

Molecular differentiation of Peroxysome proliferator activated receptor coactivator-1 among different breeds of *Bubalus bubalis*

Anil Kumar¹, Ravinder Kumar¹, Vikas Beniwal¹, Sheo Narain Kala², Anamika Mishra³, Ashwin Ashok Raut³, Pradeep Kumar Naik⁴ & Vinod Chhokar^{1*}

¹Department of Bio & Nano Technology, Guru Jambheshwar University of Science & Technology, Hisar-125001, Haryana, India; ²Central Institute for Research on Buffalo, Hisar-125001, Haryana, India; ³High Security Animal Disease Laboratory, Bhopal-462021, Madhya Pradesh, India; ⁴Department of Bioinformatics & Biotechnology, Jaypee University of Information Technology, Waknaghat, Solan-173215, Himachal Pradesh, India; Vinod Chhokar – Email: vinodchhokar@yahoo.com; Phone: +91-1662-263-355; Fax: +91-1662-276-240; *Corresponding author

Received June 16, 2012; Accepted June 26, 2012; Published July 06, 2012

Abstract:

Peroxisome proliferator activated receptor coactivator-1 gene (PPARGC1A) is a positional and functional candidate gene for milk fat yield. It has key role in energy, fat and glucose metabolism. Single nucleotide polymorphisms (SNPs) in Exon-8 of PPARGC1A are reported to be associated with milk fat yield in dairy cattle. In the present investigation PPARGC1A was partially amplified (around 767bp) by designing gene specific primer and confirm by sequencing the amplicon and its comparison with the PPARGC1A gene of bovine. Comparative study of PPARGC1A among different breeds of buffaloes reveals different level of mutations with respect to its gene sequence 0.013-1.69% and protein sequence 0.42% to 2.99%, Similarly the protein structures modeled from their sequences were compared by structural superposition that shows variations (RMSD) from 0.736 to 1.507. Furthermore, the sequences were used to generate a dendrogram. It reveals that Murrah and reference are very close to each other, similarly Toda, Bhadawari and Surti are closely related, whereas Pandharpuri is separated from both the cluster. Especially the variations are more at the binding site of this protein that may be the cause that different breeds have different percentage of milk fat. Further study is underway to detect polymorphism and associate them with milk fat related traits in buffalo.

Keywords: PPARGC1A, *Bubalus bubalis*, Murrah, Surti, Pandharpuri, Bhadawari, Toda, Molecular differentiation

Background:

Milk fat percentage is the major factor that determines the quality of milk and hence affects the economics of milk production. There is a significant variation in average milk fat percentage across the buffalo breeds. Buffalo contribute 57% of the total milk produced in India. For example the breed, Murrah consists of average 6.7% milk fat while Bhadawari contains average 13% milk fat. There are about ten breeds of buffalo: Murrah, Pandharpuri, Bhadawari, Surti, Toda, Nili-Ravi, Jaffarabadi, Mehsana, Marathwada, Nagpuri reported in India. Out of them Murrah gives maximum amount of milk (~

18-20 ltr), however, the quality of the milk is not good because of low fat content. This difference in milk fat percentage across the different breeds clearly indicates that there is an involvement of a genetic factor. It also implies that over the evolutionary period the genes affecting milk fat percentage have segregated and fixed differently in the various breeds of buffaloes. Since within a species the basic physiological processes are more or less the same this difference in production potential can be attributed only small variation. These small differences which are probably the differences in the allelic distribution of the genes involved in the milk fat

synthesis are resulting in remarkable differences of production potential. The knowledge of these small genetic differences would be invaluable as they can be used to enhance the rate of genetic improvement.

With the development of refined analytical tools in the genome analysis, the identification of causal genes and their sequence variation underlying complex traits has become feasible in humans, animals, and plant [1]. The dissection of quantitative trait loci (QTL) with impact on complex, economically important traits in livestock and the identification of the underlying genetic variation will help to gain insight in to the metabolic pathways and associated genes involved [2]. The detection of single nucleotide polymorphism (SNPs) is important in various fields of biological sciences. Genotyping is widely used in association studies aimed at characterizing genetic factors underlying the inherited traits [3, 4]. In dairy cattle and other live stock species, there is an increasing interest in using a positional candidate gene approach for the identification of the actual gene that control economically important traits. At present, candidate quantitative trait genes are chosen for association studies because of previous linkage mapping studies and comparative biological functions in the same or other species [5].

