

## ARTICLE

# Unravelling selection signatures in a single dog breed suggests recent selection for morphological and behavioral traits

Juliane Friedrich<sup>1</sup>  | Andrea Talenti<sup>1</sup>  | Per Arvelius<sup>2</sup> | Erling Strandberg<sup>3</sup>  | Marie J. Haskell<sup>4</sup>  | Pamela Wiener<sup>1</sup> 

<sup>1</sup>Division of Genetics and Genomics, The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Midlothian, UK

<sup>2</sup>Swedish Armed Forces Dog Training Center, Mårsta, Sweden

<sup>3</sup>Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, Uppsala, Sweden

<sup>4</sup>Animal & Veterinary Sciences, Scotland's Rural College (SRUC), Edinburgh, UK

## Correspondence

Pamela Wiener, Division of Genetics and Genomics, The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Midlothian, EH25 9RG, Scotland. Email: pam.wiener@roslin.ed.ac.uk

## Funding information

Dogs Trust

## Abstract

Strong selection has resulted in substantial morphological and behavioral diversity across modern dog breeds, which makes dogs interesting model animals to study the underlying genetic architecture of these traits. However, results from between-breed analyses may confound selection signatures for behavior and morphological features that were coselected during breed development. In this study, we assess population genetic differences in a unique resource of dogs of the same breed but with systematic behavioral selection in only one population. We exploit these different breeding backgrounds to identify signatures of recent selection. Selection signatures within populations were found on chromosomes 4 and 19, with the strongest signals in behavior-related genes. Regions showing strong signals of divergent selection were located on chromosomes 1, 24, and 32, and include candidate genes for both physical features and behavior. Some of the selection signatures appear to be driven by loci associated with coat color (Chr 24; *ASIP*) and length (Chr 32; *FGF5*), while others showed evidence of association with behavior. Our findings suggest that signatures of selection within dog breeds have been driven by selection for morphology and behavior. Furthermore, we demonstrate that combining selection scans with association analyses is effective for dissecting the traits under selection.

## KEYWORDS

association analysis, coat color, dog behavior, German Shepherd dog, positive selection

## 1 | INTRODUCTION

The development of current dog breeds can be viewed as a unique long-term selection experiment to study the process of domestication<sup>1</sup> as well as short-term evolutionary change as a consequence of intensive breeding.<sup>2</sup> While the domestication of the modern dog (*Canis lupus familiaris*) from wolves took place at least 15 000 years ago,<sup>3</sup> with some estimates considerably earlier (eg, 20 000 to 40 000 years ago<sup>4</sup>), the popularity of dogs has led to ongoing strict selection according to breeding schemes and standards imposed by

breed associations and national kennel clubs. The establishment of genetically and phenotypically distinctive breeds by this intense artificial selection pressure has resulted in high intraspecies variation for physical and physiological features, disease susceptibility and behavior traits,<sup>5–7</sup> which makes dogs powerful models to investigate the underlying genetic architecture and signatures of selection for various traits.

Genetic manifestation of the development of dog breeds can be seen as selection signatures, genomic regions targeted by natural or artificial selection that exhibit various characteristics, including population differentiation, extreme linkage disequilibrium (LD) and patterns of the

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2020 The Authors. *Advanced Genetics* published by Wiley Periodicals LLC.

haplotype structure (eg, long-range haplotypes) or mutations in coding region.<sup>8</sup> Accordingly, selection signatures between dog breeds have been reported for physical traits, domestication-related traits and some specific behaviors and have led to the identification of candidate genes, for example, *IGF1* for body size, *FGF5* for coat length and *HAS2* for skin wrinkling,<sup>2</sup> *AMY2B*, *MGAM* and *SGLT1* for adaptation to a starch-rich diet<sup>9</sup> and *TRPM3* and *ROBO1* for athletic success in sport-hunting.<sup>10</sup> In a recent whole-genome sequence study of 144 modern dog breeds, positive human-imposed selection was implicated in the fixation or high prevalence within breeds of a range of morphological characteristics (eg, ear shape, height, weight).<sup>11</sup> These recent studies for selection signatures in dogs have focused on between-breed or dog-wolf comparisons and while such studies have allowed detection of signatures related to notable physical features, signatures for more subtle traits like behavior characteristics may be confounded with or masked by signals for the physical features, which might complicate the interpretation of these signatures as appears to be the case for association signals.<sup>12</sup>

In this study, we analyzed a single dog breed, the German Shepherd dog (GSD), to detect signals of selection. The breed was established in the late 19th century by crossing multiple breeds, with the initial purpose of creating a sheep-herding dog<sup>13</sup> and later use as a general working dog within the military or police. GSDs used in this study originated from two populations, the UK and Sweden; while the UK population represented a random sample of pet, show and working dogs, the Swedish dogs were bred within a breeding program of the Swedish Armed Forces (SAF) and only dogs that pass a behavior test can become working dogs or be used for breeding. Accordingly, in a previous study<sup>14</sup> we showed that there were significant differences between the two GSD populations for various behavior traits as measured in a questionnaire, for example, aggression against strangers or dogs, chasing, and playfulness. In contrast, morphological differences between populations were reduced compared to between-breed studies. We hypothesize that by comparing populations of the same breed but with different behavior-related selection strategies, we may be able to identify selection signatures for behavior as well as those for physical traits. Furthermore, by applying multiple statistical tests for the detection of selection signatures, we have increased the power to detect true signals of selection. Nonetheless, despite the within-breed approach, one of the main difficulties that remains is the identification of the actual trait(s) under selection. We addressed this issue by characterizing the relationship between selection signatures and statistical associations between genotype and phenotype (behavior and morphological traits) from the same populations. We suggest that this approach, combining population genetics and quantitative genetics methods, may also be applicable in other contexts.

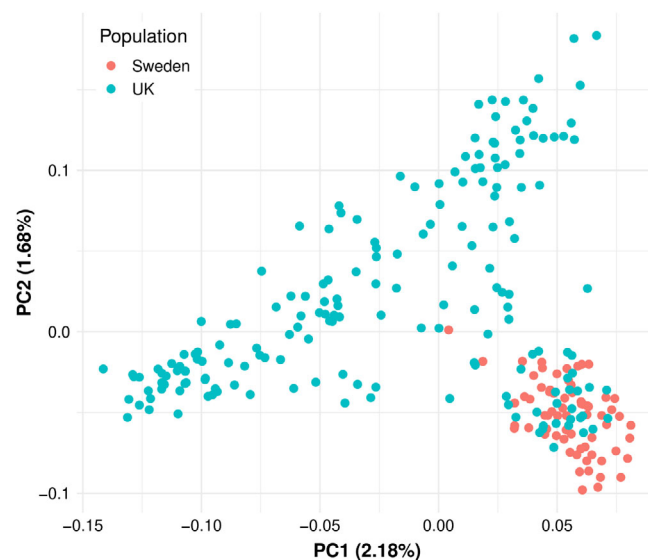
## 2 | RESULTS AND DISCUSSION

### 2.1 | Genomic structure of populations

Characterizing the genetic relationships between individual dogs is a valuable tool to evaluate the genetic structure of GSDs in this study.

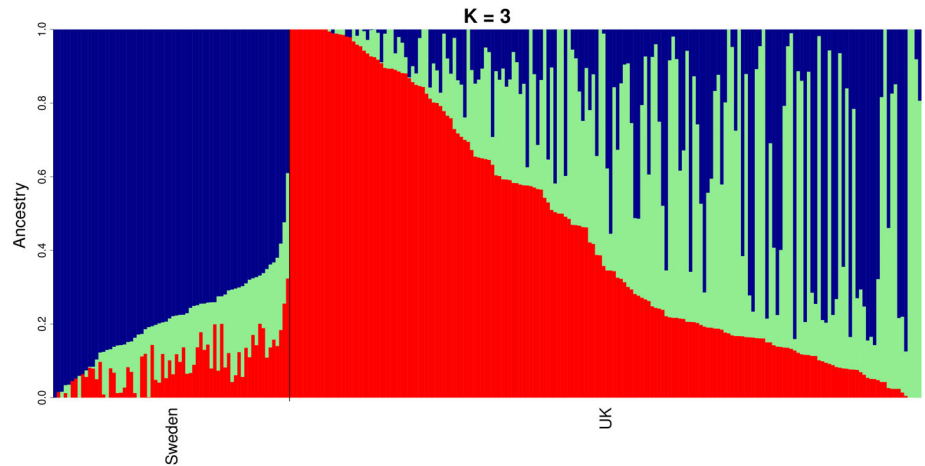
The underlying population structure in the two GSD populations (250 dogs in total) was explored by applying a principal component analysis (PCA) and ancestry estimation on a pruned SNP data set. The PCA indicated a separation between the UK and Swedish populations based on the first two principal components (PCs), which explained 2.8% and 1.9% of the genetic variance, respectively (Figure 1). With respect to PC1 and PC2, the UK dogs had a broader distribution than the Swedish GSDs, suggesting a stronger founder effect in the Swedish cohort. However, some of the UK GSDs clustered with the Swedish GSDs. The overall separation of the two populations is likely due to the geographical separation and thus primarily independent pedigrees but may also reflect the more recent origins of the Swedish population, with the SAF as the only breeder and the primary goal to breed good working dogs. The partial overlap between the two populations is likely due to the use of external dogs in the SAF breeding program, leading to some shared ancestry. A visual assessment of the ancestry estimation based on the ADMIXTURE program<sup>15</sup> (Figure 2) also revealed a clear discrimination between the UK and Swedish populations. The lowest cross-validation error of 0.55 was identified for  $K = 3$  clusters ( $K = 3$ ), with the blue cluster primarily associated with the Swedish population and the red and green clusters primarily associated with the UK population.

