**ABSTRACT**

Sperm motility is directly related to the ability of sperm to move through the female reproductive tract to reach the ovum. Sperm motility is a complex trait that is influenced by environmental and genetic factors and is associated with male fertility, oocyte penetration rate, and reproductive success of cattle. In this study we carried out a GWAS in Italian Holstein bulls to identify candidate regions and genes associated with variations in progressive and total motility (PM and TM, respectively). After quality control, the final data set consisted of 5,960 records from 949 bulls having semen collected in 10 artificial insemination stations and genotyped at 412,737 SNPs (call rate >95%; minor allele frequency >5%). (Co)variance components were estimated using single trait mixed models, and associations between SNPs and phenotypes were assessed using a genomic BLUP approach. Ten windows that explained the greatest percentage of genetic variance were located on *Bos taurus* autosomes 1, 2, 4, 6, 7, 23, and 26 for TM and *Bos taurus* autosomes 1, 2, 4, 6, 8, 16, 23, and 26 for PM. A total of 150 genes for TM and 72 genes for PM were identified within these genomic regions. Gene Ontology enrichment analyses identified significant Gene Ontology terms involving with energy homeostasis, membrane functions, sperm-egg interactions, protection against oxidative stress, olfactory receptors, and immune system. There was significant enrichment of quantitative trait loci for fertility, calving ease, immune response, feed intake, and carcass weight within the candidate windows. These results contribute to understanding the architecture of the genetic control of sperm motility and may aid in the development of strategies to identify subfertile bulls and improve reproductive success.

**Key words:** candidate genes, dairy bulls, genome-wide association, semen quality

**INTRODUCTION**

In 2020, the world market for AI was valued ~$3.96 billion with an expected annual growth of ~5.94% by 2028. Among other factors, the quantity, quality, and sanitary status of the semen will determine the reproductive success of a bull. Although fertility of semen is highly influenced by age, season, nutrition, management, and the interval between ejaculations (Fuerst-Waltl et al., 2006; Suchocki and Szyda, 2015), it is also known that genetics and biological processes involved in sperm production and sperm-ovum interactions play a significant role (Druet et al., 2009; Yin et al., 2019b; Sweett et al., 2020).

Sperm motility (SM) affects the ability of the sperm to reach and penetrate the zona pellucida, fertilize the ovum, and initiate embryogenesis (Alves et al., 2020). Sperm motility is routinely measured by insemination centers using computer-assisted sperm analysis. However, it has been shown that SM is a complex phenotype which is affected by the environmental (Druet et al., 2009) and both additive and nonadditive genetic effects (VanRaden et al., 2011; Peñagaricano et al., 2012). Consequently, a better understanding of the genomic architecture of semen-related complex traits is expected to help improve fertility through the assisted selective breeding of bulls. The heritability estimates for SM vary widely among studies on dairy and beef cattle, from very low values of 0.01–0.13 (Gredler et al., 2007; Corbet et al., 2013; Suchocki and Szyda, 2015) to...
moderate and high values of 0.43–0.60 (Hering et al., 2014).

A large number of genes, some having pleiotropic effects on different semen traits and cell types, are involved in the production, maturation, and sperm viability and can affect semen quality and fertility (Ballow et al., 2006; Zhang et al., 2016). Here, we identify novel genomic regions and candidate genes putatively involved in total and progressive motility (PM) in Italian Holstein bulls using a GWAS approach.

**MATERIALS AND METHODS**

**Phenotypic and Genotypic Data**

We used 5,960 records of total and progressive SM from 949 Italian Holstein bulls with ages between 1 and 11 yr from 10 AI centers (Figure 1a), collected between 1995 and 2014 (Figure 1b). The number of measurements per sire varied between 3 and 78 for both total motility and PM traits. The procedures for semen collection are standardized across AI centers. Frozen-thawed sperm are routinely analyzed in the collecting centers using a CASA Integrated Visual Optical System (IVOS; CASA System, Hamilton Thorne Inc.) and a standardized protocol to assess SM. Total motility (TM) is the percentage of sperm moving irrespective of direction, and PM is the percentage of sperm that exhibit linear movement. Sperm phenotypic data were retrieved from the database of the Istituto Lazzari Spallanzani, in charge of official sperm quality controls of national bulls in Italy. The DNA extracted from semen was already available at the Biobank of Università Cattolica (see Mancini et al., 2014).

Illumina BovineSNP50 genotypes for 636 animals were imputed to high-density (777,962) with Beagle 4.1 (Browning and Browning, 2016) using a reference set of 1,006 animals, genotyped with the Illumina BovineHD array. Both data sets were previously updated to the ARS-UCD1.2 (Rosen et al., 2020) bovine assembly version using Plink 1.9 (Chang et al., 2015). Quality control excluded indels, SNPs with minor allele frequency lower than 0.05 (Wiggans et al., 2009), and duplicated markers from both data sets. For the BovineSNP50 data set, only SNPs in common with HD data set were used. Finally, the filtered reference data set, comprising 569,630 SNPs, was phased and missing genotypes imputed through Beagle 5.2 (Browning et al., 2021). Low quality imputed variants, with DR2 (dosage R², which is the estimated squared correlation between the estimated allele dose and the true allele dose), lower than 0.3 were excluded (Wu et al., 2019; Chen et al., 2021), giving a final imputed data set of 417,256 SNPs. As 91 animals were genotyped both with 50k and HD panels, these were removed from the imputed data set and their HD data were considered for the downstream analysis. Therefore, the GWAS analysis was performed using 949 animals with phenotype and genotype information (404 animals genotyped with BovineHD and 545 animals imputed from BovineSNP50). A further quality control was carried out to exclude SNPs with minor allele frequency <0.05. After quality control, 412,737 SNPs and 949 animals were used for further analyses.
**Genome-Wide Association Study**

