, Pilot *et al.* 2015). This suggests that these two groups of breeds represent two distinct episodes of breed formation from ancestral FBD populations in East Asia and Europe, respectively.

Here we analysed signatures of selection between FBDs and each of these two pure-breed groups. All breeds, irrespective of their origin, are characterised by the lack of free mate choice, which is expected to result in relaxation of sexual selection compared with FBDs. In

contrast, many traits in FBDs may be under natural and sexual selection conditions similar to wild canids. Therefore, we hypothesised that the analysis of diversifying selection between FBDs and each of the two breed groups will identify shared candidate genes with functions relevant to reproductive success and survival in a non-domestic environment. Our results are consistent with this hypothesis, but we also found that the shared candidate genes have pleiotropic phenotypic effects and are involved in the same genetic and developmental pathways.

## **Materials and Methods**

### **Datasets**

We used a SNP genotype dataset of 234 free-breeding domestic dogs from 14 sites across Eurasia (Supplementary Table 1), available from our earlier study (Pilot *et al.* 2015). These samples were genotyped with the CanineHD BeadChip (Illumina) at 167,989 autosomal SNPs and 5,660 X chromosome SNPs. The presence of close relatives was assessed using the software Cervus (Marshall *et al.* 1998) and Kingroup (Konovalov *et al.* 2004), and all but one individual from each kin group was eliminated from the dataset, which resulted in a sample of 200 individuals. Further pruning for individuals with over 10% of missing data gave a final sample size of 190 individuals.

We also used two datasets of SNP genotypes of pure-breed dogs. The first dataset (“UK dataset”) consisted of 88 pure-breed dogs collected from across the United Kingdom, available from our earlier study (Pilot *et al.* 2015). These dogs represented 30 breeds, with 1–9 individuals per breed (Supplementary Table 2). The second dataset derived from the LUPA project (Vaysse *et al.* 2011) and contained 446 pure-breed dogs representing 30 different breeds, with 10–26 individuals per breed (Supplementary Table 2). All three datasets were

generated using CanineHD BeadChip (Illumina), and therefore could be merged without reduction of the usable SNP set.

From the combined dataset of pure-breed dogs, we selected the East Asian and Arctic breeds that branch from basal nodes in pure-breed dogs phylogenies (Parker *et al.* 2004, vonHoldt *et al.* 2010, Larson *et al.* 2012, Pilot *et al.* 2015). This included Shar Pei, Shiba Inu, Siberian Husky, Alaskan Malamute and Greenland Sledge Dog - they will be referred to as East Asian breeds henceforth. All breeds of European origin except for spitz-type breeds (which were excluded due to their possible relatedness to Asian spitz-type breeds) were selected to represent modern European breeds, resulting in a dataset of 40 different breeds (Supplementary Table 2). Pruning for individuals with over 10% of missing data gave the final sample sizes of 29 individuals for Asian breeds and 356 for European breeds.

### **Control for “batch effect”**

Each of the three datasets described above was generated independently, and this could have potentially led to a “batch effect”, i.e. incompatibilities between genotypes from different datasets. Such incompatibilities result from strand flips, and therefore we used the “TOP/BOT” genotype calling method that was specifically designed to ensure that different datasets are reported in a uniform way in terms of strand designation and orientation (Illumina, Inc. 2006). This method calls strands as top (TOP) and bottom (BOT) based on the polymorphism itself, or in ambiguous cases based on the surrounding sequence. Briefly, in unambiguous cases of A/C or A/G SNPs, adenine (A) is designated as Allele A on TOP strand, with cytosine (C) or guanine (G) being allele B on TOP strand. Thymine (T) being complementary to adenine is designated as Allele A on BOT strand. These rules cannot be applied for A/T or C/G SNPs, and in such case an algorithm is applied to designate the strand

(TOP/BOT) and allele (Allele A or B) based on DNA sequence surrounding the SNP (for details, see Illumina, Inc. 2006).

The publicly available LUPA dataset (Vaysse *et al.* 2011) was called using the "TOP/BOT" method, and we used the same method for the two other datasets. The LUPA and UK datasets shared 9 dog breeds (Supplementary Table 2), and individuals representing the same breed, independent of whether they originated from the UK or LUPA dataset, clustered together in an individual-based dog phylogeny (Pilot *et al.* 2015), which testified that these datasets were correctly merged. For all these reasons, it is very unlikely that the merged dataset contained any strand flips, and that any outlier SNPs we detected when looking for signatures of selection resulted from incompatibilities between the datasets.

### **Identification of candidate loci under diversifying selection between free-breeding and pure-breed dogs**

We analysed signatures of diversifying selection between FBDs and each of the two groups of pure-breed dogs by identifying  $F_{ST}$ -outlier SNPs using BAYESCAN (Foll and Gaggiotti 2008). This program calculates locus-specific pair-wise  $F_{ST}$  between each population and a common gene pool of all populations. These  $F_{ST}$  coefficients are then decomposed into two components:  $\alpha$ -component, which is locus-specific and shared by all populations considered, and  $\beta$ -component, which is population-specific and shared by all loci. If the  $\alpha$ -component significantly differs from 0 for a particular locus, this implies that selection is necessary to explain the population differentiation at this locus. Positive values of  $\alpha$ -component indicate diversifying selection, while negative values indicate balancing or purifying selection (Foll and Gaggiotti 2008).

We carried out this analysis in order to detect signatures of selection between: (a) East Asian dog breeds (N=29) and FBDs (N=190), (b) modern European dog breeds (N=356) and



FBDs, and (c) East Asian and European dog breeds. We removed X chromosome SNPs from the dataset, because differences in the mode of inheritance of this chromosome as compared with autosomal chromosomes could have led to biased results. We also pruned the datasets “a” “b” and “c” from SNPs with  $MAF < 0.01$  and those with missing data for more than 10% of individuals, which reduced them to around 146K SNPs (with small differences in the exact number of SNPs between the datasets). Dataset pruning was carried out in PLINK (Purcell *et al.* 2007).

To test whether BAYESCAN may produce false outliers due to the large numbers of loci compared, a control dataset was created, consisting of two groups of individuals with a very similar composition. Each group consisted of 94 FBDs, with matching number of individuals from each geographic region, and 187 pure-breed dogs, with matching number of individuals per breed. This dataset was pruned and analysed in the same way as the other datasets described above.

Sample size for East Asian breeds was relatively small (29 individuals). However, BAYESCAN accounts for the decreased accuracy in estimates of allele frequencies for small sample sizes, and therefore can be used for small datasets without bias, but at the expense of reduced power (Foll and Gaggiotti 2008). Therefore, this analysis has a low risk of detecting false positives, but some loci under selection may remain undetected. To further minimise the occurrence of false positives, we set up the prior odds for the neutral model at 100. Threshold values for a locus to be considered as being under diversifying selection were set at  $\alpha > 1.2$  and  $q < 0.15$ . The q-value is the minimum False Discovery Rate (the expected proportion of false positives) at which deviation of a given locus from the neutral model becomes significant. The q-value is defined in the context of multiple testing and cannot be directly compared with a p-value in classical statistics (Foll and Gaggiotti 2008).

The reduced power for the smaller sample size is demonstrated by lower q-values for the same outlier SNPs in the analysis comparing two large datasets (modern European breeds vs FBDs) relative to the analysis involving the small dataset of East Asian breeds (Table 1). The observed consistency in detecting the same outlier SNPs in different datasets, provides further support that their outlier status indeed reflects selection. Moreover, many of these shared outlier SNPs were located in introns and exons of coding genes, while most SNPs genome-wide are located in non-coding regions.

We used UCSC Genome Browser to search for the protein-coding genes within a pre-defined distance to the outlier SNPs detected in BAYESCAN (50 or 100 kb), and also identified the closest gene. This search was based on the CanFam3.1 dog genome assembly. Location of each gene identified this way was further confirmed via a search in Ensembl Release 82 database. For SNPs located within protein-coding genes, we checked whether they are located within introns or exons. Ensembl was also the starting point to obtain information on gene function, followed by searches in the NCBI database and primary literature.

### **Gene ontology analysis**

The gene ontology (GO) term enrichment analysis was carried out for genes located within 100 kb upstream and downstream from the outlier SNPs identified in the BAYESCAN analysis. We selected the distance of 100 kb following a study on genomic signatures of artificial selection during dog domestication (Axelsson *et al.* 2013), where the choice of a relatively large distance was justified by the need to take into account potential effects of mutations in regulatory elements located at some distance from genes, and to minimise the risk of excluding the outermost parts of haplotypes under selection (Axelsson *et al.* 2013). We also carried out the GO analysis including only genes located within 50 kb upstream and

downstream from the outlier SNPs to assess whether considering a shorter distance will considerably change the results.

We identified significantly overrepresented (at  $P < 0.05$ ) gene ontology (GO) terms using the Gostat program (Beissbarth and Speed 2004). We used the GOA-Human gene ontology annotation (Camon *et al.* 2004) to assign GO terms to the dog candidate genes based on their orthology with human genes. The candidate genes for which the orthology could not be determined were excluded from this analysis. Significance of overrepresentation for each GO term was tested using either a  $\chi^2$  test, or Fisher's Exact Test if the number of appearances of given GO term was below 5, and Benjamini & Hochberg correction was used to control the false discovery rate (Beissbarth and Speed 2004).

