frontiers
in Genetics

Whole Genome Sequencing Reveals Signatures for Artificial Selection for Different Sizes in Japanese Primitive Dog Breeds

Guangqi Lyu[1], Chunyu Feng[1], Shiyu Zhu[1], Shuang Ren[1], Wanyi Dang[1], David M. Irwin[2], Zhe Wang[1] and Shuyi Zhang[1]*

[1] College of Animal Science and Veterinary Medicine, Shenyang Agricultural University, Shenyang, China, [2] Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, ON, Canada
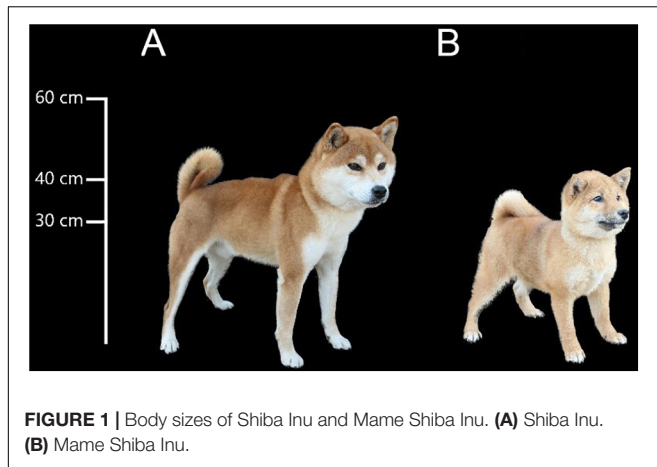
Body size is an important trait in companion animals. Recently, a primitive Japanese dog breed, the Shiba Inu, has experienced artificial selection for smaller body size, resulting in the "Mame Shiba Inu" breed. To identify loci and genes that might explain the difference in the body size of these Shiba Inu dogs, we applied whole genome sequencing of pooled samples (pool-seq) on both Shiba Inu and Mame Shiba Inu. We identified a total of 13,618,261 unique SNPs in the genomes of these two breeds of dog. Using selective sweep approaches, including F_{ST}, H_p and XP-CLR with sliding windows, we identified a total of 12 genomic windows that show signatures of selection that overlap with nine genes (*PRDM16, ZNF382, ZNF461, ERGIC2, ENSCAFG00000033351, CCDC61, ALDH3A2, ENSCAFG00000011141,* and *ENSCAFG00000018533*). These results provide candidate genes and specific sites that might be associated with body size in dogs. Some of these genes are associated with body size in other mammals, but 8 of the 9 genes are novel candidate genes that need further study.

Keywords: selective sweep, body size, dog, pool-seq, genome

INTRODUCTION

As our best friend, dogs are crucial to modern human society and are involved in many aspects of our life. During domestication, many traits found in dogs diverged greatly from their ancestors. The Shiba Inu ("inu" means dog in Japanese) is an ancient and native dog breed in Japan, and it is a basal spitz breed (Parker et al., 2017). Shiba Inus typically range in height from 34 to 41 cm (**Figure 1A**)[1] although in recent years, artificial selection has been imposed to yield smaller individuals, which are called mini or Mame Shiba Inu (Mame is a Japanese word for "bean," representing "small"). It is bred only from Shiba Inu to keep "pure ancestry," and shares some traits with Shiba Inu, such as coat color. The body height of an adult Mame Shiba Inu should be not more than 34 cm (**Figure 1B**). Although this new breed, the Mame Shiba Inu, has not yet been approved by the

---

[1] http://www.akc.org

FIGURE 1 | Body sizes of Shiba Inu and Mame Shiba Inu. **(A)** Shiba Inu. **(B)** Mame Shiba Inu.

American Kennel Club (AKC), it is approved by Kennel Club of Japan and well recognized as a breed in Japan and China.

