Single-marker and haplotype-based genome-wide association studies for the number of teats in two heavy pig breeds

S. Bovo*[,1] , M. Ballan*[,1] , G. Schiavo* , A. Ribani* , S. Tinarelli*[,†] , V. J. Utzeri* , S. Dall'Olio* , M. Gallo[†] and L. Fontanesi*

*Department of Agricultural and Food Sciences, Division of Animal Sciences, University of Bologna, Viale Fanin 46, Bologna 40127, Italy.
†Associazione Nazionale Allevatori Suini (ANAS), Via Nizza 53, Roma 00198, Italy.

Summary

The number of teats is a reproductive-related trait of great economic relevance as it affects the mothering ability of the sows and thus the number of properly weaned piglets. Moreover, genetic improvement of this trait is fundamental to parallelly help the selection for increased litter size. We present the results of single-marker and haplotypes-based genome-wide association studies for the number of teats in two large cohorts of heavy pig breeds (Italian Large White and Italian Landrace) including 3990 animals genotyped with the 70K GGP Porcine BeadChip and other 1927 animals genotyped with the Illumina PorcineSNP60 BeadChip. In the Italian Large White population, genome scans identified three genome regions (SSC7, SSC10, and SSC12) that confirmed the involvement of the *VRTN* gene (as we previously reported) and highlighted additional loci known to affect teat counts, including the *FRMD4A* and *HOXB1* gene regions. A different picture emerged in the Italian Landrace population, with a total of 12 genome regions in eight chromosomes (SSC3, SSC6, SSC8, SSC11, SSC13, SSC14, SSC15, and SSC16) mainly detected via the haplotype-based genome scan. The most relevant QTL was close to the *ARL4C* gene on SSC15. Markers in the *VRTN* gene region were not significant in the Italian Landrace breed. The use of both single-marker and haplotype-based genome-wide association analyses can be helpful to exploit and dissect the genome of the pigs of different populations. Overall, the obtained results supported the polygenic nature of the investigated trait and better elucidated its genetic architecture in Italian heavy pigs.

Keywords *ARL4C*, *FRMD4A*, *HOXB1*, Landrace, Large White, morphological trait, single nucleotide polymorphism, *Sus scrofa*, *VRTN*

Introduction

The number of piglets weaned per sow and per year is one of the most important indicators of economic sustainability of pig farming. This is a multifactorial and complex parameter that is affected, in part, by the mothering ability of the sows, which in turn, is related to the number of their teats that have to supply the needed nutrients to all piglets (Kim et al. 2005; Andersen et al. 2011). Therefore, an adequate number of teats is required to parallelly help the selection

for increased litter size and that, together, can maximise the number of weaned piglets.

The number of teats in pigs is considered a quantitative trait (but with discrete and countable values) that shows a considerable variability among and within breeds and lines (Borchers et al. 2002; Lopes et al. 2014; Verardo et al. 2016; Rohrer & Nonneman 2017; Dall'Olio et al. 2018; van Son et al. 2019). The medium/high level of heritability of this parameter can facilitate selection to improve overall sow reproductive performances (e.g. Willham & Whatley 1962; Toro et al. 1986; McKay & Rahnefeld 1990; Borchers et al. 2002; Chalkias et al. 2013; Felleki & Lundeheim 2015; Balzani et al. 2016).

Several studies have identified quantitative trait loci (QTL) for the number of teats by using reference populations constituted by crossing pigs of different breeds or lines, including some highly hyper-prolific Chinese breeds (e.g. Wada et al. 2000; Hirooka et al. 2001; Rodríguez et al.

2005; Bidanel et al. 2008; Ding et al. 2009; Hernandez et al. 2014) or by applying genome-wide association studies (GWAS) within breeds (e.g. Arakawa et al. 2015; Rohrer & Nonneman 2017; Tang et al. 2017; Lee et al. 2019; van Son et al. 2019). Results of these studies evidenced many different QTL affecting this trait and confirmed its polygenic nature. Porcine chromosome (SSC) 7 has been reported to harbor one of the most important QTL affecting this trait segregating in several populations (Mikawa et al. 2007, 2011; Duijvesteijn et al. 2014; Rohrer & Nonneman 2017; Dall'Olio et al. 2018; van Son et al. 2019; Moscatelli et al. 2020). This QTL has pleiotropic effects on the number of vertebrae due to variability in the *vertebrae development associate*d gene known also as *vertnin* (*VRTN*) (Mikawa et al. 2011; Arakawa et al. 2015), which encodes for a DNA binding factor (Duan et al. 2018). A few other studies proposed that polymorphic sites close to the *VRTN* gene and variability in other genes located on SSC7 (*latent transforming growth factor binding protein 2* or *LTBP2* and *BRMS1-like transcriptional repressor* or *BRMS1L*) could affect the number of teats and vertebrae in pigs (Zhang et al. 2016; Park et al. 2018).

GWAS may benefit from utilising haplotypes instead of SNPs to establish marker-phenotype associations (Lorenz et al. 2010; Barendse 2011). We recently demonstrated that by using haplotypes in GWAS it is possible to detect genomic regions affecting targeted phenotypes that could not be detected with a single-marker approach and vice versa (Bovo et al. 2019, 2020a). The additional information that could be extracted might be dependent on the particular mutational and recombination history between QTL alleles and the genotyped markers, which in turn, might be affected by possible ascertainment biases derived by the construction of the SNP genotyping panels used in the studies. Therefore, the use of both approaches has been recommended to take advantage of the full information content of the genotype data (Lorenz et al. 2010; Barendse 2011).

In our previous studies, we described the variability of the number of teats in the Italian Large White pig breed and evaluated the effect of DNA polymorphisms (in candidate genes and from single-marker genome-wide analyses) on this phenotype (Dall'Olio et al. 2018; Moscatelli et al. 2020).

In this study, we carried out GWAS in a much larger cohort of Italian Large White pigs and in the Italian Landrace breed (not previously investigated) that made it possible to identify additional candidate gene regions associated with the number of teats. Genomic analyses were first based on linear mixed models fitting single SNPs and then on haplotypes, that further refined the results and identified additional loci associated with this important trait.

Materials and Methods

Animals

All animals included in this study were from the national selection programme of heavy pig breeds that is run by the Italian Pig Breeders Association (ANAS). Pigs investigated were 3974 of the Italian Large White (866 castrated males and 1688 gilts and 1420 sows) and of 1943 Italian Landrace (106 castrated males and 263 gilts and 1574 sows) breeds. Castrated males and gilts were from the sib-testing programme and sows were from the female selection programme. Italian Large White pigs were born in the 1985–2018 whereas Italian Landrace pigs were born in the 2006–2018.

The number of teats on these animals was routinely recorded by direct counting at the beginning of the performance testing period for the pigs included in the sib-testing programme and at the end of the puberty for the sows that were not performance tested. Pigs having fewer than 14 teats are then discarded from the herd books of the two breeds. However, in this study, to maximise variability for this trait we included a few animals with a lower number of teats for which blood could be available in the ANAS biobank.