### **Analysis of potential transcription factors binding sites**

Twenty six outlier SNPs identified in the BAYESCAN comparisons between FBDs and either East Asian or European breeds were located outside gene coding sequences and their 5'- and 3'-UTRs. Although these SNPs can be located in sequences that have no functional role, some of them may be located in enhancers, matrix/scaffold attachment regions, miRNA target sites and other gene regulatory elements. Therefore, we analysed the effect of nucleotide substitution in outlier SNP sites on transcription factors (TF) binding. The analysis of putative TF binding sites was carried out for sequences located in 5'-upstream, 3'-downstream or intron sequences within 20 bp from the outlier SNPs identified in the BAYESCAN analysis. We identified putative TF binding sites using LASAGNA2 program (Lee and Huang 2013), based on TRANSFAC database matrices for vertebrates.

For each putative binding site, the program assessed the score and the probability of observing a higher or equal score by chance (P-value). To take into account the length of the putative promoter sequence in which a hit is found, an E-value is calculated according to the

formula:  $E\text{-value} = P\text{-value} \times (L - l + 1)$ , where  $L$  is the length of the promoter sequence and  $l$  is the length of the putative binding site (Lee and Huang 2013). Putative TF binding sites were identified assuming a threshold of  $E < 0.001$ .

### **Data availability**

SNP genotypes generated in this study are available from Dryad: doi:10.5061/dryad.078nc.

## **Results**

### **BAYESCAN analysis**

BAYESCAN analysis for the control dataset (comparing two groups composed of very similar sets of dog breeds and FBDs each) did not identify any outlier SNPs. All the SNPs analysed had  $q$ -values ranging from 0.9836 to 0.9901, alpha values ranging from -0.0078 to 0.0065, and  $F_{ST}$  values  $5.20\text{-}5.42 \times 10^{-6}$ . This supports our interpretation that outliers found in the main analyses reflect real deviations from neutrality, and are not simply the results of analysing large numbers of loci.

BAYESCAN analysis for the main dataset identified a small number of  $F_{ST}$ -outlier SNPs (Supplementary Tables 3 and 4), which was expected given the strict criteria we used to minimise the occurrence of false positives (e.g. setting the prior odds for the neutral model at 100; see Methods). Twelve SNPs, located on ten different chromosomes, were identified as candidate loci under diversifying selection between East Asian breeds and FBDs (Figure 1A, Supplementary Table 3A). Importantly, all these SNPs were also identified as candidate loci under diversifying selection either between modern European breeds and FBDs, or between East Asian and European breeds. Seven of these SNPs were located within genes, including one SNP located within an exon of *PKD1L1* gene and one within a 5'UTR of *CALCB* gene.

One SNP was located in a 5'-upstream sequence at a distance of 191,715 bp from the start site of transcription. Four SNPs were located in a 3'-downstream sequence at a distance between 652 and 365,573 bp from the transcription end point.

Sixty SNPs, located in 52 different chromosomal regions, were identified as candidate loci under diversifying selection between modern European dog breeds and FBDs (Figure 1B, Supplementary Tables 3A and 4). From among top 20 outlier SNPs with lowest q-values ( $<0.005$ ), six were located within genes, including one SNP located within an exon of PKD1L1 gene and one within a 5'UTR of CALCB gene (Supplementary Table 3B). Of the remaining 14 SNPs, eight were located in a 5'-upstream sequence at a distance between 18,237 and 437,119 bp from the start site of transcription, and six in a 3'-downstream sequence at distance from 652 to 365,573 bp from the transcription end point (Supplementary Table 3B).

The above results were obtained from the BAYESCAN analyses using all Eurasian FBDs independent of their sampling place, given that we found weak genetic differentiation among FBDs from different regions (Pilot *et al.* 2015). However, it may be argued that pure-breeds should be compared only with the regional FBD populations they derive from. Therefore, we repeated the BAYESCAN analysis comparing East Asian breeds with East Asian FBDs only (sampled in Thailand, China and Mongolia), and European breeds with European FBDs only (sampled in Poland, Slovenia and Bulgaria). Both analyses gave very consistent results with the earlier ones that included all FBDs, with 11 out of 12 outlier SNPs confirmed for the East Asian breeds vs FBDs and 17 out of top 20 outlier SNPs confirmed for European breeds vs FBDs; this included all the SNPs identified as shared outliers between the two sets. This high consistency further confirms that the outlier SNPs identified in this analysis accurately reflect the diversifying selection among the populations studied.

## **Outlier SNPs shared by the BAYESCAN analyses comparing FBDs to either East Asian or European breeds**

Eight outlier SNPs were shared by the comparison between FBDs vs East Asian breeds and FBDs vs European breeds (Table 1). Four of these SNPs were located within coding genes:

- (i) Two outlier SNPs were located in the Polycystic Kidney Disease 1-Like (PKD1L1) gene, one in an exon and another in an intron (a third outlier SNP was located 11 kb downstream of the gene; Figure 2).
- (ii) One outlier SNP was located in a 5'UTR of the calcitonin-related polypeptide beta gene (CALCB), also known as CRSP1 (calcitonin-receptor stimulating peptide 1; Figure 3). This SNP has also been mapped to CRSP3 gene - another member of CALCA gene family, located in the same chromosomal region as CALCB (Figure 3).
- (iii) One outlier SNP was located in an intron of the membrane-associated ring finger 7 (MARCH7) gene.

In addition, another outlier SNP shared by the comparisons between FBDs and both groups of breed was located 652 bp downstream from a gene homologous to a member of the vomeronasal 1 receptor (V1R) gene family, which has been annotated in several mammalian species (Supplementary Table 5).

The outlier SNP located in an exon of the PKD1L1 gene (BICF2G630842219) has been mapped to exon 54 of transcript PKD1L1-201 or exon 47 of transcript PKD1L1-202 (Figure 2). This SNP represents a synonymous substitution, but it does not necessarily imply functional neutrality (e.g. Pagani *et al.* 2005). The outlier SNP located in CALCB gene (BICF2S23653049) has been mapped to 5'-UTR (untranslated region), which is involved in the regulation of translation of a transcript (Araujo *et al.* 2012). In addition, this SNP has also been mapped to the promoter of CRSP3 gene (176 bp upstream; Figure 3). Double mapping is

probably a result of gene duplication, as indicated by high similarity of the genes from the calcitonin family (Osaki *et al.* 2008).

All FBD genotypes were heterozygous at these eight outlier SNPs (Supplementary Table 6). Heterozygous genotypes were also observed in all representatives of several dog breeds, which were genotyped in the LUPA project (Vaysse *et al.* 2011), showing that this result was not due to genotyping errors in our data, as it is unlikely that such errors would be breed-specific and repeated in independent datasets. The occurrence of a heterozygous genotype in all individuals from a particular group (FBDs or a particular breed) can be explained by a segmental duplication, with the duplicated copy fixed for an alternative SNP variant allele as compared with the original copy (Dorshorst *et al.* 2015). The occurrence of such segmental duplication resulting in a heterozygous SNP genotype in all individuals from a particular breed has been unambiguously shown in the chicken *Gallus domesticus* (Dorshorst *et al.* 2015). This suggests that diversification between dog breeds and FBDs in the candidate genes listed above may be based on segmental duplication and associated copy number variation, which constitutes a frequent source of genetic variation in pure-breed dogs and wolf-like canids (Alvarez and Akey 2012, Berglund *et al.* 2012, Axelsson *et al.* 2013, Ramirez *et al.* 2014) and may have important phenotypic effects (e.g. Salmon Hillbertz *et al.* 2007). This requires experimental verification by assessing the genomic copy number of the regions that were putatively duplicated.

### **Outlier SNPs shared by the BAYESCAN analyses comparing East Asian breeds to either FBDs or European breeds**

We carried out a BAYESCAN analysis comparing East Asian vs European breeds, and found eight outlier SNPs associated with eight different genes (Supplementary Table 3C). None of these SNPs corresponded to the eight shared outlier SNPs for the analyses comparing

FBD with either East Asian or European breeds. However, four of these SNPs also occurred as outliers in the comparison between FBDs and East Asian breeds (Table 2). One of these four outlier SNPs is located in an intron of MKKS gene (also known as BBS6), associated in humans with developmental anomaly syndromes, McKusick-Kaufman and Bardet-Biedl syndromes (Slavotinek *et al.* 2002). In each of the four outlier SNPs, one allele was fixed or nearly fixed in East Asian breeds, while it occurred in low to moderate frequencies in both FBDs and European breeds (Supplementary Table 7). In all four cases the same allele was also fixed or occurring in high frequency in grey wolves, and was fixed in two other canid species, Eurasian golden jackal *Canis aureus* and black-backed jackal *C. mesomelas*, suggesting that it represents an ancestral state for the wolf/dog lineage.

### **Gene Ontology analysis**

The Gene Ontology (GO) analysis carried out for all candidate genes located within 100 kb from outlier SNPs, showed good consistency in enriched GO terms (significant at  $P < 0.05$ ) for comparisons between (1) East Asian breeds versus FBDs and (2) modern European breeds versus FBDs. Twenty significantly enriched GO terms were shared between these two analyses, including: *metabolic process*, *developmental process*, *biological regulation*, *signalling*, *response to stimulus*, *synapse* and *behaviour* (Figure 4). The GO analysis including only genes located within 50 kb from outlier SNPs gave a similar, although smaller, set of enriched GO terms (significant at  $P < 0.05$ ). Seventeen significantly enriched GO terms were shared between these two analyses, including: *developmental process*, *biological regulation*, *signalling*, *response to stimulus*, and *behaviour* (Supplementary Figure 1).

