Genome-wide association study for semen production traits in Holstein-Friesian bulls

T. Suchocki¹ and J. Szyda

Wrocław University of Environmental and Life Sciences, Department of Genetics, Biostatistics Group, Kozuchowska 7, 51-631 Wrocław, Poland

ABSTRACT

Identifying genomic regions, particularly individual genes associated with semen quality traits, may be very important for improving sire fertility via selective breeding. The aim of the study was to estimate (co) variance components and effects of single nucleotide polymorphisms (SNP) from the Illumina BovineSNP50 BeadChip (Illumina, San Diego, CA) on semen production traits and to find candidate genes for these traits. The analyzed data set originates from the Polish Holstein-Friesian dairy cattle population and consists of 1,212 bulls kept at 4 artificial insemination stations. For each bull, 5 semen production traits were collected: sperm concentration, semen volume, number of spermatozoa, motility, and motility score. A multitrait mixed model was used to estimate genetic parameters. The parameters obtained were used to estimate SNP effects for each trait separately by the mixed model, which is used in the Polish direct genomic value project. Additionally, genes located in the vicinity of significant SNP were selected as candidate genes. For motility, 20 genome-wide significant SNP, located on 12 autosomes, were identified. For sperm concentration, we found 7 significant SNP: 3 on chromosome X, and 1 on chromosomes 1, 6, 23, and 24. For semen volume and motility score, 3 and 1 significant SNP were detected, respectively. All these SNP were located on chromosome X. For the number of spermatozoa, 12 significant SNP were observed. Six SNP were located on chromosome X, 3 on chromosome 8, and 1 on chromosomes 2, 7, and 16. This study clearly indicated a key role of the X chromosome in the determination of semen quality and emphasized that including such traits into genetic evaluation should be strongly considered.

Key words: covariance components, genome-wide association study, semen quality, single nucleotide polymorphism chip

INTRODUCTION

In several countries, female fertility is included in breeding selection programs (Miglior et al. 2005; Jamrozik et al., 2005; Weigel 2006), but male reproductive performance has received less consideration. However, recently more attention is being given to semen quality traits as, unlike female-derived traits, they can be directly measured as bull phenotypes, have higher heritabilities, and thus may be improved by selection (Berry et al. 2014). Consequently, in recent years, a few studies (Druet et al., 2009; Feugang et al. 2009; Blaschek et al. 2011; Fortes et al. 2012, 2013; Lyons et al. 2014) were performed that showed the large effect of bull fertility on breeding and selection programs. The data set analyzed by Druet et al. (2009) contained only 148 microsatellite markers (ranging from 2 to 10 per chromosome), and none were located on *Bos taurus* X (BTX), which is known to have an effect on bull fertility (Blaschek et al. 2011). Feugang et al. (2009) and Peñagaricano et al. (2012) used more informative data sets of 10K and 50K SNP chips respectively, but applied relatively simple statistical approaches for detection the association between SNP and semen-quality measures. Only the studies of Blaschek et al. (2011) and Fortes et al. (2012, 2013) showed a combination of dense SNP chips (Illumina BovineSNP50 and Illumina Bovine HD BeadChips; Illumina, San Diego, CA) and robust statistical models. Such a combination allowed for detection of candidate genes associated with semen traits on both autosomes and the X chromosome. The bioinformatics analysis of Lyons et al. (2014) was carried out for a sequenced region of chromosome X and 16 individuals, but it just confirmed the presence of the 2 QTL associated with fertility for Brahman and Tropical Composite bulls already indicated by Fortes et al. (2012). Based on the previous findings, the main aim of our study was (1) to estimate the (co)variance components for semen production traits, (2) to estimate the effects of SNP from the microarray Illumina BovineSNP50 BeadChip on semen production traits using multiple SNP mixed model, and (3) to identify candidate genes for these traits.

Received October 8, 2014.

Accepted April 15, 2015.

