Indian Journal of Animal Sciences **85** (6): 578–583, June 2015/Article

Homology modeling of single nuleotide polymorphisms in candidate genes controlling embryonic growth of buffalo

JEROME A[1], A K PANDEY[2] and SUSHEEL KUMAR SARKAR[3]

ICAR-Central Institute for Research on Buffaloes, Hisar, Haryana 125 001 India

Received: 24 November 2014; Accepted: 6 January 2015

ABSTRACT

Pregnancy involves interactions of numerous growth factors, proteins and hormones exerting their biological functions in cellular growth, migration, differentiation and signal transduction. FGF2, STAT5A and UTMP are important mediators of intra-cellular signals transduction and transcription functions during pregnancy. Mutations in these genes will eventually disrupt their biological functions leading to embryonic death. The present study was designed to analyze *in silico* the SNPs in buffalo FGF2, STAT5A and UTMP genes by homology modeling. In the present study genomic DNA was isolated from the blood of 75 adult female buffaloes which was subsequently used for the amplification of FGF2, STAT5A and UTMP gene specific regions. PCR products of 167 bp, 429 bp and 279 bp were obtained for specific FGF2, STAT5A and UTMP gene regions, respectively. Sequenced PCR products showed 96–97% similarity with bovine sequences on BLAST analysis for all the 3 gene segments. Sequence analysis showed 9, 3 and 9 distinct nucleotide differences in the regions of FGF2, STAT5A, UTMP genes, respectively. Furthermore, based on the nucleotide difference 3 variants for FGF2 and UTMP genes were deduced in comparison with the bovine sequence. Promotor region analysis of FGF2 and homology modeling of STAT5A and UTMP gene revealed modification in the protein structure arising due to the presence of nucleotide changes. In the present study single nucleotide polymorphism were deduced in FGF2, STAT5A and UTMP gene region of buffalo and homology modeling of the studied gene portions were carried out.

Key word: Buffalo, FGF2, Homology modeling, Pregnancy, STAT5A, UTMP

Fertility indices such as age at first calving, service per conception, calving interval, and calving rate determine the reproductive efficiency. Among these service per conception is an important factor of fertility (Barile 2005) and this is affected by improper heat detection, improper AI timing, luteal insufficiency etc. (Agarwal *et al*. 2005). Pregnancy involves cascade of cytokines and their putative interactions for varied biological functions. These cellular functions involve various interactions with macromolecules such as DNA, RNA and protein. Interaction of protein with macromolecule plays critical role in controlling gene expression pattern. They include several growth factors viz. angiogenic factors like vascular endothelial growth factor (VEGF), basic fibroblast growth factor (FGF2), Pregnancy associated glycoproteins (PAGs) (Xie *et al.*1991, 1994, Garbayo *et al*. 2000, Szafranska *et al*. 1995), Signal Transduction and Activators of Transcription 5A (STAT5A), Uterine Milk Protein (UTMP) and several hormones along with their on uterine cells thereby affecting the vascular and cellular functions during embryonic development (Schäfer-Somi 2003). These interactions induce or inhibit other genes/proteins essential for uterine receptivity signaling and eventually facilitating pregnancy (Spencer and Bazer 2004). Perturbation in these interaction lead to embryonic mortality as recent studies have shown that various cytokines viz. Fibroblast Growth Factor 2 (FGF2), POU class 1 homeobox 1 (POU1F1) 1, Growth hormone (GH), Growth hormone receptor (GHR), Prolactin (PRL) Prolactin receptor (PRLR) Signal transducer and activator 5A (STAT5A), Osteopontin (OPN) and Uterine milk protein (UTMP) play key role during embryogenesis having placentogenic and embryo-protective role (Khatib *et al.* 2008, 2009, 2010). Studies by Oikonomou *et al*. (2011) showed presence of mutations in cattle. Such studies are lacking in buffaloes thus the present study was designed to analyze the SNPs in FGF2, STAT5A and UTMP genes, if any through homology modeling.

Present address: [1]Scientist (jerome210982@gmail.com), Division of Animal Physiology and Reproduction, [2]Principal scientist, Division of Animal Genetics and Breeding. [3]Scientist, ICAR-Indian Agricultural Statistics Research Institute, New Delhi.