### **Potential transcription factor binding sites**



Twenty five SNP sites with 50 alleles were identified in LASAGNA2 as located in putative TF binding sites ( $E < 0.001$ ). This analysis showed that the alleles are located in putative binding sites of 44 transcription factors (Supplementary Table 8). For six alleles no putative TF binding sites were identified. The other 44 alleles are located in putative binding sites for one (e.g. BICF2P1103910, T allele) to five transcription factors (e.g. BICF2S2298493, A allele). Nucleotide substitution in 19 out of 25 SNP sites had an effect on TF binding (Supplementary Table 8).

In three out of eight BAYESCAN outlier SNPs shared by the comparisons between FBDs vs East Asian breeds and FBDs vs European breeds, nucleotide substitutions had an effect on TF binding:

(i) A SNP site located 11 kb downstream of PKD1L1 gene (BICF2S23454833) represented A/C substitution changing the type of a transcription factor bound.

(ii) A SNP site located 652 bp downstream from a gene homologous to a member of the V1R gene family (BICF2P1363919) represented A/G substitution changing the type of a transcription factor bound.

(iii) A SNP site located -191 kb upstream from matrix metalloproteinase 16 (MMP16) gene (TIGRP2P367127) represented A/G substitution changing the DNA sequence from TF-binding to not TF-binding (Supplementary Table 8).

Two other outlier SNPs shared by both comparisons, one located in an intron of PKD1L1 gene and one downstream of SETBP1 gene (see Table 1), were also located in putative TF binding sites, but the nucleotide substitution did not change binding of TF. A SNP located in an intron of MARCH7 gene was not located within a TF-binding site.

## **Discussion**

*Genes and biological processes affected by breed formation events*

The GO enrichment analysis for candidate genes suggests that FBDs differ from pure-breed dogs in the selective pressures acting on traits associated with development, metabolism, nervous system, and behaviour. These GO terms were shared between the comparisons of FBDs with pure-breed dogs of either East Asian or European origin, suggesting that the two breed formation events affected the same biological processes. Interestingly, a similar set of GO terms was identified in an analysis of signatures of selection between domestic dogs and grey wolves (Axelsson *et al.* 2013, Wang *et al.* 2016), suggesting that the process of breed formation affected the same body systems that had been subject to selection at the onset of domestication.

The strict criteria applied to minimise the false discovery rate resulted in a small number of outlier SNPs between FBDs and both breed groups. Despite the small overall number of outlier SNPs, eight were shared between the comparisons of FBDs with either European and East Asian pure-breed dogs. This could result from artificial selection targeting the same traits in both breed formation events, relaxation of natural/sexual selection on particular traits (e.g. related to mate choice) in pure-breed dogs relative to FBDs, or a combination of both mechanisms.

Importantly, four of these eight shared outlier SNPs were located within coding genes, and a fifth in close proximity (652 bp) of a gene. The majority of SNPs within a genome occur in non-coding regions, and therefore the fact that 50% of common outlier SNPs fall within coding genes supports our interpretation that their outlier status truly reflects deviations from neutrality. This also provides an unambiguous link between the outlier SNPs detected in BAYESCAN and particular genes that were a target of the inferred selection process.

In addition, of the six shared outlier SNPs not located within exons, five were located within putative TF binding sites, and in three of them the nucleotide substitution had an effect

on TF binding. Mutations in TF binding sites can be deleterious, as shown by their involvement in human disease. For instance, of 2931 disease-associated SNPs located within regulatory DNA, 93.2% fall within TF binding sites (Maurano *et al.* 2012).

Among the shared candidate genes was the Polycystic Kidney Disease 1-Like gene (PKD1L1), involved in calcium regulation in primary cilia (Delling *et al.* 2013) – a function that may affect a wide range of phenotypic traits. Among known functions of this gene is regulation of testosterone production in humans (Yuasa *et al.* 2002), sperm-egg recognition in the sea urchin *Strongylocentrotus purpuratus* (Moy *et al.* 1996), and male stereotyped mating behaviour in the nematode *Caenorhabditis elegans* (Barr and Sternberg 1999).

The second shared candidate was the MARCH7 gene, belonging to a gene family of E3 ubiquitin-protein ligases, involved in regulation of cell trafficking, signalling, and cell cycle (Teixeira and Reed 2013). MARCH7 participates in regulating the development of spermatids (Zhao *et al.* 2013). Other known functions include the regulation of neuronal stem cells and T-cell-mediated immunity (Metcalf *et al.* 2005, Szogyarto *et al.* 2010).

Another common candidate gene was a vomeronasal 1 receptor (V1R) gene homologue (which has not yet been annotated in the dog). Although only nine functional V1R genes have been identified in dogs and wolves (Young *et al.* 2010), pheromones play an important role in canine mating behaviour (Goodwin *et al.* 1979, Tirindelli *et al.* 2009).

Finally, the fourth shared candidate was calcitonin-related polypeptide beta gene (CALCB, also known as CRSP1), an isomorph of the better studied calcitonin-related polypeptide alpha gene (CALCA). Both genes are preferentially expressed in the brain and share similar functions (Rezaeian *et al.* 2009). High similarity of the genes from the calcitonin family probably results from duplications of the CALCA gene during mammalian evolution (Osaki *et al.* 2008). CALCA encodes calcitonin and other peptide hormones involved in calcium regulation by alternative RNA splicing in specific tissues (Amara *et al.* 1982). These

hormones influence a wide range of physiological functions in the endocrine, nervous, immune, respiratory, gastrointestinal and cardiovascular systems (Poyner *et al.* 2002).

PKD1L1, MARCH7 and VIR genes are specifically involved in reproduction and/or sexual behaviour, and therefore their differentiation between FBDs and breed dogs is consistent with relaxation of sexual selection in breed dogs. In FBDs, a direct correlation between sperm count/quality and reproductive success may be expected, while in breed dogs reproduction and offspring survival may depend on other factors (e.g. on specific morphological characteristics preferred in each breed). Similarly, displaying certain mating behaviours may be essential for reproductive success in FBDs, but non-essential in pure-breed dogs, where mating partners are selected by humans.

The role of MARCH7 gene in regulation of activated T lymphocytes (Metcalf *et al.* 2005) suggests that its differentiation between FBDs and pure-breed dogs could also be due to relaxation of selection pressures on the immune system in pure-breed dogs, typically living within human households and benefiting from veterinary care. Importantly, the immune system may be subject to both natural and sexual selection (Hamilton and Zuk 1982, Zuk 1996). Genes targeted by positive selection in different mammalian species are commonly enriched for roles in immunity and defense, reproduction and chemosensory perception (Kosiol *et al.* 2008), showing that these systems are frequent targets of natural and sexual selection.

Of the genes discussed above, only PKD1L1 was identified earlier as a gene under diversifying selection between different breeds of dogs (Vaysse *et al.* 2011). However, most studies looking for signatures of selection in domesticated mammals were focused either on comparisons of a domesticated species with its wild ancestor, or on comparisons between breeds. Genes under selection between dog breeds are typically associated with morphological traits such as body size, ear shape, tail shape, coat colour and fur type (Vaysse

*et al.* 2011, Schlamp *et al.* 2016), implying that artificial selection in breeds targets different phenotypic traits than natural and sexual selection (Vaysse *et al.* 2011).

*Candidate genes under diversifying selection between FBDs and pure-breed dogs belong to the Hedgehog signalling pathway*

The shared candidate genes under selection between FBDs and two groups of pure-breed dogs: PKD1L1, CALCB and MARCH7, are involved in regulatory processes in the cell. Therefore, each of these genes may have multiple other functions besides those that have already been described, and they may be involved in common regulatory pathways. Indeed, we found that the regulatory functions of these genes are linked through the Hedgehog (HH) signalling pathway, one of the key regulators of development in all metazoans (Ingham *et al.* 2011). Members of the HH family play essential roles in a wide range of developmental processes, including morphogenesis of bones and skull, muscles, brain, gonads and external genitalia, as well as development of neurons and olfactory pathways (reviewed in Ingham and McMahon 2001, Briscoe and Thérond 2013). In mammals HH gene family consists of three members having different roles. Sonic hedgehog (SHH) regulates patterning of many systems during the embryonic development, including the limbs, notochord and neural tube, and controls cell division in adult stem cells. Indian hedgehog (IHH) is involved in development of bones and cartilage, and its function partially overlaps with SHH. Desert hedgehog (DHH) regulates peripheral nerve sheath formation and development of testis germ cells (Briscoe and Thérond 2013). The HH signal transduction pathway is very complex and involves multiple gene families, including these coding for Patched 1, Smoothed and GLI proteins (Briscoe and Thérond 2013).

In mammals, HH signalling is regulated via the primary cilium, a non-motile sensory organelle protruding from the cell surface (Rohatgi *et al.* 2007, Goetz and Anderson 2010,

Mukhopadhyay and Rohatgi 2014). Primary cilia constitute specialized domains for calcium signalling within cells that regulate established HH pathways through a heteromeric PKD1L1-PKD2L1 ion channel (DeCaen *et al.* 2013, Delling *et al.* 2013). HH signalling can occur via a calcitonin receptor-like receptor CRLR (Wilkinson *et al.* 2012), which is a receptor for the calcitonin family of peptide hormones (Barwell *et al.* 2012), encoded by the CALCA gene (Amara *et al.* 1982). MARCH7 is involved in the ubiquitination reaction (Szigyarto *et al.* 2010), which is an important mechanism regulating the activity, stability and location of the HH signalling components (Hsia *et al.* 2015).