¹Corresponding author: tomasz.suchocki@up.wroc.pl

MATERIALS AND METHODS

The analyzed data set originated from the Polish Holstein-Friesian dairy cattle population and consisted of 1,212 bulls kept in 4 AI stations. For each bull, 5 semen production traits were collected: sperm concentration (million/mL), semen volume (mL), number of spermatozoa (million, calculated by multiplying sperm concentration and semen volume), motility (percentage of motile sperms), and motility score (1–3 points). Individual motility and motility score were evaluated under the microscope by a single technician. All traits were recorded from the first ejaculate on each collection day (2 ejaculates were attempted, the second was successful in only a few cases). The process of collecting the data was described in Kondracki et al. (2012). For each individual, traits were recorded several times throughout the year. Average bull age when the test was performed was 32.6 mo and varied between 10 and 137 mo. Semen-collection procedures were homogeneous because all stations belonged to the same AI center. The number of records per bull varied between 3 and 144, with an average distance between measurements of 23 ± 131 d. Descriptive statistics for these traits are presented in Table 1.

Each individual was genotyped using the Illumina BovineSNP50 BeadChip, which consists of 54,001 SNP markers (Version 1) or 54,609 (Version 2). In the final analysis, only common SNP included in both panels were used. Also, SNP outside the pseudoautosomal region (the pseudoautosomal region of on the X chromosome ranges from 29,245 to 147,387,980 bp, the most upstream SNP in this region is rs41657269 and the most downstream SNP is rs111025096) were removed from the analysis. Additionally, the SNP selection criteria were applied comprising the minor allele frequency of at least 0.01 and the technical quality of the chip, with a minimum call rate of 98%. After editing, 45,678 SNP were used in the final approach. The average minor allele frequency was 0.27 and the average call rate amounted to 99.64%.

Genetic Parameters

The (co)variance components for the 5 semen-production traits were estimated using the following multivariate mixed model:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\alpha} + \mathbf{e}, \quad [1]$$

where $\mathbf{y}$ is a matrix with all considered trait values of dimension $n \times t$, where n (21,906) is the number of all bull-trait observations and t (5) is the number of traits; $\boldsymbol{\beta}$ is a vector of fixed effects comprising a general mean, bull age in months when the test was performed, season

of the test (1 = between October and March, 2 = between April and September), and an insemination station (1–4); $\boldsymbol{\alpha}$ is a vector of random additive polygenic effects of a bull; and $\mathbf{e}$ is a vector of residuals. It is assumed that $\boldsymbol{\alpha} \sim N(0, \mathbf{A} \otimes \mathbf{G})$ and $\mathbf{e} \sim N(0, \mathbf{D} \otimes \mathbf{R})$, where $\mathbf{A}$ is an additive relationships matrix among individuals; $\mathbf{G}$ is a polygenic (co)variance matrix between traits; $\mathbf{D}$ is a diagonal matrix containing the reciprocal of the number of records per bull on the diagonal; and $\mathbf{R}$ denotes a residual (co)variance matrix between traits. $\mathbf{X}$ and $\mathbf{Z}$ are design matrices for $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ respectively. For estimation of (co)variance components, ASReml software (version 3.0) was used (Gilmour et al., 2009).

SNP Effect Estimation

The following mixed model (Szyda et al., 2011) was used to estimate the additive effects of SNP:

$$\hat{\boldsymbol{\alpha}} = \mathbf{X}^* \mathbf{b} + \mathbf{Z}^* \mathbf{g} + \mathbf{e}^*, \quad [2]$$