Regulation of dog body size, or weight, in the dog has been extensively researched, with many studies conducted at the genomic level (Ostrander et al., 2017). Through these works, many genes and loci have been identified as associated with the regulation of body size in the dog. For instance, *IGF1* is a well-studied body size gene. Using a genome-wide scan in Portuguese water dogs, a single *IGF1* SNP haplotype was linked as a major contributor to small body size (Sutter et al., 2007). Consistently, results reported from a study involving 93 dog breeds found that individuals with different *IGF1* SNP and SINE insertion alleles have significant differences in body weight (Rimbault et al., 2013). An *IGF1* SNP was also reported in a study that applied both selective sweep and GWAS analyses in 722 canids, including purebred dogs and wolfs (Plassais et al., 2019). SNPs not only in *IGF1*, but also its receptor, *IGF1R* (e.g., a non-synonymous SNP at chr3: 44,706,389) was found to be associated with reduced size in dogs (Hoopes et al., 2012). GWAS studies have also implicated other genes with SNPs linked to body size. *HMGA2* was reported to be associated with body size where individuals with the ancestral SNP allele in the *HMGA2* 5′UTR have higher body weight than those with the derived allele (Rimbault et al., 2013), as well as another variation between exon 1 and 2 (Chr10: 8351907) was also reported to have an effect on body weight (Plassais et al., 2019). A genomic region within *CDK4* was reported to be associated with size through an across-breed GWAS analysis that involved 8 different dog breeds (Vaysse et al., 2011). With a genome-wide scan on two German shepherd dog families with microsatellite markers identified a contracted DNA repeat in intron 5 of *LHX3* that was associated with dwarfism (Voorbij et al., 2011). A quantitative GWAS was also used on 1,873 dogs, from 158 breeds, which identified an interval on Chromosome X that is upstream of the *ARHGAP36*, *IGSF1,* and *OR5AK2* genes, which was strongly associated with body size (Hayward et al., 2016). Using a GWAS analyses of 690 dogs, three genes (*IRS4, IGSF1,* and *ACSL4*) were found to be associated with body weight, where *IRS4* and *IGSF1* are both involved in the GH/IGF1 and thyroid hormonal pathways involved in body size regulation (Plassais et al., 2017).

While many studies have focused on the variation of body size in dogs, it is still possible that other novel genes and loci have been involved in the selection for body size in dogs. The change in body size between Shiba Inu and Mame Shiba Inu is relatively mild, but distinct, making them good source for investigating mechanisms that change body size. In addition, these breeds are highly divergent from the other domesticated dogs. Here we performed whole-genome pool-sequencing (pool-seq) on Shiba Inu and Mame Shiba Inu dogs to try to identify key regions that experienced selection and may regulate their different body sizes.

MATERIALS AND METHODS

Ethics Statement
Owners of the dogs studied here provided informed consent for the use data of their dogs in this scientific research. All procedures used in this study was approved by the Animal Care and Use Committee of Shenyang Agricultural University.

Sample Preparation and Sequencing
Blood samples from a total of 94 dogs (59 Shiba Inu and 35 Mame Shiba Inu dogs) were collected by venipuncture of their forelimb. For genome sequencing, four pooled samples were generated. The first pooled sample was from 35 Mame Shiba Inu dogs, while the other three were from 29, 22, and 8 Shiba Inu dogs. For each pool, an equivalent volume of blood from each individual of the group was mixed to yield a total volume of 1 mL. Genomic DNA was then extracted from each pool sample using a whole blood genomic DNA extraction kit (TIANGEN, DP318). Genomic DNA were checked for concentration and purity, and then sent to the Beijing Genome Institute (Beijing, China) for library construction (insert size ~ 300 bp) and genome sequencing. After adapter ligation and DNA cluster preparation, the libraries were then subjected to Illumina HiSeq X Ten for sequencing (150 bp, paired-end). To yield similar sequencing depths for the two dog breeds, the Mame Shiba Inu pool sample was sequenced to 30× depth while each of the three Shiba Inu pool samples were only sequenced to a 10× depth (thus generating 30× for the breed).

Mapping and SNP Calling
To find loci that might be associated with the body size difference between the two breeds, reads from the three Shiba Inu dog pool-seq samples were merged into one pair of fastq files. Raw reads were processed using the PoolParty pipeline (Micheletti and Narum, 2018). Briefly, raw reads were filtered using BBDuk[2], with thresholds of 20 for base quality and 25 for minimum read length. Clean reads were mapped to the reference dog genome (assembly CanFam 3.1)[3] using BWA 0.7.12-r1039 (Li and Durbin, 2009), and the BAM (Binary Alignment Map) files were sorted by coordinate using the Picard tools[4]. To eliminate bias caused by PCR during library preparation, SAMBLASTER

---

[2]http://sourceforge.net/projects/bbmap/

[3]https://www.ncbi.nlm.nih.gov/assembly/GCA_000002285.2

[4]http://broadinstitute.github.io/picard/

(Faust and Hall, 2014) was used to mark duplicates. Subsequently, SAMtools (Li et al., 2009) and BCFtools (Li, 2011) were used to call variants and generate mpileup files, which were used in the downstream analyses. SNPs with minor allele frequency (MAF) < 0.05, depth < 10, quality < 20 and located within 15 bp of an indel were discarded.