The average inbreeding coefficient calculated based on runs of homozygosity ( $F_{ROH}$ ) was  $0.29 \pm 0.02$  (SD; SD) for Swedish GSDs and  $0.31 \pm 0.05$  for UK GSDs. The significantly lower inbreeding estimate ( $P < .05$ ) in the Swedish population might be a consequence of a strategic breeding scheme by the SAF. The average nucleotide diversity ( $\mu$ ) was  $0.30 \pm 0.16$  for both populations.



**FIGURE 1** Principal Component Analysis of the pruned genomic data. Eigenvectors for the first two principal components are plotted and individuals are colored according to the population of origin. The variances explained by the principal components are given in parentheses

**FIGURE 2** Ancestry proportions of studied GSDs based on the pruned genomic data assuming three underlying ancestries ( $K = 3$  clusters) as revealed by ADMIXTURE. Each cluster is represented by a color and the length of the specific colored segment indicates the dog's proportion of membership in that cluster

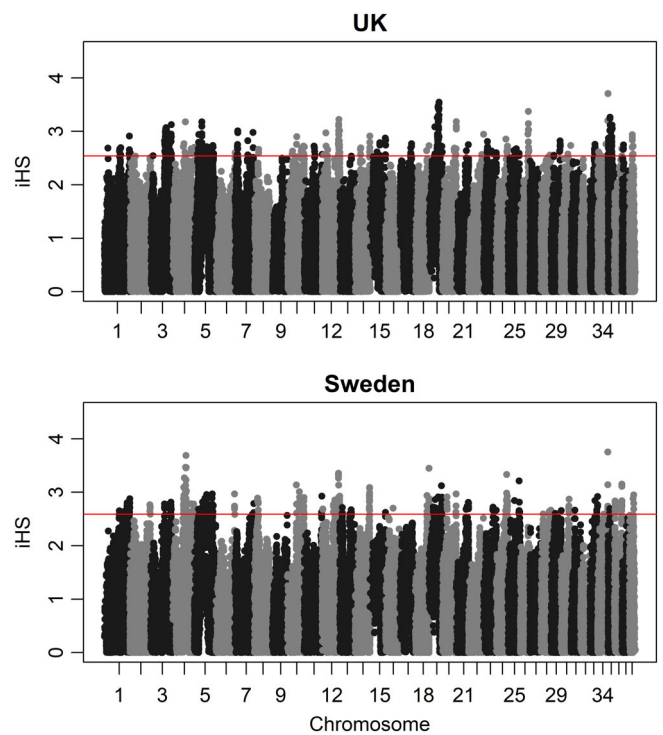


## 2.2 | Selection signatures within populations

Selection signatures can be detected within populations by identifying distinctive patterns of LD. In the event of selective sweeps, favorable genetic variants increase in frequency and form extended haplotypes with neighboring genomic regions due to LD, as reviewed in Reference 16. We computed the integrated haplotype score (iHS), which is a variation of the extended haplotype homozygosity (EHH) statistic that aims to detect recent and incomplete selective sweeps within populations.<sup>17</sup> In total, 197 and 142 regions with extreme EHH were detected within the UK and Swedish GSD population, respectively. A list of SNPs belonging to the top 0.5% of the iHS statistic in the UK and Swedish populations is given in Table S2. The iHS statistic identified similar selection signatures in both populations, but the most extreme values differed between populations, as shown by the 10 regions with the highest iHS statistics (Figure 3, Table 1). Regions with the highest iHS for the UK population were located on Chr 19 at 36.0 to 36.5 Mb and 37.5 to 37.7 Mb. A single marker on Chr 4 at 52.5 Mb showed the highest iHS in the Swedish population, followed by a region on Chr 18 at 54.9 to 55.3 Mb. The SNPs identified by iHS were further tested for their association with different traits (coat color, coat length, and behavior) separately for each population to identify the putative trait under selection.

The genes located within or closest to the 10 most extreme values of iHS (positional candidate genes) identified within populations (Table 1) have been previously associated with behavior. Regarding those on Chr 19, variants in *TMEM163* (transmembrane protein 163) were associated with active behavior in an open-field test involving cattle.<sup>18</sup> However, *TMEM163* is also a functional candidate for physical features, for example, for eye width and depth<sup>19</sup> and hair color<sup>20</sup> in humans. *NCKAP5* (NCK associated protein 5) was also identified as candidate gene for temperament in cattle<sup>21</sup> and has been associated with numerous neurological conditions in humans.<sup>22–24</sup>

The iHS peak on Chr 4 in the Swedish population points to the *CLINT1* (Clathrin Interactor 1) gene. This gene is reported to be among the top risk genes for the susceptibility to schizophrenia in humans<sup>25</sup> and markers near *CLINT1* were suggestive peaks associated with



**FIGURE 3** Distribution of integrated haplotype score (iHS) in the UK (upper plot) and Swedish population (lower plot). The red line indicates the threshold for the top 0.5% iHS

barking tendency in a genome-wide association study of behavior traits in Labrador retrievers.<sup>26</sup>

We conducted a gene list enrichment analysis with Enrichr<sup>27,28</sup> of the 256 and 338 genes that were located in and close to (within 40 kb of) the regions of the top 0.5% iHS in the UK and Swedish populations, respectively. No pathways were significantly enriched after accounting for multiple testing, however, Panther pathway analyses indicated nominally significant ( $P < .05$ ) functional enrichment of several pathways for the UK population: “heterotrimeric G-protein signaling -Gi alpha and Gs alpha mediated” ( $P = .01$ ; genes: *GRK4*,

**TABLE 1** Top selection signatures within the UK and Swedish GSD populations, showing the 10 highest integrated haplotype score (iHS) statistics

Chr	Start (Mb)	Stop (Mb)	Distance (Mb)	N <sub>SNPs</sub> <sup>a</sup>	iHS peak <sup>b</sup>	iHS mean <sup>c</sup>	Gene(s) <sup>d</sup>	Phenotypic association <sup>e</sup>
<i>UK population</i>								
5	29.2	29.8	0.62	16	3.18	2.84	ENSCAFG00000015899; MMP20; MMP27; MMP7; ENSCAFG00000030873; BIRC2; BIRC3; YAP1; C11orf70; CEP126; ANGPTL5	-
12	68.1	68.2	0.06	2	3.22	2.96	TRAF3IP2	-
19	33.0	33.1	0.04	4	3.26	2.84	n.a.	-
19	36.0	36.5	0.51	10	3.46	2.93	NCKAP5	-
19	36.8	37.0	0.19	5	3.18	2.90	n.a.	-
19	37.5	37.7	0.20	6	3.48	3.19	TMEM163	-
19	38.3	38.6	0.31	9	3.19	2.79	ZRANB3 ; ENSCAFG00000005064; R3HDM1; UBXN4	-
19	39.5	39.5	0.03	2	3.23	2.91	n.a.	-
20	57.6	57.7	0.07	3	3.18	3.10	ENSCAFG000000031730; ENSCAFG000000023991; ARHGAP45; ATP5F1D; CIRBP; MIDN; STK11; SBNO2; POLR2E	-
35	7.9	8.1	0.14	4	3.26	3.09	BMP6; TXNDC5; BLOC155; ENSCAFG000000009583; ENSCAFG000000024482	-
<i>Swedish population</i>								
4	44.3	n.a.	n.a.	1	3.09	n.a.	ENSCAFG000000017171	-
4	46.9	n.a.	n.a.	1	3.27	n.a.	ENSCAFG000000028841	-
4	50.0	50.2	0.15	4	3.09	2.90	ATP10B	-
4	52.5	n.a.	n.a.	1	3.47	n.a.	CLINT1	-
12	66.7	67.2	0.47	10	3.36	3.13	GPR6; WASF1; CDC40; METTL24; DDO; SLC22A16; CDK19	-
12	67.7	n.a.	n.a.	1	3.13	n.a.	SLC16A10	-
18	54.9	55.3	0.36	7	3.45	2.99	LRRC10B; PPP1R32; SYT7; PGA; DDB1; VWCE; ENSCAFG000000016314; SLC15A3; CD5; VPS37C; CD6	-
19	50.6	n.a.	n.a.	1	3.12	n.a.	KIF5C	-
24	42.4	42.5	0.05	3	3.33	3.05	RBM38; CTCFL	-
36	30.1	30.6	0.05	6	3.11	2.82	GULP1; COL3A1; COL5A2	-

Note: SNPs within 200 kb were summarized into selection signature regions.

<sup>a</sup>Number of top SNPs in region.

<sup>b</sup>Standardized absolute iHS of the peak SNP (in that region).

<sup>c</sup>Average standardized absolute iHS across the SNPs of a region.

<sup>d</sup>Genes located within and +/- 40 kb around selection signatures. Genes highlighted in bold include a SNP that belongs to the top 0.5% of the test statistic; all others are located within the region or +/- 40 kb around selection signatures.

<sup>e</sup>There were no phenotypic associations (behavior, coat color, or coat length) with FDR-adjusted *P*-value <.1 for markers located within the top 10 selection signatures within populations.