MATERIALS AND METHODS

Experimental animals and sampling: The animals included in the present study were from the herd of Murrah Buffaloes maintained at cattle yard of ICAR-Central institute for research on buffaloes, Hisar, Haryana, India.

Based on the reproductive maintained by the animal farm section, CIRB, Hisar, blood samples were collected from 75 buffalo females which included normal animals: Normal (50), and repeat breeding animals (25). Animals which were inseminated more than 3 inseminations/conception were grouped as repeaters. Blood genomic DNA isolation protocol was standardized using the kit method (Invitrogen, Ltd). Integrity of the extracted DNA was checked by agarose gel (1%) electrophoresis and visualization of the gel under UV light after staining with ethidium bromide.

PCR amplification: On the basis of available sequences from cattle from NCBI database, gene specific primers were designed using 'Primer select' programme of 'DNA star' software. The primers sequence for the study is shown in Table 1. Optimal PCR conditions were standardized for the 3 genes i.e. FGF2, STAT5A and UTMP. PCR amplification was carried out in 25 µl of reaction mixture containing approximately 100 ng of genomic DNA and PCR master mix (0.05 u/µl *Taq* DNA Polymerase, 100 mM Tris-HCl Reaction buffer, 4 mM $MgCl_2$ and 0.4 mM of each dNTP - dATP, dCTP, dGTP, dTTP).The PCR protocol involved an initial denaturation at (94°C for 2 min) followed by 30 cycles of denaturation (94°C for 15 sec), annealing [49.5°C (FGF2);56°C (STAT5A) and 55°C (UTMP) for 30 sec] and extension (74°C for 1 min 30 sec) followed by one cycle of final extension (74°C for 10 min). The PCR product was checked by 1% agarose gel electrophoresis. The PCR products were sequenced by di-deoxy method to deduce the presence of variants in these segments.

Sequence analysis and homology modeling: The obtained sequence was analyzed with BLAST of NCBI Database to retrieve the similar sequences of other species. The retrieved sequences were subjected to multiple sequence alignment along with buffalo sequences using MegAlign program of Lasergene software. The allele frequencies were calculated by simple allele counting according to the Hardy-Weinberg equilibrium (Falconer and Mackay 1996). Genotypic and allelic frequencies were determined using POPGENE software (Version 1.32) (Yeh *et al.* 2000). Secondary structure prediction was done. Promoter analysis software viz. BDGP and TFSEARCH were used to predict the potential promoter and transcription sites in the amplified FGF2 gene intron 1 region (Heinemeyer *et al.* 1998, Reese 2001) Using PHYRE2 (Kelley and Sternberg 2009) and SABLE server (http://sable.cchmc.org/) (Adamczak *et al.* 2004) homology modeling of buffalo STAT5A exon 8 and UTMP exon 4 region was constructed and visualized using PyMol software.

Statistical analysis: The presence of variants was correlated with reproductive traits viz. service or inseminations per conception which correlates with non-return rates. Statistical analysis was done using PROCGLM model of SAS program (Base SAS® 9.3 Procedures Guide, NC, USA) with the mixed linear model given below

$$y_{ijk} = \mu + \delta_i + \varphi_j + \gamma_{ij} + e_{ijk}$$

where y_{ijk} : is the general mean; μ, is reproductive record kth animal with gene j and ipc i; δ_i, fixed effect (ipc) 'insemination per conception'; φ_j, random effect (gene); γ_{ij}, interaction between fixed and random effect; e_{ijk}, error term.

RESULTS AND DISCUSSION

Isolated blood genomic DNA showed $A_{260}/_{280}$ between 1.6–1.9. Moreover the yield of genomic DNA was between 90–100 ng/µl. Using the obtained geneomic DNA in PCR reactions, specific PCR products of 167 bp, 429 bp and 279 bp were obtained for FGF2 intron 1, STAT5A exon 8 and UTMP exon 4 regions, respectively (Fig. 1). The PCR products for all 3 genes were purified and sequenced. Following sequencing the obtained sequences were subjected to BLAST analysis which retrieved *Bos taurus* breed Holstein FGF2 gene intron 1 sequence (HM597774.1) showing 92% identity and phylogeny showing more closeness with bovine as compared to other species (sheep, pig and mouse). Sequenced PCR products showed 97% similarity with bovine STAT5A (AJ237937.1) and UTMP gene (HM597774.1) on BLAST analysis.