Detection of Selective Sweeps

To identify genomic regions affected by selection, we applied three types of selective sweep analyses, the fixation index (F_{ST}), the pooled heterozygosity (H_p) and the cross-population composite likelihood ratio test (XP-CLR) approaches. F_{ST} in 25-kb non-overlapped sliding windows between the two breeds was calculated using the "fst-sliding.pl" module in Popoolation2 (Kofler et al., 2011), according to the Weir and Cockerham method (Weir and Cockerham, 1984). In the case that we had two populations and their different pooling strategies, the F_{ST} estimator is:

$$\hat{F}_{ST}^{WC} = 1 - \frac{2\frac{n_1 n_2}{n_1+n_2}\frac{1}{n_1+n_2-2}[n_1\tilde{p}_1(1-\tilde{p}_1)+n_2\tilde{p}_2(1-\tilde{p}_2)]}{\frac{n_1 n_2}{n_1+n_2}(\tilde{p}_1-\tilde{p}_2)^2+\left(2\frac{n_1 n_2}{n_1+n_2}-1\right)\frac{1}{n_1+n_2-2}[n_1\tilde{p}_1(1-\tilde{p}_1)+n_2\tilde{p}_2(1-\tilde{p}_2)]} \quad (1)$$

where n_i is the sample size and $\tilde{p}_i$ is the sample allele frequency of the populations (Bhatia et al., 2013). H_p and negative Z-transformed Hp ($-ZH_p$) of the Mame Shiba Inu were calculated using a custom python3 script from 25-kb non-overlapped sliding windows using the following formulas:

$$Hp = \frac{2\sum n_{MAJ}\sum n_{MIN}}{\left(\sum n_{MAJ}+\sum n_{MIN}\right)^2}$$

$$-ZHp = -\frac{Hp-\mu Hp}{\sigma Hp}$$

Z transformation allowed us to compare the two breeds at the same level, since ZH_p values indicate the number of standard deviations by which the H_p value deviates from the mean (Rubin et al., 2010). XP-CLR statistics (Chen et al., 2010) between Mame Shiba Inu and Shiba Inu was calculated using *xpclr*[5] in 25-kb non-overlapped windows. Windows containing < 10 SNPs were discarded to prevent spurious signals. Windows that were all in the top 1% of the F_{ST}, -ZH_p and normalized XP-CLR score distributions were considered to be candidate selective sweep regions. Genes overlapping with these regions were identified using the ENSEMBL dog genome gene annotations (CanFam 3.1).

RESULTS

Sequencing Data, Mapping, and SNPs Calling

In total, we obtained 192.5 Gb of clean sequence data with an average sequencing depth of $39.98\times$ for the two breeds, which

was sufficient for the downstream analyses (**Table 1**). The average coverage against the reference genome was 98.65%, indicating a good quality for the data. We identified 13,618,261 unique variants in the two breeds, of which 11,552,094 (84.8%) were retained after quality control that included 9,242,589 SNPs and 2,309,505 indels. We compared these variants to dbSNP database, 28.22% of these variants matched 58.91% of the records in dbSNP. The Ts/Tv ratios were 1.848 and 1.856 for Shiba Inu and Mame Shiba Inu, while the ratio was 2.09 when computing with dbSNP dataset.

Detection of Selective Sweeps

After discarding windows containing < 10 SNP in the F_{ST} sliding window approach, 86,886 windows were retained. The mean F_{ST} value was 0.058, with the highest and lowest F_{ST} values being 0.311 and 0.008, respectively. A total of 148 gene colocalized in the 869 windows in the top 1% of the F_{ST} distribution (F_{ST} > 0.146, **Figure 2A**). Co-localized genes included *FGFRL1, GBX1, IQCA1L, RNF212* and *FAM189A1*, as well as some previously well-studied body size genes such as *IGF1, SMAD2* and *LCORL* (**Supplementary Table 1**).

Similarly, the mean $-ZH_p$ value for the windows was -0.029, with the highest and lowest $-ZH_p$ values being 3.767 and -2.019, respectively. We identified 411 genes (**Supplementary Table 2**) in the 869 windows in the top 1% of the $-ZH_p$ distribution of the Mame Shiba Inu ($-ZH_p$ > 2.86, **Figure 2B**). *SLTM* and *SRGAP1* were found in windows with the greatest -ZH_p divergence.

