Unravelling the genomic architecture of bull fertility in Holstein cattle

Francisco Peñagaricano

Department of Animal Sciences, University of Florida, Gainesville FL 32611, USA

Introduction

Improving reproductive efficiency of dairy cattle has become one of the major challenges of the dairy industry worldwide. The intense selection for production traits in the last decades has led to a decrease in fertility. Fertilization failure and early embryonic loss have been identified as the two main factors contributing to this decline. For instance, fertilization rate in high-producing dairy cows is about 75%, and only 65% of the fertilized eggs are considered viable at 5-6 days post-fertilization [1]. It is no surprise that conception rates are only 35-45% [1]. Many reasons may account for this decline in reproductive performance, including physiological, nutritional, environmental, and genetic factors. In this sense, several studies have recognized that there is substantial genetic variation underlying reproductive success in dairy cattle [2,3].

Reproduction is a very complex process that involves numerous consecutive events, including gametogenesis, fertilization, and early embryo development, that should be accomplished in a well-orchestrated manner in order to achieve a successful pregnancy. The relative importance of the parental effects on the reproductive success, i.e., maternal versus paternal contribution to the zygote, is still largely unknown [4]. Most studies in dairy cattle have focused on female fertility, while male fertility has received much less attention. It is worth noting that the service sire has a direct influence not only in the fertilization process but also on the viability of the preimplantation embryo. In fact, previous studies have reported that the service sire represents an important source of variation for conception rate in dairy cattle [5].

There is a growing evidence that bull fertility is influenced by genetic factors. For instance, Fortes et al. reported that genomic regions in BTA2, BTA14, and BTAX are associated with testicular development, sperm quality, and hormone levels in young Brahman bulls [6]. Moreover, Li et al. showed that *MAP1B* and *PPP1R11*, two highly conserved spermatogenesis genes, are significantly associated with bull fertility in Holsteins [7]. Likewise, Peñagaricano et al. have identified a panel of genetic markers and also a set of biological pathways associated with sire conception rate in Holsteins [8,9]. Altogether, these studies suggest that it is possible to identify genomic regions and individual genes related to bull fertility in cattle.

The main objective of this study was to unravel the genomic architecture underlying sire fertility in Holstein dairy cattle. Sire Conception Rate (SCR) was used as a measured of bull fertility. Alternative methods were performed to identify genomic regions, individual genes, functional gene terms, and biological pathways associated with sire fertility. These findings can contribute to a better understanding of the genetics underlying this complex trait and may point out opportunities for improving bull fertility via selective breeding.

Sire Conception Rate

The Animal Improvement Programs Laboratory of the United States Department of Agriculture (AIPL-USDA) implemented in 2008 a national phenotypic evaluation of bull fertility called Sire Conception Rate (SCR). The model that is being used in the U.S. bull fertility evaluation includes both factors related to the service sire under evaluation (including age of the bull and AI organization) and also factors (nuisance variables) associated with the cow that receives the unit of semen (including herd-year-season, cow age, parity, and milk yield) [10,11]. The trait SCR is defined as the expected difference in conception rate of a given bull compared to the mean of all other evaluated bulls; in other words, a bull with an SCR value of +5.0% is expected to achieve a conception rate of 37% in a herd that normally averages 32% and uses average SCR bulls. The U.S. bull fertility evaluation, in contrast to evaluations for other traits such as production, is intended as a phenotypic rather that a genetic evaluation, because the estimates include not only genetic but also some (permanent) environmental effects.

Phenotypic and Genotypic Data

The entire evaluation of U.S. Holstein bull fertility was used in this study. Specifically, a total of 44,449 SCR records were available from a total of 10,884 Holstein bulls. These SCR records were obtained from 23 consecutive evaluations provided to the U.S. dairy industry between August 2008 and April 2016. Genotype data for 60,671 single nucleotide polymorphism (SNP) markers were available for 7,447 out of the 10,884 Holstein bulls with SCR evaluation. SNP markers that mapped to BTAX or had minor allele frequency less than 1% were removed. After data editing, a total of 58,029 SNP markers were retained for subsequent genomic analysis.

Statistical Methods: genome-wide association mapping

The association analysis between phenotypes and genotypes was performed using the single-step genomic best linear unbiased prediction (ssGBLUP). The ssGBLUP method is one of a

group of statistical methods that were originally developed for genomic prediction and later were extended for performing gene mapping. Indeed, ssGBLUP model is a modification of the classical BLUP model where the pedigree relationship matrix **A** is replaced by **H** which combines pedigree and genotypic information [12]. Candidate regions associated with sire fertility were identified based on the amount of genetic variance explained by 1.5 Mb window of adjacent SNPs evaluated across the entire bovine genome.