For FGF2 gene multiple alignments of the obtained sequences with cattle sequence *in silico* using MegAlign program of Lasergene software revealed 9 nucleotide variations which consisted of 4 transitions and 5 transversions. *In silico* analysis showed nucleotide variations at specific position in comparison to the cattle sequence (HM597774.1). These nucleotide variations were at position 293 (T→C), 294 (G→C), 310 (A→G), 314 (G→A), 319 (A→C/G), 320 (T→G), 327 (T→G), 335 (G→C), 366 (A→T). In addition, nucleotide variations were

Table 1. Primers used for amplifying FGF2, STAT5A and UTMP gene regions

Gene	Primer sequence (5´–3´)	Size (bp)	Annealing temperature (ºC)
FGF2	F: TCACCTCTCTGTGCACA R: CATACACTGAAGCCTGA	167	49.5
STAT5A	F: TGAGAAGTTGGCGGAGA R: GGGCCACTGTGAGTCA	429	56
UTMP	F: AAGTCCCAGTGAAGGCAA R: AGTTCCTGTGCTG CACA	279	55

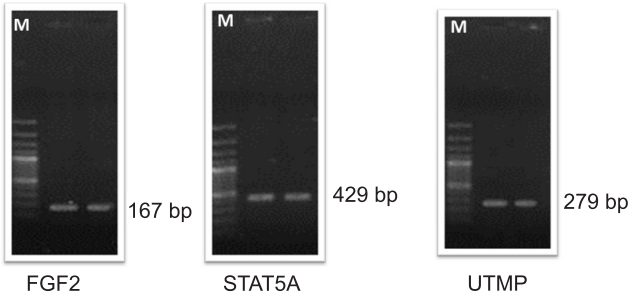

Fig. 1. Amplified PCR product of STAT5A and UTMP segment of buffalo (M:100 bp marker; A: FGF2; B: STAT5A; C:UTMP amplified PCR product).

Table 2. Variation identified in FGF2, STAT5A, and UTMP gene in buffalo

Gene	Nucleotide position	Cattle	Buffalo	Genotypic frequency	Allele frequency	Type of change	Amino acid change
FGF2	293	T	C	1	-	Transition	—
	294	G	C	1	-	Transversion	
	310	A	G	1	-	Transition	
	314	G	A	1	-	Transition	
	319	A	C/G	CC:0.426GG:0.574	C-0.714G-0.286	Transition /Tranversion	
	320	T	T/G	TT:0.32TG:0.68	T-0.66G-0.34	Transversion	
	327	T	G	1	-	Transversion	
	335	G	C	1	-	Transversion	
	366	A	T	1	-	Transversion	
STAT5A	914	G	T	1	-	Transversion	L→F *
	923	T	C	1	-	Transition	-
	1043	A	C	1	-	Transversion	-
UTMP	1100	G	A	1	-	Transition	S→N*
	1107	C	T	1	-	Transition	V→M*
	1116	C	T	1	-	Transition	K→N*
	1156	A	A/G	AA:0.306AG:0.694	A-0.654G-0.346	Transition	A→V*
	1158	T	T/C	TT:0.52TC:0.48	T-0.76C-0.24	Transition	V→M*
	1276	T	C	1	-	Transition	K→N*
	1302	T	A	1	-	Transversions	A→V*
	1303	C	T	1	-	Transition	V→D*
	1346	A	G	1	-	Transition	K→R*

(*Non-Synonymous).