The genomic BLUP using individuals with repeated measurements were analyzed with a repeatability single trait animal model:

$$y = X\beta + Z_u + W_{pe} + e,$$

where $y$ is a vector of observed phenotypes (TM or PM) for all sires, $\beta$ is the vector of fixed effects (AI center, contemporary group: year-season, quadratic, and linear effects of age), $u$ is the random vector of additive genetic effects, $pe$ is the random vector of permanent environmental effects, and $e$ is the random vector of residual effects. The matrices $X$, $Z$, and $W$ are the incidence matrices of fixed effects, additive genetic effects, and permanent environmental effects, respectively. Random effects were assumed to follow a multivariate normal distribution, as follows:

$$
\begin{bmatrix}
    u \\
    pe \\
    e
\end{bmatrix}
\sim N
\begin{pmatrix}
    G\sigma_u^2 & 0 & 0 \\
    0 & I_{pe} & 0 \\
    0 & 0 & I_{e}
\end{pmatrix},
$$

where $\sigma_u^2$, $\sigma_{pe}^2$, and $\sigma_e^2$ are the additive genetic variance, the permanent environment variance, and the residual variance, respectively; $I$ is an identity matrix. $G$ is a genomic relationship matrix derived from marker information (VanRaden, 2008). The heritability was estimated as the ratio of additive genetic variance to the total phenotypic variance ($\sigma^2_u + \sigma^2_{pe} + \sigma^2_e$) and the repeatability was calculated as sum of individual variance ($\sigma^2_u + \sigma^2_{pe}$) divided by phenotypic variance.

The percentage of additive genetic variance explained by the $i$th SNP window was then calculated as

$$\frac{Var(u_i)}{\sigma^2_u} \times 100 = \frac{\sum_{j=1}^{x} Z_{ij}^2 u_j}{\sigma^2_u} \times 100,$$

where $u_i$ is the genetic value of the $i$th genomic region under consideration, $\sigma^2_u$ is the total additive genetic variance, $x$ is the total number of adjacent SNPs within the 1 Mb region, $Z_{ij}$ is the vector of gene content of the $j$th SNP for all windows, and $u_j$ is the effect of the $j$th SNP within $i$th window. GWAS results are reported as the percentage of genetic additive variance explained by windows of 1 Mb and plotted using CMplot package in R v.1.3.1073 (R core team 2020). Additionally, we calculated the $P$-value of individual SNP according to Aguilar et al. (2019). All procedures were performed using BLUPF90 family programs (Misztal et al., 2014). To control false positives, we applied the false discovery rate (FDR) method for multiple testing according to Benjamini and Hochberg (1995) as follows:

$$FDR = m \times \frac{P_{max}}{n},$$

where $m$ is the number of times to be tested, $P_{max}$ is the genome-wide significance level empirical $P$-value of FDR adjusted, and $n$ is the number of significant SNPs at the assigned FDR level (i.e., 0.05).

**Gene Annotation and Functional Enrichment**

Our results indicated that single SNPs with the smallest $P$-values did not explain a large percentage of genetic variance. Therefore, we used the percentage of variance explained by SNPs within consecutive 1 Mb windows to identify genomic regions of interest for further investigation. Genes within the 10 windows explaining the largest amount of variance were annotated using the GALLO v.0.99 R package (Fonseca et al., 2020) and the ENSEMBL annotation of the *Bos taurus* ARS-UCD 1.2 genome assembly (http://bovinegenome.org). The genes identified from ENSEMBL were collected to perform a functional annotation by DAVID v6.8 (https://david.ncifcrf.gov/home.jsp) and REVIGO (Supek et al., 2011) using Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway (https://www.genome.jp/kegg) databases for *B. taurus*. The annotation and enrichment analyses of QTL in adjacent windows (1 Mb) were assessed for the effects of neighboring SNPs, which can be captured due to linkage disequilibrium. The QTL enrichment was evaluated for all the annotated QTL and the $P$-values were calculated and corrected for multiple testing, using a value of FDR <0.05 to identify the enriched QTL. The annotation and enrichment analyses were performed using the GALLO v0.99, package and the Animal QTLdb for the ARS-UCD 1.2 bovine genome (http://bovinegenome.org).

**RESULTS**

**Descriptive Statistics and Genetic Parameters**

The bulls had a mean of 55.33% (12.70 SD) for TM and 43.21% (11.04 SD) for PM. The fixed effect of the contemporary group and the quadratic and linear effect of age (covariable) were highly significant in the model for both traits ($P < 0.001$). The effect of the AI center was significant ($P < 0.01$) and the estimates of herita-
Table 1. Additive genetic \( (\hat{\sigma}^2_a) \), permanent environmental \( (\hat{\sigma}^2_{pe}) \), and residual \( (\hat{\sigma}^2_e) \) variance and genetic parameters heritability \( \hat{h}^2 \), genetic correlation \( \hat{r}_g^2 \), and repeatability (± SE) for total motility (TM) and progressive motility (PM) in Italian Holsteins bulls.

| Trait | \( \hat{\sigma}^2_a \) | \( \hat{\sigma}^2_{pe} \) | \( \hat{\sigma}^2_e \) | \( \hat{h}^2 \) | \( \hat{r}_g^2 \) | r |
|-------|-----------------|-----------------|-----------------|--------------|-----------------|---|
| TM    | 14.96 ± 2.02    | 20.85 ± 2.27    | 78.80 ± 1.13    | 0.13 ± 0.0062| 0.62 ± 0.010    | 0.31 ± 0.0027 |
| PM    | 13.05 ± 2.28    | 24.36 ± 2.51    | 82.43 ± 1.17    | 0.10 ± 0.0045| 0.33 ± 0.0019   | --- |

Repeatability: \( r = \sigma_a^2 + \sigma_{ye}^2 / \sigma_a^2 + \sigma_{pe}^2 + \sigma_e^2 \).

QTL Regions and Candidate Genes

As previously noted, our results indicated that the SNPs with the smallest \( P \)-values did not explain a large percentage of genetic variance (Figure 2–3), contrary to expectation. \( P \)-values are influenced by sample size (Greenland et al., 2016) and, in particular, by allele frequencies (Aguilar et al., 2019).

The 10 windows that explained the greatest percentage of genetic variance were located on BTA 1, 2, 4, 6, 7, 23, and 26 for TM and BTA 1, 2, 4, 6, 8, 16, 23, and 26 for PM (Figure 4). These regions explained 6.91% and 6.07% of the additive genetic variance for TM and PM, respectively (Table 2). A total of 150 genes for TM and 107 genes for PM were annotated by GALLO within these significant windows (Table 2). Among them, 130 genes for TM and 90 genes for PM were found in the *Bos taurus* DAVID database.