In XP-CLR approach, 861 putative selection windows were detected using the threshold of 1% cutoff of normalized XP-CLR scores (**Figure 3**), with normalized XP-CLR scores ranged from 3.07 to 40.58. We identified 442 genes in these windows (**Supplementary Table 3**), and the windows with highest XP-CLR scores co-localize with genes *ACSBG2, UBA5,* and *NPHP3*.
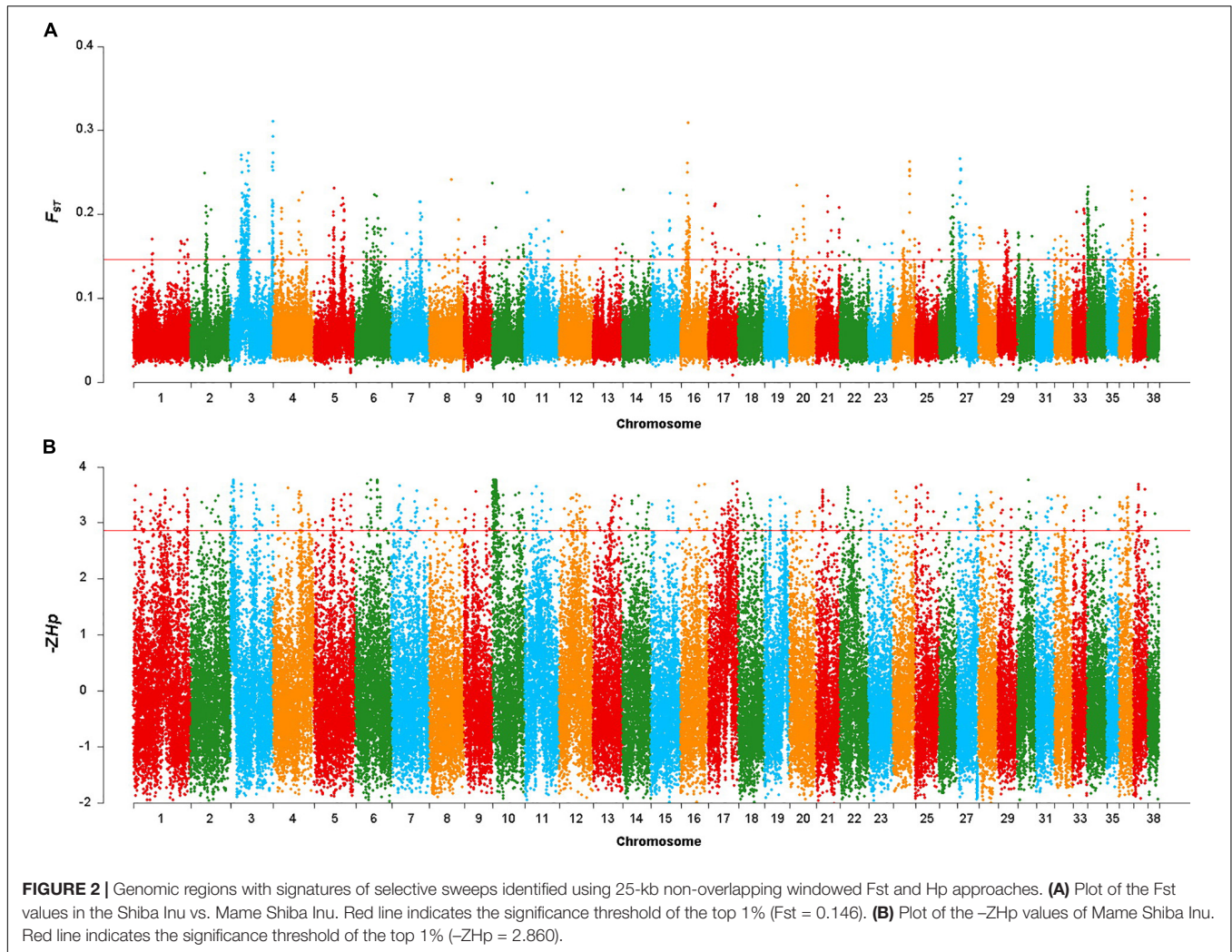
Notably, only 12 windows were located all in the top 1% of the F_{ST}, -ZH_p, and XP-CLR score distributions. Within these 12 windows with selective signatures, we identified nine co-localizing genes (*PRDM16, ZNF382, ZNF461, ERGIC2, ENSCAFG00000033351, CCDC61, ALDH3A2, ENSCAFG00000011141,* and *ENSCAFG00000018533*) that are strong candidate selected genes. These windows also included genomic regions, including on Chromosome 3, 8, and 36, which did not overlap with any annotated gene (**Table 2**).