Statistical Methods: gene set analysis

The gene set analysis consists basically in three different steps [9]: (i) the assignment of SNPs to genes, (ii) the assignment of genes to functional categories [Gene Ontology (GO) and Medical Subject Headings (MeSH)], and finally (iii) the association analysis between each functional category and the phenotype of interest (using Fisher's exact test).

Whole genome association analysis

Figure 1 displays the results obtained with ssGLUP method in terms of the proportion genetic variance explained by 1.5 Mb SNP windows across the entire bovine genome. A total of six different genomic regions, distributed on chromosomes BTA5, BTA13, BTA21 and BTA25, explained more than 0.50% of the genetic variance for sire conception rate.

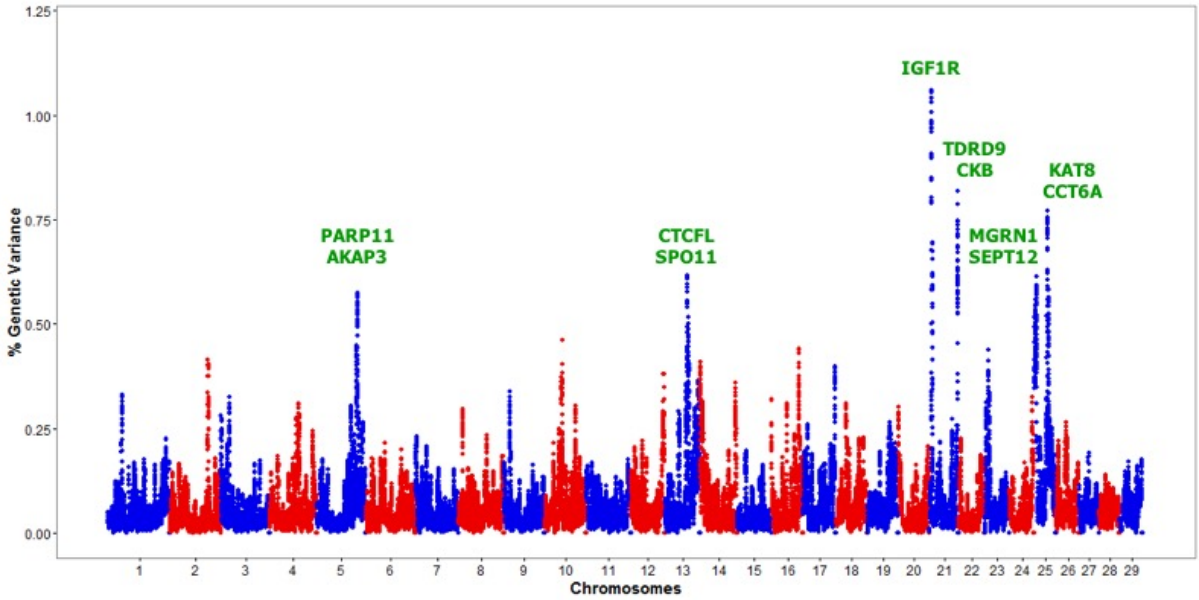

Figure 1. Results of the genome-wide association mapping for ***Sire Conception Rate*** in ***Holstein bulls***. The plot shows the percentage of genetic variance explained by 1.5 Mb SNP windows across the bovine genome. Candidate genes are shown in green. [Adapted from Han and Peñagaricano, 2016 [13]].

The region that explained the highest percentage of genetic variance (1.06%) was located on chromosome 21 (21:8031396-9528223). Interestingly, this region harbors *IGF1R*, an insulin-like growth factor receptor that plays critical roles in different reproductive events, including testis development and spermatogenesis. Another SNP-window on BTA21 (21:68,846,429-70,294,301) explained also a substantial amount of genetic variance (0.82%); this regions harbors two genes, *TDRD9* and *CKB*, which are implicated in sperm development and sperm quality, respectively. Moreover, two different regions on BTA25 (25:3148958-4647188, and 25:26736589-28233820) explained together almost 1.50% of the genetic variance. Notably, these regions harbor several putative candidate genes for bull fertility, including *KAT8*, *MGRN1* and *SEPT12*, which are directly involved in spermatogenesis, and *CCT6A* that is implicated in the fertilization process. Finally, two genomic regions on BTA5 and BTA13 were also identified; each of these windows explains roughly 0.60% of the genetic variance. The region located on BTA5 (5:105357507-106813133) harbors two genes, *PARP11* and *AKAP3*, that are involved in sperm maturation and motility. In addition, at least two putative genes related to male infertility, *CTCFL* and *SPO11*, are located in the middle of the region detected on BTA13 (13:58456868-59951247).