Multiple functions of CALCA in different body systems, involvement of MARCH7 in neuronal stem cells regulation and PKD1L1 in germ cells regulation are all consistent with the regulatory role of the HH pathway. However, these genes are highly pleiotropic and may also affect other pathways. For example, mutations in ciliary genes such as PKD1L1 may affect signaling in other pathways at the cilium, including Wnt and platelet derived growth factor pathways (Lee and Gleeson 2011). This leads to the question of how mutations in such pleiotropic genes could contribute to major phenotypic effects without causing a strongly deleterious (i.e. lethal) pleiotropy. Similar questions may be posed regarding non-lethal genetic syndromes in humans, such as e.g. ciliopathy disease spectrum. According to our knowledge, a precise answer to this question is still unknown. However, the common involvement of several candidate genes in the HH pathway suggests that the diversifying selection between pure breed dogs and FBDs did not act on each of these genes independently. The differentiation between dog breeds and FBDs involves multiple phenotypic traits, and therefore it may be expected that it has a common genetic and developmental mechanism.

*Hedgehog signalling pathway and “domestication syndrome”*

Wilkins *et al.* (2014) similarly suggested that the “domestication syndrome”, i.e. a distinctive set of heritable morphological, physiological and behavioural traits typical of domesticated mammals (e.g. altered estrous cycle, docility, shorter muzzles, smaller teeth, floppy/reduced ears) results from a common underlying mechanism. They argued that mild neural crest cell (NCC) deficits during embryonic development can account for most of these apparently unrelated traits (Wilkins *et al.* 2014).

Wilkins *et al.* (2014) predicted that the “domestication syndrome” may result from an interaction of multiple and diverse NCC genes. Our results suggest that HH genes may be among them. Specifically, the SHH inhibits adhesion and migration of NCCs from neural tube (Testaz *et al.* 2001), and HH signaling in NCCs has been shown to regulate craniofacial development in vertebrates (Jeong *et al.* 2004, Wada *et al.* 2005, Tobin *et al.* 2008). The influence of HH pathway through NCCs on a wide range of phenotypic traits may explain the complex genetic nature of some canine traits, including skull shape variation (reviewed in Schoenebeck and Ostrander 2013). Importantly, the CALCA gene also belongs to neural crest genes (Martinez-Morales *et al.* 2007). NCC genes (MITF and KITLG) and HH pathway genes (PKD1L1 and SHH gene) were also found to be under diversifying selection between different breeds of dogs (Vaysse *et al.* 2011).

Importantly, one of the shared candidate genes under selection between East Asian dog breeds and either modern European breeds or FBDs, the MKKS/BBS6 gene, is associated in humans with McKusick-Kaufman syndrome (MKKS) and Bardet-Biedl syndrome (BBS; Slavotinek *et al.* 2002). These diseases result from primary cilia dysfunction, similarly as the Polycystic Kidney Disease caused by a mutation in PKD1 gene. One of the secondary features associated with BBS in humans is premaxillary and maxillary hypoplasia, resulting in mid-facial flattening and dental crowding (Ross *et al.* 2008, Tobin *et al.* 2008). BBS6-null mice showed similar changes in craniofacial morphology as BBS-affected humans, including snout

shortening (Tobin *et al.* 2008). These craniofacial changes were shown to result from aberrant cranial NCC migration caused by disrupted SHH signaling which provides positional cues to NCCs (Tobin *et al.* 2008). Snout shortening and dental crowding are characteristic traits of the “domestication syndrome” and are frequently used by zooarchaeologists to distinguish early domestic dogs from grey wolves (e.g. Moray 1992). This provides additional support to our hypothesis that mild modifications of SHH signaling in NCCs may contribute to the “domestication syndrome”.

In the present study, we compared different groups of domestic dogs rather than comparing the dog with its wild ancestor. Therefore it may be argued that all individuals we studied should express the same set of traits related to “domestication syndrome”, which were acquired early during the domestication process. However, if genetic changes induced by domestication cause alterations in embryonic development process (such as mild neural crest cell deficits; Wilkins *et al.* 2014), the extent of these alterations may differ between different groups of dogs depending on the strength of artificial versus natural selection. Extreme values of “domestication syndrome” traits (such as shorter muzzle, smaller teeth, smaller brain, neotenus behaviour) are likely selected against in free-ranging and free-breeding populations subject to natural and sexual selection.

A recent study demonstrated that the sensitivity and accuracy of selective sweeps detection in domestic dogs based on SNP chip data varies between loci and depends on the method applied (Schlamp *et al.* 2016). Due to the limitations of this approach some important genes could have been missed or some false positives could have been detected. However, identification of common biological pathways in which multiple outlier genes are involved is likely to be more robust than identification of individual genes. The integration of population genomics methods with systems biology may provide a powerful tool for such analysis (see Benfey and Mitchell-Olds 2008).



## Conclusions

The analysis of signatures of diversifying selection between FBDs and either European or East Asian breeds, identified candidate genes shared between these two comparisons. The presence of such shared candidate genes suggests either selection for the same traits during the two breed formation events, relaxation of natural and sexual selection in pure-breed dogs as opposed to FBDs, or combination of both these processes.

The association of these candidate genes with the HH signalling pathway suggests that they could affect dog phenotypes through their influence on common developmental processes regulated by this pathway. HH inhibits adhesion and migration of neural crest cells from neural tube, and minor deficits of these cells during embryonic development have been proposed as the underlying cause of the “domestication syndrome” (Wilkins *et al.* 2014). This suggests that the process of breed formation could have involved the same genetic and developmental mechanism as the initial process of domestication. As the set of canid whole-genome sequences continues to grow (e.g. Decker *et al.* 2015, Wang *et al.* 2016), it will be interesting to revisit these questions in future based on more powerful data.

## Acknowledgments

This project was funded by the National Science Centre in Poland (grant No. 2011/01/B/NZ8/02978). Additional funding was provided by the University of Lincoln, UK (Returners Research Fund) and the Deanship of Scientific Research at the King Saud University, Saudi Arabia (project number IRG\_15-38). We thank colleagues who helped with collecting dog samples: Malik H. Ali, Sergei V. Aramilev, Mika Bagramyan, Nikolay Baskakov, Sergey Belokobylskij, Mateusz Golan, Alexander V. Gromov, Maria Hołyńska, Grzegorz Kłys, Daniel S. Mills, Thanapol Nongbua, Ksenija Praper, Akylbek Ryspaev, La-

orsri Sanoamuang, Elena A. Sazhenova, Jacek Szewo, Mikhail P. Tiunov, Odbayar Ts, and Elena Tsingarska. We are grateful to Anna Ruć and Ewa Suchecka for their excellent technical assistance. We also thank Matthew Webster for his advice regarding the LUPA dataset. We are grateful to Danika Bannasch and three anonymous reviewers for their helpful comments on the manuscript.

### **Supplementary files:**

Supplementary File 1 is provided in PDF format and includes one figure and eight tables:

- Supplementary Figure 1. Enriched Gene Ontology terms for candidate genes under selection between (A) East Asian and Arctic breeds and FBDs, (B) modern European breeds and FBDs.
- Supplementary Table 1. A list of FBDs used in this study and their sampling sites.
- Supplementary Table 2. A list of dog breeds used in this study, and their regions of origin.
- Supplementary Table 3. Outlier SNPs inferred in BAYESCAN analysis comparing: (A) East Asian dog breeds and FBDs, (B) Modern European dog breeds and FBDs (only top 20 SNPs are listed), (C) East Asian dog breeds and modern European dog breeds.
- Supplementary Table 4. Outlier SNPs inferred in BAYESCAN analysis comparing modern European dog breeds and FBDs. This table lists 40 outlier SNPs with q-values between 0.006 and 0.086, while the top 20 outlier SNPs are listed in Supplementary Table 3B.
- Supplementary Table 5. Gene Ontology terms for candidate genes under diversifying selection between pure-breed dogs and FBDs.
- Supplementary Table 6. Shared outlier SNPs inferred in two BAYESCAN analyses comparing FBDs with either East Asian or European dog breeds. Eurasian golden jackal *Canis aureus* and black-backed jackal *C. mesomelas* were genotyped for only 3 individuals each. An allele fixed in black-backed jackal is likely an ancestral allele for the wolf/dog lineage. In FBDs and

some European breeds, a heterozygous genotype is fixed at these SNPs, suggesting segmental duplication with a different allele fixed at each gene copy.

- Supplementary Table 7. Shared outlier SNPs inferred in two BAYESCAN analyses comparing East Asian breeds with either FBDs or European breeds. Eurasian golden jackal *Canis aureus* and black-backed jackal *C. mesomelas* were genotyped for only 3 individuals each. An allele fixed in both golden jackal and black-backed jackal is likely an ancestral allele for the wolf/dog lineage.

- Supplementary Table 8. Potential transcription factors binding sites identified using LASAGNA2, assuming a threshold of  $E < 0.001$ .

## References

- Alvarez, C. E., Akey J. M. 2012 Copy number variation in the domestic dog. *Mamm. Genome* 23:144-163.
- Amara, S. G., V. Jonas, M. G. Rosenfeld, E. S. Ong, and R. M. Evans, 1982 Alternative RNA processing in calcitonin gene expression generates mRNAs encoding different polypeptide products. *Nature* 298:240-244.
- Araujo, P. R., K. Yoon, D. Ko, A. D. Smith, M. Qiao *et al.*, 2012 Before it gets started: regulating translation at the 5' UTR. *Comp. Funct. Genomics*. doi:10.1155/2012/475731.
- Axelsson, E., A. Ratnakumar, M.-L. Arendt, K. Maqbool, M. T. Webster *et al.*, 2013 The genomic signature of dog domestication reveals adaptation to a starch-rich diet. *Nature* 495:360-364.
- Barr, M. M., and P. W. Sternberg, 1999 A polycystic kidney-disease gene homologue required for male mating behaviour in *C. elegans*. *Nature* 401:386-389.