Genotyping and SNP quality

Blood was collected routinely from all pigs included in the national selection programme. DNA was extracted from these samples using the Wizard Genomic DNA Purification kit (Promega Corporation, Madison, WI, USA). A total of 1943 Italian Landrace and 2047 Italian Large White pigs were genotyped with the 70K GGP Porcine BeadChip (GeneSeek, Lincoln, NE), which interrogates 68516 SNPs. The remaining 1927 Italian Large White pigs were genotyped with the Illumina PorcineSNP60 BeadChip v.2 (which interrogates 61565 SNPs). Standard genotyping protocols, based on the supplier's recommendations, were used. PLINK v.1.09 (Chang et al. 2015) was used to discard SNPs presenting a call rate <0.95, a minor allele frequency (MAF) < 0.01 and that were not in Hardy–Weinberg equilibrium ($P < 0.001$). SNPs shared between the two genotyping platforms were used in the analysis of the Italian Large White pigs. BLAST+ v.2.7.1 (Camacho et al. 2009) was used to map SNPs to the Sscrofa11.1 reference genome and markers assigned to more than one position or assigned to sex chromosomes were discarded. The genomic dataset was supplemented with the genotype status of the *VRTN* g.20311_20312ins291 mutation analyzed in 778 Italian Landrace pigs already genotyped with 70K GGP Porcine BeadChip. The genotyping protocol of this mutation was as previously described (Fontanesi et al. 2014).

Phasing and haplotype detection

Genotypes were phased using SHAPEIT v.2 (Delaneau et al. 2011) considering a window size of 2 Mb, an effective population size (N_e) estimated with SNeP v.1.1 (Barbato et al. 2015) and a chromosome specific recombination rate given by Tortereau et al. (2012). Haplotypes were further called with the R package GHap 1.2.2. Following the study by Veroneze et al. (2013), a genome window of 400 kb with a sliding block of 100 kb was used to call haplotypes. Haplotypes were exported in the *tped* file format, where haplotype allele counts 0, 1, and 2 are recoded as NN, NH, and HH genotypes (H = haplotype allele and N = NULL = all other N alleles), as if haplotypes were bi-allelic markers. A regular ped file was obtained with PLINK, filtering out haplotypes presenting a MAF < 0.01.

Genome-wide association analyses

After filtering, the Italian Large White dataset counted 3888 animals (106 castrated males, 262 gilts and 1573 sows) for which 36,243 SNPs and 196,394 haplotypes (corresponding to 21231 haploblocks) were analysed. The Italian Landrace dataset was based on 1941 animals (106 castrated males, 262 gilts and 1573 sows), 50,453 SNPs and 237087 haplotypes (corresponding to 21,707 haploblocks). Breed specific GWAS were based either on SNPs or on haplotypes. Table S1 summarises the datasets used in the GWAS in the two breeds.

Association analyses were performed using an additive genetic model assuming a trend per copy of the minor allele that specify the dependency of the number of teats on genotype categories. The following linear mixed effect model was specified:

$$y = W\alpha + x\beta + g + e \qquad (1)$$

where y ($n \times 1$) is a vector containing the phenotype (the number of teats) for the nth animal, W ($n \times k$) is a covariate matrix with $k = 2$ (a column of 1s and sex) and α is the k-dimensional vector of covariates effects, x ($n \times 1$) is the vector containing genotypes for the ith DNA marker (SNP or haplotype), β is the additive fixed effect of the ith DNA marker on the phenotype, $g \sim N(0, \sigma^2_g K)$ is a multivariate Gaussian polygenic effect, with covariance matrix proportional to the relatedness matrix K ($n \times n$) and $e \sim N(0, \sigma^2_e I)$ is a multivariate Gaussian vector of uncorrelated residuals. The assessment of the association between each DNA marker and the total number of teats was obtained by testing the null hypothesis $H_0: \beta = 0$. Significance was tested by using the Wald test. All the models were fitted with GEMMA v.0.98 (Zhou & Stephens 2012) after computing the relatedness matrices K_1 and K_2 as centred genomic matrices, for SNPs and haplotypes respectively. To account for multiple comparisons, we opted for the Bonferroni correction, which considered the total number of DNA markers or haplotypes and a value of $\alpha = 0.05$. SNPs and haplotypes presenting the lowest P in chromosome regions separated by at least 5 Mb were considered as tag DNA markers. For each trait, GEMMA was used to estimate the genomic (chip) heritability (h^2_G). Genomic control inflation factors (λ_{GC}), defined as the median of the resulting chi-squared test statistics divided by the expected median of the chi-squared distribution, were computed in R v.3.6.0 (R Core Team 2018). Quantile–quantile plotslots) and Manhattan plots were generated in R by using the *qqman* package (Turner 2018).

Haploblocks and annotation of associated SNPs and haplotypes

QTL regions were analysed using HaploView v.4.2 (https://www.broadinstitute.org/haploview/haploview) to compare haploblock structures between the two pig breeds. Only markers shared by the two datasets were retained. Protein coding genes annotated in the Sscrofa11.1 genome version spanning the significantly associated haplotypes and a region of ± 500 kb around the significantly associated SNPs were retrieved from the Sscrofa11.1 NCBI's GFF file by using Bedtools v.2.17.0 (Quinlan & Hall 2010). Comparative QTL mapping analysis across studies was obtained using the Pig QTL database (Pig QTLdb; release 39; Hu et al. 2019). QTLs were downloaded, checked and manually curated as previously reported (Bovo et al. 2020c). The final dataset comprised a total of 295 traits and 1978 QTL regions.

Results

Descriptive statistics on the number of teats in the two heavy pig breeds

Descriptive statistics on the number of teats in Italian Large White and Italian Landrace breeds are reported in Tables S2 and S3 respectively. The number of teats ranged from 12 to 20 and from 10 to 18, for the Italian Large White and Italian Landrace pigs respectively. About 46.5% (Italian Large White) and 50.1% (Italian Landrace) of the investigated pigs had 14 teats (Fig. 1), which is the lower limit considered for the registration of the animals to the Herd Book of these Italian heavy pig breeds. The percentage of pigs with more than 14 teats was 53.4% and 43.5% respectively. Within the Italian Large White population, a statistically significant difference in teat counts ($P = 0.002$, Wilcoxon rank sum test) was observed between males and females: males had on average 14.88 teats (SD = 0.92; median = 15), whereas females had on average 14.77 teats (SD = 0.87; median = 15).