The DAVID annotations and their relative \( P \)-values were grouped using REVIGO into biological processes (BP), cellular components (CC), and molecular functions (MF), each comprising 13, 23, and 45 genes for TM and 24, 18, and 61 for PM (Table 3). A total of 42 genes are in common between TM and PM.

Some QTL within the genomic windows explaining the greatest percentage of genetic variance were associated with reproductive traits: these included BTA 1 and 6 for TM, and BTA 1 and 23 for PM. The enrichment analysis of the QTL (FDR-corrected \( P < 0.05 \)) for TM identified 28 QTL related to the ease of calving and 6 QTL associated with the immune response on BTA 6 and BTA 23. In addition, QTL involved with feed intake (7), carcass weight (24), and fat thickness (10) were associated with TM on BTA 6. For PM, 2 QTL related to reproductive efficiency and with milk yield were identified on BTA 23 (Table 4).
DISCUSSION

Genetic Parameters

The heritability for TM and PM estimated in this study was low (TM: 0.14 ± 0.0062 and PM: 0.10 ± 0.0045) but comparable to estimates obtained by others using the pedigree relationship matrix A (0.13, Druet et al., 2009; 0.12, Yin et al., 2019a) for PM. High heritability estimates, of up to 0.43 have been reported for motility in Holstein bulls (Al-Kanaan et al., 2015; Butler et al., 2020); however, estimates for PM in this breed are low (0.13–0.15, Butler et al., 2020). The differences in heritability estimates may be the result of environmental and management differences that we found to have a significant effect on these traits. In our study estimated repeatability was moderate (0.31 and 0.33 for TM and PM, respectively) which suggests an effect of the permanent environment and an even higher effect of the residual variance.

Candidate Genes and QTL Annotation

In the present study, no loci with significant P-values were found using a single SNP analysis. The SNPs with the smallest P-values would be expected to explain a relatively higher proportion of genetic variation. However, this is not always the case, as P-values are affected by sample size, the magnitude of the effect detected (Greenland et al., 2016), and in particular, by allele frequencies (Aguilar et al., 2019). Liu et al. (2021), in a single-step GWAS in swine, found no overlap between the first 20 SNPs with lowest P-values and those explaining the largest variance. These authors attribute this discrepancy to linkage disequilibrium between SNPs and the causal variants and interaction between SNPs. Therefore, we used the variance explained by sliding window intervals to identify genomic regions involved in sperm motility.

The ten 1-Mb windows with highest additive genetic variance for TM and PM explained, respectively, 6.91% and 6.07% of the additive genetic variance, indicating the highly polygenic nature of these traits. Other studies have also found that several regions across the genome each make small contributions to the total genetic variation for SM. For example, Yin et al. (2019b) found that the most significant regions, located on BTA 1, 3, 4, 5, 11, 17, and 19, explained less than 0.1% of the genetic variance for SM in Holstein bulls. In beef cattle, Sweett et al. (2020) identified 10 windows on BTA 9, 13, 20, and 24 that explained 7.17% of the genetic variance of this trait. There is little consistency in the regions identified across studies, even for signals on the same chromosomes; Rezende et al. (2018) identified genes associated with SM on BTA 5, 11, 22, and 25 in Jersey cattle, with no regions shared with other studies.

The window with highest additive genetic variance we identified was on BTA1 (102,482,169–103,484,977), which explained 1.29% of the genetic variance for TM and 1.62% for PM. This window included the gene E1BHJ0, which codes for profilin. Profilins are ubiquitous proteins present in mammals that regulate actin polymerization and are essential for cell viability (Witke et al., 2001), fertilization, and normal embryo development in cattle (Rawe et al., 2006). Four profilin isoforms are expressed in mammalian testis, including spermatogenic cells (Behnen et al., 2009). Profilin-1 and profilin-3 play an important role in mouse sperm

Figure 3. Manhattan plot displaying the association of all SNP for progressive motility (PM: P-values of individual SNP effects) and quantile-quantile plots (right), which show the observed distribution of P-values against the expected P-values under the null hypothesis of no association. The horizontal axis shows −log_{10} transformed expected P-values and the vertical axis indicates −log_{10} transformed observed P-values.
metabolism and motility (Vicens et al., 2014) and the loss of profilin-3 has been associated with impairment of spermiogenesis (Umer et al., 2021). The second most important region (0.98% variance) for PM was on BTA26 (46,055,585–47,048,342) and included cytokinesis dedicator 1 (DOCK1) which encodes a member of the migration and invasion dedicator of the cytokinesis protein superfamily which is involved in cell motility (Valente et al., 2016). With regard to TM, the second most important region (0.83% variance) was on BTA 23 (24,376,529–25,372,486) and included IL17A and IL17F which have an immunoregulatory role in the male reproductive tract (Bagri et al., 2022). Many of the genes in the other significant windows that fell into BP, CC, and MF categories (Table 3) are associated with spermatogenesis and male fertility [e.g., the peptidylarginine deiminase family (PAD), on BTA 2]. The PAD are a class of Ca$^{+2}$ dependent enzymes that are widely expressed in mammalian tissues including ovules, ovaries, testes, and in early embryos (Wang and Wang, 2013). The PAD enzymes perform the citrullination of arginine residues, which affects the stability and degradation of proteins, and play a crucial role during cell development, differentiation, and fertility (Fichtner et al., 2020). The enzyme PADI6 is involved in decondensation of sperm chromatin (Wright et al., 2003; Alghamdi et al., 2019). In humans, sperm chromatin organization has been shown to affect sperm function during fertilization and early embryo development (Miller et al., 2010) and has been implicated