present at position 319 where A was replaced either with C or G and at position 320 where T showed no variation or replaced with G. This resulted in transition or transversion at position 319 as compared to position 320 showing only transversion due to nucleotide variation. Genotype frequency of CC and GG were 0.426 and 0.574, respectively and the allelic frequency was 0.714 and 0.286 for C and G, respectively. Genotype frequency of TT was 0.32 and TG 0.68 with allelic frequency 0.66 and 0.34 for T and G allele, respectively. Variation at the position at 319 and 320 resulted in 3 variants in FGF2 gene which were designated as I, II, and III variants as shown in Table 3. For STAT5A gene, analysis revealed 3 nucleotide variations in buffalo at position 914 (G→T), 923 (T→C), 1043 (A→C) which include 1 transition and 2 transversion which resulted a single amino acid change from leucine to phenyalanine (L→F). Buffalo UTMP gene showed nine nucleotide differences at position 1100 (G→A), 1107 (C→T), 1116 (C→T), 1156(A→G), 1158 (T→C), 1276 (T→C), 1302 (T→A), 1303(C→T), 1346 (A→G) which include 8 transitions and 1 transversions as compared to bovine sequence. Genotype frequency of AA and AG were 0.306 and 0.694, respectively and the allelic frequency was 0.654 and 0.346 for A and G, respectively. Genotype frequency of TT was 0.52 and TC 0.48 with allelic frequency 0.76 and 0.24 for T and C allele, respectively (Table 2). Variation at the position at 1156 and 1158 resulted in 3 variants in UTMP gene which were designated as I, II, and III variants as shown in Table 3. These variant were submitted to the European Nucleotide Archive (KC582611–13). Analysis with TFSEARCH software to predict transcriptional binding

Table 3. Variants detected in buffalo FGF2 and UTMP gene

Gene	Nucleotide position	Cattle	Variant		
			I	II	III
FGF2	319	A	C	C	G
	320	T	T	G	T
UTMP	1156	A	A	G	G
	1158	T	C	T	C

factors potential revealed potential 28 transcription factor binding sites with a high threshold (85–100%) in the obtained buffalo FGF2 intron 1 nucleotide sequence (Fig. 2; Tables 4, 5). Therefore by these analyses presence of SNPs flanking the promoter region was deduced in the intronic region of FGF2 gene. It can be speculated that any nucleotide variations in the promoter region will eventually disrupt the expression pattern of FGF2 gene or alter the secretion of interferon–tau, the prime signal for maternal recognition of pregnancy resulting in embryonic loss. Secondary structure prediction of STAT5A and UTMP gene threw some light the alteration in the protein secondary structure arising due to nucleotide variation and amino acid change (Figs 3, 4). STAT5A gene secondary structure showed single amino acid changes and showed less change in secondary and homology model (Fig. 5). In contrast UTMP gene revealed profound changes in the secondary structure and homology model structure arising due to amino acid changes as compared the cattle model (Fig. 6). This alteration specific domain in the protein structure will affect its receptor binding property and eventually its

Table 4. Predicted transcriptional sites for FGF2 gene of buffalo

Start	End	Promoter sequence	Score
6	56	AAGGATAAAGTCAAAATGGGGCTGTTATAGAGTGGGCGCTAACCACATCT	0.83
21	71	ATGGGGCTGTTATAGAGTGGGCGCTAACCACATCTGACCGGTGTCCTCAC	0.92

Table 5. Strength of binding for transcription binding factors sites in FGF2 gene of buffalo

Region	Strength of binding	Region	Strength of binding
1	95.3	15	88.2
2	95.3	16	87.2
3	93.8	17	86.9
4	93.7	18	86.7
5	93.7	19	86.4
6	92.2	20	86.2
7	92.2	21	85.4
8	90.6	22	96.7
9	88.6	23	95.4
10	87.8	24	95.3
11	86.2	25	93.7
12	85.8	26	90.9
13	100	27	88.5
14	95.4	28	87.2

Fig. 2. Location of transcription factors binding sites in FGF2 intron 1 region of buffalo.

biological action. In the present study 3 candidate genes have been selected, viz. Fibroblast growth factor 2, Signal transducer and activator 5A and Uterine milk protein on investigation presence of variant type were present in FGF2 and UTMP gene. In the present study there was no significant association between the deduced SNPs and trait under study. Though in the present study different gene loci were studied but the results obtained were in accordance with the reports of Oikonomou *et al.* (2011) (Table 6). Furthermore the study of FGF2 and UTMP gene SNPs reported in cattle were from *in vitro* studies on embryonic survival and fertilization rates (Khatib *et al.* 2008, 2009, 2010) which did not have significant effect under farm conditions (Oikonomou *et al.* 2011). Present study results differ from that of Ahmed *et al.* (2011) due to the difference in the gene loci and the candidate genes under study.