Gene Set Analysis

The whole-genome association analysis was complemented with a gene set enrichment analysis in order to detect potential functional categories and molecular mechanisms associated with sire fertility. **Figure 2** displays a set of GO Biological Process terms that were significantly enriched with genes associated with SCR. Noticeably, some of these terms are closely associated with male fertility, such as *reproduction process* (GO:0022414) and *fertilization* (GO:0009566). These two categories, highly related in the GO hierarchy, had four significant genes in common, namely *BSP3*, *BSP5*, *SLC22A16*, and *ZP2*, all of them directly involved in the process of spermatogenesis and subsequent ovum fecundation. Furthermore, many significant GO terms were associated with ion transport and homeostasis, including *cation transport* (GO:0006812), *zinc II ion transport* (GO:0006829), *regulation of sodium ion transport* (GO:0002028), *zinc ion homeostasis* (GO:0055069), and *cellular metal ion homeostasis* (GO:0006875). Moreover, terms related to developmental biology (e.g. GO:0048588), small GTPase mediated signal transduction (e.g. GO:0032482), and mRNA processing (e.g. GO:0050685) were also enriched with significant genes.

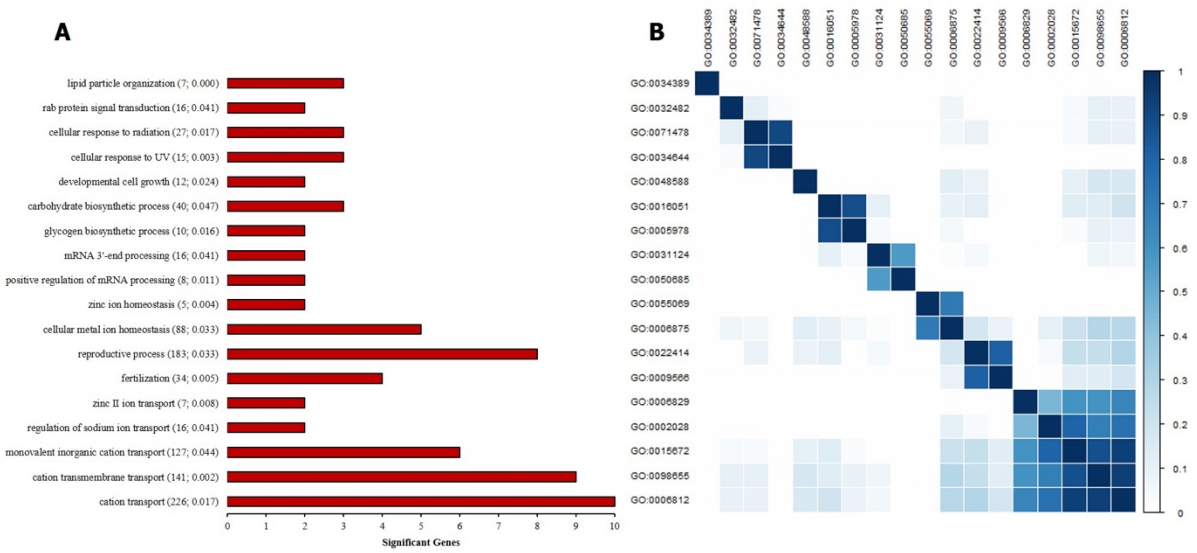

Figure 2. Gene Ontology Biological Process terms significantly enriched with genes associated with Sire Conception Rate: (A) Name, total number of genes, *P*-value, and total number of significant genes per functional term, and (B) Semantic similarity among functional terms.

Several GO terms classified into the Molecular Function domain showed an overrepresentation of genes associated with sire fertility. Especially, functional terms related to channel regulation [e.g., *calcium channel regulator activity* (GO:0005246) and *sodium channel regulator activity* (GO:0017080)], and transmembrane transporter activity [e.g., *inorganic cation transmembrane transporter activity* (GO:0022890) and *ion transmembrane transporter activity* (GO:0015075)] showed an overrepresentation of significant genes. Of particular interest, two closely related terms, *SNARE binding* (GO:0000149) and *SNAP receptor activity* (GO:0005484), which involve a group of membrane-associated proteins that participate in different reproductive events including spermatogenesis and acrosome reaction, were significantly enriched with at least three genes, *STX1A*, *STX1B* and *STX8,* associated with sire conception rate.