where $\hat{\boldsymbol{\alpha}}$ represents a vector of additive genetic effects estimated by equation [1]; $\mathbf{b}$ is a vector of fixed effects representing a general mean with a design matrix $\mathbf{X}^*$; $\mathbf{Z}^*$ is a design matrix for SNP genotypes, which is parameterized as -1 , 0 , or 1 for a homozygous, heterozygous, and an alternative homozygous SNP genotype, respectively; $\mathbf{g}$ is a vector of random additive SNP effects; and $\mathbf{e}^*$ is a vector of residuals with $\mathbf{e}^* \sim N(0, \mathbf{D} \hat{\sigma}_a^2)$, where $\mathbf{D}$ has the same form as in equation [1]. The covariance structure of $\mathbf{g}$ was assumed to be $\mathbf{g} \sim N\left(0, \mathbf{I} \frac{\hat{\sigma}_a^2}{N_{\text{SNP}}}\right)$, with $\mathbf{I}$ being an identity matrix, $\hat{\sigma}_a^2$ representing the additive genetic variance of a given trait estimated by equation [1], and N_{SNP} is the number of SNP used (here 45,678).

The estimation of parameters of the above model was based on solving the mixed model equations (Henderson, 1984):

$$\begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{g}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}^{*T} \mathbf{R}^{*-1} \mathbf{X}^* & \mathbf{X}^{*T} \mathbf{R}^{*-1} \mathbf{Z}^* \\ \mathbf{Z}^{*T} \mathbf{R}^{*-1} \mathbf{X}^* & \mathbf{Z}^{*T} \mathbf{R}^{*-1} \mathbf{Z}^* + \mathbf{G}^{*-1} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{X}^{*T} \mathbf{R}^{*-1} \mathbf{y} \\ \mathbf{Z}^{*T} \mathbf{R}^{*-1} \mathbf{y} \end{bmatrix}, \quad [3]$$

where $\mathbf{R}^* = \mathbf{D} \hat{\sigma}_e^2$ and $\mathbf{G}^* = \mathbf{I} \frac{\hat{\sigma}_a^2}{N_{\text{SNP}}}$. and T is the matrix transposition. The iteration on data technique was based on the Gauss-Seidel algorithm with residuals update (Legarra and Misztal, 2008). Consequently, the variance of $\hat{\boldsymbol{\alpha}}$ is then given by $\mathbf{Z}^* \mathbf{G}^* \mathbf{Z}^{*T} + \mathbf{R}$.

For testing the significance ($H_0: g_i = 0$ vs. $H_1: g_i \neq 0$, where H is the hypothesis) of i th SNP effects in $\mathbf{g}$, the

Table 1. Descriptive statistics for semen production traits

Trait	Mean	SD	Minimum	Maximum
Sperm concentration (million/mL)	1,471.21	480.85	18.00	8,157.00
Sperm volume (mL)	4.18	2.21	0.60	25.00
Number of spermatozoa (million)	6,125.83	3,880.03	72.00	38,960.00
Motility (%)	72.40	5.43	10.00	90.00
Motility score (1–3 points)	2.16	0.57	1.00	3.00

Wald test (W) was used. The multiple testing problem was addressed by calculating a false discovery rate following (Benjamini and Hochberg, 1995) for each nominal P -value. The test statistic follows under H_0 a normal distribution with mean 0 and variance 1. Test statistics

have the following form: $W = \frac{\hat{g}_i}{SE(\hat{g}_i)}$, where $SE(\hat{g}_i)$ is a

standard error of i th estimated SNP effect $\hat{g}_i$. Additionally, for significant SNP (false discovery rate ≤ 0.05), linkage disequilibrium between each pair of SNP was calculated as an r^2 coefficient of linkage disequilibrium using Plink software (Purcell et al., 2007). In this study we did not use any threshold to claim a candidate gene based on SNP significance. We only identified the closest gene for each significant SNP marker. The SNP positions were assigned following the UMD 3.1 assembly (Zimin et al., 2009). Ensembl (Cunningham et al. 2015) bovine genome build release 78 was used for assignment of the gene positions.