Genomic heritability of the number of teats, estimated using SNP and haplotype data, was 0.25 (SE = 0.02) and 0.31 (SE = 0.03) in the Italian Large White breed and 0.30
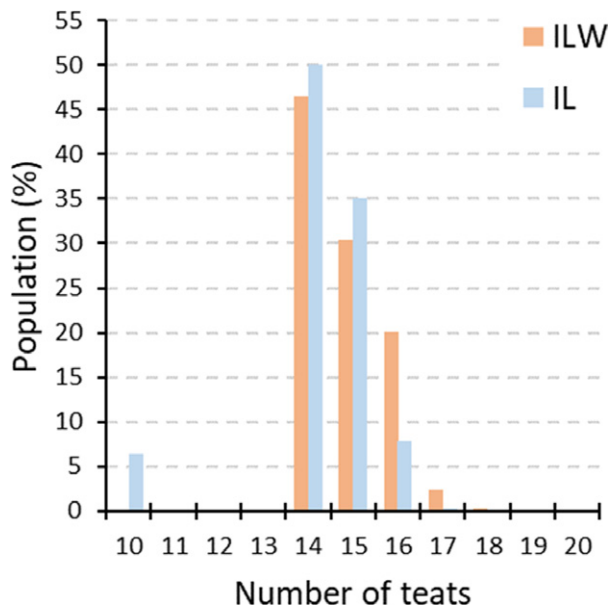
Figure 1 Distribution of the number of teats within the Italian Large White (ILW) and Italian Landrace (IL) populations.

(SE = 0.03) and 0.43 (SE = 0.04) in the Italian Landrace pigs.

Single-marker and haplotype-based genome-wide analyses in the two heavy pig breeds

Results of the genome-wide association studies are summarised in Table 1 that reports the top associated DNA markers (SNPs and haplotypes) for each genomic region. This table also includes information on QTLs for teat count, vertebrae count, and reproductive traits reported in the Pig QTLdb that overlapped the significant markers/regions detected in our study. Table S4 reports the full set of associated DNA markers (SNPs and haplotypes). Fig. 2 shows the Manhattan plots for the number of teats in the Italian Large White and Italian Landrace pig populations. Quantile–quantile plots are available as Fig. S1. Box plots showing the effect of the haplotype/SNP alleles at all identified QTLs are included in Figs S2 and S3 for the Italian Large White and the Italian Landrace breeds, respectively.

Genome scans in the Italian Large White breed – Three genomic regions in three different chromosomes (SSC7, SSC10, and SSC12) were associated with the analysed trait in this breed (Table 1 and Fig. 2a). The most significant SNP was MARC0038565 ($P = 2.65 \times 10^{-12}$), located on SSC7 at position 97 652 632 bp, in the region of the *VRTN* gene. This peak was also detected by the haplotype analysis ($P = 1.61 \times 10^{-15}$), confirming the involvement of the *VRTN* gene region in affecting the number of teats in the Italian Large White pig population (Dall'Olio et al. 2018; Moscatelli et al. 2020). Fig. 3a shows the allelic effects of this top associated haplotype region. Genome scans

confirmed the direction of the effects of the MARC0038565 SNP alleles already reported by Moscatelli et al. (2020) and of the haplotypes including the alleles of this SNP.

The second major peak was observed on SSC10, in the region of the *FERM domain containing 4A* (*FRMD4A*) gene (position 47 385 718-48 049 718 bp), with the most significant SNP (M1GA0014145; $P = 2.83 \times 10^{-8}$) located at position 47 751 164 bp. Haplotype analysis ($P = 1.56 \times 10^{-8}$) confirmed the results obtained in the single-marker analysis and further supported the role of the *FRMD4* gene region in affecting the variability of the number of teats in this pig breed.

A third region affecting the investigated trait was detected on SSC12 with the single-marker analysis. The significant marker MARC0031045 ($P = 1.03 \times 10^{-6}$), located at position 24 723 142 bp, is positioned near a cluster of Hox genes, including the *homeobox B1* (*HOXB1*) gene (position 24 491 486–24 492 906 bp). According to the key role of this gene (and other genes of the same family) in the developmental regulatory system, *HOXB1* (and/or other Hox genes) could be considered the candidate gene(s) of this region. For comparative analysis between breeds, Table S5 reports the effect of the top associated SNPs when they were investigated in Italian Landrace pigs.

Genome scans in the Italian Landrace breed – A total of 12 genomic regions (11 detected with the haplotype-based analysis and one with the single-marker analysis) in eight different chromosomes (SSC3, SSC6, SSC8, SSC11, SSC13, SSC14, SSC15, and SSC16) were associated with the number of teats in this breed (Table 1 and Fig. 2). Haplotype or SNP alleles determining a reduced number of teats were the less frequent in the population (Table 1, Fig. S3). Among all these QTL regions, only five were reported to be located in chromosome regions in which other studies already reported QTL directly or indirectly related to teat number (Table 1). The most significant region, detected with the haplotype-based analysis ($P = 8.1 \times 10^{-16}$), was on SSC15 (positions 134 200 001–134 600 001 bp). This is the only region detected in the Italian Landrace breed that overlaps with a QTL region reported in a previous study to affect teat numbers (van Son et al. 2019). Fig. 3b shows the allelic effects of this top associated haplotype region. Other related QTL were already reported in this region (Table 1).

The second and third most significantly associated haplotypes were located on SSC16 (position 68.7 Mb; $P = 5.08 \times 10^{-9}$) and on SSC13 (position 189.1 Mb; $P = 5.43 \times 10^{-9}$). Only the second region was previously reported to overlap a QTL related to maternal behaviour (Table 1).

The single-marker analysis identified a significant peak on SSC14 (position 23 582 019 bp; marker WU_10.2_14_25047530), located near the *adhesion G protein-coupled receptor D1* (*ADGRD1*) gene (also known as *GPR133*) and proposed to be a gene associated with adult height (Tönjes et al. 2009).

Table 1 Top genome regions associated with the number of teats in Italian Large White and Italian Landrace pigs. Results are stratified by population and chromosome.

Genome scan[1]	SSC[2]	Pos[3]	Marker[4]	Min/Maj[5]	MAF[6]	β[7]	P[8]	QTLs[9]	Genes[10]
Italian Large White									
Haplotype-based	7	97435001	CHR7_B973	H/N	0.426	0.188	1.61×10^{-15}	NNF, TVN, VN	ENTPD5, FAM161B, **VRTN**, ABCD4, ALDH6A1, SYNDIG1L, VSX2, BBOF1, COQ6, LIN52, ZNF410
Single-marker	7	97652632	MARC0038565	A/G	0.482	−0.163	2.65×10^{-12}	TVN	**VRTN**, SYNDI1GL
Single-marker	10	47751164	M1GA0014145	A/G	0.477	−0.135	2.83×10^{-8}	–	**FRMD4A**
Haplotype-based	10	48035001	CHR10_B479	H/N	0.367	−0.129	1.56×10^{-8}	LTN	**FRMD4A**, PRPF18, BEND7
Single-marker	12	24723142	MARC0031045	G/A	0.358	0.117	1.03×10^{-6}	THVN	SKAP1, **HOXB1**
Italian Landrace									
Haplotype-based	3	120600001	CHR3_B1205	H/N	0.014	−1.048	8.40×10^{-8}	THVN	RAD51AP2, FAM49A
Haplotype-based	6	113000001	CHR6_B1129	H/N	0.062	−0.466	1.95×10^{-7}	–	CDH2
Haplotype-based	8	4500001	CHR8_B44	H/N	0.011	−1.142	9.30×10^{-9}	NFN	C8H4orf50, JAKMIP1, WFS1, LOC102165510, PPP2R2C
Haplotype-based	11	18900001	CHR11_B188	H/N	0.034	−0.529	1.55×10^{-7}	–	FNDC3A, CYSLTR2, RCBTB2
Haplotype-based	13	189100001	CHR13_B1890	H/N	0.016	−0.957	5.43×10^{-9}	MI	ATP5J, GABPA, JAM2, MRPL39
Haplotype-based	14	132300001	CHR14_B1322	H/N	0.014	−0.953	1.73×10^{-7}	–	C14H10orf120, LOC100620521, AWN, LOC100519221, FAM24B, SPMI, DMBT1, PSP-II, AQN-1, HTRA1, PSP-I
Haplotype-based	14	114800001	CHR14_B1147	H/N	0.051	−0.547	1.79×10^{-7}	–	COL17A1, SH3PXD2A, STN1, SLK
Single-marker	14	23582019	WU_10.2_14_25047530	A/G	0.078	−0.445	3.18×10^{-7}	–	SFSWAP
Haplotype-based	15	134400001	CHR15_B1343	H/N	0.030	−0.951	8.08×10^{-16}	BW, BL, LW, TNBA, LS	TRPM8, SPP2, Close to **ARL4C**
Haplotype-based	15	19300001	CHR15_B192	H/N	0.013	−0.877	6.29×10^{-8}	–	NCKAP5
Haplotype-based	16	57400001	CHR16_B573	H/N	0.024	−0.806	6.15×10^{-8}	–	TENM2
Haplotype-based	16	68700001	CHR16_B686	H/N	0.017	−0.993	5.08×10^{-9}	–	SAP30L, HAND1, GALNT10