Figure 4. Circular Manhattan plot for the proportion of genetic variance explained by nonoverlapping windows of 1 Mb for (a) total motility (TM) and (b) progressive motility (PM) in Italian Holstein Bulls. The scale bar represents the percentage of variance explained by each window. Chr = chromosome.
Table 2. Genomic regions and genes in the 10 windows explaining the highest percentage of genetic variance for total motility (TM) and progressive motility (PM) in Italian Holstein Bulls

| Window region | Chr¹ | Start | End | TM variance (%) | PM variance (%) | Gene
|---------------|------|-------|-----|-----------------|-----------------|------
| 1             | 65,216,449 | 66,211,180 | 0.55 | * | ENSBTAG00000048371, FBXO10, GOLGB1, GTF2E1, HCLS1, HGD, NDUFB4, POLQ, RABL3, STXBPF5L |
|               | 68,304,713 | 69,303,189 | 0.48 | * | CDC414, ENSBTAG00000053239, ITGB5, KALRN, ROPN1, UMPS |
|               | 102,624,551 | 103,622,454 | 1.29 | 1.62 | E1B1H0 |
|               | 148,523,333 | 149,497,374 | 0.36 | * | CHAF1B, CLDN14, HLC5, MORC3, PIGP, RIPPLY3, SIM2, TTC3, VPS26C |
| 2             | 134,98,1487 | 135,980,997 | 0.50 | 0.43 | ARHGEF10L, ARHGEF19, ATP1A2, CPLANE2, CROCC, EPHA2, FBXO42, MFAP2, NECAP2, PADI1, PADI2, PADI3, PADI4, PADI6, RCC2, SDHB, SPATA21, SZRD1 |
| 4             | 69,538,518 | 70,533,461 | 0.78 | * | CBFX3, HNRNPAP2B1, NFE2L3, SNX10 |
|               | 90,956,171 | 91,949,079 | 0.39 | 0.44 | ARF5, FSCN3, GCC1, GRM8, PAX4, SND1, ZNF800 |
| 6             | 37,145,522 | 38,144,195 | 0.38 | * | DCAF16, FAM184B, LAP3, LCORL, MED28, NCAKG |
|               | 102,406,084 | 103,403,193 | 0.41 | 0.42 | CRMP1, CTD, DSSP, ENSBTAG0000006508 |
|               | 102,624,551 | 103,622,454 | 1.29 | 1.62 | CBFX3, HNRNPAP2B1, NFE2L3, SNX10 |
| 7             | 44,213,448 | 45,212,096 | 0.44 | * | CBFX3, HNRNPAP2B1, NFE2L3, SNX10 |
| 8             | 54,914,798 | 55,908,878 | 0.52 | * | TLE1 |
| 16            | 70,458,024 | 71,454,118 | 0.49 | * | ANGEL2, ATF3, BATF3, DTL, FLVR1, NENF, NSL1, PACC1, PPP2R5A, RPS6KC1, SPATA45, TATDN3, VASH2 |
| 23            | 336,748 | 1,330,072 | 0.49 | * | CBFX3, HNRNPAP2B1, NFE2L3, SNX10 |
|               | 24,376,529 | 25,372,486 | 0.83 | * | CBFX3, HNRNPAP2B1, NFE2L3, SNX10 |
|               | 27,974,488 | 28,973,510 | 0.74 | * | ABCF1, ATAT1, BOLA, BOLA-NC1, C23H6orf136, C23H6orf15, CHCR1, CDSN, DDR1, DHHX16 |
|               | 28,014,882 | 29,014,696 | 0.56 | * | ABCF1, ATAT1, BOLA, BOLA-NC1, C23H6orf136, C23H6orf15, CHCR1, CDSN, DDR1, DHHX16 |
| 26            | 46,055,585 | 47,048,342 | 0.52 | 0.98 | DOCK1, INSYN2 |

¹Chr = chromosome.

*Regions not identified for the trait; "genes associated only with TM. bgenes associated only PM.

*Regions not identified for the trait; "genes associated only with TM. ggenes associated with TM and PM.
Table 3. Gene Ontology terms significantly associated with biological processes (BP), cellular components (CC), and molecular functions (MF) related to total motility (TM) and progressive motility (PM)

| Category                  | Group                                      | GO: Term                              | Gene name | BTA | TM FDR | PM FDR |
|---------------------------|--------------------------------------------|---------------------------------------|-----------|-----|--------|--------|
| Biological process        | Alpha-AA biosynthetic process              | GO: 0036414 Histone citrullination    | PADI1, PADI2 | 2   | 9.74 × 10⁻⁹ | 1.5 × 10⁻⁶ |
|                           | GO: 0018101 Protein citrullination          | PADI6, PADI4 PADI3                     |           |     |        |        |
|                           | GO: 0018195 Peptidyl-arginine modification | GSTA3, GSTA4, GSTA1, GSTA5             |           | 23  | 2.23 × 10⁻³ |        |
|                           | Glutathione metabolic process             |                                       |           |     |        |        |
| Antigen processing and presentation | GO: 0002428 antigen processing and presentation | JSP.1, BOLA BOLA-NC1                  |           | 23  | 0.0016 | 0.0046 |
| Immune system process     | GO: 0002376 immune system process          | EPHA2, PADI2                          |           | 2   | *      | 0.0192 |
|                           |                                             | SNX10                                 |           |     |        |        |
|                           |                                             | BOLA, BOLA-NC1, FLOT1, JSP.1, BOLA    |           |     |        |        |
|                           |                                             | MGC10, RNF29, TRIM10, TRIM15, TRIM81  |           |     |        |        |
|                           |                                             | TRIM40 DOCK1                          |           |     |        |        |
|                           |                                             | OR12D2, OR10C1, OR2H1, OR5V1, OR12D3 |           | 23  | *      | 4.8 × 10⁻¹⁰ |
| SENSORY PERCEPTION OF CHEMICAL STIMULUS | GO: 0003008 system process                | OR12D2, OR12D3                         |           | 23  | *      | 5.1 × 10⁻⁶ |
|                           |                                             | GABBR1, OR12D2, OR10C1, OR2H1, OR5V1 |           |     |        |        |
|                           |                                             | OR12D3, OR12D3                         |           |     |        |        |
| Cellular components       | Microtubule cytoskeleton                   | CCHCR1, CIKL1, EFHC1, MCM3, PKHD1     |           | 23  | 0.0037 | *      |
|                           |                                             | CBX3, SNX10                            |           | 4   | *      | 0.0058 |
|                           |                                             | EVC                                   |           | 6   |        |        |
|                           |                                             | PPP2R5A                               |           | 16  |        |        |
|                           |                                             | CRCC                                  |           | 2   | 0.0037 | 0.0058 |
|                           |                                             | CRMP1, JAKMIP1                        |           | 6   |        |        |
|                           |                                             | ATAT1, FLOT1, TUBB                     |           | 23  |        |        |
|                           |                                             | ATAT1, EFHC1                          |           | 23  | 0.0002 | *      |
|                           |                                             | PDKHD1                                |           |     |        |        |
|                           |                                             | CCHCR1, EFHC1, MCM3, PKHD1            |           | 23  | 0.0320 | *      |
|                           |                                             | CBX3                                  |           | 4   | *      | 0.0310 |
|                           |                                             | EVC                                   |           | 6   |        |        |
|                           |                                             | PPP2R5A                               |           | 16  |        |        |
|                           |                                             | CRMP1, JAKMIP1                        |           | 6   |        |        |
|                           |                                             | ATAT1, FLOT1                          |           | 23  |        |        |
|                           |                                             | PPP1R10, PPP1R11, PPP2R5C             |           | 23  | *      | 0.0045 |
| FOCAL ADHESION            | Focal adhesion                             | EPHA2, NECAP2                         |           | 2   | 0.0713 | *      |
|                           |                                             | ATP13A2                               |           |     |        |        |
|                           |                                             | MDC1, ATAT1, FLOT1                    |           | 23  |        |        |