RESULTS

Genetic Parameters

The additive polygenic (lower diagonal) and residual (upper diagonal) correlations and heritabilities (diagonal) for the analyzed traits are shown in Table 2. Only 2 genetic correlation coefficients were negative (between sperm volume and motility score, as well as between sperm volume and sperm concentration), but these correlations were not statistically significantly different from zero. High genetic correlation was estimated between the number of spermatozoa and sperm volume (0.82). All residual correlations were positive,

the highest value was estimated between motility and motility score (0.61). Generally high heritabilities were estimated, which varied between 0.34 (± 0.068) for semen concentration, through 0.31 (± 0.064) for motility, 0.27 (± 0.060) for the number of spermatozoa, 0.26 (± 0.054) for motility score, and 0.26 (± 0.062) for semen volume. Age effect (in months) estimated as 0.001849 ± 0.0001385 was significant with a P -value of $1.18e-40$.

Candidate Genes

Nominal P -values corresponding to SNP effect estimates for the analyzed traits are presented on Figure (1 A–E). Additionally, QQ plots between observed and expected P -values are presented on Figure (2 A–E). Deviation of P -values from the expected distribution on QQ-plots can be interpreted as P -values that are very likely to be true positive. We obtained such a situation for each considered trait. Significant SNP, genes located close to them, and the explained percentage of total genetic variation are summarized in Table 3. For motility, 20 significant SNP were identified. The most significant SNP with the additive effect of $0.013\% \hat{\sigma}_a^2$ was located on chromosome 10 in the neighborhood of *7SK* gene (small nuclear RNA coding gene; ENS-BTAG00000043728). For sperm concentration, the 2 most significant SNP identified were located on chromosome X. The highest effect of 1.09 million/mL was observed for the marker located 12,537 bp from the *HS6ST2* (heparan sulfate 6-O-sulphotransferase 2) gene. For semen volume, 3 SNP with significant effects were identified (all on BTX). The SNP with the highest additive effect, amounting to 0.005 mL, was located in the vicinity of *MAGEB10* (melanoma antigen family B

Table 2. Additive polygenic (below diagonal) and residual (above diagonal) correlations between traits with SE; on the diagonal are heritabilities with SE

Item	Semen volume	Motility score	Motility	Sperm concentration	Number of spermatozoa
Semen volume	0.26 $\pm$ 0.062	0.04 $\pm$ 0.029	0.07* $\pm$ 0.029	0.01 $\pm$ 0.029	0.82* $\pm$ 0.016
Motility score	–0.01 $\pm$ 0.029	0.26 $\pm$ 0.054	0.61* $\pm$ 0.023	0.51* $\pm$ 0.025	0.29* $\pm$ 0.028
Motility	0.12* $\pm$ 0.029	0.64* $\pm$ 0.022	0.31 $\pm$ 0.064	0.33* $\pm$ 0.027	0.23* $\pm$ 0.028
Sperm conc.	–0.06 $\pm$ 0.029	0.59* $\pm$ 0.023	0.56* $\pm$ 0.024	0.34 $\pm$ 0.068	0.50* $\pm$ 0.025
Number of spermatozoa	0.82* $\pm$ 0.016	0.36* $\pm$ 0.027	0.41* $\pm$ 0.026	0.49* $\pm$ 0.025	0.27 $\pm$ 0.060

* $P < 0.05$.

Figure 1. P -values for each SNP-trait combination. A dotted line represents a nominal 0.05 P -value $[-\log(0.05)]$ threshold and a solid line a 0.05 false discovery rate threshold.

10). For the number of spermatozoa, 12 significant SNP were observed, but the 2 most significant ones (rs110685046 and rs109349108) were in high linkage disequilibrium ($r^2 = 0.57$) and had a relatively short physical distance of 34,406 bp between each other. Both polymorphisms show a similar additive effect of 10.74 million and 10.53 million sperm and are located close to the *MAGEB10* gene. For motility score, 1 significant SNP located on BTX was detected. This SNP is located in the vicinity of the *KLH13* gene (kelch-like protein 13).