DISCUSSION

Due to human imposed artificial selection, some Shiba Inus now have body sizes that are smaller than their ancestors. In this

---

[5]https://github.com/hardingnj/xpclr

TABLE 1 | Sequencing statistics of the clean data from the pooled samples.

Sample (individuals)	Bases (G)	Reads (M)	Q20 rate (%)	Q30 rate (%)
Shiba Inu (29)	35.70	238.03	98.16	94.82
Shiba Inu (22)	33.48	223.17	98.39	95.37
Shiba Inu (8)	40.13	267.54	98.43	95.46
Mame Shiba Inu (35)	83.19	554.61	98.45	95.51

FIGURE 2 | Genomic regions with signatures of selective sweeps identified using 25-kb non-overlapping windowed Fst and Hp approaches. **(A)** Plot of the Fst values in the Shiba Inu vs. Mame Shiba Inu. Red line indicates the significance threshold of the top 1% (Fst = 0.146). **(B)** Plot of the –ZHp values of Mame Shiba Inu. Red line indicates the significance threshold of the top 1% (–ZHp = 2.860).

study, we conducted whole-genome pool sequencing on two breeds of primitive Japanese dogs, Shiba Inu and Mame Shiba Inu, to investigate the molecular mechanisms responsible for the differences in their body size, and to identify candidate genes that may regulate body size in all dogs. In the case that only allele frequencies were needed, pool-seq is well suited for detecting selective sweeps (Boitard et al., 2012). The Ts/Tv ratios were relatively low, probably due to selective forces (Wakeley, 1996). 28.22% of the variants matched dbSNP records, suggesting that there might be breed-specific variants for the two breeds. Using selective sweep analyses, we identified several genomic regions with signatures consistent with selection. When compared to the reference dog genome, we found several genes that are possibly involved in the change in body size of Shiba Inus.

With the F_{ST}, H_p and XP-CLR with sliding window approaches, nine genes (*PRDM16, ZNF382, ZNF461, ERGIC2, ENSCAFG00000033351, CCDC61, ALDH3A2, ENSCAFG00000011141,* and *ENSCAFG00000018533*) were found to show selection signatures. Among these genes, *PRDM16* was reported to be associated with growth traits in some species, such as bovine and chicken (Wang et al., 2010;