Peroxisome proliferator-activated receptor- γ coactivator-1 α (PPARGC1A termed PGC-1 α) plays a key role in the activation of various nuclear hormone receptors and transcription factors regulating energy balance. It has also been demonstrated that PPARGC1A mediates the expression of genes involved in oxidative metabolism, adipogenesis and gluconeogenesis [3]. PPARGC1A plays a significant role in many aspects of glucose and fat metabolism and energy balance and is able to coordinate the metabolic processes of the liver, fat and muscle tissue in humans and mice. The bovine PPARGC1A gene acts as a plausible positional and functional candidate gene for a QTL for milk fat yield on BTA6 because of its chromosomal position and its key role in energy, fat and glucose metabolism. To analyze the role of bovine PPARGC1A gene in regulation of milk fat synthesis in dairy cattle, Weikard *et al.* [2] determine its cDNA sequence, genomic organization, chromosomal localization and expression pattern. In this study attempt has been made to study the molecular differences in PPARGC1A among different breeds of buffalo that have different percentage of milk fat.

Methodology:

Samples collection and DNA isolation

We have used five different breeds of buffaloes such as Murrah, Pandharpuri, Bhadawari, Surti, and Toda that are reared in different agroclimatic regions of India. Murrah is the major buffalo breed from northern India whereas Pandharpuri is from western India. Bhadawari is the major breed of Agra and Etawa (U.P) Bhind and Morena (M.P). Surti is major breed from north Gujarat and Rajasthan. Toda is major breed of Nilgiris hills (Tamilnadu). High milk fat yields and breed-specific traits were the main selection among these breeds—Bhadawari composed of high milk fat (13%) followed by Toda (8.2%) Surti (8.10%), Pandharpuri (7.0%), Murrah (6.7%). Only healthy animals with the characteristic traits of each breed were selected. Approximately 10 ml venous blood was collected from each breed with 0.5 ml of 2.7% EDTA as anticoagulant. The samples

were brought to the laboratory in a double walled ice box containing ice with cool pack and stored at -20 °C till the isolation of DNA. Genomic DNA was isolated from buffy coat of venous blood sample of buffalo using phenol chloroform extraction. The quality of DNA was checked by spectrophotometer taking ratio of optical density (OD) value at 260 and 280 nm. Good quality of DNA having OD ratio between 1.80 and 1.85 were used after an appropriate dilution for further work.

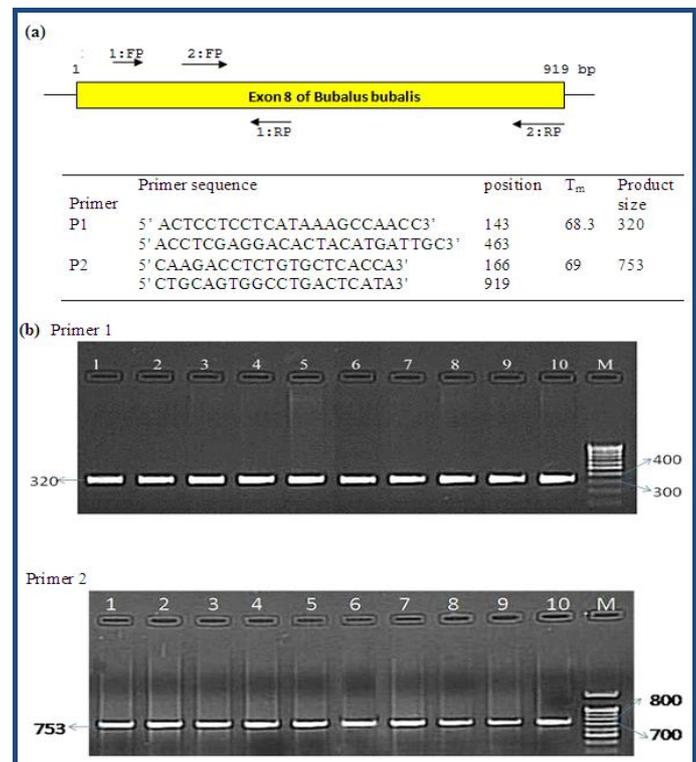


Figure 1: Primer and partially amplified PPARGC1A from different breeds of buffaloes. (a) A gene specific primer was designed based on Exon 8 sequence from bovine with a product size of 320 and 752 bp out of total length of 919 bp; (b) A representative gel picture showing the amplified product of PPARGC1A gene from different breeds of buffaloes (B): 1 & 2, Murrah; 3 & 4, Pandharpuri; 5 & 6, Bhadawari; 7& 8 Surti; 9 & 10 Toda.

