Global Journal of Animal Breeding and Genetics

ISSN: xxxx-xxxx Vol. 1 (1), pp. 048-057, November, 2013. © Global Science Research Journals

Full Length Research Paper

Molecular phylogenetic analysis of the sequences of candidate genes involved in milk production traits in riverine buffalo (*Bubalus bubalis*)

Aditi Sharma[1,2]*, S. S. Kanwar[1], M. S. Tantia[2] and R. K. Vijh[2]

[1]Himachal Pradesh University, Summer Hill-05, Shimla, Himachal Pradesh, India.
[2]National Bureau of Animal Genetic Resources, Baldi Bypass, Karnal, Haryana, India.

Abstract

Domestic buffalo and cattle are two extremely important livestock species in worldwide agricultural production. Despite some similarities with respect to morphologic and genetic characters, cattle and buffalo are divergent evolutionarily and are classified as different genera within the subfamily of Bovinae (Bos and Bubalus). The present study aimed at partial bayesian phylogenetic reconstruction of bovini tribe (Bovidae, Bovinae) from cDNA of 7 autosomal genes. Divergence times between cattle and buffalo were estimated using a relaxed molecular clock using calibration points based on best estimates of divergence times in the fossil record for Suina-Ruminantia and Laurasiatheria-Euarchontoglires splits. In the present analysis two calibration points were accessed. The two bovine subtribes consistently resolved themselves as a dichotomous group with strong support for a bifurcation between representatives of bovina and bubalina subtribes. Based on the molecular calibrations divergence time of buffalo and cattle was estimated to be 10.4 MYA. The mouse and rat split was estimated to be 36.6 MYA. The results are in agreement with the previous studies being carried out different research groups.

Key words: Phylogenetics, molecular calibration, bovidae, bovinae, divergence time, buffalo, cattle, laurasiatheria, ruminantia, euarchontoglires.

INTRODUCTION

Molecular systematics occupies one of the central stages in biology in the genomic era, ushered in by unprecedented progress in DNA technology. The inference of organismal phylogeny is now based on many independent genetic loci, a widely accepted approach to assemble the tree of life (Chenhong et al., 2007). The evolutionary connections between organisms are represented graphically through phylogenetic trees. Due to the fact that evolution takes place over long periods of time that cannot be observed directly, we must reconstruct phylogenies by inferring the evolutionary relationships among present-day organisms.

In the present work candidate gene approach was utilized, with an aim to identify their coding regions from the tissue of expression, which affect the milk attributes Alpha-S2-casein precursor, Epidermal growth factor, Phospholipase A2, group XIIA, ATP-binding cassette, sub-family G, solute carrier family 34 (sodium phosphate), member 2, Transmembrane protein 165 and Kappa-casein precursor genes were the seven candidate genes screened from cgQTL db for cattle. cgQTL is a web tool for candidate genes for QTL that allows navigation between the map of bovine milk production QTL, potential candidate genes and their level of expression in mammary gland arrays and in Gene-Atlas. The synteny of cattle with respect to buffalo (Goldammer et al., 2007) was analyzed and hereby used for studying the aforementioned genes in buffalo by sequencing and targeting the sequence using mRNA. Molecular data, which includes sequencing the coding regions, the cDNA sequences, were utilized to reconstruct phylogenetic trees in order to study the relationship of the selected candidate genes in ruminants on the basis of 7 gene trees the species tree was generated and calibrated to calculate the divergence time of buffalo and cattle. The two bovine subtribes consistently resolved themselves as

*Corresponding author. E-mail: adi2827@gmail.com

Table 1. List of primers utilized to amplify the seven cDNA sequences of the candidate genes taken up for the study.

S/N	Primer name	Left primer	Right primer	Length	Product size
1.	Cask	TACTGCCAAGCAAGAGCTGA	CTGGCCAAAGGAAAGTTTGA	40	692
2.	Casa2	TTCCATTGCCTGGACTACTTG	RCCACTCCAAACATKCTGGT	41	831
3.	Abcg1	TTGMGAAGTGAGGCTGAAGT	AACGAGGCTGATGAATGGAG	40	885
4.	Abcg2	TCCATCTTGTTCCTGGATGA	AGCTTCTTCTCCACCACCAG	40	752
5.	Abcg3	GCAGGAATYCAGAACAGAGC	TCCCTTTTTAATTGGGGGTAA	41	751
6.	Egf1	CGCACTTGAACTGTCCTGAA	TGGGACTTAGGCTCTTGCTC	40	940
7.	Egf2	TCGTATAGCCCAAGGCAGAG	ATCTCTTGGCCATTGGATGA	40	936
8.	Egf3	GTTCATCCAATGGCCAAGAG	TCGGCACTCATCAATATCCA	40	954
9.	Egf4	GAGCTACCAGGGAGATGGAA	GAGCTGATGGGAGAGGACAG	40	887
10.	Slc34a2- 1	GGCTGTCTTTKGRGCTAMAG	ACGGASAGCCAGTTGAAGAA	40	716
11.	Slc34a2- 2	GCCAACATTGGGACTTCAAT	GGTAGCCAGTCACCCAGCTA	40	691
12.	Slc34a2- 3	GTGGCTGCTGTCATCAAGAA	CCTGAAATGCAGGAAGAACC	40	983
13.	Tmem165-1	CTTGCTGGTTCCTTTGCTGT	CCGTTTCAACATCTCCTGGT	40	599
14.	Tmem165- 2	TCGGGAAGGCTTAAAGATGA	TGGGTCAAACTCACAAATCAG	41	571
15.	Pla2g12a	GCMGTCCTCCTGGTCCTCTT	GCAGCAAYGTCTGGGAAGAT	40	840