[1]Genome scans performed within each pig population. SNPs and Haplotypes indicate which DNA markers have been used to carry out the genome scans.

[2]*Sus scrofa* chromosome.

[3]Position, in bp, on the *Sus scrofa* reference genome (v.11.1).

[4]DNA marker identifier reported in the chip panels. For haplotypes, the haploblock identifier (chromosome specific) is reported.

[5]Minor/Major alleles. Haplotypes have been treated as bi-allelic variants (H = haplotype allele and N = other N alleles).

[6]Minor allele frequency.

[7]Regression coefficient. A positive value indicates that the no. of teats increases with the increasing of the number of copies of the minor allele. A negative value indicates that the no. of teats decreases with the increasing of the number of copies of the minor allele.

[8]*P* at the Wald test (GEMMA).

[9]QTLs related to reproductive traits overlapping the association peaks. Short names for QTL are: BL, body length; BW, body width; LS, litter size; LTN, left teat number; LWT, litter weight; MI, maternal infanticide; NFN, non-functional nipples; NNF, number of non-viable foetuses; THVN, thoracolumbar vertebra number; TNBA, total number of born alive; TVN, thoracic vertebra number; VN, vertebra number.

[10]Genes overlapping the haplotype or closed to a SNP. The candidate genes are reported in bold.
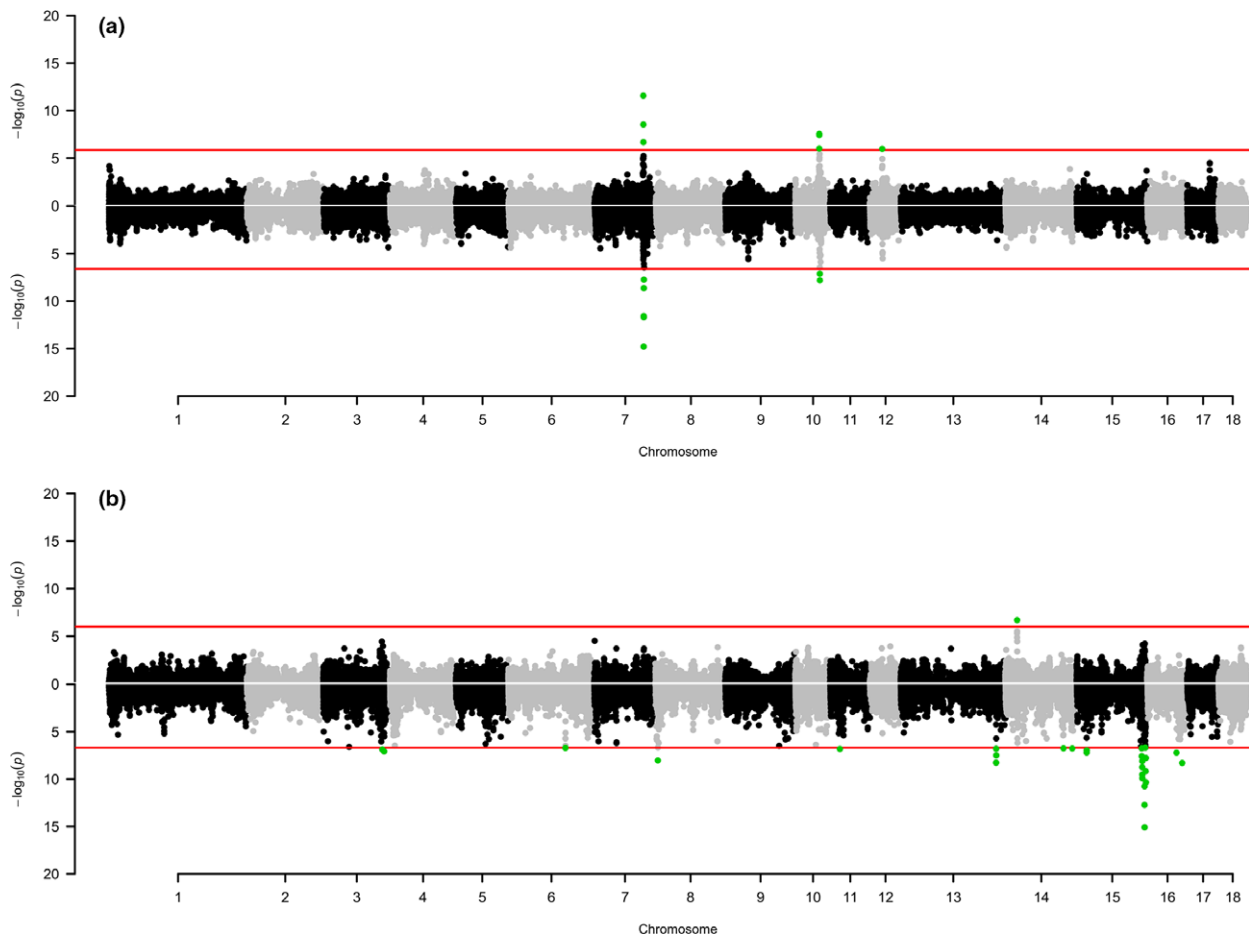
Figure 2 Manhattan plots of the genome-wide association studies on the number of teats in the Italian Large White (a) and in the Italian Landrace (b) populations. Single-marker analysis is on the top part of each plot and haplotype-based analysis is on the bottom part of each plot. Each dot represents a DNA marker or a haplotype. The red lines identify the significance thresholds (Bonferroni correction; $\alpha = 0.05$).