Continued
Table 3 (Continued). Gene Ontology terms significantly associated with biological processes (BP), cellular components (CC), and molecular functions (MF) related to total motility (TM) and progressive motility (PM).

| Category | Group | GO: Term | Gene name |
|----------|-------|----------|-----------|
| Molecular Functions | Hydrolase activity acting on carbon-nitrogen (but not peptide) bonds | GO: 0004668 protein-arginine deiminase activity | PADI2, PADI1, PADI6, PADI4, PADI3 |
| | | GO: 0016830 hydrolase activity acting on carbon-nitrogen (but not peptide) bonds | |
| Binding | GO: 0043167 ion binding | FBOX10, GTF2E1, HGD, KALRN, MORC1, LAP3, CSNK1G2, FSTL4, HSPA4, CILK1, EFHC1, GCM1, MCM3, TCF19 |
| | GO: 0005488 binding | CROCC, MFAP2, RCC2, CBX3, FSCN3, HNRNPA2B1, NFE2L3, PAX4, SND1, SNX10, CRMP1, DSPP, JAKMIP1, WFS1, TLE4, PPP2R5A, RPS6KC1, CDSN, FLOT1, GTF2H1, KHDRBS2, MDC1, MOG, NMR, PPP1R11, PPP1R18, RPP21 |
| | GO: 0043167 ion binding | ATP13A2, CPLANE2, EPHA2, PADI1, PADI2, PADI3, PADI4, PADI6, SDHB, SPATA21, ARF5, SPARCL1, ABCF1, DDR1, DHX16, GNL1, POLR1H, PPP1R10, PRR3, RNF39, TRIM10, TRIM15, TRIM31, TRIM40, TUBB, VAR52, ZFP57 |
| | GO: 0004364 glutathione transferase activity | GSTA5, GSTA3, GSTA4, GSTA2, GSTA3 |
| | Olfactory reception activity | GO: 0040984 olfactory reception activity | OR12D12, OR12C1, OR2H1, OR5V1, OR12D3, EPHA2, OR12D12, OR12C1, OR2H1, OR5V1, OR12D3 |

1FDR = false discovery rate.
*Genome Ontology not identified for the trait.
in subfertility (Souza et al., 2018). The BoLA class 1 genes and JSP.1 are associated with both PM and TM. These genes map on BTA 23 and are part of the major histocompatibility complex involved in antigen presentation (Tijjani et al., 2019). The major histocompatibility complex antigens play a key role in immune response and affect production, diseases, and fertility traits, in particular the sperm-ovum interaction (Ellis and Hammond, 2014). BoLA-NC1 is essential for the establishment and maintenance of pregnancy and in the maternal immune response to the fetus (Shu et al., 2012).

Five genes in the most significant GO biological processes term (GO:BP), were associated exclusively with TM. Glutathione S-transferase alpha (GSTA), which is within one of the regions with the highest effect on BTA 23, belongs to the GST superfamily of multigenic and multifunctional isoenzymes that catalyze glutathione-dependent reactions (GO:BP Glutathione metabolic process and GO:MF Glutathione transferase activity). Expression of the GST has been reported in male reproductive tissues in cattle (Juyena and Stelletta, 2012), pigs (Pérez-Patiño et al., 2019), and humans (Dutta et al., 2021) including sperm, seminal plasma, and in testicular somatic cells, particularly in Leydig and Sertoli cells (Benbrahim-Tallaa et al., 2002). GSTA also is expressed in spermatogonia, Sertoli cells, spermatids, and spermatozoa of pig testes (Luo et al., 2019) and is present in the cytoplasm of Sertoli cells, Leydig cells, spermatogonia, primary spermatocytes, secondary spermatocytes, and spermatids of seminiferous tubules (Kalebic et al., 2013). Tubulin plays a role in microtubule formation in sperm cells of buffaloes (Fu et al., 2019) and is expressed in spermatogonia, Sertoli cells, spermatids, and spermatooza of pig testes (Luo et al., 2010). In buffalo sperm, α-tubulin has been shown to provide structural support, facilitate cell motility, and to participate in oocyte binding and activation (Patil et al., 2020). JAKMIP1 and PKDH1 are related to sperm function and cell polarization (Hatzirodotos et al., 2014), embryo development and survival (Gigarel et al., 2008), respectively. MDC1, EPHAH2, ATAT1, and FLOT1 code for proteins that are associated with focal adhesion, a process involved in fertilization, ovum transport, and sperm-ovum interaction (Talbot et al., 2003). Additionally, PPP1R10, PPP1R11, PPP2R2C, and FSCN3 genes were identified and code for proteins related to spermatogenesis, SM, and fertility (Han et

