IґIS AᴘᴇʀTO

UNIVERSITÀ DEGLI STUDI DI TORINO

This is the author's final version of the contribution published as:

The publisher's version is available at:

https://onlinelibrary.wiley.com/doi/full/10.1111/rda.12747

When citing, please refer to the published version.

Link to this full text:

http://hdl.handle.net/2318/1687002

1 **Genome-wide analysis of bull sperm quality and fertility traits**

2

3 R. Puglisi[1], G. Gaspa[2], D. Balduzzi[1], A. Severgnini[1], R. Vanni[1], N.P.P. Macciotta[2], A. Galli[3]

4

5 [1] Istituto Sperimentale Italiano Lazzaro Spallanzani, Loc. La Quercia, 26027, Rivolta d'Adda (CR),

6 Italy

7 [2] Dipartimento di Agraria, Università di Sassari, 07100 Sassari, Italy

8 [3] CREA - Research Centre for Fodder Crops and Dairy Productions, 26900, Lodi, Italy

9

10 Corresponding author: Phone: +39 0363 78883; fax: +39 0363 371021.

11 E-mail address: roberto.puglisi@istitutospallanzani.it (R. Puglisi)

12

13 Running title: GWAS of bull fertility traits

14

16

Abstract

18 Because the priority of AI industry is to identify sub fertile bulls, a predictive model that allowed for

19 the prediction of 91% bulls of low fertility was implemented based on seminological (motility)

20 parameters and DNA status assessed both as DNA fragmentation index (DFI) and by TUNEL assay

21 using sperm of 105 Holstein Friesian bulls (4 batches per bull) selected based on *in vivo* estimated

22 relative conception rates (ERCR). Thereafter, sperm quality and male fertility traits of bulls were

23 explored by GWAS using a high density (777K) Illumina chip.

24 After data editing, 85 bulls and 591,988 SNPs were retained for GWAS. Out of 12 SNPs with false

25 discovery rate < 0.2, four SNPs located on BTA28 and BTA18 were significantly associated (LD

26 adjusted Bonferroni < 0.05) with the non-compensatory sperm parameters DFI and TUNEL. Other

27 SNPs of interest for potential association with TUNEL were found on BTA3, in the same

28 chromosome where associations with non-compensatory *in vivo* bull fertility were already reported.

29 Further suggestive SNPs for sperm membrane integrity were located on BTA28, the chromosome

30 where QTL studies previously reported associations with sperm quality traits. Suggestive SNPs for

31 ERCR were found on BTA18 in the vicinity of a site already associated with *in vivo* bull fertility.

32 Additional SNPs associated with ERCR and sperm kinetic parameters were also identified. In contrast

33 to other, but very few GWAS on fertility traits in bovine spermatozoa, which reported significant

34 SNPs located on BTX, we have not identified SNPs of interest in this sexual chromosome.

35

Introduction

37

38 Numerous authors investigating the genetic basis of fertility suggest that genome-wide association

39 studies (GWAS) are more effective in detecting causal variants associated to complex fertility traits

40 when compared to traditional quantitative trait loci (QTL) mapping (Zhang et al. 2012). However,

41 only few GWAS have focused on bull fertility (Fortes et al. 2013). Although the accurate estimate of

42 fertility allows to identify the critical number of viable sperm required to obtain adequate pregnancy

43 rates, there are some uncompensable characteristics, such as the state of nuclear chromatin, which

44 cannot be overcome by simply increasing the sperm number (Evenson and Wixon 2006). Because

45 the priority of AI industry is to identify hypofertile bulls, which require more sperm in the dose to

46 reach maximum fertility, a predictive model for the low level of fertility as estimated *in vivo* was

47 developed in the present work based on standard seminological and DNA status assessments.