DISCUSSION

Recently, 2 genome-wide association studies for fertility traits were conducted based on the same 50K Illumina SNP chip as the one used in our study. Interestingly, results of all 3 genome-wide association studies

(including the current study) indicate a large number of significant SNP located on BTX. The study of Blaschek et al. (2011) was devoted to a more descriptive fertility trait, bull conception rate in the Holstein-Friesian breed, and clearly indicated BTX as the chromosome with the most significant effect. Comparing specific regions, a very close resemblance occurs. The *FMR1* gene, indicated as a candidate for the number of spermatozoa and sperm concentration in our study, is located close to the region marked by rs41585028 polymorphism, as reported by Blaschek et al. (2011). Also, the *GLCCI1* gene, marked by a significant SNP for motility in our study, is located close to rs110984522 polymorphism, which was significant for noncompensatory fertility in bulls in the study of Blaschek et al. (2011). Fortes et al. (2012) performed a genome-wide association study for fertility-related traits in *Bos indicus* using the 50K Illumina SNP chip enhanced by SNP

Figure 2. QQplots for nominal P -values and trait.

genotypes imputed from the HD Illumina SNP chip. For traits directly related to semen production (scrotal circumference, ability to produce sperm, percentage of normal sperm), those authors found the X chromosome to contain many significant hits, the same as in our study. Moreover, some of genes proposed as candidate loci in our study are located within the genomic regions of BTX indicated as significant by Fortes et al. (2012). In particular, LOC100848828, which was identified as a possible candidate gene for the number of spermatozoa in our study, overlaps with the region significant for scrotal circumference in their study.

Some of the genes indicated by significant SNP in our study were not reported elsewhere as directly related to bull fertility, but reveal related physiological background. In particular, the significant SNP in the vicinity of the *7SK* gene on BTA10 relates to small RNA

that are important for the regulation of transcription. Our significant result stays in line with experimental findings by Okamura et al. (2012), who demonstrated a role of small RNA in the proliferation of germ cells in mouse embryos. The gene *PAPPA* on BTA8, with a significant effect on the number of spermatozoa estimated for rs109336730 located within an intron of this gene, codes for a pregnancy-associated plasma protein A and is thus related to female fertility. For *FSHR* on BTA11, which is located 0.2 Mbp from an SNP significant for motility, evidence of its role in female estrogen levels has been recently reported by Yang et al. (2014).

CONCLUSIONS

Agreement between the results of our study and previous studies clearly indicates a key role of the X

Table 3. Significant SNP selected by the physical approach and candidate genes for semen production traits; *P*-values were transformed into false discovery rate to account for multiple testing