Primer design and amplification of PPARGC1A

PPARGC1A from bovine has been well characterized but not in buffalo. It consists of 14 numbers of exons with 6324 total base pairs, exon 8 is the second largest one consisting of 919 base pairs. The amino acid sequence of PPARGC1A spanning within 10 Å of active site of the enzyme was extracted and its corresponding gene sequence was used for the primer designing. This region composed of approximately 767 bp of Exon 8 and covered 143 bp to 919 bp as mentioned in (Figure 1a), two sets of primers were used forward primer of first primer pair consisting of 22bp of P1f5' ACTCCTCCTCATAAAGCCAACC 3'reverse primer consisting of 24 bp P1r5' TGGAGCTCCTGTGATGTAACG 3'and forward primer of second primer pair consisting of 20bp P2f5' GCTTGGGACCAGACCTTAATTTG 3' and reverse primer consisting of 20bp P2r5' TGTACTTGGGCTTGTGACGTC 3'

were designed by using the *Bos taurus* sequence available on the NCBI.

For amplification of these partial sequences of PPARGC1A, 50 µl of PCR reaction was prepared by adding 10 pM of each primer, 0.2 µM of each dNTPs, 1.5 mM MgCl₂, 10x PCR assay buffer, 130 ng DNA template and 1 unit Taq DNA polymerase. The amplification was carried by using a programmable thermal controller (PTC-100, MJ Research) with the following conditions for first primer: initial denaturation of 5 minute at 94°C followed by 30 cycles of denaturation at 94°C for 30 second, annealing at 53°C for 30 second and extension at 72°C for 30 second and lastly the extension of 10 min. at 72°C and for second primer with the following conditions: initial denaturation of 5 minute at 94°C followed by 30 cycles of denaturation at 94°C for 45 second, annealing at 64°C for 45 second and extension at 72°C for 45 second and lastly the extension of 10 min. at 72°C. For sequencing, the PCR products were run in 1.5% agarose gel and product band was eluted using gel elution kit for purification. The amplified products were sequenced by Sanger method by outsourcing (Xclaris lab. Ahmadabad, India).

Comparison of DNA and protein sequences

The gene sequences obtained based on forward and reverse primers from different breeds of buffaloes were assembled utilizing overlapping sequences (consisting of 320 and 753 bp) and were compared. These sequences were aligned using multiple alignment program, Clustal W (<http://www.ebi.ac.uk/Tools/msa/clustalw2/>) with DNA weight matrix and multiple parameters like gap opening 10.0, gap extension 0.20, transition weight 0.50. A phylogenetic tree based on similarity coefficients generated by neighborhood-joining method was performed. Furthermore, all DNA sequences obtained were translated to protein sequence using EMBL trans seq (<http://www.ebi.ac.uk/Tools/st/>). These protein sequences were further aligned using multiple alignment program, Clustal W (<http://www.ebi.ac.uk/Tools/msa/clustalw2/>) with protein weight matrix and default parameters for gap opening and extension. A phylogenetic tree based on similarity coefficients was also generated.

Modeling of protein structure

The protein sequences were BLAST with PDB data base that gives only one similar protein with sequence similarity of 68% (PDB ID: 3U9Q). This homolog protein structure was used as template for homology model building of PPARGC1A using Prime accessible through the Maestro interface (Schrodinger, Inc.). During the homology model building Prime keeps the backbone rigid for the cases in which the backbone does not need to be reconstructed due to gaps in the alignment. The model was screened for unfavorable steric contacts and remodeled using a rotamer library database of Prime. Explicit hydrogen was added to the protein and the protein model was subjected to energy minimization using the Macromodel (Prime version 1.5) force-field OPLS-2005. Energy minimization and relaxation of the loop regions were performed using 300 iterations in a simple minimization method. The steepest descent energy minimization was carried out until the energy showed stability in the sequential repetition. Model evaluation was performed in PROCHECK v3.4.4 [6] producing plots which

were analyzed for the overall and residue-by-residue geometry. Ramachandran plot [7] provided by the program PROCHECK assured very good confidence for the predicted protein. Nevertheless, PROCHECK assured the reliability of the structure and the protein was subjected to VERIFY3D [8] available from NIH MBI Laboratory Servers.