- Barwell, J., J. J. Gingell, H. A. Watkins, J. K. Archbold, D. R. Poyner *et al.*, 2012 Calcitonin and calcitonin receptor-like receptors: common themes with family B GPCRs? *Br. J. Pharmacol.* 166:51-65.
- Beissbarth, T., and T. P. Speed, 2004 GOstat: find statistically overrepresented Gene Ontologies within a group of genes. *Bioinformatics* 20:1464-1465.
- Benfey, P.N., and T. Mitchell-Olds, 2008 From genotype to phenotype: systems biology meets natural variation. *Science* 320:495-497.
- Berglund, J., E. M. Nevalainen, A.-M. Molin, M. Perloski, C. André *et al.*, 2012. Novel origins of copy number variation in the dog genome. *Genome Biol.* 13:R73.
- Briscoe, J. and P. P. Thérond, 2013 The mechanisms of Hedgehog signalling and its roles in development and disease. *Nature Rev. Mol. Cell Biol.* 14:416-429.
- Brown, S. K., C. M. Darwent, and B. N. Sacks, 2013 Ancient DNA evidence for genetic continuity in Arctic dogs. *J. Archaeol. Sci.* 40:1279-1288.
- Brown, S. K., C. M. Darwent, E. J. Wictum, and B. N. Sacks, 2015. Using multiple markers to elucidate the ancient, historical and modern relationships among North American Arctic dog breeds. *Heredity*. doi:10.1038/hdy.2015.49.
- Boyko, R. H., and A. R. Boyko, 2014 Dog conservation and the population genetic structure of dogs, p. 185-210 in *Free-ranging dogs and wildlife conservation*, edited by M. E. Gompper. Oxford University Press, Oxford.
- Camon, E., M. Magrane, D. Barrell, V. Lee, E. Dimmer *et al.*, 2004 The Gene Ontology Annotation (GOA) Database: sharing knowledge in Uniprot with Gene Ontology. *Nucl. Acid. Res.* 32:(Database issue): D262–D266.
- Carneiro, M., C.-J. Rubin, F. Di Palma, F. W. Albert, J. Alföldi *et al.*, 2014. Rabbit genome analysis reveals a polygenic basis for phenotypic change during domestication. *Science* 345:1074-1079.

- Chen, R., D. M. Irwin, and Y.-P. Zhang, 2012 Differences in selection drive olfactory receptor genes in different directions in dogs and wolf. *Mol. Biol. Evol.* 29:3475-84.
- Coppinger, R., and L. Coppinger 2016 Why do village dogs all look alike?, pp. 35-42 in *What is a dog?*, edited by R. Coppinger and L. Coppinger. The University of Chicago Press, Chicago.
- DeCaen, P. G., M. Delling, T. N. Vien, and D. E. Clapham, 2013 Direct recording and molecular identification of the calcium channel of primary cilia. *Nature* 504:315-318.
- Decker, B., B. W. Davis, M. Rimbault, A. H. Long, E. Karlins *et al.*, 2015 Comparison against 186 canid whole-genome sequences reveals survival strategies of an ancient clonally transmissible canine tumor. *Genome Res.* 25:1646-1655.
- Delling, M., P. G. DeCaen, J. F. Doerner, S. Febvay, and D. E. Clapham, 2013 Primary cilia are specialized calcium signalling organelles. *Nature* 504:311–314.
- Dorshorst, B., M. Harun-Or-Rashid, A.J. Bagherpoor, C.-J. Rubin, C. Ashwell *et al.*, 2015 A genomic duplication is associated with ectopic eomesodermin expression in the embryonic chicken comb and two duplex-comb phenotypes. *PLoS Genet.* 11:e1004947.
- Drake, A. G., and C. P. Klingenberg, 2010 Large-scale diversification of skull shape in domestic dogs: disparity and modularity. *Am. Nat.* 175:289-301.
- Duleba, A., K. Skonieczna, W. Bogdanowicz, B. Malyarchuk, and T. Grzybowski, 2015 Complete mitochondrial genome database and standardized classification system for *Canis lupus familiaris*. *Forensic Sci. Int. Genet.* 19:123-129.
- Foll, M., and O. Gaggiotti, 2008 A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective. *Genetics* 180:977-993.

- Freedman, A. H., I. Gronau, R. M. Schweizer, D. Ortega-Del Vecchyo, E. Han *et al.*, 2014  
Genome sequencing highlights the dynamic early history of dogs. *PLoS Genet.*  
10:e1004016.
- Freedman, A. H., R. M. Schweizer, D. Ortega-Del Vecchyo, E. Han, B.W. Davis *et al.*, 2016  
Demographically-based evaluation of genomic regions under selection in domestic dogs.  
*PLoS Genet.* 12:e1005851.
- Goetz, S.C. and K.V. Anderson 2010 The primary cilium: a signalling centre during  
vertebrate development. *Nat. Rev. Genet.* 11:331-344.
- Gompper, M. E. 2014 The dog-human-wildlife interface: Assessing the scope of the problem,  
pp. 9-54 in *Free-ranging dogs and wildlife conservation*, edited by M. E. Gompper.  
Oxford University Press, Oxford.
- Goodwin, M., K. M. Gooding, and F. Regnier, 1979 Sex pheromone in the dog. *Science*  
203:559-561.
- Gou, X., Z. Wang, N. Li, F. Qiu, Z. Xu *et al.*, 2014 Whole genome sequencing of six dog  
breeds from continuous altitudes reveals adaption to high-altitude hypoxia. *Genome Res.*  
24:1308-1315.
- Hamilton W.D. and M. Zuk 1982 Heritable true fitness and bright birds: A role for parasites?  
*Science* 218:384-387.
- Hsia, E. Y., Y. Gui, and X. Zheng, 2015 Regulation of Hedgehog signaling by ubiquitination.  
*Front. Biol.* 10(3):203-220.
- Illumina, Inc., 2006 Technical note: “TOP/BOT” Strand and “A/B” Allele. A guide to  
Illumina’s method for determining Strand and Allele for the GoldenGate® and Infinium™  
Assays. [http://www.illumina.com/documents/products/technotes/technote\\_topbot.pdf](http://www.illumina.com/documents/products/technotes/technote_topbot.pdf).  
Pub. No. 370-2006-018 27Jun06.
- Ingham, P. W., and A. P. McMahon, 2001 Hedgehog signaling in animal development:

- paradigms and principles. *Genes Dev.* 15:3059-3087.
- Ingham, P. W., Y. Nakano, and C. Seger, 2011 Mechanisms and functions of Hedgehog signalling across the metazoa. *Nat. Rev. Genet.* 12:393-406.
- Jeong, J., J. Mao, T. Tenzen, A. H. Kottmann, and A. P. McMahon, 2004 Hedgehog signaling in the neural crest cells regulates the patterning and growth of facial primordia. *Genes Dev.* 18:937-951.
- Kononov, D. A., C. Manning, and M. T. Henshaw, 2004 KINGROUP: a program for pedigree relationship reconstruction and kin group assignments using genetic markers. *Mol. Ecol. Notes* 4:779-782.
- Kosiol, C., T. Vinar, R. R. da Fonseca, M. J. Hubisz, C. D. Bustamante *et al.*, 2008 Patterns of positive selection in six mammalian genomes. *PLoS Genet.* 4:e1000144.
- Larson, G., and D. Q. Fuller, 2014 The evolution of animal domestication. *Annu. Rev. Ecol. Evol. Syst.* 45: 115–136.
- Larson, G., E. K. Karlsson, A. Perri, M. T. Webster, S. Y. W. Ho *et al.*, 2012 Rethinking dog domestication by integrating genetics, archeology, and biogeography. *Proc. Natl. Acad. Sci. USA* 109:8878–8883.
- Lee, C., and C.H. Huang, 2013 LASAGNA-Search: an integrated web tool for transcription factor binding site search and visualization. *BioTechniques* 54(3):141-153.
- Lee, E.J., and J.G. Gleeson 2011 A systems-biology approach to understanding the ciliopathy disorders. *Genome Med.* 3:59.
- Li, M., S. Tian, L. Jin, G. Zhou, Y. Li *et al.*, 2013a Genomic analyses identify distinct patterns of selection in domesticated pigs and Tibetan wild boars. *Nat. Genet.* 45:1431-1438.
- Li, Y., B. M. vonHoldt, A. Reynolds, A. R. Boyko, R. K. Wayne *et al.*, 2013b Artificial selection on brain-expressed genes during the domestication of dog. *Mol. Biol. Evol.* 30:1867-1876.

- Li, Y., D.-D. Wu, A. R. Boyko, G.-D. Wang, S.-F. Wu *et al.*, 2014 Population variation revealed high-altitude adaptation of Tibetan mastiffs. *Mol. Biol. Evol.* 31:1200-1205.
- Lindblad-Toh, K., C. M. Wade, T. S. Mikkelsen, E. K. Karlsson, D. B. Jaffe *et al.*, 2005 Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature* 438:803-819.
- Marsden, C. D., D. Ortega-Del Vecchyo, D. P. O'Brien, J. F. Taylor, O. Ramirez *et al.*, 2015 Bottlenecks and selective sweeps during domestication have increased deleterious genetic variation in dogs. *Proc. Natl. Acad. Sci. USA* 113:152-157.
- Marshall, T.C., J. Slate, L. E. B. Kruuk, and J. M. Pemberton, 1998 Statistical confidence for likelihood-based paternity inference in natural populations. *Mol. Ecol.* 7:639-655.
- Martinez-Morales, J.-R., T. Henrich, M. Ramialison, and J. Wittbrodt, 2007 New genes in the evolution of the neural crest differentiation program. *Genome Biol.* 8(3): R36.
- Maurano, M.T., R. Humbert, E. Rynes, R.E. Thurman, E. Haugen *et al.*, 2012 Systematic localization of common disease-associated variation in regulatory DNA. *Science* 337:1190-1195.
- Metcalf, S. M., P. A. Muthukumarana, H. L. Thompson, M. A. Haendel, and G. E. Lyons, 2005 Leukaemia inhibitory factor (LIF) is functionally linked to axotrophin and both LIF and axotrophin are linked to regulatory immune tolerance. *FEBS Lett.* 579:609-614.
- Montague, M. J., G. Li, B. Gandolfi, R. Khan, B. L. Aken *et al.*, 2014 Comparative analysis of the domestic cat genome reveals genetic signatures underlying feline biology and domestication. *Proc. Natl. Acad. Sci. USA* 111:17230-17235.
- Moray, D., 1992 Size, shape, and development in the evolution of the domestic dog. *J. Archaeol. Sci.* 19:181-204.