SNP identification	Chromosome	Position (bp)	Effect	<i>P</i> -value	Closest gene	Distance from gene (bp)	Total variation (%) explained by the significant SNP
Motility							
rs41255421	1	8,963,404	0.01112	0.03865	<i>ADAMTS1</i>	0	0.05260
HAPMAP38887-BTA-22281	1	9,955,276	0.01087	0.03865	<i>APP</i>	34,272	0.03253
rs42359025	1	89,358,747	0.01097	0.03865	<i>U6</i>	244,835	0.03502
rs42736798	1	104,926,538	0.01091	0.03865	<i>U6</i>	89,926	0.04994
rs43304396	2	32,939,824	0.01277	0.00394	<i>FIGN</i>	87,396	0.05067
BTA-07805-NO-RS	3	55,442,801	0.01087	0.03966	<i>PKN2</i>	77,065	0.04495
rs29026870	4	17,109,984	0.01162	0.02256	<i>GLCH1</i>	0	0.04686
rs29023533	7	73,549,865	0.01127	0.03617	<i>ADRA1B</i>	61,252	0.05387
rs43581982	8	102,123,125	0.01125	0.03617	<i>SNORA71</i>	5,915	0.05446
rs42300887	10	96,065,191	0.01334	0.00284	<i>7SK</i>	337,733	0.07162
rs110137262	10	98,519,683	0.01087	0.03865	<i>FLRT2</i>	185,642	0.04944
rs42214341	11	31,519,537	0.01106	0.03865	<i>FSHR</i>	214,340	0.05228
rs110931056	11	36,931,716	0.01094	0.03865	<i>C2orf73</i>	18,155	0.04507
BTA-93012-NO-RS	11	38,544,853	0.01094	0.03865	<i>Bta-mir-216b</i>	33,651	0.04329
rs109658408	21	29,923,689	0.01287	0.00394	<i>CHRNA7</i>	155,854	0.05850
rs109762328	21	31,938,171	0.01265	0.00394	<i>C15orf27</i>	0	0.06677
rs41987007	21	49,048,548	0.01183	0.01875	<i>CLEC14A</i>	202,926	0.05966
rs109293969	24	56,967,385	0.01328	0.00284	<i>WDR7</i>	150,006	0.07296
Hapmap49238-BTA-60825	26	21,372,449	0.01313	0.00284	<i>HIF1AN</i>	72,793	0.06964
HAPMAP47897-BTA-25453	28	26,021,025	0.01096	0.03865	<i>TSPAN15</i>	8,442	0.04843
Sperm concentration							
rs29013410	1	93,892,947	0.99285	0.04638	<i>NLGN1</i>	260,718	0.05133
HAPMAP23494-BTC-046433	6	105,572,119	1.01950	0.03897	<i>STK32B</i>	40,404	0.05635
BTA-121413-NO-RS	23	47,795,608	1.01099	0.03897	<i>RIOK1</i>	0	0.05626
rs109293969	24	56,967,385	1.03711	0.03897	<i>WDR7</i>	150,006	0.05784
HAPMAP44537-BTA-30597	X	30,662,406	1.07051	0.02933	<i>FMRI</i>	260,368	0.05978
rs109349108	X	122,341,165	1.00880	0.03897	<i>MAGEB10</i>	361,656	0.05078
rs110685046	X	122,375,570	1.08908	0.02933	<i>HS6ST2</i>	12,537	0.06596
Semen volume							
rs110966399	X	4,995,135	0.00449	0.01289	<i>MGC142908</i>	202	0.06352
rs109349108	X	122,341,165	0.00453	0.01289	<i>MAGEB10</i>	361,656	0.05995
rs110685046	X	122,375,570	0.00453	0.01289	<i>MAGEB10</i>	396,061	0.06667
Motility score							
rs29018061	X	2,108,758	0.00120	0.00195	<i>KLHL13</i>	70,734	0.08169
Number of spermatozoa							
rs110131051	2	99,881,355	7.87675	0.00711	<i>ERBB4</i>	167,086	0.06077
ARS-BFGL-NGS-106895	7	77,261,597	7.11893	0.03745	<i>NUDCD2</i>	3,131	0.03888
rs460741022	8	30,596	7.49891	0.01830	<i>H1ATL1</i>	0	0.05668
rs43577039	8	101,446,930	7.21227	0.03098	<i>C8H9orf152</i>	8,673	0.05499
rs109336730	8	107,243,184	7.26078	0.03098	<i>PAPPA</i>	0	0.05504
rs111027013	16	75,262,436	7.71161	0.00979	<i>SYT14</i>	49,792	0.06303
rs110966399	X	4,995,135	7.16358	0.03247	<i>MGC142908</i>	202	0.05248
HAPMAP44437-BTA-30597	X	30,662,406	8.58277	0.00131	<i>FMRI</i>	260,368	0.07287
rs29010277	X	66,714,338	7.89719	0.00711	<i>LOC100848828</i>	77,253	0.06272
rs110419531	X	122,102,151	7.93819	0.00711	<i>MAGEB10</i>	122,642	0.05840
rs109349108	X	122,341,165	10.52952	<0.00001	<i>MAGEB10</i>	361,656	0.10504
rs110685046	X	122,375,570	10.73611	<0.00001	<i>MAGEB10</i>	396,061	0.12154

chromosome in the determination of semen quality. From a practical, industry-oriented perspective, it is obvious that semen quality traits play a key role in dairy cattle breeding. Based on results of this and other recent studies, it is now evident that semen quality reveals a considerable genetic component, both through moderate heritability estimated by models assuming an infinitesimal model of inheritance as well as through several candidate genes identified through genome-wide association studies. As many traits describing semen quality undergo routine recording, it is tempting to consider including the phenotypes into genetic evaluation.