Prediction and characterization of binding site

For each protein structure that has been modelled the binding sites were defined as a set of site points that were predicted using 'SiteMap' (Schrodinger Inc., version 2.4). A SiteMap calculation has three stages. In the first, relevant site points are selected based on geometric and energetic properties and the points are grouped in sets to define the binding sites. Next, hydrophobic, hydrophilic, and other key properties are computed at grid points and contour maps are prepared. Finally, binding site properties such as site score, size, volume, enclosure, exposure, contact, hydrophobic, hydrophilic and ratio of hydrogen donor to acceptor are computed. Finally, each binding site is scored by an expression that uses just three terms: (1) the square-root of the number of site points, capped at 100 site points to avoid overly rewarding large sites; (2) the enclosure score, and (3) the hydrophilic score. The 'SiteScore' function is represented by the following equation.

$$\text{SiteScore} = 0.0733 n^{1/2} + 0.6688 e - 0.20 p \quad (1);$$

where n is the number of site points (up to 100), e is the enclosure score, and p is the hydrophilic score, capped at 1.0 (the average for the submicromolar sites) to limit the impact of hydrophilicity in charged and highly polar sites.

Discussion:

Genomic DNA from blood samples was isolated as per the method of Sambrook and Russel (2001). The quality of genomic DNA was checked in 0.8% submarine agarose gel electrophoresis. Most of the DNA samples were of good quality and exhibited single band, where as few samples of poor quality DNA showed a smear throughout the lanes. Smearing of the DNA was found either due to the shearing of DNA or presence of protein in the sample. Those samples were re-extracted by phenol-chloroform method. Spectrophotometric reading of DNA samples also exhibited purity of DNA. The observed ratio of OD₂₆₀ to OD₂₈₀ was between 1.7 and 1.9 in most of the DNA samples indicating their purity. Samples with OD ratio beyond this range were reprocessed by phenol: chloroform extraction method. The samples were diluted in nuclease free water in 1.5 ml micro-centrifuge tube so that 1 µl solution contains approximately 80-100 ng of double stranded DNA. This diluted DNA was used for PCR amplification of all the experiments.

Amplification of PPARGC1A gene

PPARGC1A gene from different breeds of buffaloes was partially amplified using the primer specifically designed for PPARGC1A. Primer 1 gives an amplified product of ~ 320 bp, whereas the primer 2 produced an amplified fragment of ~753 bp. A clear, high reproducible and distinct band of size 320 and 753 bp was obtained (**Figure 1b**). The amplified products were confirmed by nucleotide sequencing using the same set of primers and by comparing with the reference gene of

PPARGC1A from bovine. The amplified fragments of PPARGC1A correspond to part of Exon 8.