Wu et al., 2021). It is known that *PRDM16* controls a brown fat/skeletal muscle switch (Seale et al., 2008), and it might take a role in the growth of Mame Shiba Inu. Additionally, there are genes identified by two approaches, for instance, *NCAPG* located in a window that had both high F_{ST} and low H_p values (**Supplementary Tables 1, 2**). *NCAPG-LCORL* region, has previously been reported to be associated with body size or weight in dogs and several other species. The most significant SNP, located on Chromosome 6, identified in a GWAS of 1,781 sheep associated with body weight is in a genomic region that contains the *LAP3*, *NCAPG,* and *LCORL* genes (Al-Mamun et al., 2015). An indel in *LCORL* (Chr3: 105,755,416) and a SNP (Chr3: 105,363,241) 246 kb upstream of *NCAPG-LCORL* were identified to be associated with body size in the Tennessee Walking Horse (Staiger et al., 2016). A SNP (Chr3: 105,547,002) in the *NCAPG-LCORL* region was reported to be associated with the sizes of horses (Makvandi-Nejad et al., 2012). In dogs, a genomic region in *LCORL* (Chr3: 93,933,450-93,944,095) was previously reported to be under selection in dogs (Vaysse et al., 2011). To date, however, there has been no reports linking *PRDM16* or *NCAPG* with body size in dogs. We identified a genomic region
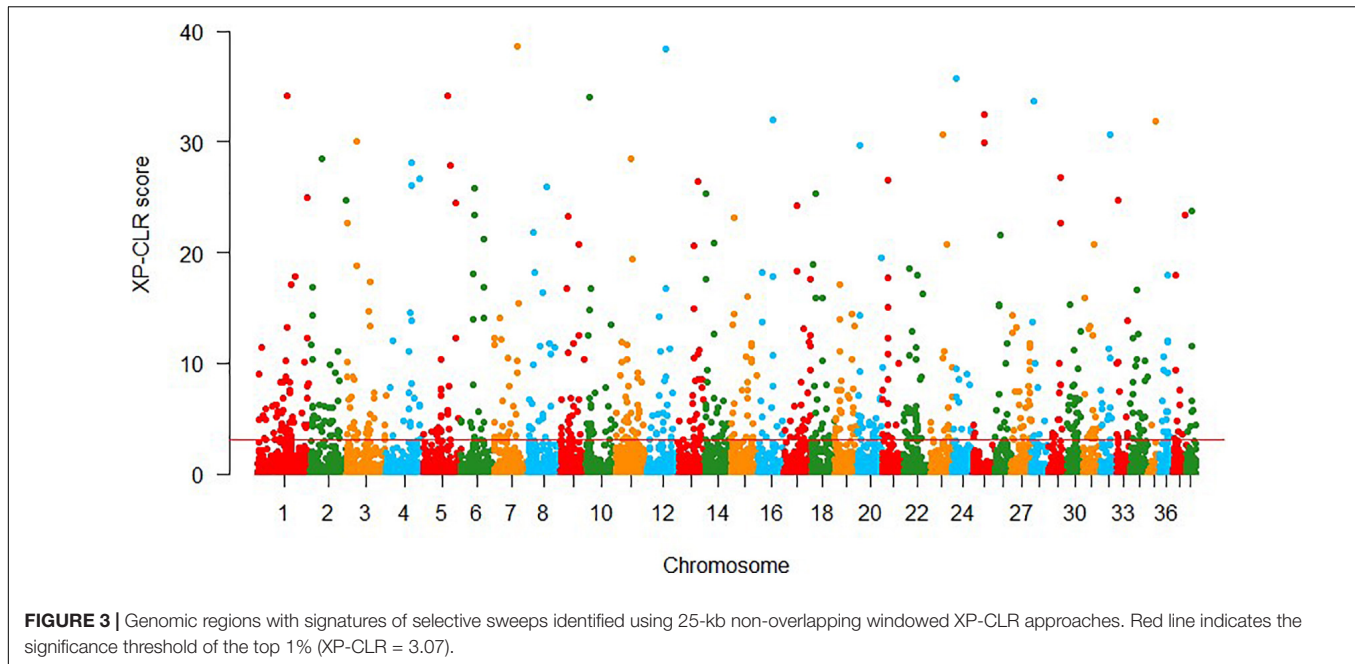
FIGURE 3 | Genomic regions with signatures of selective sweeps identified using 25-kb non-overlapping windowed XP-CLR approaches. Red line indicates the significance threshold of the top 1% (XP-CLR = 3.07).

TABLE 2 | Regions showing signals of selection with the F_{ST}, Hp and XP-CLR approaches.

Chr[a]	Start[b]	End[c]	F_{ST}	Hp[d]	−ZHp[e]	XP-CLR[f]	Gene Symbol[g]
1	109,625,000	109,650,000	0.149	0.076	2.873	3.305	*CCDC61*
1	116,300,000	116,325,000	0.149	0.040	3.292	6.477	*ZNF382, ZNF461*
3	52,950,000	52,975,000	0.196	0.008	3.677	15.727	–
5	40,575,000	40,600,000	0.199	0.045	3.232	3.143	*ALDH3A2*
5	42,875,000	42,900,000	0.184	0.060	3.060	9.002	*ENSCAFG00000018533*
5	57,800,000	57,825,000	0.170	0.031	3.406	8.475	*PRDM16*
5	57,825,000	57,850,000	0.211	0.040	3.297	20.152	*PRDM16*
8	34,250,000	34,275,000	0.147	0.044	3.244	24.152	*ENSCAFG00000033351*
8	62,250,000	62,275,000	0.149	0.064	3.015	4.824	–
27	18,625,000	18,650,000	0.217	0.041	3.278	19.421	*ERGIC2*
33	23,675,000	23,700,000	0.169	0.055	3.121	31.624	*ENSCAFG00000011141*
36	7,975,000	8,000,000	0.162	0.040	3.294	25.154	–

[a] Chromosome.
[b] Start position of the region.
[c] End position of the region.
[d] H_p value of Mame Shiba Inus.
[e] $−ZH_p$ value of Mame Shiba Inus.
[f] Normalized XP-CLR score of the region.
[g] Genes that overlap with the region. Multiple genes are separated with comma. "–" means no overlapping annotated genes in this region.

(Chr3: 91,275,000–91,300,000) that has both a high F_{ST} value (0.212) and low H_p value (0.045), thus showing a selective signal, between two breeds of Shiba Inu. We also identified selective signals in three well-studied body size genes (*IGF1*, *SMAD2*, and *LCORL*) by the F_{ST} approach (**Supplementary Table 1**), however they did not show extremely low H_p values, which might indicate that they involved in the change in body size of Shiba Inus during artificial selection, but not in key roles.