GRK7, RGS12, ADCY2, ADRA2C, DRD2), "Alzheimer disease-presenilin" (*P* = .02; TRPC6, MMP7, MMP27, RBPJ, MMP20), "heterotrimeric G-protein signaling -Gq alpha and Go alpha mediated" (*P* = .02; GRK4, GRK7, CACNA1A, RGS12, DRD2), "ionotropic glutamate receptor" (*P* = .03; CACNA1A, SLC17A8, GRIA4), and "axon guidance mediated by semaphorins" (*P* = .03; CRMP1, FYN). All of these functions have been shown to be relevant for behavior among other functions, for example, heterotrimeric G proteins in mood disorders, as reviewed in Reference 29, ionotropic glutamate receptors for long-term synaptic plasticity, as reviewed in References 30, 31 and semaphorins in neuronal structure, as reviewed in Reference 32. Nominally significant pathways for the Swedish population were "5-Hydroxytryptamine

degradation" (*P* = .003; ALDH3A2, ALDH3A1), "apoptosis signaling" (*P* = .01; MAP2K3, CASP9, DAXX, BAK1, BIRC2, BIRC3), and "Thyrotropin-releasing hormone receptor signaling" (*P* = .03; PLCE1, STX3, TRHR). 5-hydroxytryptamine (serotonin) is an important neurotransmitter and plays a key role in numerous behavioral disorders and characteristics, for example, depression<sup>33</sup> and aggressiveness.<sup>34</sup>

## 2.3 | Selection signatures between populations

Another approach to identify signatures of selection is the comparison of genetic variation (eg, allele frequencies or haplotype structure)

between different populations. Accordingly, signatures of differential selection between the two GSD populations were analyzed employing three different tests: the fixation index ( $F_{ST}$ ), the cross-population extended haplotype homozygosity (XP-EHH), and differences between ROH ( $\Delta ROH_{prop}$ ).  $F_{ST}$  was calculated to determine genetic differentiation between UK and Swedish GSD populations. Low genome-wide genetic differentiation was detected for the single SNP-based statistic ( $F_{ST} = 0.021 \pm 0.029$ ) and for the SNP window-based statistic ( $F_{ST} = 0.021 \pm 0.016$ ), consistent with previous within-dog-breed estimates.<sup>35</sup>

We scanned the genome for regions of genetic differentiation within overlapping 1 Mb windows and found 17 distinctive peaks that comprise the top 1% window-based  $F_{ST}$  values on Chr 1, 9, 20, 22, 24, 29, 30, and 32, with values ranging from 0.07 to 0.16 (Table S3). The highest  $F_{ST}$  value (0.16) was found for a region on Chr 24 (22.0–24.5 Mb), which contains 46 genes. Among these genes are several with functions in physical characteristics and behavior, for example, *SPAG4* and *SUN5* involved in cytoskeletal anchoring, *NCOA6* involved in glucocorticoid and corticosteroid receptor signaling and *ASIP* and *RALY* associated with skin and fur pigmentation. Furthermore, seven members of the bactericidal/permeability-increasing (BPI) fold-containing (BPIF) superfamily of genes are located in this region (*BPIFB2*, *BPIFB6*, *BPIFB3*, *BPIFB4*, *BPIFA2*, *BPIFA3*, *BPIFA1*, and *BPIFB1*). It was shown that these genes play a role in the innate immune system and lipoprotein metabolism, but also in the brain's response to oxidative stress (aging), relevant for neuropsychiatric diseases.<sup>36</sup> Interestingly, high  $F_{ST}$  for Labrador retriever populations differentiated based on their coat color and function (gundog and show dog) was also detected in the same region on Chr 24 (22.4–22.8 Mb) in a previous study.<sup>37</sup>

While the  $F_{ST}$  statistic detects differences in allele frequencies between populations, the XP-EHH test, an approach based on LD, is designed to detect regions that are fixed (or nearly fixed) in one population but remain segregating in the other population. Extreme high (positive) and low (negative) scores are indicators of a region under strong positive selection in the UK and Swedish population, respectively. The region including the SNP with the highest score (3.4) for the UK population was located on Chr 35 (11.0–11.5 Mb) and contains three genes (*NEDD9*, *ADTRP*, and *TMEM170B*) (Table S3). The *NEDD9* (Neural Precursor Cell Expressed, Developmentally Down-Regulated 9) gene has been shown to be associated to cognitive impairment in mice,<sup>38</sup> *ADTRP* is important for vascular development and function in mouse and zebrafish<sup>39</sup> and *TMEM170B* has been reported to be downregulated in TCGA human breast cancer data.<sup>40</sup> The region with the highest absolute score (3.8) for the Swedish population was located on Chr 12 (3.6–7.5 Mb). This region contains 59 genes; *RNF8* and *TBC1D22B* are closest to the SNP with the most extreme score. The ubiquitin gene *RNF8* (ring finger protein 8) plays a role in the immune system and has also been linked to autism; a recent study in *RNF8* knockout mice indicated a role of this gene in synapse formation and cerebellar-dependent learning abilities.<sup>41</sup> The function of *TBC1D22B* is largely unknown but it may encode a GTPase-activating protein.

As a third approach to identifying differential selection between the populations, we identified the regions showing differences in extended homozygosity. To identify these selection signatures, we calculated the between-population differences in runs of homozygosity ( $\Delta ROH_{prop}$ ), which describes the difference in the proportion of dogs with an ROH of a specified length at a given SNP. The average  $\Delta ROH_{prop}$  value across the genome was low ( $0.07 \pm 0.06$ ), indicating considerable overlap of ROH between the UK and Swedish populations. However, some regions with ROH were predominantly present in only one population (Table S3). The highest absolute  $\Delta ROH_{prop}$  indicating selection signatures in the UK population were found on Chr 17 and 32: the ROH mapped to Chr 17 (8.3–8.4 Mb) and Chr 32 (13.3–13.4 Mb) were present in over 70% of the UK dogs but less than 40% of the Swedish dogs. The genes located in these regions are *GREB1*, *NTSR2*, and *LPIN1* on Chr 17, with no characterized genes in the Chr 32 region. The neurotensin gene *NTSR2* is involved in dopamine modulation and a SNP in this gene has been tested in a polygenic model of highly sensitive personality in humans.<sup>42</sup> *LPIN1* plays a prominent role in lipid metabolism regulating adipocyte differentiation and coregulating other genes involved in lipid metabolism. The highest absolute  $\Delta ROH_{prop}$  indicating selection signatures in the Swedish population was found on Chr 1: a ROH mapped to Chr 1 (24.7–25.5 Mb) was present in 90% of the Swedish dogs but only in 42% of the UK dogs and contains the genes *LDLRAD4*, *MOXD1*, and *CTGF* (see below).

### 2.3.1 | Target regions for divergent selection signatures between populations

In the detection of selection signatures, the application of multiple approaches is recommended to reduce the rate of false positive signals.<sup>16</sup> To identify target regions under differential selection in the two GSD populations, we selected regions from the 99th percentile (top 1%) of each score distribution (SNP window-based  $F_{ST}$ ,  $\Delta ROH_{prop}$ , and XP-EHH) and searched for intersecting signals between two or three of the approaches. Using this criterion, we identified 433 SNPs (Table S3), with the greatest overlap between the SNP window-based  $F_{ST}$  and  $\Delta ROH_{prop}$  statistics (374 SNPs). No SNPs were detected by all three approaches. The 433 SNPs were located in 16 candidate selected regions on Chr 1, 9, 12, 22, 24, 32, and 34, which harbor 114 genes in total (Table 2; Figure 4). One Panther pathway was nominally significantly ( $P < .05$ ) enriched by these 114 genes: “p53 pathway feedback loops” ( $P = .03$ ; *CDKN1A*, *RBL1*). The SNPs identified as under divergent selection by these analyses were further tested for their association with different traits (coat color, coat length, and behavior) separately for each population to identify the putative trait under selection.

A visual inspection of the Circos plot (Figure 4), which illustrates the results for the three approaches, indicates regions on Chr 1, 24, and 32 where peaks can be seen based on all three methods, although not belonging to the top 1% for XP-EHH. Linear plots for these three regions illustrate the results from association analyses for traits with

**TABLE 2** Selection signatures that belonged to the top 1% of the distribution of at least two methods used to detect signatures of different selection between the GSD populations

Chr	Start	Stop	N <sub>SNPs</sub> <sup>a</sup>	Population	F <sub>ST</sub> <sup>b</sup>	ΔROH <sub>Prop</sub> <sup>c</sup>	XP-EHH <sup>d</sup>	Gene(s)	Phenotypic association <sup>e</sup>
1	24 024 856	25 483 783	61	Sweden	0.12	0.46	NA	ME2; MRO; MC2R; MC5R; ENSCAFG000000000172; ENSCAFG000000029562; ENSCAFG000000029833; FAM210A; LDLRAD4; ENSCAFG000000023012; MOXD1; ENSCAFG000000031561; CTGF	Chasing*(UK)
9	16 472 361	16 493 753	4	UK	0.09	NA	2.81	KCNJ16; KCNJ2	-
12	5 349 354	6 130 868	44	Sweden	NA	0.27	3.44	BRPF3; PNPLA1; C12H6orf222; ETV7; PXT1; ENSCAFG00000001396; KCTD20; STK38; SRSF3; CDKN1A; ENSCAFG0000000001418; ENSCAFG0000000001419; CPNE5; PPIL1; C12H6orf89; MTCH1; PI16; FGD2	Stranger-directed fear**(UK)
12	6 466 863	6 554 339	7	Sweden	NA	0.27	3.46	FGD2; CMTR1; ENSCAFG000000030835	Separation anxiety* (Sweden)
22	1 027 334	1 140 100	6	UK	0.08	0.26	NA	RNASEH2B	-
22	1 683 950	2 496 568	46	UK	0.12	0.26	NA	KCNRG; TRIM13; SPRYD7; KPNA3; ENSCAFG000000031710; EBPL; ENSCAFG000000010362; RCBTB1; PHF11; SETDB2; CAB39L; CDADC1; ENSCAFG000000028525; MLNR; FNDC3A	-
24	22 002 778	22 463 326	24	UK	0.07	0.29	NA	COMMD7; DNMT3B; MAPRE1; EFCAB8; SUN5; BPIFB2; BPIFB6; BPIFB3; BPIFB4; ENSCAFG000000032553; BPIFA2; ENSCAFG000000007369; BPIFA3; BPIFA1	Coat color**(UK)
24	22 908 179	23 816 844	37	UK	0.14	0.28	NA	ENSCAFG000000029918; ENSCAFG000000007430; ENSCAFG000000007435; ENSCAFG000000029879; NECAB3; PXMP4; ZNF341; CHMP4B; EIF2S2; RALY; ASIP; ENSCAFG000000007508; AHCY; ITCH; DYNLRB1; PIGU; MAP1LC3A; NCOA6; TP53INP2	Coat color**(UK)
24	24 867 975	25 952 679	64	UK	0.13	0.28	NA	CNBD2; EPB41L1; AAR2; DLGAP4; MYL9; TGIF2; SLA2; TGIF2-C20orf24; NDRG3; DSN1; SOGA1; TLDC2; SAMHD1; RBL1; MROH8; RPN2; GHRH; MANBAL; SRC	Coat color**(UK)
32	4 172 082	4 455 360	7	UK	0.09	0.27	NA	ANTXR2; PRDM8	Coat length**(UK)
32	5 350 389	5 399 877	4	UK	0.13	0.26	NA	PRKG2	Coat length**(UK) and * (Sweden) Stranger-directed aggression** (Sweden)
32	5 609 507	5 667 788	4	UK	0.12	0.26	NA	ENSCAFG000000008928; RASGEF1B	Coat length** (UK and Sweden)
32	13 000 437	14 125 551	44	UK	0.11	0.37	NA	SNCA; MMRN1; CCSER1	Coat color* (UK) Separation anxiety*(UK) Stranger-directed aggression* (Sweden)
32	14 527 559	14 597 957	4	UK	0.11	0.38	NA	ENSCAFG000000009954	-