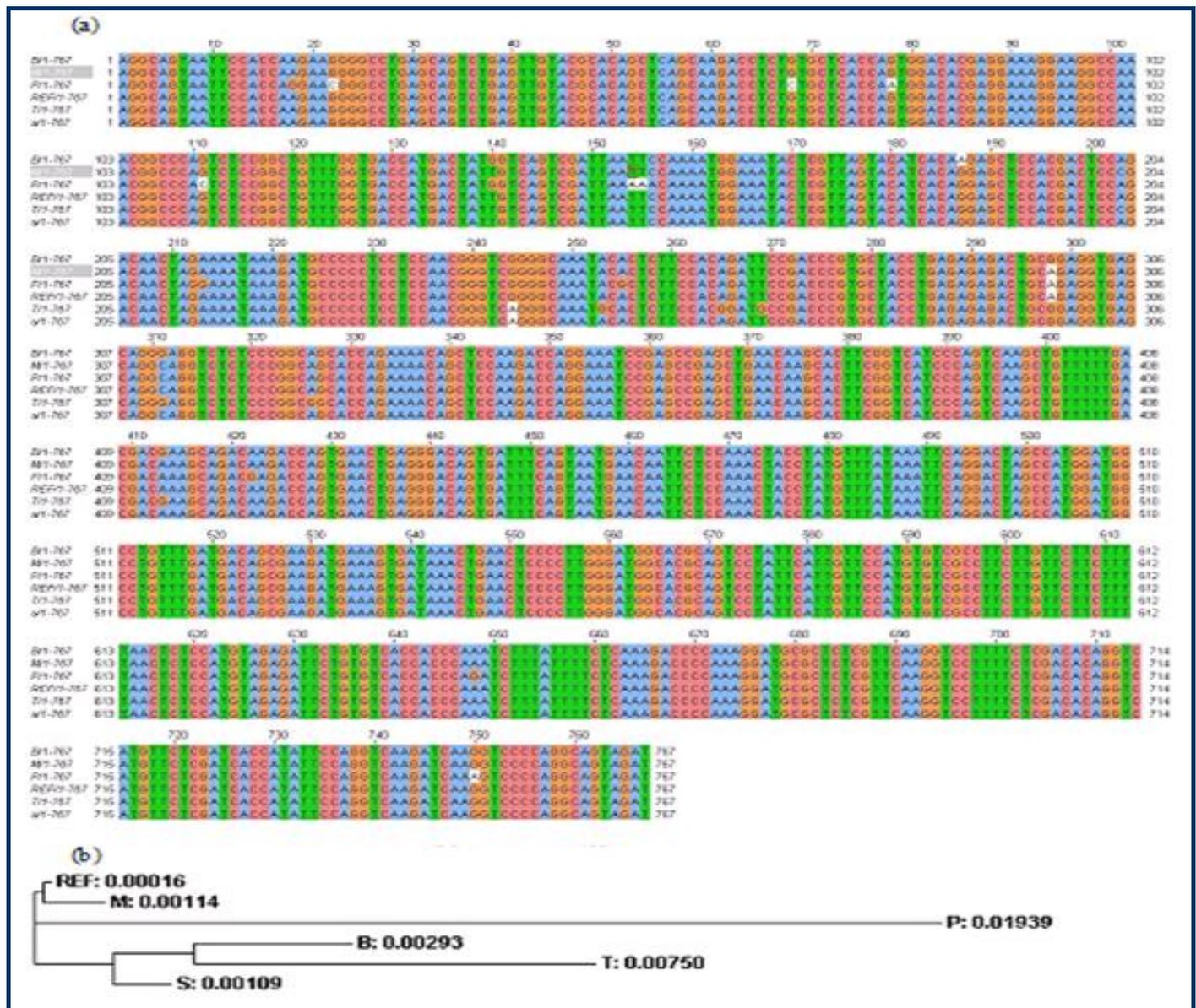


Figure 2: Comparison of PPARGC1A gene sequence. **(a)** The PPARGC1A sequences obtained from different breeds of buffaloes were aligned using multiple sequence alignment program, the changed nucleotide bases were represented in white box. Only few mutations were detected between different breeds; **(b)** The dendrogram representing the similarity between PPARGC1A gene sequences obtained from different breeds.

Sequence comparison of PPARGC1A

A relatively small variation was detected among the PPARGC1A gene sequences obtained from different breeds of buffaloes. All the gene sequences were compared using multiple sequence alignment program and level of variation in gene sequences is represented in **(Figure 2a)**. Genomic variation among various breeds ranges from 0.013-1.69% in comparison to PPARGC1A reference sequence obtained from bovine. The breed, Murrah has the minimum percentage of variation (0.013%) followed by Surti (0.26%), Bhadawari (0.65%), Toda (1.04%) and Pandharpuri (1.69%). Furthermore, the sequences were used to generate a dendrogram **(Figure 2b)**. It reveals that Murrah and reference are very close to each other, similarly Toda, Bhadawari and Surti are closely related, whereas

Pandharpuri is separated from both the cluster. Based on the comparison of protein sequences of PPARGC1A, a relatively small variation was detected among the different breeds of buffaloes **(Figure 3a)**. Overall the rate of variation varies in between 0.42% to 2.99%, with Toda having the highest variation (2.99%), followed by Pandharpuri (2.56), Bhadawari (1.70), and Surti (0.85) Murrah (0.42) in comparison to the reference sequence of bovine. The dendrogram **(Figure 3b)** obtained from pairwise comparison of protein sequence also reveals similar findings as mentioned above: Pandharpuri and reference are very close to each other, similarly Toda, Bhadawari and Surti are closely related, whereas Murrah is separated from both the cluster. The variation in gene as well as protein sequences of GPAM might be associated with variation in milk fat content. It

is previously reported that all the breeds; Toda, Bhadawari and Surti are having high milk fat content (Bianca Moioli and Antonio Borghese).