48 Furthermore, sperm quality and male fertility traits were explored by GWAS using a high density

49 Illumina chip.

50

51 **Materials and methods**

52

53 **Estimate of *in vivo* bull fertility**

54 Four batches of commercial frozen sperm (years 2002-2014; 13 AI centres) of 105 Holstein Friesian

55 bulls were selected according to their fertility, based on 56-day non-return to oestrus adjusted for

56 environmental effects, calculated as the random effect of service sire (estimated relative conception

57 rates, ERCR; 90% reliability) using the model described in Puglisi et al. (2012). Fifteen bulls were of

58 low fertility (ERCR < -2.46; mean = -3.8 ± 0.8) and 90 bulls were of middle-high fertility (ERCR >

59 -2.46; mean = +0.4 ± 1.7), based on the threshold fixed at 3 standard deviations below the mean

60 ERCR calculated on a dataset of 4989 bulls (mean ERCR = 0.0005 ± 0.82).

61

62 **Sperm analysis**

63 Sperm quality parameters of the 105 bulls were assessed as follows: membrane integrity (MI) was

64 evaluated by the NucleoCounter SP100 (ChemoMetec A/S, Allerød, Denmark); motility (total, TM;

65 progressive, PM; average path velocity, VAP) was evaluated by CASA System-HTM IVOS v.12

66 (Hamilton Thorne); DNA status, assessed both as DNA fragmentation index (DFI) implemented in

67 the sperm chromatin structure assay (SCSA$^{©}$) and by the TUNEL assay, was determined using the

68 flow cytometer Guava EasyCyte Plus® (IMV Technologies, l'Aigle, France) as described (Evenson

69 and Wixon 2006).

70

71 **Genomic analysis**

72 Sperm genomic DNA of bulls was genotyped with the Illumina BovineHD chip (777K) (Illumina,

73 San Diego, CA). Both SNPs and bulls with call rate < 95% and < 97.5%, respectively, were

74 discharged. SNPs were removed if the Minor Allele Frequency (MAF) was lower than 0.02, or if they

75 statistically deviated from the Hardy Weimberg equilibrium (p < 0.0001).

76

77 **Statistical analysis**

78 Statistical analysis was implemented by R procedures (R Core Team, 2012).

79 At first, seminological data were evaluated by general linear mixed model (GLMM) using bull and

80 batch as random effects and semen production centres as fixed effect. Thereafter, in order to

81 implement the model for the identification of the bulls of low fertility, the variable LowFERT was

82 defined as follows: LowFERT = 1 for ERCR < -2.46 and LowFERT = 0 for ERCR > -2.46.

83 A first logistic model was implemented with the continuous seminal variables and the discrete

84 variable BATCH, as follows:

85 *LowFERT = ß0 + ß1 TM +ß2 PM + ß3 VAP + ß4 MI + ß5 DFI + ß6 TUNEL + BATCHi + e*

86 A second model was, then, implemented including only the effects identified in the first model as

87 significant, as follows:

88 *LowFERT = ß0 + ß1 PM + ß2 MI + ß3 DFI + ß4 TUNEL + e*

89 Results were validated by bootstrapping with nonparametric resampling (1000 trials) using package

90 "boot".

91

92 For GWAS, sperm parameters were pre-corrected for the effect of production batch. The GWAS was

93 carried out with the Grammar genomic control (GC)approach, that account for genetic substructure

94	in the population (Aulchenko et al. 2007), implemented in the GenABLE R package (*polygenic* and

95	*grammar* functions).

96	At first, data were analysed with the following linear mixed model:

97	$$y_{jk} = AI\_cent_j + a_k + e_{jk}$$

98	where y_{jk} is the sperm parameter for the k-th bull; AI_cent is the fixed effect of the j-th AI centre; a_k

99	is the random polygenic additive effect of the k-th bull$\sim$N(0, $\mathbf{G}\sigma^2_a$); e_{jk} is the random residual$\sim$N(0,

100	Iσ^2_e), where G and I are the genetic (co)variance and identity matrices, respectively. The genetic

101	(co)variance between animals was structured using the genomic relationship matrix. Residual of the

102	model were, then, analysed with a linear model that included the fixed effect of the SNP genotype.

103	Given that Bonferroni is the simplest and more conservative correction for multiple testing assuming

104	independence of performed test, and that its application largely ignores the correlation between

105	markers due to linkage disequibrium, the genome wide significance was assessed by LD adjusted

106	Bonferroni (Sun et al. 2014; Wu et al. 2014). To discover SNPs potentially associated to seminal

107	parameters the threshold was fixed at $8.06*10^{-7}$ (0.05/N), where N is the number of haplotype blocks

108	estimated with –blocks flag in plink (N = 62,062). False discovery rate (FDR) and q-values were also

109	calculated: SNPs with FDR < 0.20 are discussed.