a dichotomous group and showed no evidence of producing a viable hybrid offspring. Based on the molecular calibrations divergence time of buffalo and cattle was estimated to be 10.4 MYA.

MATERIALS AND METHODS

Gene candidates

Candidate genes were selected for sequencing, using a priori knowledge of the gene's position in relation to suspected milk QTL on chromosomes 6 (BTA6). This potential list of genes was further reduced by selecting genes that might have a role in milk production from examination of the most recent annotation of the cattle genome.

Tissue sample collection

25 to 50 mg of tissue from mammary gland of buffalo was collected from slaughter house which was immediately flash freezed and immersed in tubes containing RNA later. Tissue was transported to the laboratory and stored at -80°C for subsequent use.

Isolation of the total RNA

Isolation of total RNA was carried out using Roche high pure RNA tissue kit.

First-strand cDNA synthesis

First strand cDNA synthesis was carried out by oligo (dT) method from invitrogen kit.

Primer designing for the seven candidate genes

Utilizing cDNA sequence of *Bos Taurus* available at NCBI and Ensemble, fifteen pairs of primers were designed to amplify cDNA of the selected candidate genes in buffalo. Primer 3 software (Rozen and Skaletsky, 2000) was used for primer designing. Degenerate primers were designed where necessary. Primer sequences and estimated product size with respect to the reference sequence are depicted in Table 1.

Polymerase chain reaction amplification (PCR)

The PCR conditions were standardized for all 15 primer pairs designed for the amplification of seven candidate gene chosen for study (Table 1) in buffalo.

PCR was performed using 10% of the first strand reaction (2 µl) with 10 mM dNTPs, 10X PCR buffer with 50 mM $MgCl_2$, 10 mM of each forward and reverse primers and 1 U of *Taq* DNA polymerase in 50 µl of reaction mixture. The amplification was carried out in thermal cycler (Biorad / Eppendorf) at suitable annealing temperatures of respective primers.

The reaction mixture was initially dentured for 5 min at 95°C, followed by 35 cycles of denaturation at 94°C for 1 min, annealing at 55°C for 1 min, extension at 72°C for 1 min; followed by final extension at 72°C for 10 min and held at 4°C. However, annealing temperature varied based on melting temperatures (Tm) of the primers. The PCR products were checked for amplification by loading on 2% agarose gel and electrophoresing at constant voltage of 10 V/cm. 100 bp ladder was loaded alongside as molecular size marker. The gels were checked for amplification in Gel documentation system (Gel Doc, Alpha Innotech). The amplified products without non specific amplification were further used for sequencing and a consensus sequence was obtained by aligning the sequences of different PCR products covering different regions of the candidate genes.

Sequence analysis

The chromatograms and sequences obtained for the seven candidate genes were edited using Chromas software to avoid ambiguity and their fasta files were prepared. The nucleotide sequences of the seven genes undertaken for study were searched

Table 2. Species chosen for the study.

Species	Super order
Rat	Eurochontoglires
Mouse	
Buffalo	Laurasiatheria
Cow	
Pig	
Avians (Birds, Outgroup)	Coelurosauria

Table 3. The name of seven genes that were sequenced along with their accession number.

S/N	Gene	Accession number
1.	Alpha-S2-casein precursor or casa2	HQ840513
2.	Kappa-casein precursor or cask	HQ677596
3.	Epidermal growth factor or egf	HQ823615
4.	Phospholipase A2, group XIIA or plag12a	HQ439434
5.	ATP-binding cassette, sub-family G or ABCG2	HQ713395
6.	Solute carrier family 34 (sodium phosphate), member 2 or Slc34a2	HQ709136
7.	Transmembrane protein 165 or tmem 165	HQ698271

for homology using *blastn* available on line at NCBI website. The comparison of the sequence with other species was obtained. Nucleotide sequences of various species aligning with the sequence of buffalo were selected. The details of the selected species and their super order are shown in Table 2. Fasta files of the cDNA sequences of the selected candidate genes were prepared and aligned using Clustal W.

Phylogenetic analysis

Genewise estimates of diversity between species were calculated using maximum composite likelihood method. Several statistical methods exist that can be used for reconstruction of the phylogenetic trees. A small bias was detected for transitions to transversions, therefore Kimura-2 parameter was utilized for reconstructing gene trees. Maximum likelihood tree was constructed and tested with 10000 Bootstrap replicates under the Kimura-2 parameter model of DNA substitution. Mega 5 (Tamura et al., 2011) was utilized.