TABLE 2 (Continued)

Chr	Start	Stop	N <sub>SNPs</sub> <sup>a</sup>	Population	F <sub>ST</sub> <sup>b</sup>	ΔROH <sub>Prop</sub> <sup>c</sup>	XP-EHH <sup>d</sup>	Gene(s)	Phenotypic association <sup>e</sup>
32	14 952 127	15 194 499	4	UK	0.10	0.28	NA	ENSCAFG00000009965	-
34	33 480 270		1	UK	NA	0.27	2.80		-

Note: SNPs within 200 kb were summarized into selection signature regions. NA indicates that this selection signature was not present in the top 1% of the test distribution. Genes highlighted in bold include a SNP that belongs to the top 1% of the test distribution; all others are located within the region or +/- 40 kb around selection signatures.

<sup>a</sup>Number of top SNPs in region.

<sup>b</sup>Fixation index.

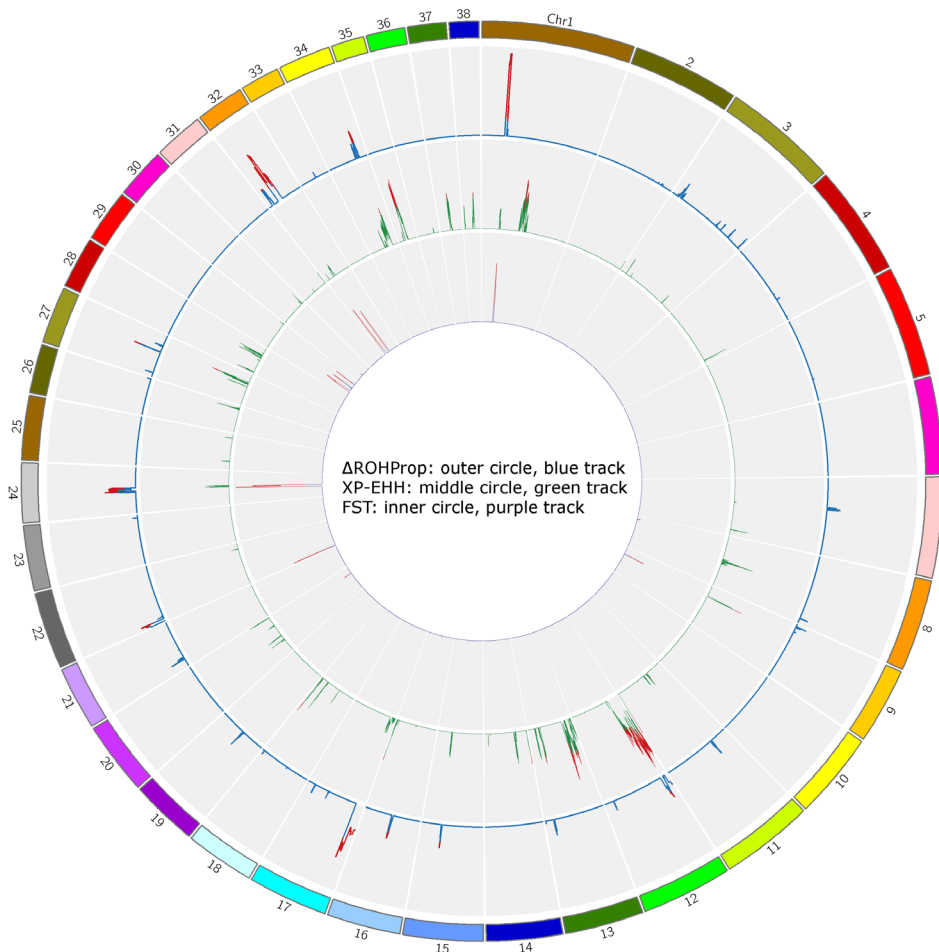
<sup>c</sup>Differences between runs of homozygosity.

<sup>d</sup>Cross-population extended haplotype homozygosity.

<sup>e</sup>Significant phenotypic associations (behavior, coat color, coat length) for the UK and Swedish population within selection signature region. P-values were adjusted using false discovery rate (FDR), with significant associations determined as adjusted P-values < .05 (\*\*) and suggestive associations as adjusted P-values < .1 (\*). The population for which the phenotypic association was identified is specified in parentheses.

SNPs located in that region that have adjusted  $P < .1$  ("Regional association") and the selection signature test statistics ("Selection signatures") (Figure S2). The specific population showing evidence of selection can be determined by the  $\Delta\text{ROH}_{\text{Prop}}$  or XP-EHH score. Three regions showing evidence of selection in the Swedish population are located on Chr 1 (24.0–24.1, 24.4–25.1, and 25.3–25.9 Mb; 17 genes), each harboring several interesting candidate genes. The *LDLRAD4* (low-density lipoprotein receptor class A domain containing 4) gene inhibits transforming growth factor- $\beta$  signaling<sup>43</sup> and is a putative schizophrenia-related gene.<sup>44</sup> Another growth factor-related gene in this region is *CTGF* (connective tissue growth factor). Other candidates for genes under selection in this region are the G-protein-associated melanocortin receptor genes *MC2R* and *MC5R*. *MC2R* (also known as the adrenocorticotrophic hormone receptor gene, *ACTHR*) is a major modulator of glucocorticoid secretion regulation. *MC5R* has been associated with a range of phenotypes, including shedding and fur length in dogs,<sup>45</sup> fatness in pigs, reviewed by Reference 46, and psychiatric disorders in humans.<sup>47</sup> It was also differentially expressed in the brains of aggressive and tame foxes.<sup>48</sup> These reported associations with different traits highlight one of the difficulties in identifying phenotypic targets of selection. In our analysis, we found no significant associations (FDR-adjusted  $P < .05$ ) between any of the selection signatures on Chr 1 with behavior traits, coat color, or coat length, but there was a suggestive association (FDR-adjusted  $P < .1$ ) with chasing behavior in the UK population (Table 2). Regarding fur shedding, GSDs as a breed are considered to be shedders, making it unlikely that there are large differences between the two populations for this trait.

Regions showing evidence of selection in the UK population are located on Chr 24 and 32. The Chr 24 candidate region under selection (22.9–23.8 Mb; 18 genes) in the UK population comprises well-known genes associated with black-and-tan and saddle-tan coat color in dogs (*ASIP*, *RALY*).<sup>49,50</sup> We found highly significant associations in between coat color and SNPs in this region showing evidence of selection (Table 2, Figure S2). The saddle and tan/black and tan coat color was the dominant coat color in the UK GSDs while sable was predominant in the Swedish population (Table S1). The region on Chr 32 (5.4–5.7 Mb; 3 genes) encompasses two behavior- and growth-related candidate genes: *PRKG2* and *RASGEF1B*. *RASGEF1B* (RasGEF domain family member 1B) has been identified as a positional candidate gene for dog rivalry in a genome-wide association study across multiple dog breeds.<sup>51</sup> Several case studies have been carried out in humans on chromosomal diseases related to a microdeletion of loci homologous to the region on Chr 4 comprising the *PRKG2* and *RASGEF1B* genes.<sup>52–54</sup> The loss of these genes leads to growth restriction, aggression, self-injurious behaviors and mental retardation in affected individuals. The association analysis revealed a significant association between SNPs in this region and aggressive behavior toward strangers in the Swedish GSD population and *PRKG2* has previously been reported as a top candidate gene for anxiety in mice.<sup>55</sup> However, the region on Chr 32 is in close proximity to the *BMP3* gene associated with skull morphology<sup>56</sup> and the *FGF5*<sup>2</sup> gene associated with coat length in dogs. Regarding *BMP3*, differences in skull morphology have not previously been identified in GSDs nor have they



**FIGURE 4** Circos plot for signatures of selection between GSD populations. The plot shows the three statistics used to identify regions under differential selection: differences between runs of homozygosity ( $\Delta\text{ROH}_{\text{Prop}}$ , outer circle, and blue track), cross-population extended haplotype homozygosity (XP-EHH, middle circle, green track), and the fixation index ( $F_{\text{ST}}$ , inner circle, purple track). The plot indicates concordant evidence in regions on Chr 1, 24, and 32, where peaks can be seen based on all three methods (although not within the top 1% of SNPs for XP-EHH, shown in red for the three methods)

been shown to carry a derived allele in this gene previously associated with brachycephaly,<sup>56</sup> thus selection on skull morphology seems unlikely. However, we also found a highly significant association with coat length in both populations (Table 2, Figure S2), suggesting that this trait drives the selection signature on Chr 32 (via *FGF5*).