### Table 4. Quantitative trait loci annotated and significant in the enrichment analysis for total (TM) and progressive (PM) motility

| Trait          | QTL type       | QTL class                       | BTA | Observed number of QTL | Expected number of QTL | P-value   | FDR 1 |
|----------------|----------------|---------------------------------|-----|------------------------|------------------------|-----------|-------|
| TM            | Production     | Average daily feed intake       | 6   | 7                      | 9                      | 2.75 × 10^{-9} | 6.48 × 10^{-9} |
|               | Reproduction   | Calving ease (direct)           | 6   | 28                     | 37                     | 2.83 × 10^{-11} | 1.33 × 10^{-7} |
|               | Meat and carcass| Carcass weight                  | 6   | 24                     | 137                    | 3.01 × 10^{-9} | 3.53 × 10^{-4} |
|               | Health         | Fat thickness at the 12th rib    | 6   | 10                     | 21                     | 5.15 × 10^{-7} | 8.07 × 10^{-3} |
|               | Health         | Antibody-mediated immune response| 23  | 4                      | 10                     | 3.76 × 10^{-4} | 3.53 × 10^{-3} |
| PM            | Milk           | Milk yield, and percentage of protein and fat in milk (EBV) | 1   | 1                      | 7                      | 0.00674 | 0.00549 |
|               | Reproduction   | Reproductive efficiency         | 23  | 1                      | 7                      | 0.01097 | 0.00974 |

1FDR = false discovery rate.
al., 2007, 2008; Cheng et al., 2009; Gao et al., 2019), PPP2R5A codes for protein phosphatase 2 regulatory subunit B alpha and is correlated with spermatogonial mitosis and it is related to sphingolipid signaling pathway which is associated with SM (Sararakul et al., 2018; Cai et al., 2021). Genes within the region of BTA 23, which showed an association with PM in the current study, belong to the OR family and KHDRBS2. The OR are involved in chemotaxis and determine speed, strength, and direction of movement of the spermatozoa toward the ovum (Ali et al., 2021). Of the 20 to 66 testicular OR known in mammals, at least 8 are expressed in sperm cells, testis, and epididymus (Milardi et al., 2018). In humans, the OR influence spermatogenesis, epididymal growth, acrosome reaction, initiate flagella movement and facilitate cell-to-cell communication (Olaniany et al., 2021). The OR in the sperm cell activate proteins that increase intracellular calcium ions and hence motility (Vanderhaeghen et al., 1997). KHDRBS2, has been associated with fertility in San Martinero cattle (De León et al., 2019) and pregnancy status in Brahman beef cattle (Reverter et al., 2016).

The QTL enrichment analysis for TM revealed QTL associated with feed intake, calving ease, carcass weight, and fat thickness on BTA 6, and antibody-mediated immune response and cell-mediated immune response on BTA 23. Previous studies have identified other QTL on BTA 23 related to genes such as IL4 (interleukin-4), antibody-mediated immune response, and DMI that may be associated with SM (Wang et al. 2012).

CONCLUSIONS

In this study, we identified a large number of genes associated with SM and involved with sperm and sperm-related traits in Italian Holsteins bulls. These findings suggest that many genes controlling SM could have pleiotropic effects and be functionally correlated with other traits related to fertility. Our results show the highly polygenic nature of SM, with each of numerous candidate genes responsible for a small proportion of the genetic variance. SM influences male fertility, a trait that has probably experienced strong selective pressure even before cattle domestication, about 10,000 years ago. As such, beneficial alleles of genes with major effects on this trait are likely to have reached fixation, leaving only variants with small effect in the domesticated populations, which were identified by GWAS as presented here. The characterization of selection signatures across different timescales, as well as the analysis of ancient genomes from domestic and wild animals, may help in identifying variants related to fertility that have experienced selection before and after domestication events.

ACKNOWLEDGMENTS

J.L.W. was supported by the project LEO Livestock Environment Open-data, 16.2 – PSRN 2014-2020 funded through Fondo Europeo Agricolo per lo Sviluppo Rurale (FEASR). A.A. and P.A.M. received support from the Italian Ministry of Education, University and Research (MUR; Rome, Italy) for Progetti PRIN2017 20174BTC4R. The authors have not stated any conflicts of interest.

REFERENCES

Aguiar, L. A., Legarra, F., Cardoso, Y., Masuda, D., Lourenco, and I. Misetal. 2019. Frequentist p-values for large-scale-single step genome-wide association, with an application to birth weight in American Angus cattle. Genet. Sel. Evol. 51:28. https://doi.org/10.1186/s12711-019-0469-9.

Al-Kanaan, A., S. König, and K. Brügemann. 2015. Effects of heat stress on semen characteristics of Holstein bulls estimated on a continuous phenotypic and genetic scale. Livest. Sci. 177:15–24. https://doi.org/10.1161/j.vasc.1389.005003.

Alghamdi, M., D. Alasmari, A. Assiri, E. Mattar, A. A. Aljaddawi, S. G. Alattas, and E. M. Redwan. 2019. An overview of the intrinsic role of circullitation in autoimmune disorders. J. Immunol. Res. 2019:7592851. https://doi.org/10.1155/2019/7592851.

Ali, M. A., Y. Wang, Z. Qin, X. Yuan, Y. Zhang, and C. Zeng. 2021. Odorant and Taste Receptors in sperm chemotaxis and cryopreservation: Roles and implications in sperm capacitation, motility, and fertility. Genes (Basel) 12:488. https://doi.org/10.3390/genes12040488.

Alves, M. B. R., E. C. C. Celeghini, and C. Belleannée. 2020. From sperm motility to sperm-borne microRNA signatures: New approaches to predict male fertility potential. Front. Cell Dev. Biol. 8:791. https://doi.org/10.3389/fcell.2020.00791.

Bagri, P., V. C. Anipindi, and C. Kaushic. 2022. The role of IL-17 during infections in the female reproductive tract. Front. Immunol. 13:861444. https://doi.org/10.3389/fimmu.2022.861444.

Ballow, D., M. L. Meistrich, M. Matzuk, and A. Rajkovic. 2006. Sdhhl1 is essential for spermatogonial differentiation. Dev. Biol. 294:161–167. https://doi.org/10.1016/j.ydbio.2006.02.027.

Bhade, M. K., Murk, P. Kursula, H. Cappallo-Obermann, M. Rothkugel, A. L. Kierszenbaum, and C. Kirchhoff. 2009. Testis-expressed profilins 3 and 4 show distinct functional characteristics and localize in the acrosplaxome-manchette complex in spermatids. BMC Cell Biol. 10:34. https://doi.org/10.1186/1471-2229-10-34.