- Moy, G. W., L. M. Mendoza, J. R. Schulz, W. J. Swanson, C. G. Glabe *et al.*, 1996 The sea urchin sperm receptor for egg jelly is a modular protein with extensive homology to the human polycystic kidney disease protein, PKD1. *J. Cell Biol.* 133:809-817.
- Mukhopadhyay, S., and R. Rohatgi, 2014 G-protein-coupled receptors, Hedgehog signaling and primary cilia. *Semin. Cell Dev. Biol.* 33:63-72.
- Olsen, K. M., and J. F. Wendel, 2013 A bountiful harvest: genomic insights into crop domestication phenotypes. *Annu. Rev. Plant. Biol.* 64:47-70.
- Osaki, T., T. Katafuchi, and N. Minamino, 2008 Genomic and expression analysis of canine calcitonin receptor-stimulating peptides and calcitonin/calcitonin gene-related peptide. *J Biochem.* 144:419-430.
- Pagani, F., M. Raponi, and F. E. Baralle, 2005 Synonymous mutations in CFTR exon 12 affect splicing and are not neutral in evolution. *Proc. Natl. Acad. Sci. USA* 102:6368-6372.
- Parker, H. G., L. V. Kim, N. B. Sutter, S. Carlson, T. D. Lorentzen *et al.*, 2004 Genetic structure of the purebred domestic dog. *Science* 304:1160-1164.
- Pilot, M., T. Malewski, A. E. Moura, T. Grzybowski, K. Oleński *et al.*, 2015 On the origin of mongrels: Evolutionary history of free-breeding dogs in Eurasia. *Proc. R. Soc. Lond. B* 282:20152189. <http://dx.doi.org/10.1098/rspb.2015.2189>.
- Poyner, D., P. M. Sexton, I. Marshall, D. M. Smith, R. Quirion *et al.*, 2002 The mammalian calcitonin gene-related peptides, adrenomedullin, amylin, and calcitonin receptors. Paper presented at International Union of Pharmacology XXXII. *Pharmacol. Rev.* 54:233-246.
- Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M. A. R. Ferreira *et al.*, 2007 PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81:559-575.

- Ramirez, O., I. Olalde, J. Berglund, B. Lorente-Galdos, J. Hernandez-Rodriguez *et al.*, 2014 Analysis of structural diversity in wolf-like canids reveals post-domestication variants. *BMC Genomics* 15:465.
- Rezaeian, A. H., T. Isokane, M. Nishibori, M. Chiba, N. Hiraiwa *et al.*, 2009 alphaCGRP and betaCGRP transcript amount in mouse tissues of various developmental stages and their tissue expression sites. *Brain Dev.* 31:682-693.
- Rohatgi, R., L. Milenkovic, and M. P. Scott, 2007 Patched1 regulates hedgehog signaling at the primary cilium. *Science* 317:372-376.
- Ross, A., Beales, P. L. and Hill, J. 2008 The Clinical, Molecular and Functional Genetics of Bardet-Biedl Syndrome, pp. 147-186 in *Genetics of Obesity Syndromes*, edited by P. L. Beales, S. Farooqi, and S. O'Rahilly. Oxford Univ. Press, New York.
- Rubin, C.-J., M. Zody, J. Eriksson, and J. Meadows, 2010 Whole-genome resequencing reveals loci under selection during chicken domestication. *Nature* 464:587-591.
- Rubin, C.-J., H.-J. Megens, A. M. Barrio, K. Maqbool, S. Sayyab *et al.*, 2012 Strong signatures of selection in the domestic pig genome. *Proc. Natl. Acad. Sci. USA* 109:19529-19536.
- Salmon Hillbertz, N. H. C., M. Isaksson, E. K. Karlsson, E. Hellmén, G. R. Pielberg *et al.* 2007 Duplication of FGF3, FGF4, FGF19 and ORAOV1 causes hair ridge and predisposition to dermoid sinus in Ridgeback dogs. *Nat. Genet.* 39:1318-1320.
- Schoenebeck, J. J., and E. A. Ostrander, 2013 The genetics of canine skull shape variation. *Genetics* 193:317-325.
- Shannon, L. M., R. H. Boyko, M. Castelhana, E. Corey, J. J. Hayward *et al.*, 2015 Genetic structure in village dogs reveals a Central Asian domestication origin. *Proc. Natl. Acad. Sci. USA* 112:13639-13644.

- Skoglund, P., E. Ersmark, E. Palkopoulou, and L. Dalén, 2015 Ancient wolf genome reveals an early divergence of domestic dog ancestors and admixture into high-latitude breeds. *Curr. Biol.* 25:1515-1519.
- Slavotinek, A. M., C. Searby, L. Al-Gazali, R. C. M. Hennekam, C. Schrandt-Stumpel *et al.*, 2002 Mutation analysis of the MKKS gene in McKusick-Kaufman syndrome and selected Bardet-Biedl syndrome patients. *Hum. Genet.* 110:561-567.
- Smadja, C. M., and R. K. Butlin, 2011 A framework for comparing processes of speciation in the presence of gene flow. *Mol. Ecol.* 20:5123-5140.
- Szigyarto, C. A., P. Sibbons, G. Williams, M. Uhlen, and S. M. Metcalfe, 2010 The E3 ligase axotrophin/MARCH-7: protein expression profiling of human tissues reveals links to adult stem cells. *J. Histochem. Cytochem.* 58:301-308.
- Tapadia, M.D., D.R. Cordero, and J.A. Helms, 2005 It's all in your head: New insights into craniofacial development and deformation. *J. Anat.* 207:461-477.
- Teixeira, L. K., and S. I. Reed, 2013 Ubiquitin ligases and cell cycle control. *Annu. Rev. Biochem.* 82:387-414.
- Testaz, S., A. Jarov, K. P. Williams, L. E. Ling, V. E. Koteliansky *et al.*, 2001 Sonic hedgehog restricts adhesion and migration of neural crest cells independently of the Patched-Smoothed-Gli signaling pathway. *Proc. Natl. Acad. Sci. USA* 98:12521-12526.
- Tirindelli, R., M. Dibattista, S. Pifferi, and A. Menini, 2009 From pheromones to behavior. *Physiol. Rev.* 89:921-956.
- Tobin, J. L., M. Di Franco, E. Eichers, H. May-Simera, M. Garcia *et al.*, 2008 Inhibition of neural crest migration underlies craniofacial dysmorphology and Hirschsprung's disease in Bardet-Biedl syndrome. *Proc. Natl. Acad. Sci. USA* 105:6714-6719.
- van Asch, B., A.-B. Zhang, M. C. R. Oskarsson, C. F. C. Klütsch, A. Amorim *et al.*, 2013 Pre-Columbian origins of Native American dog breeds, with only limited replacement by

European dogs, confirmed by mtDNA analysis. *Proc. R. Soc. B.* 280:20131142.

<http://dx.doi.org/10.1098/rspb.2013.1142>.

Vaysse, A., A. Ratnakumar, T. Derrien, E. Axelsson, G. Rosengren Pielberg *et al.* 2011

Identification of genomic regions associated with phenotypic variation between dog breeds using selection mapping. *PLoS Genet.* 7:e1002316.

doi:10.1371/journal.pgen.1002316.

vonHoldt, B. M., J. P. Pollinger, K. E. Lohmueller, E. Han, H. G. Parker *et al.*, 2010 Genome-wide SNP and haplotype analyses reveal a rich history underlying dog domestication.

*Nature* 464:898-902.

Wada, N., Y. Javidan, S. Nelson, T. J. Carney, R. N. Kelsh *et al.*, 2005 Hedgehog signaling is required for cranial neural crest morphogenesis and chondrogenesis at the midline in the zebrafish skull. *Development* 132:3977-3988.

Wang G.-D., W. Zhai, H.-C. Yang, L. Wang, L. Zhong *et al.*, 2015 Out of southern East Asia: the natural history of domestic dogs across the world. *Cell Res.* 26:21–33.

Wilkins, A. S., R. W. Wrangham, and W. T. Fitch, 2014 The “domestication syndrome” in mammals: a unified explanation based on neural crest cell behavior and genetics.

*Genetics* 197:795-808.