## 2.4 | Which traits are under selection?

One of the main difficulties in interpreting genomic selection signatures is the identification of the actual trait(s) under selection. In dogs, the traits under selection are assumed to be primarily related to physical traits (eg, skull shape, coat color, body size) and/or behavior.<sup>57</sup> While between-breed studies have greatly contributed to the understanding of the genetic control of physical traits,<sup>11,58</sup> addressing behavior genetics by performing across-breed selection signature analyses is likely to be challenging because breeds differ in multiple characteristics, including both behavior and these physical traits, many of which show Mendelian inheritance and thus tend to show very strong signals.

We employed several approaches to characterize the relationship between the detected selection signatures and phenotypic traits that were recorded for these populations. First, we repeated the ADMIXTURE analysis using only genotypes from SNPs identified as selection

signatures (Figure S1) and fitted the ancestry assignment probabilities to the three individual clusters that were detected as factors in linear models for the phenotypes. We observed significant associations between UK (primarily associated with cluster 1) and Swedish (cluster 3) ancestries and some behavior traits (Stranger-directed interest, Dog-directed fear) (Table S4). Furthermore, highly significant associations were identified between the ancestries and other dog characteristics, including the function of the dog (working, pet or show dog), coat length, and coat color (Table S4). These results demonstrate a statistical association between these phenotypes and the dog's genotypes in the selection signature regions.

We then performed association analyses for behavior traits, coat length and coat color within each population only for markers within selection signature regions. We identified 87 SNPs with FDR-adjusted  $P < .05$  associated with coat length, coat color, human-directed playfulness, stranger-directed aggression, stranger directed fear, and dog-directed fear (Table S5) in at least one of the populations. The striking significant associations for coat color (lowest FDR-adjusted  $P = 3.37 \times 10^{-14}$ ) and coat length (lowest FDR-adjusted  $P = 1.13 \times 10^{-25}$ ), comprising regions on Chr 24 and 32, respectively, have previously been identified for these traits<sup>49,59-61</sup> (Table 2).

As discussed above, previous studies on selection signatures in dogs have generally focused on inter-breed or dog-wolf comparisons and primarily detected selection signatures (and thus candidate genes)

for physical features, for example, body size, coat characteristics, and skeletal morphology.<sup>2,11,58</sup> Some studies, however, also identified signatures for neural crest development<sup>1</sup> or brain function and nervous system development,<sup>9</sup> which might be relevant for behavior especially in regard to domestication. We compiled a list of candidate genes reported in previous genomic analyses of phenotype associations and selection signatures in canids (dogs, wolves, foxes) focused on morphology and behavior and compared them with genes located in regions showing evidence of selection in our study (Table S6, note that the number of overlapping genes is not informative for identifying the trait under selection because the number of reported candidate genes differs substantially between studies). The biological functions of genes in common between the two lists are diverse and include a number of genes that have been associated with behavior. Major candidate genes for physical features in dogs, for example, *IGF1*, *SMAD2*, *FGF5*, and *BMP3*, as reviewed in Reference 7, were not detected within selection signatures in our study. However, *FGF5*, which has previously been associated with coat length, is located in close proximity to the selection signature on Chr 32 and we detected a highly significant association with coat length for this region (*BMP3*, associated with skull morphology, is also located near this region, but as discussed above, our data does not support a signature of selection associated with this trait). We also detected well-described genes associated with coat color (Chr 24: *ASIP*, *RALY*). Together these results suggest that selection for morphological traits (coat length and coat color) has driven differences between the two populations in the genomic regions on Chr 24 and 32. In contrast, the region we detected on Chr 1 showed an association with Chasing in the UK population and comprises candidate genes with functions in behavior, but was not associated with morphological traits that we measured. Moreover, some of the selection signature regions showed associations with both morphological and behavior traits, for example, the region on Chr 32 was associated with both Stranger-directed aggression and coat length in the Swedish population (Table 2). Furthermore, genes associated with physical appearance like *ASIP* have previously been associated with behavior traits, for example, social behavior in mice.<sup>62</sup> Thus, it is possible that some of the selection signatures we detected are also associated with multiple traits.

## 2.5 | Limitations of the study

By comparing UK and Swedish GSDs, we hypothesized that we would be able to detect selection signatures for behavior because behavior was the main selection target in the Swedish population. However, we found that the geographical origin of the dogs was confounded with other attributes, for example, coat color and length. We addressed the issue of which trait(s) were under selection by characterizing the relationship between selection signatures and associations with phenotypic attributes (behavior, coat length, coat color), recognizing that the sample size for the association analyses within populations was small and therefore these results should be interpreted with caution. In addition, measurements on other

morphological traits (eg, body size and weight) were not available, but these might also be under selection and should be considered in future studies. We conclude that our study of GSD has identified selection signatures probably driven by selection for coat color and length (eg, at the *ASIP* and *FGF5* genes) as well as other signatures that may be related to differential selection for behavior between the Swedish and UK populations. Functional analyses are needed to test whether the identified candidate genes within regions showing evidence of selection do influence dog behavior characteristics.

## 3 | MATERIAL AND METHODS

### 3.1 | SNP genotyping and quality control

DNA was extracted from saliva samples collected with Performagene PG-100 swabs (UK population) or blood samples (Swedish population). The genotyping was performed using the CanineHD Whole-Genome Genotyping BeadChip<sup>63</sup> featuring 172 115 SNPs. The data were filtered for sample call rate of >90%, SNP call rate > 98%, reproducibility (GTS) > 0.6 and low or confounded signal characterized by AB R mean (mean normalized intensity of the AB cluster) > 0.3 in GenomeStudio version 2.0. Minor allele frequency filtering of >0.01 was used to include rare but informative variants, leaving a final data set of 108 817 SNPs for analyses. Genotype information was available for 741 GSDs. Following further sample-based quality control, closely related dogs were removed following the procedure described in Chen et al.<sup>64</sup> Briefly, a pruned genotype data set to remove closely related dogs was created for SNPs with MAF > 0.05 using PLINK version 1.9<sup>65</sup>: based on the variance inflation factor, a function of the multiple correlation coefficient of a given SNP regressed on all other SNPs within a window (using default parameters: window size = 50 SNPs, overlapping SNPs for shifting windows = 5, the variance inflation factor threshold = 2). Then, GCTA version 1.24.7<sup>66</sup> was used to compute the genetic relationship matrix and to remove one dog per pair with a genetic relationship higher than 0.2 (equivalent to second degree or closer relatives) leaving a final set of 182 UK and 68 Swedish GSDs for subsequent analyses.

### 3.2 | Samples and phenotypes

The GSDs used in this analysis originated from the UK and Sweden. For the UK population, GSDs that were at least 2 years old and registered with the UK Kennel Club were recruited via email to participate in a study on behavior genetics.<sup>14,67</sup> GSDs from the UK population were bred by multiple breeders and primarily were pet dogs. All GSDs from the Swedish population were bred within the breeding program of the SAF starting in 2004 with the purpose of becoming working dogs. The strongest systematic selection pressure in the SAF breeding program is for behavior traits. Briefly, puppies were raised at the SAF, weaned at the age of 8 weeks and then fostered by members of the Swedish public.<sup>68</sup> After a behavior test at the age of 15 to 18 months,

some dogs started working with the SAF, Swedish Police or other authorities and companies, and/or were selected as breeding animals, whereas others were kept as pet dogs. For the Swedish population, owners, trainers, or handlers of GSDs bred within the breeding program of the SAF were invited via email or letter to participate in the study. Several phenotypes were analyzed. Data on GSD behavior was assessed using the Canine Behavior and Research Questionnaire (C-BARQ).<sup>69</sup> The C-BARQ consists of questions related to training and obedience, aggression, fear and anxiety, separation-related behavior, excitability, attachment, and attention seeking, and miscellaneous behaviors. To calculate the behavior traits, a PCA was applied to the data to condense the questions to a smaller number of 13 components, as described in Reference 14. The dogs' scores for the 13 components, adjusted for fixed effects (excluding cohort) as described in Reference 67, were considered as adjusted behavior traits in the subsequent analyses. Other dog characteristics (eg, sex, coat color, coat length, role) were assessed using a lifestyle survey.<sup>14</sup> Summary statistics for behavior traits and other characteristics within the two GSD populations are given in Table S1.

### 3.3 | Genomic structure of populations

To characterize the genomic structure of the GSD populations, a PCA and a cluster analysis were performed. PLINK version 1.9<sup>65</sup> with default parameters was used to create a pruned SNP data set with reduced LD between SNPs, leaving a pruned data set of 9180 SNPs. This dataset was employed only to characterize the genomic structure of populations, via PCA and ADMIXTURE analyses. The PCA was performed in PLINK version 1.9<sup>65</sup> and ancestry estimation was performed using ADMIXTURE version 1.3.0.<sup>15</sup> The best number of clusters (K) was determined by comparing five-fold cross-validation (CV) errors.