Comparative analysis of the *VRTN* genome region in the two heavy pig breeds

In the Italian Landrace breed, no significant markers or haplotypes were identified on SSC7 in the region of the *VRTN* gene. The most significant marker in this region was a SNP at position 98 763 633 bp (WU_10.2_7_104470681; $P = 3.04 \times 10^{-4}$) that is far away (1.11 Mb) from the top SNP identified in the Italian Large White breed. Inclusion of the *VRTN* g.20311_20312ins291 mutation in the association analysis carried out in Italian Landrace pigs (allele frequencies of the two alleles were: insertion = 0.67, wild type = 0.33) did not produce any significant result ($P = 0.088$). The level of linkage disequilibrium (r^2) between the *VRTN* mutation and the top associated SNP identified in Italian Landrace pigs in this region (WU_10.2_7_104470681) was low ($r^2 = 0.05$).

In Italian Landrace, we also evaluated in the association analysis the top associated haplotype detected in Italian Large White pigs. No significant association emerged ($P = 0.38$) and different allelic frequencies were detected.

To further compare the genome region in the two breeds, we also analysed the level of linkage disequilibrium between key markers. For example, in our previous study in Italian Large White pigs (Moscatelli et al. 2020), the *VRTN* insertion allele and the top associated marker (MARC0038565) had similar MAF and β of association (MAF ~ 0.24 and β ~ 0.35) both explained by the high linkage disequilibrium ($r^2 = 0.58$). We therefore analysed the linkage disequilibrium between the same two markers in the Italian Landrace population. In this breed linkage disequilibrium was much lower ($r^2 = 0.07$). In Italian Landrace pigs we also further evaluated the level of linkage disequilibrium between the *VRTN* marker and the other two significant SNPs (M1GA0010653 and H3GA0022659) detected in this study in the Italian Large White population (Table S4). Again, a poor linkage disequilibrium was evidenced in the Italian Landrace breed ($r^2 = 0.06$ and $r^2 = 0.14$ respectively).

A comparative haploblock analysis of this region in both Italian Landrace and Italian Large White, considering shared DNA markers between the two datasets, showed
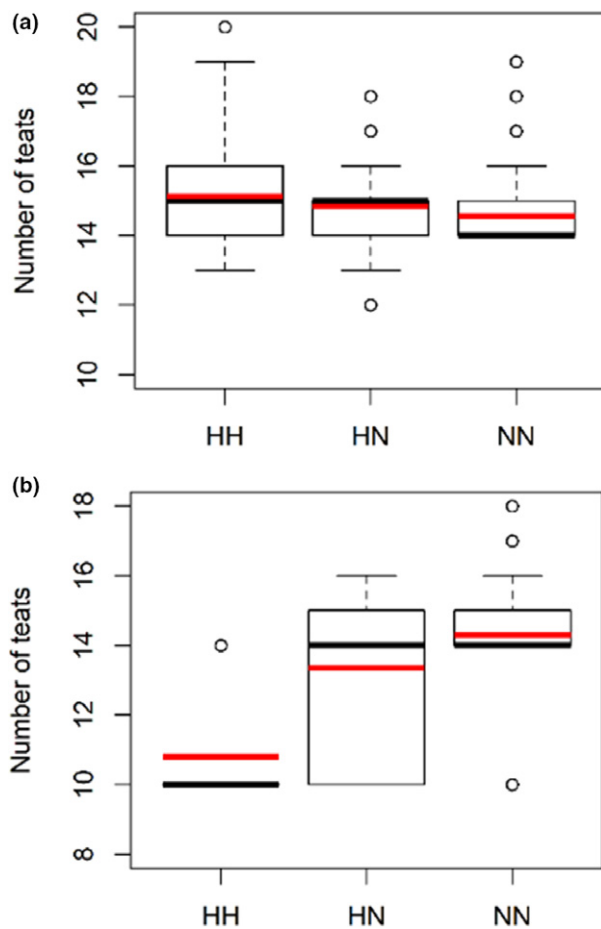
Figure 3 Boxplots showing the allelic effects of the top associated haplotype regions for the number of teats. (a) Genome region SSC7:97435001 bp in Italian Large White pigs and (b) genome region SSC15:134400001 bp in Italian Landrace pigs. Haplotypes have been treated as bi-allelic variants (H = haplotype allele and N = other N alleles). The red line indicates the average value.

some differences in the structure of linkage disequilibrium that could be potentially involved in the contrasting results between the two breeds (Fig. S4). Finally, we also investigated the whole set of haplotypes coming from the related haploblock and again no association was evidenced in the Italian Landrace breed (Table S6).

Discussion

Among the traits directly or indirectly affecting the reproduction efficiency of the sows, the number of teats in pigs is considered one of the parameters with the highest heritability. For this trait, most studies have indicated medium to high heritability (Rohrer & Nonneman 2017). Genomic heritability estimated in our analyses confirmed what was reported in a previous estimation we carried out in the Italian Large White breed, but with a lower number of pigs (Moscatelli et al. 2020). Our estimated genomic heritability obtained for the two breeds (h_G^2 ranged from 0.25 to 0.43) is

also similar to what was described by other authors in different pig breeds and populations using genomic based approaches (Duijvesteijn et al. 2014; Lopes et al. 2014; Arakawa et al. 2015; Balzani et al. 2016; Rohrer & Nonneman 2017; Lee et al. 2019; van Son et al. 2019; Moscatelli et al. 2020). It is interesting to note that in both heavy pig breeds haplotype-based estimation of h_G^2 resulted in a substantial increase of its value compared to the h_G^2 based on the single-marker genome analysis ($h_G^2 = 0.31$ obtained using haplotypes vs $h_G^2 = 0.25$ obtained using single-SNPs in Italian Large White and $h_G^2 = 0.43$ vs. $h_G^2 = 0.30$ in Italian Landrace). Therefore, a haplotype analysis extracted more information from the dataset and was able to recover a fraction of the so-called missing heritability, as already demonstrated in several other studies (e.g. Ehret et al. 2012).

It is also interesting to note that results of the GWAS in the two heavy pig breeds did not report any overlapping QTL region for teat number. This is surprising considering that the two breeds may share a quite similar genetic background as already demonstrated by whole-genome resequencing and genotyping data (Schiavo et al. 2020a,b; Bovo et al. 2020b). However, the two breeds have divergent selection histories that might have shaped their genome and fixed various genomic regions affecting the analysed trait. This consideration could be supported by the characteristics of the QTL regions reported in the two breeds: few QTL regions in the Italian Large White breed (only three significant regions) with quite high MAF, indicating balanced segregation of the alternative alleles in the population captured both using single-marker and haplotype-based analysis (with the exception for one QTL region); a larger number of QTL regions in the Italian Landrace breed (12 QTL regions), with low MAF and mainly detected with the haplotype-based analysis. In the latter breed, it seems that the haplotype approach worked better in capturing low frequency QTL alleles that could not be detected by a single-marker analysis. Low frequency QTL alleles might be more difficult to be detected using a standard SNP chip, due to the low linkage disequilibrium of the single-SNPs with causative mutations that do not have an ancestral and common origin across breeds (as a possible result of the ascertainment bias in the selection of the SNPs included in the SNP chip used in the study). It is also clear that alleles (haplotypes) with low MAF could potentially identify false-positive QTL regions. However, higher estimated genomic heritability of the haplotype analysis in this breed might indicate that most of the haplotype-detected QTL regions are actually true QTL regions, as we already reported in another study using a similar approach (Bovo et al. 2020a). Moreover, the low MAF at several QTL regions in this breed (captured by the haplotypes; Fig. S1) is also in line to the low frequency of pigs having fewer than 14 teats that we purposely included in the study (6.4%; Table S3). To maximise the variability for this trait, that in the Italian heavy pigs is in some way artificially truncated for the lowest part of the tail distribution

by the constant selection against animals with fewer than 14 teats (pigs of these breeds must have at least 14 teats to be registered to their Herd Books), we were able to genotype 124 out of 1941 pigs with a number of teats lower than 14, by selecting animals from the Italian Landrace population that had blood stored in the ANAS biobank (these animals were then eliminated from the Herd Book of the breed).