Table 2 shows a panel of MeSH terms that were enriched with genes associated with SCR. Many of these terms are closely related to male fertility, such as *spermatozoa* (D013094), *sperm capacitation* (D013075), and *sperm motility* (D013081). Five genes associated with SCR, namely *AKAP3*, *BSP3*, *BSP5*, *NTRK2* and *ZP2*, were part of these terms. Additionally, two other terms related to fertility, *follicle stimulating hormone* (D005640) and *pregnancy rate* (D018873), were also enriched with significant genes, including *AKT1*, *CTTNBP2NL*, *FSHR* and *IGF1R*. Finally, functional categories involving protein kinases (D017868) and GTPases (D020691) were also detected as significant in the MeSH-informed enrichment analysis.

Table 2. MeSH terms significantly enriched with genes associated with Sire Conception Rate

Mesh ID	MeSH Term Name	Total Genes	Significant Genes	*P*-value
D005640	Follicle Stimulating Hormone	34	4	6.4×10^{-3}
D013075	Sperm Capacitation	9	2	1.6×10^{-2}
D013081	Sperm Motility	13	4	1.4×10^{-4}
D013094	Spermatozoa	71	5	2.0×10^{-2}
D017868	Cyclic AMP-Dependent Protein Kinases	75	5	2.5×10^{-2}
D018698	Glutamic Acid	35	4	7.1×10^{-3}
D018873	Pregnancy Rate	4	2	2.8×10^{-3}
D020691	rab GTP-Binding Proteins	12	3	2.0×10^{-3}

Conclusions

In this study, a comprehensive genomic analysis was performed with the purpose of unravelling the genetic architecture underlying sire conception rate in Holstein dairy cattle. Genomic regions in BTA5, BTA13, BTA21 and BTA25 were associated with sire fertility. Most of these regions harbor genes with known roles in sperm biology, including sperm maturation, motility and fertilization. Moreover, gene set analysis revealed that many of the significant terms, such as reproductive process, calcium ion channels, and SNARE proteins, are implicated in biological processes related to male fertility. This integrative study sheds light on the genetic variants and mechanisms underlying this complex phenotype in cattle. In addition, these findings can provide opportunities for improving bull fertility via marker-assisted selection.

References

1. Santos, J.E.P., Thatcher, W.W., Chebel, R.C., Cerri, R.L.A. & Galvao, K.N. The effect of embryonic death rates in cattle on the efficacy of estrus synchronization programs. *Animal Reproduction Science* **82-3**, 513-535 (2004).
2. Shook, G.E. Major advances in determining appropriate selection goals. *Journal of Dairy Science* **89**, 1349-1361 (2006).
3. Weigel, K.A. Prospects for improving reproductive performance through genetic selection. *Animal Reproduction Science* **96**, 323-330 (2006).
4. Kropp, J., Peñagaricano, F., Salih, S.M. & Khatib, H. Invited review: Genetic contributions underlying the development of preimplantation bovine embryos. *Journal of Dairy Science* **97**, 1187-1201 (2014).

5. Amann, R.P. & DeJarnette, J.M. Impact of genomic selection of AI dairy sires on their likely utilization and methods to estimate fertility: A paradigm shift. *Theriogenology* **77**, 795-817 (2012).
6. Fortes, M.R.S., Reverter, A., Hawken, R.J., Bolormaa, S. & Lehnert, S.A. Candidate Genes Associated with Testicular Development, Sperm Quality, and Hormone Levels of Inhibin, Luteinizing Hormone, and Insulin-Like Growth Factor 1 in Brahman Bulls. *Biology of Reproduction* **87**(2012).
7. Li, G. *et al.* Comparative genomics between fly, mouse, and cattle identifies genes associated with sire conception rate. *Journal of Dairy Science* **95**, 6122-6129 (2012).
8. Peñagaricano, F., Weigel, K.A. & Khatib, H. Genome-wide association study identifies candidate markers for bull fertility in Holstein dairy cattle. *Animal Genetics* **43**, 65-71 (2012).
9. Peñagaricano, F., Weigel, K.A., Rosa, G.J.M. & Khatib, H. Inferring quantitative trait pathways associated with bull fertility from a genome-wide association study. *Frontiers in Genetics* **3**, 307 (2013).
10. Kuhn, M.T. & Hutchison, J.L. Prediction of dairy bull fertility from field data: Use of multiple services and identification and utilization of factors affecting bull fertility. *Journal of Dairy Science* **91**, 2481-2492 (2008).
11. Kuhn, M.T., Hutchison, J.L. & Norman, H.D. Modeling nuisance variables for prediction of service sire fertility. *Journal of Dairy Science* **91**, 2823-2835 (2008).
12. Aguilar, I. *et al.* Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. *Journal of Dairy Science* **93**, 743-752 (2010).
13. Han, Y. & Peñagaricano, F. Unravelling the genomic architecture of bull fertility in Holstein cattle. *BMC Genet* **17**, 143 (2016).