Inbreeding, heterozygosity and nucleotide diversity were calculated within both GSD populations on the final data set of 108 817 SNPs. To determine inbreeding coefficients based on runs of homozygosity ( $F_{ROH}$ ), runs of homozygosity (ROH) were computed in PLINK version 1.9<sup>65</sup> using the default settings of a ROH length of 1000 kb and a window size of 65 SNPs, as in Pfahler and Distl.<sup>70</sup> The inbreeding was then estimated as the individual's total ROH length divided by the total genome length. ROH-based methods have been shown to perform best in relation to the true inbreeding.<sup>71</sup> Finally, nucleotide diversity (Nei's  $\mu$ ) was calculated per SNP using the  $-pi$  specifier in VCFtools.<sup>72</sup>

### 3.4 | Identification of selection signatures

#### 3.4.1 | Within populations

Signatures of selection within the two GSD populations were identified using the iHS statistic, which measures the EHH in the genome as an indicator of selective sweeps. The iHS statistic is based on the integrated EHH (iHH<sub>i</sub>), which is the integral of the observed decay of EHH away from a specified core allele  $i$  until the EHH reaches a specified

cutoff. Phased genotypes of the final SNP data set generated by Beagle version 4.1<sup>73</sup> (the phasing in Beagle was performed without specifying a reference population) were used to compute the SNP-wise iHS statistic using hapbin,<sup>74</sup> specifying that the iHH should be calculated up to the point at which EHH drops below 0.05 (–cutoff 0.05). As in Voight et al,<sup>17</sup> the standardized iHS (iHS) for a SNP was calculated as

$$iHS = \frac{\text{unstandardized iHS} - \mu_{\text{unstandardized iHS}}}{\sigma_{\text{unstandardized iHS}}}$$

where the *unstandardized iHS* is  $\ln(iHH_i/iHH_j)$  for alleles  $i$  and  $j$ , and  $\mu$  and  $\sigma$  are the mean and the SD of the unstandardized iHS estimated from the empirical distribution of SNPs for which the derived allele frequency matches the frequency at the core SNP.

#### 3.4.2 | Between populations

To detect divergent signatures of selection between populations, three different approaches were used: the fixation index ( $F_{ST}$ ), XP-EHH, and differences between runs of homozygosity (ROH).

First, the  $F_{ST}$  analysis was performed using the script described in Talenti et al.<sup>75</sup> The  $F_{ST}$  between UK and Swedish dogs was calculated for each SNP according to the formula reported by Karlsson et al,<sup>76</sup> which is a comparison of the allele frequencies between populations:

$$F_{ST} = \frac{f_1^{UK} (f_2^S - f_2^{UK}) + f_1^S (f_2^{UK} - f_2^S)}{(f_1^{UK} * f_2^S) + (f_2^{UK} * f_1^S)}$$

where  $f_1^{UK}$  and  $f_2^{UK}$  are frequencies in the UK population for the two alleles and  $f_1^S$  and  $f_2^S$  are allele frequencies in the Swedish population. Next, the mean  $F_{ST}$  was calculated in 1 Mb sliding windows (window-based  $F_{ST}$ ) with an overlap between windows of 500 kb, resulting in each SNP being located in exactly one or two windows. To derive a SNP-based value (to select the top 1% for calculating the intersection with other methods as described below), we averaged the window-based  $F_{ST}$  for the one or two windows in which the SNP was found.

Second, the XP-EHH statistic<sup>77</sup> was calculated to compare the EHH between populations, that is, whether alleles are homozygous in one population and polymorphic in the other population. The XP-EHH statistic was calculated for the UK and Swedish populations using phased haplotypes generated by Beagle version 4.1<sup>73</sup> in hapbin,<sup>74</sup> as described above.

For the third approach, ROH were computed in PLINK version 1.9.<sup>65</sup> We ran the analysis with the default settings of a ROH length of 1000 kb and a window size of 65 SNPs, as described above.<sup>70</sup> For every SNP, a homozygosity score ( $ROH_{prop}$ ) was calculated by dividing the number of dogs with a ROH at a specific SNP by the total number of dogs, such that  $ROH_{prop}$  ranges from 0 to 1, as described in Bertolini et al.<sup>78</sup> The absolute difference between  $ROH_{prop}$  between populations ( $\Delta ROH_{prop}$ ) was used as statistic to determine which

ROH are highly represented in one population but underrepresented in the other population. Therefore, for every SNP,  $\Delta\text{ROH}_{\text{Prop}}$  values were calculated to identify ROH that are present in the majority of dogs in one population but not in the other.

### 3.4.3 | Gene identification and Gene ontology analysis

To detect putative genomic regions showing evidence of selection, the most extreme values from the test statistics were selected for both the within- and between-population analyses to define selection signatures. For iHS, SNPs belonging to the top 0.5% of the distribution were selected. For  $F_{\text{ST}}$ , XP-EHH and  $\Delta\text{ROH}_{\text{Prop}}$ , the top 1% of each test distribution was selected and the overlap between these top SNPs was determined to identify SNPs that had most extreme values for at least two of the three methods, to reduce the chance of false positive signals. We chose a less stringent threshold for top SNPs for between-population statistics to allow for greater overlap since the three approaches differ in their methodologies and thus the ranking of top SNPs will vary. For a visual representation of target regions under selection between populations, the visualization tool Circos<sup>79</sup> was used. For every SNP, the  $\Delta\text{ROH}_{\text{Prop}}$  and XP-EHH scores were plotted. Since the  $F_{\text{ST}}$  was calculated as a window-based average and Circos required a SNP-based value, we averaged the window-based  $F_{\text{ST}}$  for the one or two window in which the SNP was found, as described above.

The pairwise distances between the top SNPs were calculated and SNPs located within 200 kb were merged into a region. The distance of 200 kb was determined based on the LD in the genome. First, the squared correlation ( $r^2$ ) between all pairs of SNPs within 10 Mb was calculated in PLINK version 1.9.<sup>65</sup> The average  $r^2$  was then calculated for bins of increasing distance between SNPs to identify the distance around SNPs at which average  $r^2$  drops below 0.5. The longest bin for which average  $r^2 \geq 0.5$  was 200 kb.

To characterize functional relevance of regions showing evidence of selection, the top SNPs or regions (if multiple SNPs were found within 200 kb) were annotated for genes based on the CanFam3.1 genome assembly,<sup>80</sup> using BEDtools 2.27 software.<sup>81</sup> SNPs were annotated considering a flanking region of  $\pm 40$  kb, chosen based on the average between-marker distance of the array (20 kb), which was doubled to account for nonevenly spaced SNPs and SNPs lost through quality-control filtering. The genes detected for these selection signatures were then submitted to Enrichr<sup>27,28</sup> to perform gene set enrichment analyses. Enrichr is an integrative web-based application that compares submitted gene lists to various gene-set libraries; the standard Fisher exact test option was used to calculate  $P$ -values for this study.

### 3.5 | Characterizing trait(s) under selection

We employed two approaches to gain insights into the trait(s) under selection, as detected as genomic selection signatures: (I) we modeled

behavior traits and other dog characteristics as a function of the dog's ancestry based on selection signature regions and (II) we analyzed the association within each population between these traits and SNP markers in these regions. For both approaches, we compiled a genotype data set of SNPs within the regions showing evidence of selection; this included SNPs belonging to the top 0.5% of the iHS distribution in UK and Swedish populations and SNPs belonging to the top 1% of  $F_{\text{ST}}$ , XP-EHH, and  $\Delta\text{ROH}_{\text{Prop}}$  distributions that overlapped between at least two methods.

For (I), we repeated the ADMIXTURE analysis as described above, but only used genotypes of SNPs from putatively selected regions to estimate the ancestry. Then, a linear regression was performed, as described in Reference 82, to model the relationship between the traits and ancestry assignment probabilities.

For (II), we analyzed the association between the traits and SNP markers within the regions showing evidence of selection, separately for each population. Behavior traits were adjusted based on other fixed effects as defined in the previous study<sup>67</sup> and treated as quantitative traits, while coat color ("saddle tan," "sable," "black," "other") and coat length ("long," "short") were treated as categorical traits and not corrected for environmental factors. The association analysis was performed using GEMMA,<sup>83</sup> fitting the genomic relationship matrix (based on 108 817 genome-wide SNPs) as a random effect to account for population stratification. To correct for multiple testing,  $P$ -values were adjusted using the false discovery rate (FDR).

### ACKNOWLEDGMENTS

The authors want to thank all owners of German Shepherd dogs participating in this study for their time and effort to answer the questionnaires and send saliva samples for genotyping. Thanks are also extended to the Kennel Club, the British Association for German Shepherd Dogs, and the German Shepherd Dog Breed Council of Great Britain for assistance in participant recruitment for the UK cohort. Thanks to Zita Polgar, Carol-Anne Duthie, and Joanna Warner for assistance in contacting dog owners. We would also like to thank the SAF Dog Training Centre, in particular Lisa Rutström, for recruiting participants for the Swedish cohort, and Susanne Gustafsson and Gabriela Bottani Claros (Swedish University of Agricultural Sciences) for extracting the DNA. We are very thankful to Lee Murphy and his colleagues at the Edinburgh Clinical Research Facility (University of Edinburgh) for DNA extraction and genotyping of the UK dogs. Further thanks to Vanessa Kapsona for conducting preliminary population structure analyses. Primary funding was provided by the Dogs Trust (UK); further funding was provided by BBSRC Institute Strategic Programme Grants (to the Roslin Institute; BBS/E/D/30002276, BBS/E/D/10002070) and RESAS, Scottish Government (to SRUC).

### CONFLICT OF INTEREST

The authors declare no conflict of interest.