Wilkinson, R. N., M. J. Koudijs, R. K. Patient, P. W. Ingham, S. Schulte-Merker *et al.*, 2012

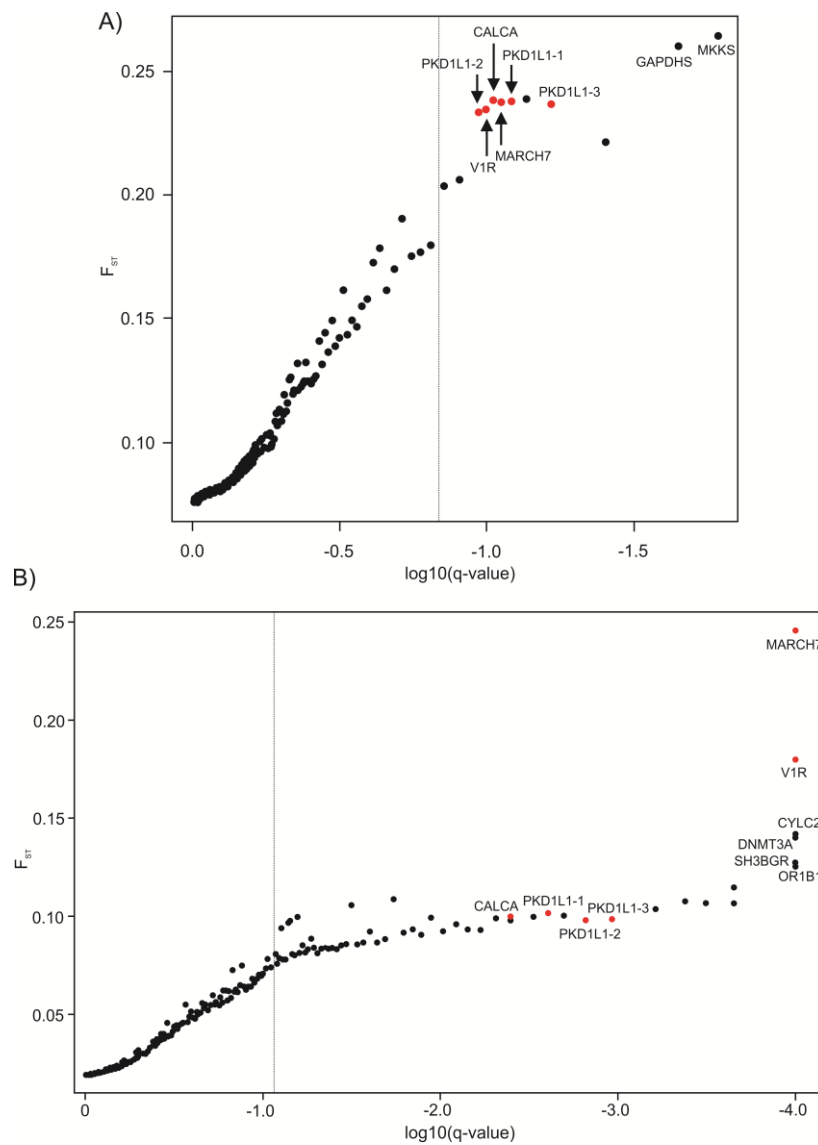
Hedgehog signaling via a calcitonin\_receptor-like receptor can induce arterial differentiation independently of VEGF signaling in zebrafish. *Blood* 120:477-488.

Young, J. M., H. F. Massa, L. Hsu, and B. J. Trask, 2010 Extreme variability among mammalian V1R gene families. *Genome Res.* 20:10-18.

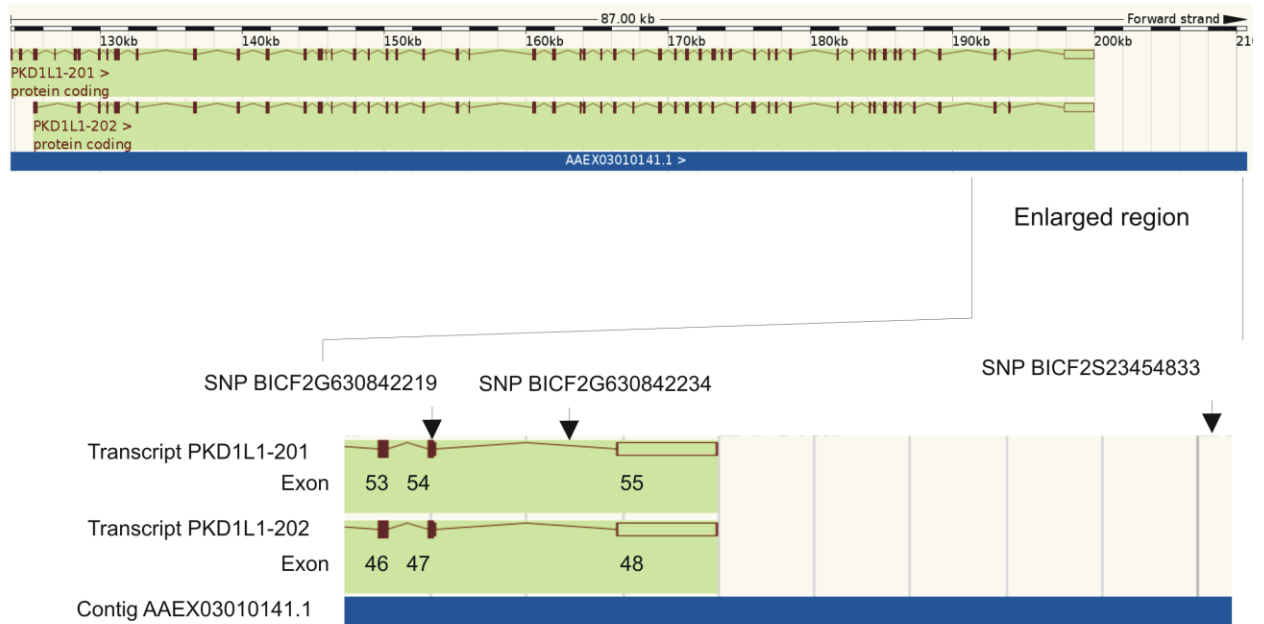
Yuasa, T., B. Venugopal, S. Weremowicz, C. C. Morton, L. Guo *et al.*, 2002 The sequence, expression, and chromosomal localization of a novel polycystic kidney disease 1-like gene, PKD1L1, in human. *Genomics* 79:376-386.

Zhao, B., K. Ito, P. V. Iyengar, S. Hirose, and N. Nakamura, 2013 MARCH7 E3 ubiquitin ligase is highly expressed in developing spermatids of rats and its possible involvement in head and tail formation. *Histochem. Cell Biol.* 139:447-460.

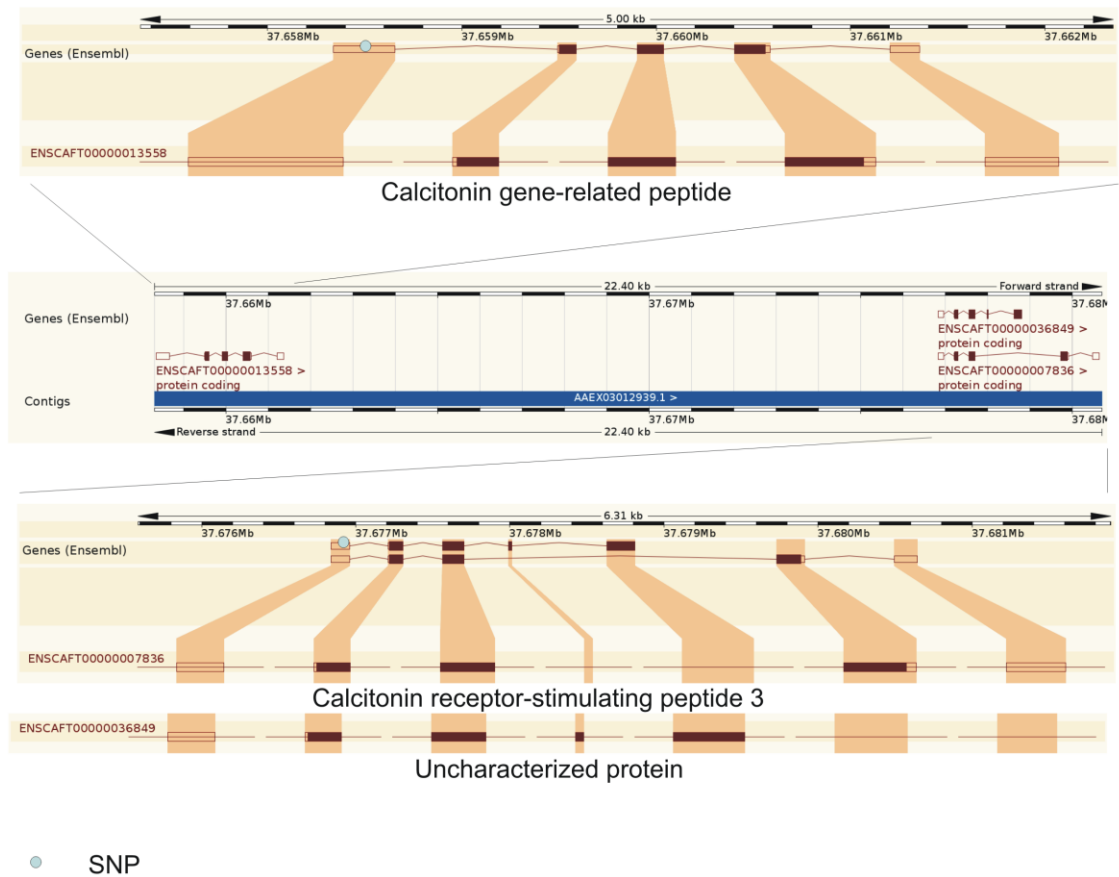
Zuk, M. 1996 Disease, endocrine-immune interactions, and sexual selection. *Ecology* 77:1037-1042.