Other interesting insights emerged by comparing the results of the QTL identified in this study in the two breeds and considering also results of previous studies in the same populations and in other populations. Particularly, in the Italian Large White pigs, we expected to identify a strong QTL signal on SSC7 in the *VRTN* gene region as our previous studies pointed out that variability in this gene might be directly involved in affecting teat numbers (Dall'Olio et al. 2018; Moscatelli et al. 2020). Both single-marker and haplotype-based GWAS showed that the most significant QTL region in the Italian Large White breed completely overlapped the *VRTN* gene region, further confirming the candidacy of the variability in this gene. However, no major QTL segregates in this region in the Italian Landrace breed. This is puzzling considering that the *VRTN* g.20311_20312ins291 alleles segregate with a quite balanced frequency also in the Italian Landrace breed, similarly to the results we already reported in a previous study (Fontanesi et al. 2014). However, the inclusion in the association study of the *VRTN* g.20311_20312ins291 mutation confirmed the lack of association between the *VRTN* gene and the number of teats in the Italian Landrace pigs. Moreover, the genome architecture of this region in the Italian Landrace and in the Italian Large White breed seems quite different as it could be inferred from the linkage disequilibrium and haploblock analyses we carried out. Therefore, it seems that there could be breed specific differences on the effect of variability in this SSC7 genome region on the number of teats, as also reported by other authors for the same trait or for the correlated number of vertebrae traits (Rohrer et al. 2015; Zhang et al. 2015, 2016; Park et al. 2017; van Son et al. 2019). Haploblock structure of this SSC7 region between the two breeds was different (Fig. S4). Overall, the two breeds showed a different *P* of association for the haplotypes within the top associated haploblock detected in Italian Large White pigs (Table S6).

Variability in other candidate genes, including the *latent transforming growth factor binding protein 2* (*LTBP2*) and the *ATP binding cassette subfamily D member 4* (*ABCD4*) genes, mapped in this region, has been proposed to explain part of the effects that cannot be explained by the *VRTN* polymorphisms (Zhang et al. 2016; Park et al. 2017; van Son et al. 2019). Considering that we did not identify any QTL signal in this region in the Italian Landrace population, pigs of this breed might be homozygous at these other QTL alleles (that could cover the *VRTN* allele effects) or other breed specific effects could counterbalance the expected effect derived by

the *VRTN* gene variants. The lack of QTL reported in this region in the Italian Landrace breed differentiate this heavy pig population from other Landrace populations in which QTL for the number of teats have been reported in this SSC7 position (van Son et al. 2019). Other studies are needed to characterise in more detail this region in which several QTL for other production traits have been also already reported (e.g. Yue et al. 2003).

Among the QTL regions identified in our study, we could mention a few that have been also detected in other pig populations. On SSC10, a QTL emerged in the Italian Large White breed in the *FRMD4A* gene region, already indicated by van Son et al. (2019) to harbour a QTL for the number of teats in another Large White commercial population and in Duroc pigs. *FRMD4A* encodes a FERM domain-containing protein that regulates epithelial polarity by connecting ADP ribosylation factor 6 (*ARF6*), which is a central player in actin cytoskeleton dynamics and membrane trafficking, with the Par protein complex (Ikenouchi & Umeda 2010).

The most significant signal obtained in the Italian Landrace breed was on SSC15 in a region that was already reported to contain a QTL region for number of teats (including the *ARL4C* gene) in another Landrace population (van Son et al. 2019) further confirming the results we obtained. The closest annotated genes were the *transient receptor potential cation channel subfamily M member 8* (*TRPM8*) and the *secreted phosphoprotein 2* (*SPP2*) genes that based on their role cannot be considered obvious candidates. *TRPM8* encodes a receptor-activated non-selective cation channel involved in detection of sensations such as coolness that also plays a role in prostate cancer cell migration whereas SPP2 is involved in retinitis pigmentosa in humans and its function is to respond to elevated platelet cytosolic Ca^{2+}.

The obtained results indicated that the Italian Large White and Italian Landrace breeds have a different structure on the segregating QTL affecting teat number. It is worth speculation that this result (if also further extended for other reproductive traits) could contribute to justify, at least in part, the common practice of producing hybrid F1 gilts by crossing Italian Large White and Italian Landrace to exploit heterosis for reproductive related traits that derives by combing different alleles segregating in the two pure breeds.

Both single marker and haplotype-based genome wide association studies captured QTL regions that were complementary in most cases. In pig populations where the haplotype structure or the level of linkage between QTL alleles and SNPs is not known, it is useful to use both methodologies to fully exploit and dissect the genome of the pigs of different populations.

Acknowledgements

Conflict of interest

The authors declare they do not have any competing interests.

Data availability statement

The datasets used and analysed during the current study are available from the corresponding author on reasonable request.

References

Andersen I.L., Nævdal E. & Bøe K.E. (2011) Maternal investment, sibling competition, and offspring survival with increasing litter size and parity in pigs (*Sus scrofa*). *Behavioral Ecology and Sociobiology* **65**, 1159–67.

Arakawa A., Okumura N., Taniguchi M., Hayashi T., Hirose K., Fukawa K., Ito T., Matsumoto T., Uenishi H. & Mikawa S. (2015) Genome-wide association QTL mapping for teat number in a purebred population of Duroc pigs. *Animal Genetics* **46**, 571–5.

Balzani A., Cordell H.J., Sutcliffe E. & Edwards S.A. (2016) Heritability of udder morphology and colostrum quality traits in swine. *Journal of Animal Science* **94**, 3636–44.

Barbato M., Orozco-terWengel P., Tapio M. & Bruford M.W. (2015) SNeP: a tool to estimate trends in recent effective population size trajectories using genome-wide SNP data. *Frontiers in Genetics* **6**, 109.

Barendse W. (2011) Haplotype analysis improved evidence for candidate genes for intramuscular fat percentage from a genome wide association study of cattle. *PLoS One* **6**, e29601.

Bidanel J.P., Rosendo A., Iannuccelli N., Riquet J., Gilbert H., Caritez J.C., Billon Y., Amigues Y., Prunier A. & Milan D. (2008) Detection of quantitative trait loci for teat number and female reproductive traits in Meishan × Large White F2 pigs. *Animal* **2**, 813–20.