Although eight of the nine identified genes (*ZNF382*, *ZNF461*, *ERGIC2*, *ENSCAFG00000033351*, *ALDH3A2*, *ENSCAFG00000011141*, and *ENSCAFG00000018533*) and the two genomic regions that do not overlap with any gene have

not previously been associated with body size, they all showed signals for selection with the F_{ST}, H_p, and XP-CLR approaches. We suggest that these genes and genomic regions might possibly have an effect on body size in dogs, they also might be involved in some other traits. More in-depth studies of these genomic regions are required to resolve this question.

In conclusion, using pool-seq data and selective analyses, using the F_{ST}, H_p, and XP-CLR approaches, we identified nine genes and two genomic regions that experienced selection, including one that were have previously been reported to be associated with body size in mammals (*PRDM16*). Our study identified eight novel genes (*ZNF382*, *ZNF461*, *ERGIC2*,

ENSCAFG00000033351, ALDH3A2, ENSCAFG00000011141, and *ENSCAFG00000018533*) that need further study. Our results are consistent with previous findings, and provide novel genes and genomic regions that potentially are associated with body size in dogs and other mammals.

DATA AVAILABILITY STATEMENT

Pool-seq data was submitted to NCBI Sequence Read Archive (accession number PRJNA636107).

ETHICS STATEMENT

The animal study was reviewed and approved by Animal Care and Use Committee of Shenyang Agricultural University. Written informed consent was obtained from the owners for the participation of their animals in this study.

AUTHOR CONTRIBUTIONS

GL participated in the design of this study. GL, SR, and WD collected the samples. GL, CF, and SZ performed the experiments. GL, SZ, and ZW analyzed the data. GL, DI, and ZW wrote the manuscript. SZ and ZW conceived, designed the study, and supervised the work. All authors approved the final version of the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.671686/full#supplementary-material

REFERENCES

Al-Mamun, H. A., Kwan, P., Clark, S. A., Ferdosi, M. H., Tellam, R., and Gondro, C. (2015). Genome-wide association study of body weight in Australian merino sheep reveals an orthologous region on OAR6 to human and bovine genomic regions affecting height and weight. *Genet. Sel. Evol.* 47:66. doi: 10.1186/s12711-015-0142-4

Bhatia, G., Patterson, N., Sankararaman, S., and Price, A. L. (2013). Estimating and interpreting FST: the impact of rare variants. *Genome Res.* 23, 1514–1521. doi: 10.1101/gr.154831.113

Boitard, S., Schlötterer, C., Nolte, V., Pandey, R. V., and Futschik, A. (2012). Detecting selective sweeps from pooled next-generation sequencing samples. *Mol. Biol. Evol.* 29, 2177–2186. doi: 10.1093/molbev/mss090

Chen, H., Patterson, N., and Reich, D. (2010). Population differentiation as a test for selective sweeps. *Genome Res.* 20, 393–402. doi: 10.1101/gr.100545.109

Faust, G. G., and Hall, I. M. (2014). SAMBLASTER: fast duplicate marking and structural variant read extraction. *Bioinformatics* 30, 2503–2505. doi: 10.1093/bioinformatics/btu314

Hayward, J. J., Castelhano, M. G., Oliveira, K. C., Corey, E., Balkman, C., Baxter, T. L., et al. (2016). Complex disease and phenotype mapping in the domestic dog. *Nat. Commun.* 7:10460. doi: 10.1038/ncomms10460

Hoopes, B. C., Rimbault, M., Liebers, D., Ostrander, E. A., and Sutter, N. B. (2012). The insulin-like growth factor 1 receptor (IGF1R) contributes to reduced size in dogs. *Mamm. Genome* 23, 780–790. doi: 10.1007/s00335-012-9417-z

Kofler, R., Pandey, R. V., and Schlötterer, C. (2011). PoPoolation2: identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). *Bioinformatics* 27, 3435–3436. doi: 10.1093/bioinformatics/btr589

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* 27, 2987–2993. doi: 10.1093/bioinformatics/btr509

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352

Makvandi-Nejad, S., Hoffman, G. E., Allen, J. J., Chu, E., Gu, E., Chandler, A. M., et al. (2012). Four loci explain 83% of size variation in the horse. *PLoS One* 7:e39929. doi: 10.1371/journal.pone.0039929