ACKNOWLEDGMENTS

This work was financially supported by the Polish National Science Centre, grant no. N N311 524940. The SNP data set was obtained within the frame of the MASinBULL project. Publication supported by the Wrocław Centre of Biotechnology, program of the National Scientific Leadership Centre.

REFERENCES

- Benjamini, 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.
- Berry, D. P., E. Wall, and J. E. Pryce. 2014. Genetics and genomics of reproductive performance in dairy and beef cattle. *Animal* 8(Suppl 1):105–121. <http://dx.doi.org/10.1017/S1751731114000743>.
- Blaschek, M., A. Kaya, N. Zwald, E. Memili, and B. W. Kirkpatrick. 2011. A whole-genome association analysis of noncompensatory fertility in Holstein bulls. *J. Dairy Sci.* 94:4695–4699. <http://dx.doi.org/10.3168/jds.2010-3728>.
- Cunningham, F., M. R. Amode, D. Barrell, K. Beal, K. Billis, S. Brent, D. Carvalho-Silva, P. Clapham, G. Coates, S. Fitzgerald, L. Gil, C. G. Girón, L. Gordon, T. Hourlier, S. E. Hunt, S. H. Janacek, N. Johnson, T. Juettemann, A. K. Kähäri, S. Keenan, F. J. Martin, T. Maurel, W. McLaren, D. N. Murphy, R. Nag, B. Overduin, A. Parker, M. Patricio, E. Perry, M. Pignatelli, H. S. Riat, D. Sheppard, K. Taylor, A. Thormann, A. Vullo, S. P. Wilder, A. Zadissa, B. L. Aken, E. Birney, J. Harrow, R. Kinsella, M. Muffato, M. Ruffier, S. M. Searle, G. Spudich, S. J. Trevanion, A. Yates, D. R. Zerbino, and P. Flicek. 2015. Ensembl 2015. *Nucleic Acids Res.* 43:D662–D669. <http://dx.doi.org/10.1093/nar/gku1010>.
- Druet, T., S. Fritz, S. Sellem, B. Basso, O. Ge, L. Salas-Cortes, P. Humblot, X. Druart, and A. Eggen. 2009. Estimation of genetic parameters and genome scan for 15 semen characteristics traits of Holstein bulls. *J. Anim. Breed. Genet.* 126:269–277. <http://dx.doi.org/10.1111/j.1439-0388.2008.00788.x>.
- Feugang, J. M., A. Kaya, G. P. Page, L. Chen, T. Mehta, K. Hirani, L. Nazareth, E. Topper, R. Gibbs, and E. Memili. 2009. Two-stage genome-wide association study identifies integrin beta 5 as having potential role in bull fertility. *BMC Genomics* 10:176–185. <http://dx.doi.org/10.1186/1471-2164-10-176>.
- Fortes, M. R., A. Reverter, R. J. Hawken, S. Bolormaa, and S. A. Lehnert. 2012. Candidate genes associated with testicular development, sperm quality, and hormone levels of inhibin, luteinizing hormone, and insulin-like growth factor 1 in Brahman bulls. *Biol. Reprod.* 87:58. <http://dx.doi.org/10.1095/biolreprod.112.101089>.
- Fortes, M. R., A. Reverter, M. Kelly, R. McCulloch, and S. A. Lehnert. 2013. Genome-wide association study for inhibin, luteinizing hormone, insulin-like growth factor 1, testicular size and semen traits in bovine species. *Andrology* 1:644–650. <http://dx.doi.org/10.1111/j.2047-2927.2013.00101.x>.
- Gilmour, A. R., B. J. Gogel, B. R. Cullis, and R. Thompson. 2009. ASReml User Guide Release 3.0 VSN International Ltd., Hemel Hempstead, UK www.vsn.co.uk.