Borchers N., Reinsch N. & Kalm E. (2002) Teat number, hairiness and set of ears in a Piétrain cross: variation and effects on performance traits. *Archives in Animal Breeding* **45**, 465–80.

Bovo S., Ballan M., Schiavo G., Gallo M., Dall'Olio S. & Fontanesi L. (2020a) Haplotype-based genome-wide association studies reveal new loci for haematological and clinical–biochemical parameters in Large White pigs. *Animal Genetics* **51**, 601–6.

Bovo S., Mazzoni G., Bertolini F., Schiavo G., Galimberti G., Gallo M., Dall'Olio S. & Fontanesi L. (2019) Genome-wide association studies for 30 haematological and blood clinical-biochemical traits in Large White pigs reveal genomic regions affecting intermediate phenotypes. *Scientific Reports* **9**, 7003.

Bovo S., Ribani A., Muñoz M. *et al.* (2020b) Whole-genome sequencing of European autochthonous and commercial pig

breeds allows the detection of signatures of selection for adaptation of genetic resources to different breeding and production systems. *Genetics Selection Evolution* **52**, 33.

Bovo S., Ribani A., Muñoz M. *et al.* (2020c) Genome-wide detection of copy number variants in European autochthonous and commercial pig breeds by whole-genome sequencing of DNA pools identified breed-characterising copy number states. *Animal Genetics* **51**, 541–56.

Camacho C., Coulouris G., Avagyan V., Ma N., Papadopoulos J., Bealer K. & Madden T.L. (2009) BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421.

Chalkias H., Rydhmer L. & Lundeheim N. (2013) Genetic analysis of functional and non-functional teats in a population of Yorkshire pigs. *Livestock Science* **152**, 127–34.

Chang C.C., Chow C.C., Tellier L.C., Vattikuti S., Purcell S.M. & Lee J.J. (2015) Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* **4**, 7.

Dall'Olio S., Ribani A., Moscatelli G., Zambonelli P., Gallo M., Nanni Costa L. & Fontanesi L. (2018) Teat number parameters in Italian Large White pigs: phenotypic analysis and association with vertnin (*VRTN*) gene allele variants. *Livestock Science* **210**, 68–72.

Delaneau O., Marchini J. & Zagury J.-F. (2011) A linear complexity phasing method for thousands of genomes. *Nature Methods* **9**, 179–81.

Ding N., Guo Y., Knorr C. *et al.* (2009) Genome-wide QTL mapping for three traits related to teat number in a White Duroc x Erhualian pig resource population. *BMC Genetics* **10**, 6.

Duan Y., Zhang H., Zhang Z. *et al.* (2018) VRTN is required for the development of thoracic vertebrae in mammals. *International Journal of Biological Sciences* **14**, 667–81.

Duijvesteijn N., Veltmaat J.M., Knol E.F. & Harlizius B. (2014) High-resolution association mapping of number of teats in pigs reveals regions controlling vertebral development. *BMC Genomics* **15**, 542.

Ehret G.B., Lamparter D., Hoggart C.J., Whittaker J.C., Beckmann J.S. & Kutalik Z. (2012) A multi-SNP locus-association method reveals a substantial fraction of the missing heritability. *American Journal of Human Genetics* **91**, 863–71.

Felleki M. & Lundeheim N. (2015) Genetic heteroscedasticity of teat count in pigs. *Journal of Animal Breeding and Genetics* **132**, 392–8.

Fontanesi L., Scotti E., Buttazzoni L., Dall'Olio S. & Russo V. (2014) Investigation of a short interspersed nuclear element polymorphic site in the porcine vertnin gene: allele frequencies and association study with meat quality, carcass and production traits in Italian Large White pigs. *Italian Journal of Animal Science* **13**, 3090.

Hernandez S.C., Finlayson H.A., Ashworth C.J., Haley C.S. & Archibald A.L. (2014) A genome-wide linkage analysis for reproductive traits in F2 Large White × Meishan cross gilts. *Animal Genetics* **45**, 191–7.

Hirooka H., de Koning D.j., Harlizius B., van Arendonk J.a., Rattink A.p., Groenen M.a., Brascamp E.w. & Bovenhuis H. (2001) A whole-genome scan for quantitative trait loci affecting teat number in pigs. *Journal of Animal Science* **79**, 2320–6.

Hu Z.-L., Park C.A., Reecy J.M. (2019) Building a livestock genetic and genomic information knowledgebase through integrative developments of Animal QTLdb and CorrDB. *Nucleic Acids Research* **47**, D701–D710.

Ikenouchi J., Umeda M. (2010) FRMD4A regulates epithelial polarity by connecting Arf6 activation with the PAR complex. *Proceedings of the National Academy of Sciences, USA* 107, 748–53.

Kim J.S., Jin D.I., Lee J.H., Son D.S., Lee S.H., Yi Y.J., Park C.S. (2005) Effects of teat number on litter size in gilts. *Animal Reproduction Science* 90, 111–6.

Lee J., Lee S., Park J.-E., Moon S.-H., Choi S.-W., Go G.-W., Lim D. & Kim J.-M. (2019) Genome-wide association study and genomic predictions for exterior traits in Yorkshire pigs. *Journal of Animal Science* 97, 2793–802.

Lopes M.S., Bastiaansen JWM, Harlizius B., Knol E.F. & Bovenhuis H. (2014) A genome-wide association study reveals dominance effects on number of teats in pigs. *PLoS One* 9, e105867.

Lorenz A.J., Hamblin M.T. & Jannink J.-L. (2010) Performance of single nucleotide polymorphisms versus haplotypes for genome-wide association analysis in barley. *PLoS One* 5, e14079.

McKay R.M. & Rahnefeld G.W. (1990) Heritability of teat number in swine. *Canadian Journal of Animal Science* 70, 425–30.

Mikawa S., Morozumi T., Shimanuki S.-I., Hayashi T., Uenishi H., Domukai M., Okumura N. & Awata T. (2007) Fine mapping of a swine quantitative trait locus for number of vertebrae and analysis of an orphan nuclear receptor, germ cell nuclear factor (NR6A1). *Genome Research* 17, 586–93.

Mikawa S., Sato S., Nii M., Morozumi T., Yoshioka G., Imaeda N., Yamaguchi T., Hayashi T. & Awata T. (2011) Identification of a second gene associated with variation in vertebral number in domestic pigs. *BMC Genetics* 12, 5.

Moscatelli G., Dall'Olio S., Bovo S. *et al.* (2020) Genome-wide association studies for the number of teats and teat asymmetry patterns in Large White pigs. *Animal Genetics* 51, 595–600.

Park H.-B., Han S.-H., Lee J.-B. & Cho I.-C. (2017) Rapid Communication: High-resolution quantitative trait loci analysis identifies *LTBP2* encoding latent transforming growth factor beta binding protein 2 associated with thoracic vertebrae number in a large F2 intercross between Landrace and Korean native pigs. *Journal of Animal Science* 95, 1957–62.