110

111	**Results and discussion**

112	Results of seminological and DNA status assessments are presented in Table 1. Statistical analysis

113	shown high variability among bulls for all the parameters, and a moderate variability among batches

114	for TM, PM and MI (Table 2). Differently, for DFI and TUNEL a negligible variability was reported

115	among sperm batches, thus confirming these parameters as intrinsic-not compensable characteristics

116	of individual bulls (Evenson 1999). The effect of the semen production centre was not significant.

117	The statistical model implemented for the identification of hypofertile bulls allowed for the prediction

118	of 91% (n = 14) bulls of low fertility and data was further validated by bootstrapping (89-91%; 95%

119	CI).

For GWAS, 85 bulls and 591,988 SNPs were retained, while 130,462 SNPs were discarded because did not reach the MAF threshold. Table 3 lists the top 12 significant SNPs with FDR < 20%, among which, four SNPs located on BTA28 and BTA18 were also significantly associated (Bonferroni-LD) with DFI and TUNEL. Further SNPs for DFI were found on BTA1-4-16-23-28. With respect to BTA28, in this chromosome QTL were previously detected for several semen quality traits (Valour et al. 2015). The complete list of suggestive SNPs for seminological traits and ERCR is presented as supplemental Table 4. Of interest, suggestive SNPs (p-value < $1.61*10^{-5}$= 1/62,062; Sun et al. 2014) for ERCR were found on BTA18 in the vicinity of the site were Peñagaricano et al. (2012) found association with *in vivo* bull fertility. Other SNPs, ranked by their values of nominal significance, were found on BTA3 for TUNEL, in the chromosome where associations with noncompensatory bull fertility were reported (Blaschek et al. 2011). Similarly, GWAS was successfully used for identifying candidate genes associated with several sperm traits in bulls (Fortes et al. 2013; Hering et al. 2014). In contrast to other studies reporting significant SNPs located on BTX (Suchocki and Szyda 2015), our work has not identified SNPs of interest on this sexual chromosome.

References

Aulchenko YS, Ripke S, Isaacs A, van Duijn CM, 2007: GenABEL: an R library for genome-wide association analysis. Bioinformatics **15**, 23(10), 1294-1296.

145   Blaschek M, Kaya A, Zwald N, Memili E, Kirkpatrick BW, 2011: A whole-genome association

146   analysis of noncompensatory fertility in Holstein bulls. J Dairy Sci **94**, 4695-4699.

147

148   Evenson DP, 1999. Loss of livestock breeding efficiency due to uncompensable sperm nuclear defects.

149   Reprod Fertil Dev **11**, 1-15.

150

151   Evenson DP, Wixon R, 2006: Clinical aspects of sperm DNA fragmentation detection and male

152   infertility. Theriogenology **65**, 979-991.

153

154   Fortes MR, DeAtley KL, Lehnert SA, Burns BM, Reverter A, Hawken RJ, Boe-Hansen G, Moore

155   SS, Thomas MG. 2013: Genomic regions associated with fertility traits in male and female cattle:

156   Advances from microsatellites to high-density chips and beyond. Anim Reprod Sci **141**, 1-19.

157

158   Hering D, Olenski K, Kaminski S, 2014: Genome-wide association study for poor sperm motility in

159   Holstein-Friesian bulls. Anim Reprod Sci **146**, 89-97.

160

161   Peñagaricano F, Weigel KA, Khatib H, 2012: Genome-wide association study identifies candidate

162   markers for bull fertility in Holstein dairy cattle. Anim Genet **43**, 65-71.

163

164   Puglisi R, Pozzi A, Foglio L, Spanò M, Eleuteri P, Maria G. Grollino, Bongioni G., Galli A, 2012:

165   The usefulness of combining traditional sperm assessments with in vitro heterospermic insemination

166   to identify bulls of low fertility as estimated in vivo. Anim Reprod Sci **132**, 17-28.