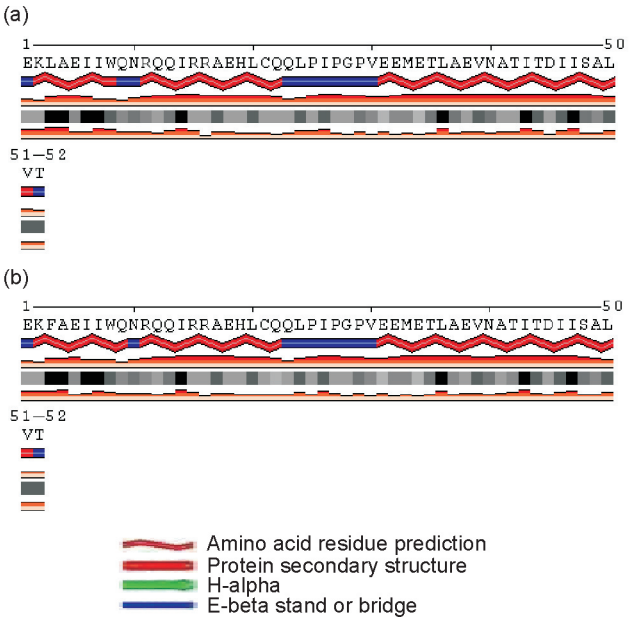

Fig. 3. Secondary structure prediction of (A) cattle; (B) Buffalo; STAT5A exon 8.

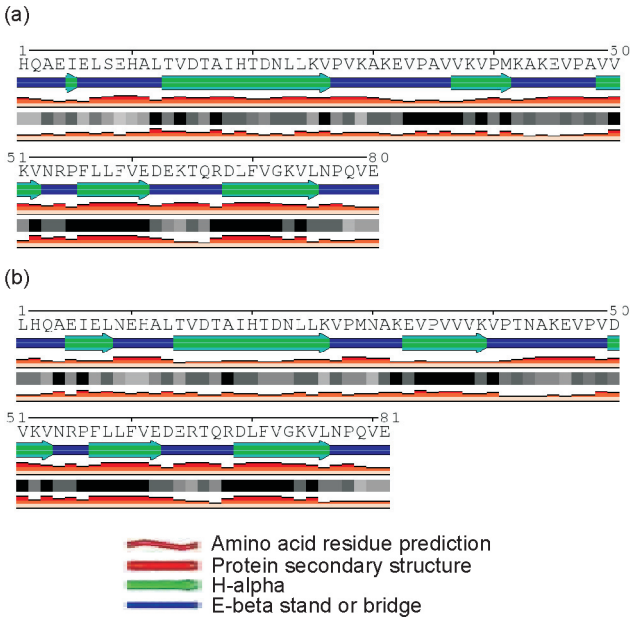

Fig. 4. Secondary structure prediction of (A) cattle; (B) buffalo UTMP exon 4.

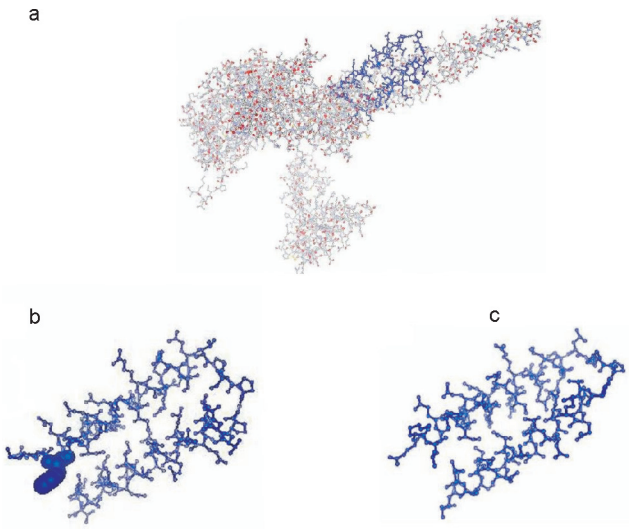

Fig. 5. Comparative modeling (A) cattle STAT5A; (B) buffalo STAT5A exon 8; (C) cattle STAT5A exon 8 (*change in amino acid indicated by space fill).