### AUTHOR CONTRIBUTIONS

**Juliane Friedrich:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Writing-original

draft, Writing-review, and editing. **Andrea Talenti**: Methodology, Software, Visualization, Writing-review, and editing. **Per Arvelius**: Data curation, Project administration, Resources, Supervision, Writing-review, and editing. **Erling Strandberg**: Data curation, Project administration, Resources, Supervision, Writing-review, and editing. **Marie J. Haskell**: Conceptualization, Funding acquisition, Project administration, Supervision, Writing-review, and editing. **Pamela Wiener**: Conceptualization, Funding acquisition, Project administration, Resources, Supervision, Writing-review, and editing.

## DATA AVAILABILITY

Genotype and phenotype data for the UK dogs is available under CC-BY license from the Dryad Digital Repository.<sup>84</sup> The data for the Swedish dogs are restricted by the SAF for reasons of national security.

## ORCID

Juliane Friedrich  <https://orcid.org/0000-0003-2253-417X>

Andrea Talenti  <https://orcid.org/0000-0003-1309-3667>

Erling Strandberg  <https://orcid.org/0000-0001-5154-8146>

Marie J. Haskell  <https://orcid.org/0000-0001-9373-0624>

Pamela Wiener  <https://orcid.org/0000-0001-6129-5795>

## REFERENCES

- Pendleton AL, Shen F, Taravella AM, et al. Comparison of village dog and wolf genomes highlights the role of the neural crest in dog domestication. *BMC Biol*. 2018;16:64.
- Akey JM, Ruhe AL, Akey DT, et al. Tracking footprints of artificial selection in the dog genome. *Proc Natl Acad Sci U S A*. 2010;107(3):1160-1165.
- Larson G, Karlsson EK, Perri A, et al. Rethinking dog domestication by integrating genetics, archeology, and biogeography. *Proc Natl Acad Sci U S A*. 2012;109(23):8878-8883.
- Botigué LR, Song S, Scheu A, et al. Ancient European dog genomes reveal continuity since the early Neolithic. *Nat Commun*. 2017;18(8):16082.
- Mehrkam LR, Wynne C. Behavioral differences among breeds of domestic dogs (*Canis lupus familiaris*): current status of the science. *Appl Anim Behav Sci*. 2014;155:12-27.
- Lewis TW, Wiles BM, Llewellyn-Zaidi AM, Evans KM, O'Neill DG. Longevity and mortality in kennel Club registered dog breeds in the UK in 2014. *Canine Genet Epidemiol*. 2018;5(1):10.
- Schoenebeck JJ, Ostrander EA. Insights into morphology and disease from the dog genome project. *Annu Rev Cell Dev Biol*. 2014;30(1):535-560.
- Nielsen R. Molecular signatures of natural selection. *Annu Rev Genet*. 2005;39(1):197-218.
- Axelsson E, Ratnakumar A, Arendt M-L, et al. The genomic signature of dog domestication reveals adaptation to a starch-rich diet. *Nature*. 2013;495(7441):360-364.
- Kim J, Williams FJ, Dreger DL, et al. Genetic selection of athletic success in sport-hunting dogs. *Proc Natl Acad Sci U S A*. 2018;115(30):E7212-E7221.
- Plassais J, Kim J, Davis BW, et al. Whole genome sequencing of canids reveals genomic regions under selection and variants influencing morphology. *Nat Commun*. 2019;10(1):1489.
- Ostrander EA, Wayne RK, Freedman AH, Davis BW. Demographic history, selection and functional diversity of the canine genome. *Nat Rev Genet*. 2017;18(12):705-720.
- Lord K, Schneider RA, Coppinger R. Evolution of working dogs. *The Domestic Dog: Its Evolution, Behavior and Interactions with People*. Cambridge, England: Cambridge University Press; 2016. /core/books/domestic-dog/evolution-of-working-dogs/CC5083D37F741470DDFA69AFBB238AB1.
- Friedrich J, Arvelius P, Strandberg E, Polgar Z, Wiener P, Haskell MJ. The interaction between behavioural traits and demographic and management factors in German Shepherd dogs. *Appl Anim Behav Sci*. 2018;211:67-76. <http://www.sciencedirect.com/science/article/pii/S0168159118303265>.
- Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res*. 2009;19(9):1655-1664.
- Vitti JJ, Grossman SR, Sabeti PC. Detecting natural selection in genomic data. *Annu Rev Genet*. 2013;47:97-120.
- Voight BF, Kudaravalli S, Wen X, Pritchard JK. A map of recent positive selection in the human genome. *PLoS Biol*. 2006;4(3):e94. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1382018/>.
- Friedrich J, Brand B, Ponsuksili S, et al. Detection of genetic variants affecting cattle behaviour and their impact on milk production: a genome-wide association study. *Anim Genet*. 2016;47(1):12-18.
- Crouch DJM, Winney B, Koppen WP, et al. Genetics of the human face: identification of large-effect single gene variants. *Proc Natl Acad Sci U S A*. 2018;115(4):E676-E685.
- Morgan MD, Pairo-Castineira E, Rawlik K, et al. Genome-wide study of hair colour in UK biobank explains most of the SNP heritability. *Nat Commun*. 2018;9(1):5271.
- Valente TS, Baldi F, Sant'Anna AC, Albuquerque LG, da Costa MJRP. Genome-wide association study between single nucleotide polymorphisms and flight speed in Nellore cattle. *PLoS ONE*. 2016;11(6):e0156956.
- Luciano M, Huffman JE, Arias-Vásquez A, et al. Genome-wide association uncovers shared genetic effects among personality traits and mood states. *Am J Med Genet B Neuropsychiatr Genet*. 2012;159B(6):684-695.
- Smith EN, Bloss CS, Badner JA, et al. Genome-wide association study of bipolar disorder in European American and African American individuals. *Mol Psychiatry*. 2009;14(8):755-763.
- Wang K-S, Liu X-F, Aragam N. A genome-wide meta-analysis identifies novel loci associated with schizophrenia and bipolar disorder. *Schizophr Res*. 2010;124(1):192-199.
- Sun J, Kuo P-H, Riley BP, Kendler KS, Zhao Z. Candidate genes for schizophrenia: a survey of association studies and gene ranking. *Am J Med Genet B Neuropsychiatr Genet*. 2008;147B(7):1173-1181.
- Ilksa J, Haskell MJ, Blott SC, et al. Genetic characterisation of dog personality traits. *Genetics*. 2017;116:192674.
- Chen EY, Tan CM, Kou Y, et al. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinform*. 2013;14:128.
- Kuleshov MV, Jones MR, Rouillard AD, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res*. 2016;44(W1):W90-W97.
- González-Maeso J, Meana JJ. Heterotrimeric G proteins: insights into the neurobiology of mood disorders. *Curr Neuropharmacol*. 2006;4(2):127-138.
- Lipsky RH, Marini AM. Brain-derived neurotrophic factor in neuronal survival and behavior-related plasticity. *Ann N Y Acad Sci*. 2007;1122(1):130-143.
- Lüscher C, Malenka RC. NMDA receptor-dependent long-term potentiation and long-term depression (LTP/LTD). *Cold Spring Harb Perspect Biol*. 2012;4(6):a005710. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3367554/>.
- Pasterkamp RJ, Giger RJ. Semaphorin function in neural plasticity and disease. *Curr Opin Neurobiol*. 2009;19(3):263-274.