Park H.-B., Lee J.-B., Han S.-H., Kim S.-G., Shin H.-S., Shin M.-C., Kang Y.-J. & Cho I.-C. (2018) Joint linkage and linkage disequilibrium mapping reveals association of *BRMS1L* with total teat number in a large intercross between Landrace and Korean native pigs. *Livestock Science* 218, 92–6.

Quinlan A.R. & Hall I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–2.

R Core Team (2018) *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing.

Rodríguez C., Tomás A., Alves E. *et al.* (2005) QTL mapping for teat number in an Iberian-by-Meishan pig intercross. *Animal Genetics* 36, 490–6.

Rohrer G.A. & Nonneman D.J. (2017) Genetic analysis of teat number in pigs reveals some developmental pathways independent of vertebra number and several loci which only affect a specific side. *Genetics Selection Evolution* 49, 4.

Rohrer G.A., Nonneman D.J., Wiedmann R.T. & Schneider J.F. (2015) A study of vertebra number in pigs confirms the association of vertnin and reveals additional QTL. *BMC Genetics* 16, 129.

Schiavo G., Bertolini F., Galimberti G., Bovo S., Dall'Olio S., Nanni Costa L., Gallo M. & Fontanesi L. (2020a) A machine learning approach for the identification of population-informative markers from high-throughput genotyping data: application to several pig breeds. *Animal* 14, 223–32.

Schiavo G., Bovo S., Bertolini F., Tinarelli S., Dall'Olio S., Nanni Costa L., Gallo M. & Fontanesi L. (2020b) Comparative evaluation of genomic inbreeding parameters in seven commercial and autochthonous pig breeds. *Animal* 14, 910–20.

Tang J., Zhang Z., Yang B. *et al.* (2017) Identification of loci affecting teat number by genome-wide association studies on three pig populations. *Asian-Australasian Journal of Animal Sciences* 30, 1–7.

Tönjes A., Koriath M., Schleinitz D. *et al.* (2009) Genetic variation in GPR133 is associated with height: genome wide association study in the self-contained population of Sorbs. *Human Molecular Genetics* 18, 4662–8.

Toro M., Dobao M.T., Rodrigáñez J. & Silio L. (1986) Heritability of a canalized trait: teat number in Iberian pigs. *Genetics Selection Evolution* 18, 173–84.

Tortereau F., Servin B., Frantz L. *et al.* (2012) A high density recombination map of the pig reveals a correlation between sex-specific recombination and GC content. *BMC Genomics* 13, 586.

Turner S. (2018) qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. *Journal of Open Source Software*.

van Son M., Lopes M.S., Martell H.J., Derks MFL, Gangsei L.E., Kongsro J., Wass M.N., Grindflek E.H. & Harlizius B. (2019) A QTL for number of teats shows breed specific effects on number of vertebrae in pigs: bridging the gap between molecular and quantitative genetics. *Frontiers in Genetics* 10, 272.

Verardo L.L., Silva F.F., Lopes M.S., Madsen O., Bastiaansen JWM, Knol E.F., Kelly M., Varona L., Lopes P.S. & Guimarães SEF (2016) Revealing new candidate genes for reproductive traits in pigs: combining Bayesian GWAS and functional pathways. *Genetics Selection Evolution* 48, 9.

Veroneze R., Lopes P.S., Guimarães SEF, Silva F.F., Lopes M.S., Harlizius B. & Knol E.F. (2013) Linkage disequilibrium and haplotype block structure in six commercial pig lines. *Journal of Animal Science* 91, 3493–501.

Wada Y., Akita T., Awata T. *et al.* (2000) Quantitative trait loci (QTL) analysis in a Meishan × Göttingen cross population. *Animal Genetics* 31, 376–84.

Willham R.L. & Whatley J.A. (1962) Genetic variation in nipple number in swine. *Zeitschrift für Tierzüchtung und Züchtungsbiologie* 78, 350–63.

Yue G., Stratil A., Cepica S. *et al.* (2003) Linkage and QTL mapping for Sus scrofa chromosome 7. *Journal of Animal Breeding and Genetics* 120, 56–65.

Zhang L., Liu X., Liang J. *et al.* (2015) Quantitative trait loci for the number of vertebrae on *Sus scrofa* chromosomes 1 and 7 independently influence the numbers of thoracic and lumbar vertebrae in pigs. *Journal of Integrative Agriculture* 14, 2027–33.

Zhang L.-C., Yue J.-W., Pu L. *et al.* (2016) Genome-wide study refines the quantitative trait locus for number of ribs in a Large White × Minzhu intercross pig population and reveals a new candidate gene. *Molecular Genetics and Genomics* 291, 1885–90.

Zhou X. & Stephens M. (2012) Genome-wide efficient mixed-model analysis for association studies. *Nature Genetics* 44, 821–4.

Supporting information

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

Figure S1 Quantile-quantile plots of the genome-wide association studies (GWAS) carried out in the Italian Large White (ILW) and Italian Landrace (IL) breeds using the single-marker- (SNP) and haplotype- (Haplotype) based approaches. Inflation factor (λ_{GC}) is reported.

Figure S2 Boxplots showing the allelic effects of the top associated SNPs and haplotype regions for the number of teats in the Italian Large White population. Haplotypes have been treated as bi-allelic variants (H = haplotype allele and N = other N alleles). In red is highlighted the average number of teats. Genome region SSC7:97435001 bp is also presented in Fig. 3.

Figure S3 Boxplots showing the allelic effects of the top associated SNPs and haplotype regions for the number of teats in the Italian Landrace population. Haplotypes have been treated as bi-allelic variants (H = haplotype allele and N = other N alleles). In red is highlighted the average number of teats. Genome region SSC15:134400001 bp is also presented in Fig. 3.

Figure S4 Pairwise linkage disequilibrium (LD) analysis of the *VRTN* gene region (SSC7) in (A) Italian Large White (ILW) and (B) Italian Landrace (IL) pigs. Only markers shared between the two populations are showed. LD was measured as R^2 and is presented in each box coloured considering the magnitude of linkage. The associated SNP and haplotypes detected in the ILW pigs are highlighted with a red star symbol whereas position of the *VRTN* gene is marked with a green arrow. DNA markers within the top associated haplotype (CHR7_B973_97200001_97670001_GAGAG) are marked with a red triangle. Frequency of the different haplotypes (MAF > 0.01) are also reported.

Table S1 Information on the datasets used in the single-marker (SNP) and haplotype-based (Haplotype) genome-wide association studies (GWAS) carried out in Italian Large White and Italian Landrace breeds.

Table S2 Teat counts in the Italian Large White pig population.

Table S3 Teat counts in the Italian Landrace pig population.

Table S4 Genome regions associated with the number of teats in Italian Large White and Italian Landrace pigs. Results are stratified by population, genome scan and chromosome.

Table S5 Comparison between pig populations in allele frequency, β and *P* of association of the top associated marker identified in the Italian Large White pigs.

Table S6 Comparison between pig populations in allele frequency, β and *P* of association of the top associated haplotype within the *VRTN* region (and all other haplotypes in the haploblock) detected in the Italian Large White pigs.