Benbrahim-Tallaa, L., E. T Bone, G. Tosser-Klopp, F. Hatey, and M. Benahmed. 2002. Glutathione S-Transferase Alpha expressed in the male reproductive tract. Front. Cell Biol. 10:34. https://doi.org/10.1016/j.cellimmunol.2001.08.005.

Benjamin, Y., and Y. Hochberg. 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. J. R. Stat. Soc. B 57:289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.

Browning, B. L., and S. R. Browning. 2016. Genotype imputation with millions of reference samples. Am. J. Hum. Genet. 98:116–126. https://doi.org/10.1016/j.ajhg.2015.11.020.

Browning, B. L., X. Tian, Y. Zhou, and S. R. Browning. 2021. Fast two-stage phasing of large-scale sequence data. Am. J. Hum. Genet. 106:1880–1890. https://doi.org/10.1016/j.ajhg.2021.08.005.

Butler, M. L., J. M. Bormann, R. L. Weber, D. M. Griefer, and M. M. Roll. 2020. Selection for bull fertility: A review. Transl. Anim. Sci. 4:423–441. https://doi.org/10.1093/tas/txz174.

Cai, X., S. Wu, T. D. Mipam, H. Luo, C. Yi, C. Xu, W. Zhao, H. Wang, and J. Zhong. 2021. Testis transcriptome profiling identified lngRNAs involved in spermatogenic arrest of cattleyak. Funct.
Saracuk, M. A., Elzo, S., Koamawoottritron, T., Siwanasopsce, D., Jattawa, and T. Laodim. 2018. Characterization of biological pathways associated with semen traits in the Thai multibreed dairy population. Anim. Reprod. Sci. 197:324–334. https://doi.org/10.1016/j.anireprosci.2018.09.002.

Shu, L., X. Peng, S. Zhang, G. Deng, Y. Wu, M. He, B. Li, C. Li, and K. Zhang. 2012. Non-classical major histocompatibility complex class II makes a crucial contribution to reproduction in the dairy cow. J. Reprod. Dev. 58:569–575. https://doi.org/10.20622/jrdev.2011-0181.

Song, X., X. He, X. Li, and Y. Qian. 2016. The roles and functional mechanisms of interleukin-17 family cytokines in mucosal immunity. Cell. Mol. Immunol. 13:418–431. https://doi.org/10.1038/cmi.2015.105.

Souza, E. C. V. Silva, B. A. N. Travançôlo, B. G. Alves, and M. E. Beletti. 2018. Sperm chromatin alterations in fertile and subfertile bulls. Reprod. Biol. 18:177–181. https://doi.org/10.1016/j.reprobiol.2018.04.001.

Suchocki, T., and J. Szyda. 2015. Genome-wide association study for semen production traits in Holstein-Friesian bulls. J. Dairy Sci. 98:5774–5780. https://doi.org/10.3168/jds.2014-9551.

Supe, F., M. Bošnjak, N. Škunca, and T. Šmuc. 2011. HEVIGO summarizes and visualizes long lists of gene ontology terms. PLoS One 6:e21800. https://doi.org/10.1371/journal.pone.0021800.

Sweett, H., P. A. S. Fonseca, A. Suárez-Vega, A. Livernois, F. Miglior, and A. Cánovas. 2020. Genome-wide association study to identify genomic regions and positional candidate genes associated with male fertility in beef cattle. Sci. Rep. 10:20102. https://doi.org/10.1038/s41598-020-75758-3.

Talbot, B. D. Shur, and D. G. Myles. 2003. Cell adhesion and fertilization: Steps in oocyte transport, sperm-zona pellucida interactions, and sperm-egg fusion. Biol. Reprod. 68:1–9. https://doi.org/10.1095/biolreprod.102.007856.

Tijiani, A., Y. T. Utsunomiya, A. G. Ezekwe, O. Nashiru, and O. Hanotte. 2019. Genome sequence analysis reveals selection signatures in endangered trypanotolerant West African Murine Cattle. Front. Genet. 10:442. https://doi.org/10.3389/fgene.2019.00442.

Uner, N., L. Arevalo, I. Kosterin, S. Phakle, K. Lohanadan, G. Kirfel, D. Sons, W. Witke, and H. Schorle. 2021. Loss of profilin3 impairs spermiogenesis by affecting acrosome biogenesis, autophagy, manchette development, and mitochondrial organization. SSRN E-Prints. https://doi.org/10.2139/ssrn.4223257.

Valente, T. S., F. Baldi, A. C. Sant’Anna, L. G. Albuquerque, and M. J. R. Paranhos da Costa. 2016. Genome-wide association study between single nucleotide polymorphisms and flight speed in Nellore cattle. PLoS One 11:e0156956. https://doi.org/10.1371/journal.pone.0156956.

Vanderhaeghen, P., S. Schurmans, G. Vassart, and M. Parmentier. 1997. Specific repertoire of olfactory receptor genes in the male germ cells of several mammalian species. Genomics 39:239–246. https://doi.org/10.1006/geno.1996.4490.

VanRaden, P. M. 2008. Efficient methods to compute genomic predictions. J. Dairy Sci. 91:4141–4143. https://doi.org/10.3390/genes10010044.

VanRaden, P. M., K. M. Olson, D. J. Null, and J. L. Hutchinson. 2011. Harmful recessive effects on fertility detected by absence of homozygous haplotypes. J. Dairy Sci. 94:6153–6161. https://doi.org/10.3168/jds.2011-4624.

Vera, O., L. A. Vásquez, and M. Gladys Muñoz. 2003. Semen quality and presence of cytokines in seminal fluid of bull ejaculates. Theriogenology 60:553–558. https://doi.org/10.1016/S0093-691X(03)00031-1.

Vicens, A., L. Lió, and E. R. S. Rolland. 2014. Proteins involved in motility and sperm-egg interaction evolve more rapidly in mouse spermatozoa. PLoS One 9:e91302. https://doi.org/10.1371/journal.pone.0091302.

Wang, H., I. Misztal, I. Aguilera, A. Legarra, and W. M. Muir. 2012. Genomic-wide association mapping including phenotypes from relatives without genotypes. Genet. Res. 94:73–83. https://doi.org/10.1017/S0093-1385(12)000274.