Micheletti, S. J., and Narum, S. R. (2018). Utility of pooled sequencing for association mapping in nonmodel organisms. *Mol. Ecol. Resour.* 18, 825–837. doi: 10.1111/1755-0998.12784

Ostrander, E. A., Wayne, R. K., Freedman, A. H., and Davis, B. W. (2017). Demographic history, selection and functional diversity of the canine genome. *Nat. Rev. Genet.* 18, 705–720. doi: 10.1038/nrg.2017.67

Parker, H. G., Dreger, D. L., Rimbault, M., Davis, B. W., Mullen, A. B., Carpintero-Ramirez, G., et al. (2017). Genomic analyses reveal the influence of geographic origin, migration, and hybridization on modern dog breed development. *Cell Rep.* 19, 697–708. doi: 10.1016/j.celrep.2017.03.079

Plassais, J., Kim, J., Davis, B. W., Karyadi, D. M., Hogan, A. N., Harris, A. C., et al. (2019). Whole genome sequencing of canids reveals genomic regions under selection and variants influencing morphology. *Nat. Commun.* 10:1489. doi: 10.1038/s41467-019-09373-w

Plassais, J., Rimbault, M., Williams, F. J., Davis, B. W., Schoenebeck, J. J., and Ostrander, E. A. (2017). Analysis of large versus small dogs reveals three genes on the canine X chromosome associated with body weight, muscling and back fat thickness. *PLoS Genet.* 13:e1006661. doi: 10.1371/journal.pgen.1006661

Rimbault, M., Beale, H. C., Schoenebeck, J. J., Hoopes, B. C., Allen, J. J., Kilroy-Glynn, P., et al. (2013). Derived variants at six genes explain nearly half of size reduction in dog breeds. *Genome Res.* 23, 1985–1995. doi: 10.1101/gr.157339.113

Rubin, C. J., Zody, M. C., Eriksson, J., Meadows, J. R., Sherwood, E., Webster, M. T., et al. (2010). Whole-genome resequencing reveals loci under selection during chicken domestication. *Nature* 464, 587–591. doi: 10.1038/nature08832

Seale, P., Bjork, B., Yang, W., Kajimura, S., Chin, S., Kuang, S., et al. (2008). PRDM16 controls a brown fat/skeletal muscle switch. *Nature* 454, 961–967. doi: 10.1038/nature07182

Staiger, E. A., Al Abri, M. A., Pflug, K. M., Kalla, S. E., Ainsworth, D. M., Miller, D., et al. (2016). Skeletal variation in tennessee walking horses maps to the LCORL/NCAPG gene region. *Physiol. Genom.* 48, 325–335. doi: 10.1152/physiolgenomics.00100.2015

Sutter, N. B., Bustamante, C. D., Chase, K., Gray, M. M., Zhao, K., Zhu, L., et al. (2007). A single IGF1 allele is a major determinant of small size in dogs. *Science* 316, 112–115. doi: 10.1126/science.1137045

Vaysse, A., Ratnakumar, A., Derrien, T., Axelsson, E., Rosengren Pielberg, G., Sigurdsson, S., 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

Voorbij, A. M., van Steenbeek, F. G., Vos-Loohuis, M., Martens, E. E., Hanson-Nilsson, J. M., van Oost, B. A., et al. (2011). A contracted DNA repeat in LHX3 intron 5 is associated with aberrant splicing and pituitary dwarfism in German shepherd dogs. *PLoS One* 6:e27940. doi: 10.1371/journal.pone.0027940

Wakeley, J. (1996). The excess of transitions among nucleotide substitutions: new methods of estimating transition bias underscore its significance. *Trends Ecol. Evol.* 11, 158–162. doi: 10.1016/0169-5347(96)10009-4

Wang, J., Li, Z. J., Lan, X. Y., Hua, L. S., Huai, Y. T., Huang, Y. Z., et al. (2010). Two novel SNPs in the coding region of the bovine PRDM16 gene and its associations with growth traits. *Mol. Biol. Rep.* 37, 571–577. doi: 10.1007/s11033-009-9816-8

Weir, B. S., and Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution* 38, 1358–1370. doi: 10.1111/j.1558-5646.1984.tb05657.x

Wu, Z., Bortoluzzi, C., Derks, M. F. L., Liu, L., Bosse, M., Hiemstra, S. J., et al. (2021). Heterogeneity of a dwarf phenotype in Dutch traditional chicken breeds revealed by genomic analyses. *Evol. Appl.* 14, 1095–1108. doi: 10.1111/eva.13183