33. Jacobsen JPR, Medvedev IO, Caron MG. The 5-HT deficiency theory of depression: perspectives from a naturalistic 5-HT deficiency model, the tryptophan hydroxylase 2Arg439His knockin mouse. *Philos Trans R Soc Lond Ser B Biol Sci*. 2012;367(1601):2444-2459.
34. de Almeida RMM, Ferrari PF, Parmigiani S, Miczek KA. Escalated aggressive behavior: dopamine, serotonin and GABA. *Eur J Pharmacol*. 2005;526(1):51-64.
35. Quignon P, Herbin L, Cadieu E, et al. Canine population structure: assessment and impact of intra-breed stratification on SNP-based association studies. *PLoS ONE*. 2007;2(12):e1324. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2129117/>.
36. Moriya S, Soga T, Wong DW, Parhar IS. Transcriptome composition of the preoptic area in mid-age and escitalopram treatment in male mice. *Neurosci Lett*. 2016;622:67-71.
37. Wiener P, Sánchez-Molano E, Clements DN, Woolliams JA, Haskell MJ, Blott SC. Genomic data illuminates demography, genetic structure and selection of a popular dog breed. *BMC Genomics*. 2017;18:609.
38. Knutson DC, Mitzey AM, Talton LE, Clagett-Dame M. Mice null for NEDD9 (HEF1) display extensive hippocampal dendritic spine loss and cognitive impairment. *Brain Res*. 2016;1632:141-155.
39. Patel MM, Silasi-Mansat R, Keshari RS, et al. Role of androgen dependent TFPI-regulating protein (ADTRP) in vascular development and function. *Blood*. 2016;128(22):556-556.
40. Li M, Han Y, Zhou H, et al. Transmembrane protein 170B is a novel breast tumorigenesis suppressor gene that inhibits the Wnt/ $\beta$ -catenin pathway. *Cell Death Dis*. 2018;9(2):91. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5833782/>.
41. Valnegri P, Huang J, Yamada T, et al. RNF8/UBC13 ubiquitin signaling suppresses synapse formation in the mammalian brain. *Nat Commun*. 2017;8(1):1271.
42. Chen C, Chen C, Moyzis R, et al. Contributions of dopamine-related genes and environmental factors to highly sensitive personality: a multi-step neuronal system-level approach. *PLoS ONE*. 2011;6(7):e21636.
43. Nakano N, Maeyama K, Sakata N, et al. C18 ORF1, a novel negative regulator of transforming growth factor- $\beta$  signaling. *J Biol Chem*. 2014;289(18):12680-12692.
44. Meerabux JMA, Ohba H, Iwayama Y, et al. Analysis of a t(18;21)(p11.1;p11.1) translocation in a family with schizophrenia. *J Hum Genet*. 2009;54(7):386-391.
45. Hayward JJ, Castelhana MG, Oliveira KC, et al. Complex disease and phenotype mapping in the domestic dog. *Nat Commun*. 2016;7:10460.
46. Switonski M, Mankowska M. Dog obesity – the need for identifying predisposing genetic markers. *Res Vet Sci*. 2013;95(3):831-836.
47. Miller CL, Murakami P, Ruczinski I, et al. Two complex genotypes relevant to the kynurenine pathway and melanotropin function show association with schizophrenia and bipolar disorder. *Schizophr Res*. 2009;113(2):259-267.
48. Wang X, Pipes L, Trut LN, et al. Genomic responses to selection for tame/aggressive behaviors in the silver fox (*Vulpes vulpes*). *Proc Natl Acad Sci U S A*. 2018;115(41):10398-10403.
49. Dreger DL, Schmutz SM. A SINE insertion causes the black-and-Tan and saddle Tan phenotypes in domestic dogs. *J Hered*. 2011;102(Suppl\_1):S11-S18.
50. Dreger DL, Parker HG, Ostrander EA, Schmutz SM. Identification of a mutation that is associated with the saddle tan and black-and-tan phenotypes in basset hounds and Pembroke welsh corgis. *J Hered*. 2013;104(3):399-406.
51. Zapata I, Serpell JA, Alvarez CE. Genetic mapping of canine fear and aggression. *BMC Genomics*. 2016;17:572.
52. Bonnet C, Andrieux J, Béri-Dexheimer M, et al. Microdeletion at chromosome 4q21 defines a new emerging syndrome with marked growth restriction, mental retardation and absent or severely delayed speech. *J Med Genet*. 2010;47(6):377-384.
53. Bhoj E, Halbach S, McDonald-McGinn D, et al. Expanding the spectrum of microdeletion 4q21 syndrome: a partial phenotype with incomplete deletion of the minimal critical region and a new association with cleft palate and Pierre Robin sequence. *Am J Med Genet A*. 2013;161A(9):2327-2333.
54. Fee A, Noble N, Valdovinos MG. Functional analysis of phenotypic behaviors of a 5-year-old male with novel 4q21 microdeletion. *J Pediatr Neuropsychol*. 2015;1(1):36-41.
55. Le-Niculescu H, Balaraman Y, Patel SD, et al. Convergent functional genomics of anxiety disorders: translational identification of genes, biomarkers, pathways and mechanisms. *Transl Psychiatry*. 2011;1(5):e9.
56. Schoenebeck JJ, Hutchinson SA, Byers A, et al. Variation of BMP3 contributes to dog breed skull diversity. *PLoS Genet*. 2012;8(8):e1002849.
57. Rimbault M, Ostrander EA. So many doggone traits: mapping genetics of multiple phenotypes in the domestic dog. *Hum Mol Genet*. 2012;21(R1):R52-R57.
58. Vaysse A, Ratnakumar A, Derrien T, et al. Identification of genomic regions associated with phenotypic variation between dog breeds using selection mapping. *PLoS Genet*. 2011;7(10):e1002316.
59. Legrand R, Tired L, Abitbol M. Two recessive mutations in FGF5 are associated with the long-hair phenotype in donkeys. *Genet Sel Evol*. 2014;46(1):65. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4175617/>.
60. Housley DJE, Venta PJ. The long and the short of it: evidence that FGF5 is a major determinant of canine 'hair'-itability. *Anim Genet*. 2006;37(4):309-315.
61. Cadieu E, Neff MW, Quignon P, et al. Coat variation in the domestic dog is governed by variants in three genes. *Science*. 2009;326(5949):150-153.
62. Carola V, Perlas E, Zonfrillo F, Soini HA, Novotny MV, Gross CT. Modulation of social behavior by the agouti pigmentation gene. *Front Behav Neurosci*. 2014;8:259. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4117936/>.
63. Illumina I: Canine HD BeadChip. In Data Sheet: DNA Genotyping; 2010. [https://www.illumina.com/documents/products/datasheets/datasheet\\_caninehd.pdf](https://www.illumina.com/documents/products/datasheets/datasheet_caninehd.pdf)
64. Chen M, Wang J, Wang Y, Wu Y, Fu J, Liu J. Genome-wide detection of selection signatures in Chinese indigenous Laiwu pigs revealed candidate genes regulating fat deposition in muscle. *BMC Genet*. 2018;19:64. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5960162/>.
65. Purcell SM, Chang CC. PLINK 1.9. [www.cog-genomics.org/plink/1.9/](http://www.cog-genomics.org/plink/1.9/)
66. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet*. 2011;88(1):76-82.
67. Friedrich J, Strandberg E, Arvelius P, et al. Genetic dissection of complex behaviour traits in German shepherd dogs. *Heredity*. 2019;14:1-13.
68. Wilsson E, Sinn DL. Are there differences between behavioral measurement methods? A comparison of the predictive validity of two ratings methods in a working dog program. *Appl Anim Behav Sci*. 2012;141(3-4):158-172.
69. Hsu Y, Serpell JA. Development and validation of a questionnaire for measuring behavior and temperament traits in pet dogs. *J Am Vet Med Assoc*. 2003;223(9):1293-1300.
70. Pfahler S, Distl O. Effective population size, extended linkage disequilibrium and signatures of selection in the rare dog breed Lundehund. *PLoS ONE*. 2015;10(4):e0122680. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4393028/>.
71. Forutan M, Ansari Mahyari S, Baes C, Melzer N, Schenkel FS, Sargolzaei M. Inbreeding and runs of homozygosity before and after genomic selection in north American Holstein cattle. *BMC Genomics*. 2018;19(1):98.

72. Danecek P, Auton A, Abecasis G, et al. The variant call format and VCFtools. *Bioinformatics*. 2011;27(15):2156-2158.
73. Browning SR, Browning BL. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am J Hum Genet*. 2007;81(5):1084-1097.
74. Maclean CA, Chue Hong NP, Prendergast JGD. Hapbin: an efficient program for performing haplotype-based scans for positive selection in large genomic datasets. *Mol Biol Evol*. 2015;32(11):3027-3029.
75. Talenti A, Bertolini F, Pagnacco G, et al. The Valdostana goat: a genome-wide investigation of the distinctiveness of its selective sweep regions. *Mamm Genome*. 2017;28(3):114-128.
76. Karlsson EK, Baranowska I, Wade CM, et al. Efficient mapping of mendelian traits in dogs through genome-wide association. *Nat Genet*. 2007;39(11):1321-1328.
77. Sabeti PC, Varilly P, Fry B, et al. Genome-wide detection and characterization of positive selection in human populations. *Nature*. 2007;449(7164):913-918.
78. Bertolini F, Gandolfi B, Kim ES, Haase B, Lyons LA, Rothschild MF. Evidence of selection signatures that shape the Persian cat breed. *Mamm Genome*. 2016;27(3):144-155.
79. Krzywinski MI, Schein JE, Birol I, et al. Circos: an information aesthetic for comparative genomics. *Genome Res*. 2009;19:1639-1645. <http://genome.cshlp.org/content/early/2009/06/15/gr.092759.109>.
80. Zerbino DR, Achuthan P, Akanni W, et al. Ensembl 2018. *Nucleic Acids Res*. 2018;46(D1):D754-D761.
81. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010;26(6):841-842.
82. Jarvis JP, Scheinfeldt LB, Soi S, et al. Patterns of ancestry, signatures of natural selection, and genetic association with stature in Western African pygmies. *PLoS Genet*. 2012;8(4):e1002641.
83. Zhou X, Stephens M. Genome-wide efficient mixed model analysis for association studies. *Nat Genet*. 2012;44(7):821-824.
84. Friedrich J, Talenti A, Arvelius P, et al. (2020), *Data from: Unravelling Selection Signatures in a Single Dog Breed Suggests Recent Selection for Morphological and Behavioural Traits*, Dryad, <https://doi.org/10.5061/dryad.g4f4qrfmr>
85. Boyko AR, Quignon P, Li L, et al. A simple genetic architecture underlies morphological variation in dogs. *PLoS Biol*. 2010;8(8):e1000451.
86. MacLean EL, Snyder-Mackler N, von Holdt BM, Serpell JA. Highly heritable and functionally relevant breed differences in dog behaviour. *Proc R Soc B Biol Sci*. 2019;286(1912):20190716.
87. Freedman AH, Schweizer RM, Vecchyo DO-D, et al. Demographically-based evaluation of genomic regions under selection in domestic dogs. *PLoS Genet*. 2016;12(3):e1005851.
88. Kukekova AV, Johnson JL, Xiang X, et al. Red fox genome assembly identifies genomic regions associated with tame and aggressive behaviours. *Nat Ecol Evol*. 2018;2(9):1479-1491.
89. Schlamp F, van der Made J, Stambler R, Chesebrough L, Boyko AR, Messer PW. Evaluating the performance of selection scans to detect selective sweeps in domestic dogs. *Mol Ecol*. 2016;25(1):342-356.
90. Saxena R, Voight BF, Lyssenko V, et al. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science*. 2007;316(5829):1331-1336.

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Friedrich J, Talenti A, Arvelius P, Strandberg E, Haskell MJ, Wiener P. Unravelling selection signatures in a single dog breed suggests recent selection for morphological and behavioral traits. *Advanced Genetics*. 2020; 1:e10024. <https://doi.org/10.1002/ggn2.10024>