**Figure 1.** Outlier SNPs inferred in BAYESCAN analysis comparing FBDs and (A) East Asian breeds or (B) European breeds. The vertical axis represents values of locus-specific  $F_{ST}$  coefficient, and the horizontal axis indicates the logarithm of q-values. The vertical line corresponds to a threshold q-value assumed in each analysis. Dots correspond to SNPs, and red dots correspond to shared candidate SNPs between the two analyses that are placed within genes or in close proximity of genes, with names of these genes being given. PKD1L1-1: BICF2G630842219, PKD1L1-2: BICF2G630842234, PKD1L1-3: BICF2S23454833. MKKS gene is also known as BBS6.

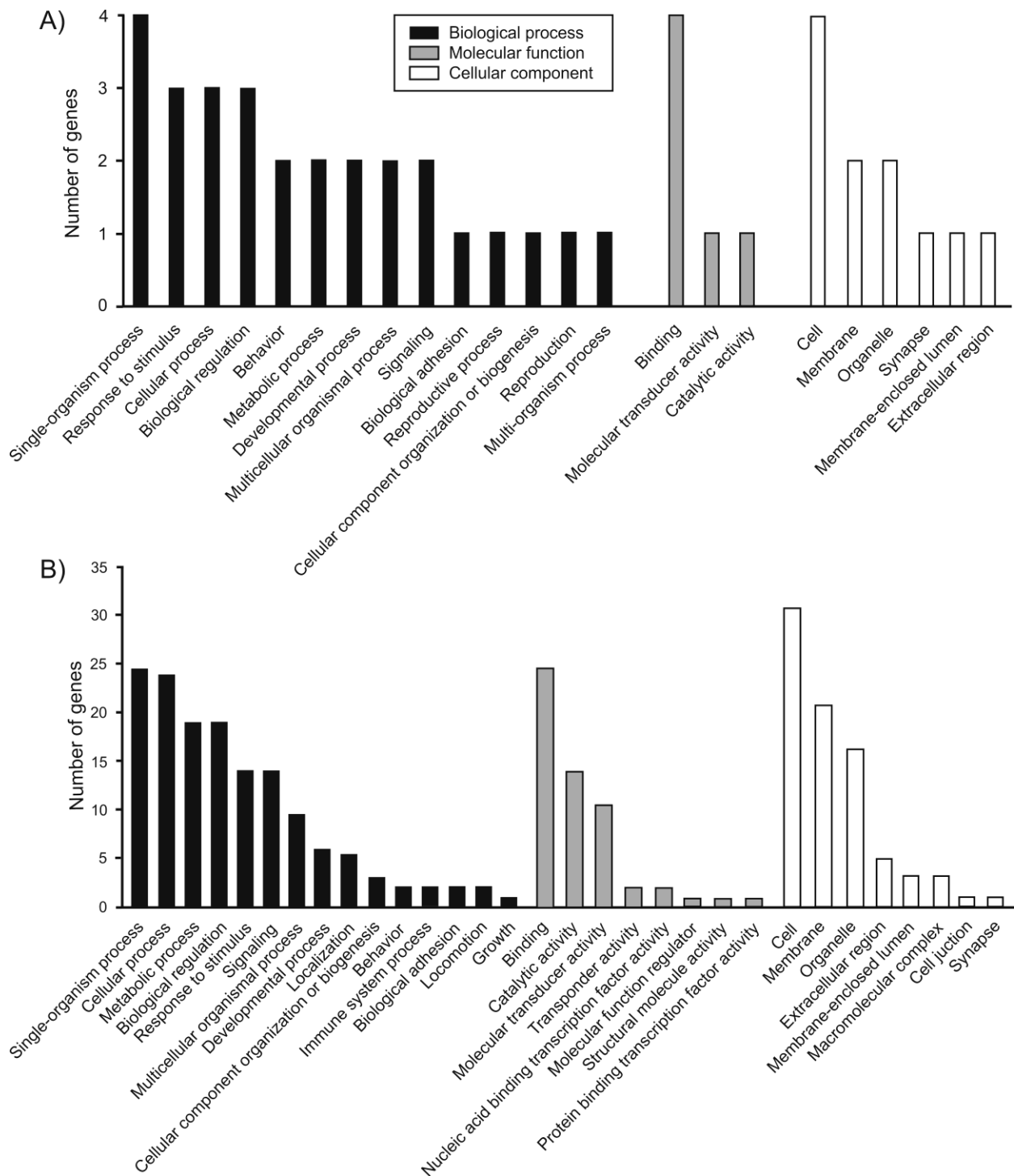


**Figure 2.** Locations of three outlier SNPs relative to PKD1L1 gene. These SNPs were identified as shared outliers in two BAYESCAN analyses comparing FBDs with either European or East Asian breeds.



**Figure 3.** Locations of an outlier SNP (BICF2S23653049) that maps to both *CALCB* and *CRSP3* genes belonging to calcitonin gene family. This SNP was identified as a shared outlier in two BAYESCAN analyses comparing FBDs with either European or East Asian breeds.





**Figure 4.** Enriched Gene Ontology terms for candidate genes under diversifying selection between (A) East Asian and Arctic breeds and FBDs, (B) modern European breeds and FBDs. The terms presented on the graphs are significantly overrepresented at  $P < 0.05$ . This analysis included genes located within 100 kb distance upstream or downstream of outlier SNPs.

**Table 1.** Shared outlier SNPs inferred in two BAYESCAN analyses comparing FBDs with East Asian (EA) or modern European (ME) breeds.

SNP ID	chr	SNP position CanFam2	SNP position CanFam3.1	Substitution type	Location relative to closest gene	Gene symbol	BayeScan FBD vs EA			BayeScan FBD vs ME			Gene function
							q-value	alpha	F <sub>ST</sub>	q-value	alpha	F <sub>ST</sub>	
BICF2G630842219	16	3,195,982	193,966	A/G	exon	PKD1L1	0.082	1.584	0.238	0.002	1.607	0.098	calcium regulation in primary cilia; associated with polycystic kidney disease in humans; plays a role in the male reproductive system
BICF2G630842234	16	3,198,732	196,716	A/G	intron	PKD1L1	0.106	1.543	0.234	0.003	1.632	0.102	calcium regulation in primary cilia; associated with polycystic kidney disease in humans; plays a role in the male reproductive system
BICF2S23454833	16	3,212,612	210,603	A/C	10,650 3'- downstream	PKD1L1	0.061	1.576	0.237	0.001	1.607	0.099	calcium regulation in primary cilia; associated with polycystic kidney disease in humans; plays a role in the male reproductive system
TIGRP2P369635_rs 8651736	36	8,528,500	5,525,355	G/T	intron	MARCH7	0.089	1.583	0.238	0.000	2.766	0.246	member of MARCH family of membrane-bound E3 ubiquitin ligases involved in regulation of diverse cellular processes; plays a role in the immune system (MHC chains retro-translocation) and spermiogenesis
BICF2S23653049	21	40,866,371	37,658,358	C/T	exon (5'UTR)	CALCB (CRSP1)	0.096	1.588	0.238	0.004	1.617	0.100	belong to CALCA gene family that encodes peptides and receptors involved in calcium regulation
BICF2P1363919	31	42,251,731	39,884,152	A/G	-756 5'- upstream	CRSP3 <sup>1</sup>							
BICF2P1363919	31	42,251,731	39,884,152	A/G	-652 3'- downstream	V1R homo- logue	0.101	1.556	0.235	0.000	2.350	0.180	homologue of vomeronasal 1 receptor gene in several mammalian species
TIGRP2P367127_rs 8543245	29	36,729,715	33,726,769	A/G	-191,715 5'- upstream	MMP16	0.073	1.596	0.239	0.003	1.610	0.100	encodes matrix metalloproteinase, involved in embryonic development, reproduction, and tissue remodelling
TIGRP2P97765_rs8 917688	7	49,723,506	46,745,071	A/G	365,573 3'- downstream	SETBP1	0.139	1.250	0.204	0.004	1.593	0.098	associated with Schinzel-Giedion midface retraction syndrome in humans

<sup>1</sup> Additional mapping

**Table 2.** Shared outlier SNPs inferred in two BAYESCAN analyses comparing East Asian (EA) dog breeds with FBDs or modern European (ME) breeds.

SNP ID	chr	SNP position CanFam2	SNP position CanFam3.1	Substitution type	Location relative to closest gene	Gene symbol	BayeScan EA vs FBD			BayeScan EA vs ME			Gene function
							q-value	alpha	F <sub>ST</sub>	q-value	alpha	F <sub>ST</sub>	
BICF2G630560144; rs24457899	7	58,926,419	55,945,622	A/G	intron	NOL4	0.040	1.456	0.222	0.076	1.359	0.279	encodes nucleolar protein 4 expressed in fetal brain, adult brain, and testis
BICF2P1348247; rs8579426	18	17,773,402	14,783,296	A/C	intron	ATXN7L1	0.124	1.274	0.204	0.059	1.602	0.317	associated with spinocerebellar ataxia type 7 in humans
BICF2G630509420; rs23187455	24	14,905,265	11,907,423	C/T	intron	MKKS	0.016	1.841	0.265	0.145	1.272	0.272	associated with McKusick- Kaufman syndrome and Bardet- Biedl syndrome type 6 in humans; among secondary symptoms are genital abnormalities and dental crowding
BICF2G630662694	13	35,179,641	32,140,606	A/G	257,806 3'- downstream	GAPDHS homo- logue	0.022	1.801	0.260	0.093	1.451	0.296	glyceraldehyde-3-phosphate dehydrogenase, spermatogenic; specifically expressed in spermatogenic cells