Wang, S., and Y. Wang. 2013. Peptidylarginine deiminases in citrullination, gene regulation, health, and pathogenesis. Biochim. Biophys. Acta. 1829:1126–1135. https://doi.org/10.1016/j.bbadmr.2013.07.003.

Wattegedera, S. R., Y. Corripio-Miyar, Y. Pang, D. Frew, T. N. McNeilly, J. Palarca-Albaladejo, C. J. McInnes, J. C. Hope, E. J. Glass, and G. Entrican. 2017. Enhancing the toolbox to study IL-17A in cattle and sheep. Vet. Res. 48:20. https://doi.org/10.1186/s13597-014-0247-4.

Wiggins, G. R., T. S. Somstegard, P. M. VanRaden, L. K. Matukumalli, R. D. Schnabel, J. F. Taylor, F. S. Schenkel, and C. P. Van Tassell. 2009. Selection of single-nucleotide polymorphisms and quality of genotypes used in genomic evaluation of dairy cattle in the United States and Canada. J. Dairy Sci. 92:3431–3436. https://doi.org/10.3168/jds.2008-1758.

Witke, W., J. D. Sutherland, A. Sharpe, M. Ari, and D. J. Kwiatkowski. 2001. Profilin I is essential for cell survival and cell division in early mouse development. Proc. Natl. Acad. Sci. USA 98:3832–3836. https://doi.org/10.1073/pnas.051515498.
Wright, P. W., L. C. Bolling, M. E. Calvert, O. F. Sarmento, E. V. Berkeley, M. C. Shea, Z. Hao, F. C. Jayses, L. A. Bush, J. Shetty, A. N. Shore, P. P. Reddi, K. S. Tung, E. Samy, M. M. Allietta, N. E. Sherman, J. C. Herr, and S. A. Coonrod. 2003. ePAD, an oocyte and early embryo-abundant peptidylarginine deiminase-like protein that localizes to egg cytoplasmic sheets. Dev. Biol. 256:74–89. https://doi.org/10.1016/S0012-1606(02)00126-4.

Wu, D., J. Dou, X. Chai, C. Bellis, A. Wilm, C. C. Shih, W. W. J. Soon, N. Bertin, C. B. Lin, C. C. Khor, M. DeGiorgio, S. Cheng, L. Bao, N. Karnani, W. Y. K. Hwang, S. Davila, P. Tan, A. Shabbir, A. Moh, E-K. Tan, J. N. Foo, L. L. Goh, K. P. Leong, R. S. Y. Foo, C. S. P. Lam, A. M. Richards, C.-Y. Cheng, T. Aiung, T. Y. Wong, H. H. Ng, J. Liu, C. Wang, M. A. Ackers-Johnson, E. Aliwarga, K. H. K. Ban, D. Bertrand, J. C. Chambers, D. L. H. Chan, C. X. L. Chan, M. L. Chee, M. L. Chee, P. Chen, Y. Chen, E. G. Y. Chew, W. J. Chew, L. H. Y. Chiam, J. P. C. Chong, I. Chua, S. A. Cook, W. Dai, R. Dorajoo, C.-S. Foo, R. S. M. Goh, A. M. Hillmer, I. D. Irwan, F. Jaufeerally, A. Javed, J. Jeyakani, J. T. Li, K. Koh, J. Y. Koh, P. KrishnaSwamy, J. L. Kuan, N. Kumari, A. S. Lee, S. E. Lee, S. Lee, Y. L. Lee, S. T. Leong, Z. Li, P. Y. Li, J. X. Liew, O. W. Liow, S. C. Lim, W. K. Lim, C. W. Lim, T. B. Lim, C. K. Lim, S. Y. Loh, A. W. Lok, C. W. L. Chin, S. Majithia, S. Maurer-Stroh, W. Y. Meah, S. Q. Mok, N. Nargarajan, P. Ng, S. B. Ng, Z. Ng, J. Y. X. Ng, E. Ng, S. L. Ng, S. Nusinovici, C. T. Ong, B. Pan, V. Pedergnana, S. Poh, S. Prabhakar, K. M. Prakash, I. Quek, C. Sabanayagam, W. Q. See, Y. Y. Sia, X. Sim, W. C. Sim, J. So, D. K. N. Soon, E. S. Tai, N. Y. Tan, L. C. S. Tan, H. C. Tan, W. L. W. Tan, M. Tandiono, A. Tay, S. Thakur, Y. C. Tham, Z. Tian, G. L.-X. Toh, P. K. Tsai, L. Veeravalli, C. S. Verma, L. Wang, M. R. Wang, W.-C. Wong, Z. Xie, K. K. Yeo, L. Zhang, W. Zhai, and Y. Zhao. 2019. Large-scale whole-genome sequencing of three diverse Asian populations in Singapore. Cell 179:736–749.e15. https://doi.org/10.1016/j.cell.2019.09.019.

Yin, H., L. Fang, C. Qin, and S. Zhang. 2019a. Estimation of the genetic parameters for semen traits in Chinese Holstein bulls. BMC Genet. 20:51. https://doi.org/10.1186/s12863-019-0752-x.

Yin, H., C. Zhou, S. Shi, L. Fang, J. Liu, D. Sun, L. Jiang, and S. Zhang. 2019b. Weighted single-step genome-wide association study of semen traits in Holstein bulls of China. Front. Genet. 10:1053. https://doi.org/10.3389/fgene.2019.01053.

Zenobia, C., and G. Hajishengallis. 2015. Basic biology and role of interleukin-17 in immunity and inflammation. Periodontol. 2000 69:142–159. https://doi.org/10.1111/per.12083.

Zhang, J., F. Liu, J-C. Reif, and Y. Jiang. 2021. On the use of GBLUP and its extension for GWAS with additive and epistatic effects. G3 Genes|Genomes|Genetics 11:7. https://doi.org/10.1093/g3journal/jkab122.

Zhang, T., J. Oostley, V. J. Bardwell, and D. Zarkower. 2016. DMRT1 is required for mouse spermatogonial stem cell maintenance and replenishment. PLoS Genet. 12:e1006293. https://doi.org/10.1371/journal.pgen.1006293.