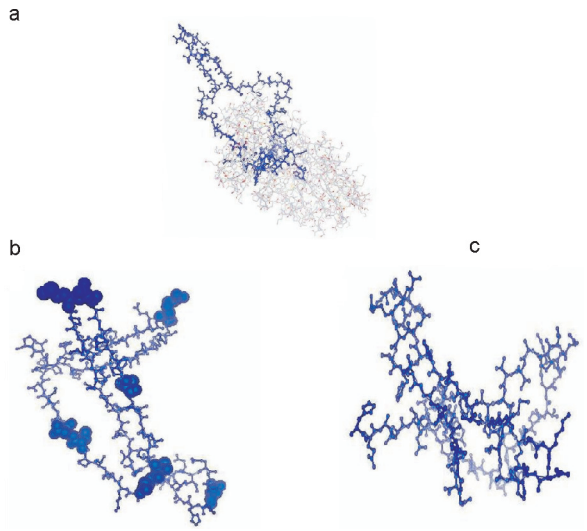

Fig. 6. Comparative modeling (A) cattle UTMP; (B) buffalo UTMP exon 4; (C) cattle UTMP exon 4.

Table 6. Association of polymorphism of FGF2, STAT5A and UTMP on inseminations/conception in buffaloes

Gene	Group	N	Inseminations/conception		F-value	p-value
			Mean	SD		
FGF2	1	9	3.22	0.57	1.10	0.3848
	2	23	2.86	0.23		
	3	43	2.90	0.18		
STAT5A	Normal	50	2.28	0.12	0.21	0.8091
	Repeater	25	4.56	0.42		
UTMP	1	17	3.11	0.24	0.58	0.7411
	2	30	3.13	0.17		
	3	28	3.00	0.17		

(P>0.05).

Furthermore this is supported with the inherent genetic difference between Murrah and Egyptian buffalo breeds. From the present study protein homology modeling of the variant gene segments of FGF2 and UTMP gene revealed less change in FGF2 protein structure due to synonymous changes. For UTMP gene loci there is considerable change in the protein structure due non-synonymous substitution altering the protein structure but these did not affect the overall fertility traits in accordance with earlier reports (Khatib *et al.* 2009, 2010). This can be due to the inherent genetic constituent of buffalo gene structure and other protein gene interactions which will counteract this effect during cellular functions. Thus the present study is the first report of homology modeling of SNPs in candidate genes controlling embryonic growth of buffalo.

ACKNOWLEDGEMENT

The authors thank the Director, ICAR-CIRB for providing the necessary facilities for conducting the experiment. The support of Dr R K Sharma and Dr S P Yadav is duly acknowledged.

REFERENCES

Adamczak R, Poroll A and Meller J. 2004. Accurate prediction of solvent accessibility using neural networks based regression. *Proteins* **56**: 753–67.

Ahmed S S, Abdel eziz K B, Hassan N A and Mabrouk D M. 2011. Genetic polymorphism of some genes related to reproductive traits and their association with calving interval in Egyptian buffalo. *Genomics and Quantitative Genetics* **3**: 1–8.

Agarwal S K, Singh S K and Rajkumar R. 2005. Reproductive disorders and their management in cattle and buffalo: A Review. *Indian Journal of Animal Sciences* **75** (7): 858–73.

Ashworth C J and Bazer F W. 1989. Changes in ovine conceptus and endometrial function following asynchronous embryo transfer or administration of progesterone. *Biology of Reproduction* **40**: 425–29.

Ayalon N. 1978. A review of embryonic mortality in cattle. *Journal of Reproduction and Fertility* **54**: 483–85.

Barile V L. 2005. Reproductive efficiency in female buffaloes. A Borghese. *Buffalo Production and Research.* FAO, Rome:77–108.

Base SAS® 93. 2011. Procedures Guide ISBN 978–1–60764–

895–6 SAS Institute Inc Cary, NC,USA.

Berglund B. 2008. Genetic improvement of dairy cow reproductive performance. *Reproduction in Domestic Animals* **43**: 89–95.