- Henderson, C. R. 1984. Applications of Linear Models in Animal Breeding. University of Guelph, Guelph, ON, Canada.
- Jamrozik, J., J. Fatehi, G. J. Kistemaker, and L. R. Schaeffer. 2005. Estimates of genetic parameters for Canadian Holstein female reproduction traits. *J. Dairy Sci.* 88:2199–2208.
- Kondracki, S., D. Banaszewska, A. Wysokinska, and M. Iwanina. 2012. The effect of sperm concentration in the ejaculate on morphological traits of bull spermatozoa. *Folia Biol. (Krakow)* 60:85–91.
- Legarra, A., and I. Misztal. 2008. Technical note: Computing strategies in genome-wide selection. *J. Dairy Sci.* 91:360–366.
- Lyons, R. E., N. T. Loan, L. Dierens, M. R. Fortes, M. Kelly, S. M. McWilliam, Y. Li, R. Bunch, B. E. Harrison, W. Barendse, S. A. Lehnert, and S. S. Moore. 2014. Evidence for positive selection of taurine genes within a QTL region on chromosome X associated with testicular size in Australian Brahman cattle. *BMC Genet.* 15:6. <http://dx.doi.org/10.1186/1471-2156-15-6>.
- Miglior, F., B. L. Muir, and B. J. Van Doormaal. 2005. Selection indices in Holstein cattle of various countries. *J. Dairy Sci.* 88:1255–1263.
- Okamura, D., I. Maeda, H. Taniguchi, Y. Tokitake, M. Ikeda, K. Ozato, N. Mise, K. Abe, T. Noce, J. C. I. Belmonte, and Y. Matsui. 2012. Cell cycle gene-specific control of transcription has a critical role in proliferation of primordial germ cells. *Genes Dev.* 26:2477–2482. <http://dx.doi.org/10.1101/gad.202242.112>.
- Peñagaricano, F., K. A. Weigel, and H. Khatib. 2012. Genome-wide association study identifies candidate markers for bull fertility in Holstein dairy cattle. *Anim. Genet.* 43:65–71. <http://dx.doi.org/10.1111/j.1365-2052.2012.02350.x>.
- Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M. A. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. de Bakker, M. J. Daly, and P. C. Sham. 2007. PLINK: A toolset for whole-genome association and population-based linkage analysis. *Am. J. Hum. Genet.* 81:559–575.
- Szyda, J., A. Żarnecki, T. Suchocki, and S. Kamiński. 2011. Fitting and validating the genomic evaluation model to Polish Holstein-Friesian cattle. *J. Appl. Genet.* 52:363–366. <http://dx.doi.org/10.1007/s13353-011-0047-z>.
- Weigel, K. A. 2006. Prospects for improving reproductive performance through genetic selection. *Anim. Reprod. Sci.* 96:323–330.
- Yang, W. C., S. J. Li, L. Chen, and L. G. Yang. 2014. FSHR genotype affects estrogen levels but not pregnancy rates in Luxi cattle subjected to embryo transfer. *Genet. Mol. Res.* 13:1563–1569. <http://dx.doi.org/10.4238/2014.March.12.8>.
- Zimin, A. V., A. L. Delcher, L. Florea, D. R. Kelley, M. C. Schatz, D. Puiu, F. Hanrahan, G. Pertea, C. P. Van Tassell, T. S. Sonstegard, G. Marçais, M. Roberts, P. Subramanian, J. A. Yorke, and S. L. Salzberg. 2009. A whole-genome assembly of the domestic cow, *Bos taurus*. *Genome Biol.* 10:R42. <http://dx.doi.org/10.1186/gb-2009-10-4-r42>.