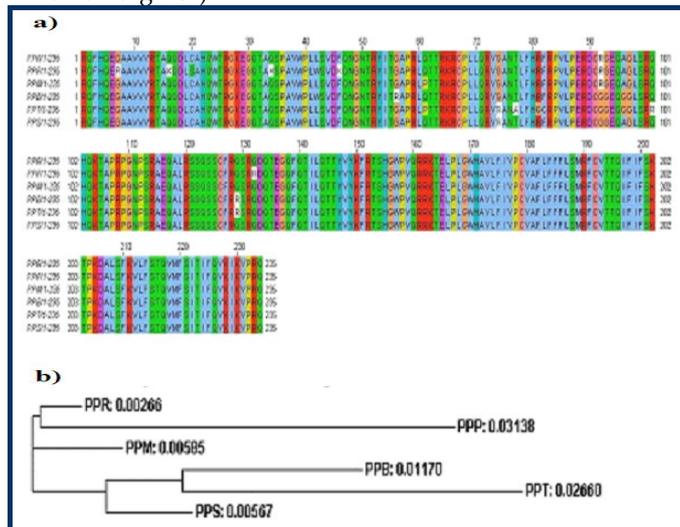


Figure 3: Comparison of PPARGC1A protein sequences. (a) The PPARGC1A sequences obtained from different breeds of buffaloes were aligned using multiple sequence alignment program, the changed amino acids were represented in white box. Only few mutations were detected between different breeds; (b) The dendrogram representing the similarity between PPARGC1A protein sequences obtained from different breeds.

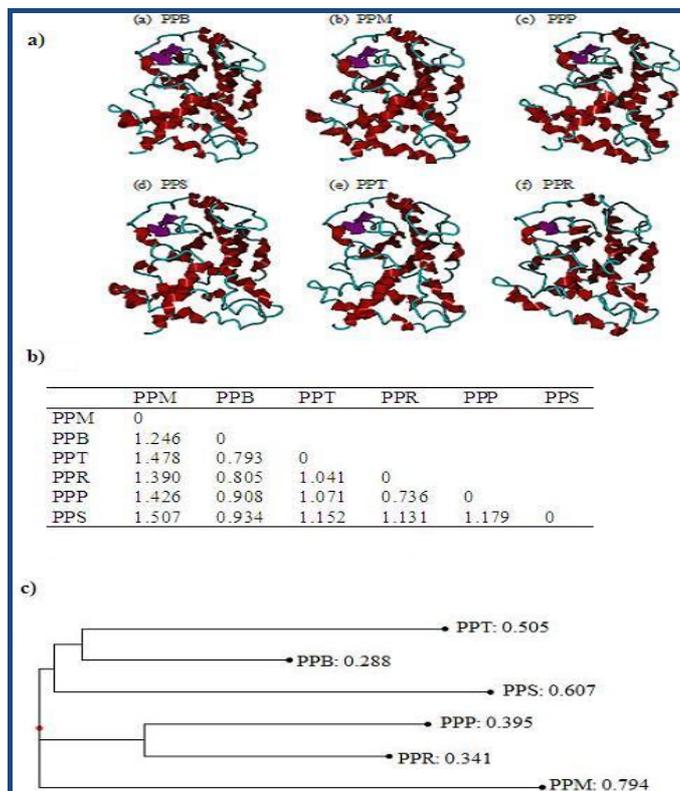


Figure 4: The modeled protein structure of PPARGC1A and its comparison. (a) The cartoon structure of PPARGC1A protein modeled using homology based technique from different breeds. The protein structures are more or less similar to each other; (b) The pairwise comparison of modeled PPARGC1A protein structure from different breeds. The value represents root mean square deviation (RMSD); (c) A dendrogram

representing the similarity between protein structures of PPARGC1A from different breeds. The tree was constructed using pairwise RMSD matrix.