167

168   R Core Team (2012). R: A language and environment for statistical computing. R Foundation for

169   Statistical Computing, Vienna, Austria.

170

171   Suchocki T, Szyda J, 2015: Genome-wide association study for semen production traits in Holstein-

172   Friesian bulls. J Dairy Sci **98**, 5774-5780.

173

174   Sun YF, Liu RR, Zhao GP, Zheng MQ, Sun Y, Yu XQ, Li P, Wen J, 2014: Genome-Wide linkage

175   analysis and association study identifies loci for polydactyly in chickens. G3-Genes Genomes

176   Genetics **4**, 1167-1172.

177

178   Valour D, Michot P, Eozenou C, Lefebvre R, Bonnet A, Capitan A, Uzbekova S, Sellem E, Ponsart

179   C, Schibler L, 2015: Dairy cattle reproduction is a tightly regulated genetic process: Highlights on

180   genes, pathways, and biological processes. Animal Frontiers **5**. doi:10.2527/af.2015-0006

181

182   Zhang H, Wang Z, Wang S, Li H, 2012: Progress of genome wide association study in domestic

183   animals. J Anim Sci Biotech **3**, 26. doi: 10.1186/2049-1891-3-26.

184

185   Wu Y, Fan H, Wang Y, Zhang L, Gao X, Chen Y, Li J, Ren H, Gao H, 2014: Genome-wide

186   association studies using haplotypes and individual SNPs in Simmental cattle. PLOS ONE **9**, e109330.

187

188   Table 1

189   Sperm quality parameters of 105 bulls.

Parameter	Low fertility		Medium-high fertility	
	Mean ± SD	Range	Mean ± SD	Range
Total Motility (%)	46.2 (14.8)[a]	11.0-72.0	53.2 (15.5)[b]	8.0-88.0
Progressive Motility (%)	34.9 (13.7)[a]	6.0-63.0	42.2 (13.8)[b]	6.0-70.0
Average Path Velocity (μm/sec)	90.5 (14.4)	59.0-120	92.5 (16.4)	11.0-138.0
Membrane Integrity (%)	50.0 (11.2)	25.0-71	52.0 (15.2)	0.0-89.0
DNA fragmentation Index (%)	9.8 (4.2)[a]	3.0-25.0	5.8 (5.5)[b]	1.0-96.0
TUNEL (%)	8.11 (4.4)[a]	2.0-22.0	5.6 (3.3)[b]	1.0-32

190   [a,b] Different superscripts within rows indicate statistical difference at the ANOVA test ($p < 0.001$)

191

192   Table 2

193   Random (bull and batch) and fixed (production centre) effects by general linear mixed model.

Parameter	Variance		Centre
	Bull	Batch	
Total Motility (%)	64.88 (27.49%)	20 (8.47%)	< 0.0001
Progressive Motility (%)	70.76 (37.69%)	15.46 (8.24%)	< 0.0001
Average Path Velocity (μm/sec)	96 (44.20%)	8 (3.50%)	< 0.0001
Membrane Integrity (%)	89.24 (46.05%)	19.4 (10.01%)	< 0.0001
DNA fragmentation Index (%)	11 (35.62%)	0.30 (1.00%)	0.0571
TUNEL (%)	9 (20.91%)	0.38 (0.89%)	< 0.0001

194

195   Table 3

196   Top significant SNPs from GWAS with false discovery rate < 0.2.

Trait	SNP	BTA	bp	p-Bonf-LD	q-value
DFI	BovineHD2800009025	28	33,677,489	0.001	0.006
DFI	BovineHD2800009027	28	33,682,118	0.008	0.020
TUNEL	BovineHD0800005232	18	16,773,834	0.016	0.149
DFI	BovineHD2800008609	28	32,601,290	0.026	0.050
DFI	BovineHD2800006900	28	26,638,772	0.100	0.135
DFI	BovineHD1600001050	16	3,700,646	0.191	0.139
DFI	ARS-BFGL-NGS-117941	16	4,095,536	0.254	0.150
DFI	BovineHD2600011042	26	40,124,425	0.327	0.150
DFI	BovineHD0400022506	4	81,577,828	0.383	0.150
DFI	BovineHD1600000982	16	3,486,636	0.428	0.150
DFI	BovineHD2300001056	23	4,581,363	0.483	0.157
DFI	BovineHD0100016322	1	57,604,927	0.712	0.192

197

198   DFI, DNA fragmentation index; p-Bonf-LD = nominal p-value/62,062.

199