Demmers K J, Derecka K and Flint A. 2001. Trophoblast interferon and pregnancy. *Reproduction* **121**: 41–49.

Falconer D S and Mackay T F C. 1996. *Introduction to quantitative genetics*, 4th edition Longman, Harlow, UK.

Garbayo J M, Green J, Manikkam M, Beckers J F, Kiesling D O, Ealy A D and Roberts R M. 2000. Caprine pregnancy-associated glycoproteins (PAGs): their cloning, expression and evolutionary relationship to other PAG. *Molecular Reproduction Development* **57**: 311–22.

Heinemeyer T, Wingender E, Reuter I, Hermjakob H, Kel A, Kel O, Ignatieva E, Ananko E, Podkolodnaya O, Kolpakov F, Podkolodny N and Kolchanov N. 1998. Databases on transcriptional regulation: TRANSFAC, TRRD, and COMPEL. *Nucleic Acids Research* **26**: 364–70.

Kelley L A and Sternberg M J E. 2009. Protein structure prediction on the web: a case study using the Phyre server. *Nature Protocols* **4**: 363 –71.

Khatib H, Maltecca C, Monson R L, Schutzkus V, Wang X and Rutledge J J. 2008. The fibroblast growth factor 2 gene is associated with embryonic mortality in cattle. *Journal of Animal Science* **86**: 2063–67.

Khatib H, Huang W, Wang X, Tran A H, Bindrim A B, Schutzkus V, Monson R L and Yandell B S. 2009. Single gene and gene interaction effects on fertilization and embryonic survival rates in cattle. *Journal of Dairy Science* **92**: 2238–47.

Khatib H, Monson R L, Huang W, Khatib R, Schutzkus V, Khateeb H and Parrish J J. 2010. Validation of *in vitro* fertility genes in a Holstein bull population. *Journal of Dairy Science* **93**: 2244–49.

Michael D D, Alvarex I M, Ocón O M, Powell A M, Talbot N C, Johnson S E and Ealy A D. 2006. Fibroblast growth factor–2 is expressed by the bovine uterus and stimulates interferon-tau production in bovine trophectoderm. *Endocrinology* **147**: 3571–79.

Oikonomou G, Michailidis G, Kougioumtzis A, Avdi M and Banos G. 2011. Effect of polymorphisms at the STAT5A and FGF2 gene loci on reproduction, milk yield and lameness of Holstein cows. *Research in Veterinary Science* **91** (2): 235–39.

Reese M G. 2001. Application of a time-delay neural network to promoter annotation in the *Drosophila melanogaster* genome. *Computers and Chemistry* **26** (1): 51–56.

Schäfer-Somi S. 2003. Cytokines during early pregnancy of mammals: a review. *Animal Reproduction Science* **75** (1–2): 73–94.

Szafranska B, Xie S, Green J and Roberts R M. 1995. Porcine Pregnancy-Associated Glycoproteins:new members of the aspartic proteinase gene family expressed in the trophectoderm. *Biology of Reproduction* **53**: 21–28.

Thatcher W W, Guzeloglu A, Mattos R, Binelli M, Hansen T R and Pru J K. 2001.Uterine-conceptus interactions and reproductive failure in cattle. *Theriogenology* **56**: 1435–50.

Veerkamp R F and Beerda B. 2007.Genetics and genomics to improve fertility in high producing dairy cows. *Theriogenology* **68**: 266–73.

Xie S, Low B G, Nagel R J, Kramer K K, Anthony R V, Zoli A P, Beckers J F and Roberts R M. 1991.Identification of the major pregnancy specific antigens of cattle and sheep asinactive members of the aspartic proteinase family. *Proceeding of National Academy of Science, USA* **88**: 10247–51.

Xie S, Low B G, Nagel R J, Beckers J F and Roberts R M. 1994. A novel glycoprotein of the aspartic proteinase gene family expressed in bovine placental trophectoderm. *Biology of Reproduction* **51**: 1145–53.

Yeh F, Yang R C and Boyle T. 2000. POPGENE - for the analysis of genetic variation among and within populations using co-dominant and dominant markers, Win 95+ Version 132.