Modeling of PPARGC1A protein structure and comparison among different breeds

The atomic coordinates of PPARGC1A protein structure for the organism *Bubalus bubalis* was not available in Protein Data Bank which necessitated for developing a protein model. The final model, which we modeled and took for further analysis, consisted of 269 amino acid residues. We used both PROCHECK and VERIFY3D softwares to check the quality of the modeled protein. Ramachandran plot obtained from the program PROCHECK, which checks the stereochemical quality of a protein structures, producing a number of postscript plots analyzing its overall and residue-by-residue geometry, assured the reliability of the modeled protein with 91% residues in most allowed region and 3.7% in additional allowed region. There were only 1.2% residues in disallowed region and 4.1% in generously allowed region. The assessment with VERIFY3D, which derives a "3D-1D" profile based on the local environment of each residue, described by the statistical preferences for, the area of the residue that is buried, the fraction of side-chain area that is covered by polar atoms (oxygen and nitrogen) and the local secondary structure, also substantiated the reliability of the three dimensional structure. The residues that deviated from the standard conformational angles of Ramachandran plot were the members of N terminal domain of the protein. This was an ignorable condition since the N-terminal end was not critical in our study. The distance of these residues to the active site residues also were found to be more than 10 Å, which suggested that those residues would interfere little with the binding of ligands in the active site region of PPARGC1A.

The structural comparison of PPARGC1A from different breeds revealed small variation among each other (Figure 4a). The pairwise comparison of protein structure (expressed as root mean square deviation, RMSD) varies from 0.736-1.507; the minimum RMSD value of 0.736 was found between Pandharpuri and the reference from bovine, whereas the maximum RMSD value of 1.507 was obtained between Murrah and Surti (Figure 4b). Furthermore, the dendrogram generated based on pairwise matrix of RMSD corroborate similar findings: Pandharpuri and reference are very close to each other, similarly Toda, Bhadawari and Surti are closely related, whereas Murrah is separated from both the cluster (Figure 4c). Various mutations detected in the gene and protein sequence of PPARGC1A among different breeds of buffaloes have different level of impact in the binding site Table 1 (see supplementary material). We have predicted the binding sites using SiteMap algorithm (Schrodinger software package). However, considered only the best binding site (based on Site Score) for the comparative study. The characteristic features of binding site analyzed using SiteMap is included in Table 1 (see supplementary material). The significant variation in binding site among different breeds of buffaloes may be associated with variation in milk content.

References:

[1] Glazier AM *et al. Science*. 2002 **298**: 2345 [PMID: 12493905].
 [2] Weikard R *et al. Physiol Genomics*. 2005 **21**: 1 [PMID: 15781588]

- [3] Puigserver P & Spiegelman BM, *Endocr Rev.* 2003 **24**: 78 [PMID: 12588810]
- [4] Liang H & Ward WF, *Adv Physiol Educ.* 2006 **30**: 145 [PMID: 17108241]
- [5] Rothschild M & Soller M, *Probe.* **8**: 13.
- [6] Laskowski RA *et al.* *J Appl Crystallogr.* 1993 **26**: 283.
- [7] Ramachandran GN *et al.* *J Mol Biol.* 1963 **7**: 95 [PMID: 13990617]
- [8] Eisenberg D *et al.* *Methods Enzymol.* 1997 **277**: 396 [PMID: 9379925]

Edited by P Kanguane

Citation: Kumar *et al.* *Bioinformation* 8(13): 600-606 (2012)

License statement: This is an open-access article, which permits unrestricted use, distribution, and reproduction in any medium, for non-commercial purposes, provided the original author and source are credited

Supplementary material:

Table 1: The binding site of PPARGC1A (predicted using Schrodinger software package) and its physico-chemical properties.

Proteins	Site Score	Size	Volume	Exposure	Enclosure	Contact	Hydrophobic	Hydrophilic	Donor/Acceptor
PPR	1.171	99	180.1	0.365	0.920	1.296	1.559	0.867	0.538
PPM	1.219	94	112.8	0.210	0.988	1.199	2.612	0.762	0.262
PPP	1.036	56	103.6	0.411	0.984	1.276	2.470	0.856	0.607
PPB	1.288	101	119.7	0.217	0.988	1.137	3.402	0.529	0.302
PPS	1.240	90	109.8	0.182	0.989	1.172	3.797	0.585	0.183
PPT	0.718	43	110.1	0.623	0.653	0.853	0.327	1.003	0.208