Bayesian evolutionary analysis by sampling trees (BEAST), (Drummond and Rambaut, 2007) version 1.4.6 was used to the bayesian analysis. The nucleotide substitution model used was HKY with invariant sites and rate heterogeneity of rates across sites. The number of generations was set to 50000 with a sampling frequency of 50. In case when convergence was not obtained the number of generations was increased to 80000 with a sampling frequency of 200. Burn-in was set to 25% of the generations and the inferred tree was estimated as a consensus of all compatible groups of the post burn-in trees. The inferred species tree was compared to the tree generated by PhyML (Guindon and Gascuel , 2003).

Bayesian evolutionary analysis utility (BEAUti) was used to create the XML file to be used by BEAST as an input and tracer was used to analyze the text files generated by BEAST. Tree annotator was used for tree concatenation and Fig tree/treeview was used for the visualization of the reconstructed species tree.

Molecular calibration

Divergence times were estimated using a relaxed molecular clock using calibration points based on best estimates of divergence times in the fossil record for Suina-Ruminantia (Gatesy, 2009) and Laurasiatheria-Euarchontoglires (Murphy et al., 2009) splits.

RESULTS

All the seven candidate genes were successfully amplified and sequenced from the cDNA sample extracted from the mammary gland of buffalo. The final cDNA sequences/consensus sequences thus obtained after sequencing, editing and aligning were submitted to Genbank and accession numbers were obtained. The genes along with their accession numbers are shown in Table 3.

Comparative analysis

The seven gene sequences were then compared with the gene sequences of cattle and other species and based on the gene sequences the phylogeny was derived. Based on the gene tree the species tree was constructed. A total of 2807 trees were generated and the best one representing the data was selected and further analyzed. The sequences were aligned using Clustal W (Thompson et al., 1994) and the changes identified between buffalo

Table 4. Number of variations found between buffalo and cattle in each cDNA sequence.

S/N	Gene	Total variations
1.	Abcg2	29
2.	Casa2	14
3.	EGF	101
4.	Cask	21
5.	Pla	10
6.	Slc	30
7.	Tmem	10

and cattle are shown in Table 4. Pairwise estimates of evolutionary divergence were calculated for all seven genes in 6 species and were plotted as graph in Figure 1(a) to (g). For all the seven genes the divergence between buffalo and cattle was found to be least as compared to other species in the data set. Based on the evolutionary distances between sequences, the phylogeny of seven genes was evaluated with traditional approach. All the seven genes were analyzed independent of each other. A small bias was detected for transition (ti) to transversion (tv) substitutions, with a ratio of ti/tv = 1.8. Therefore Kimura's -2 parameter model was used for phylogenetic reconstruction. Phylogenetic tree was constructed with maximum likelihood method with Kimura-2 parameter (Kimura, 1980).

Based on the evolutionary distances between sequences, the phylogeny of seven genes was evaluated with traditional approach. All the seven genes were analyzed independent of each other. A small bias was detected for transition (ti) to transversion (tv) substitutions, with a ratio of ti/tv = 1.8. Therefore Kimura's-2 parameter model was used for phylogenetic reconstruction. Phylogenetic tree was constructed with maximum likelihood method with Kimura-2 parameter [Figure 2(a) to (g)].

The 7 gene trees were concatenated using consensus program and species tree was generated (Figure 3). We inferred maximum likelihood and Bayesian trees for the 7 cDNA sequence datasets. For 7 of these datasets, every combination of approach and model that we investigated yielded the same topology (Figure 4). Interestingly, for these, the bootstrap consensus ML trees generated by PhyML were topologically identical to the Bayesian trees, indicating that the sequences in these datasets show a high degree of internal consistency across positions. Divergence times were estimated using a relaxed molecular clock using calibration points based on best estimates of divergence times in the fossil record for Suina-Ruminantia (Gatesy et al., 2009) and Laurasiatheria-Euarchontoglires (Murphy et al., 2009) splits.

Based on the molecular calibrations divergence time of Buffalo and cattle was estimated to be 10.4 MYA (Figure 5). The mouse and rat split was estimated to be 36.6 MYA. The results are in agreement with the previous studies being carried out across the globe.

In the present study the gene tree of the seven candidate genes namely: Alpha-S2-casein precursor, Epidermal growth factor, Phospholipase A2, group XIIA, ATP-binding cassette, sub-family G, solute carrier family 34 (sodium phosphate), member 2, Transmembrane protein 165 and Kappa-casein precursor genes were constructed. The seven gene trees were then concatenated and species tree was generated in order to reconstruct the phylogeny of bovini tribe. The species tree was found to comply with the gene trees. Incongruence among gene trees may be an artifact of the data or used stochastic factors (although utmost care has been taken in dealing with data, generating consensus sequences and choosing the stochastic factors). Based on the estimates of divergence, molecular phylogeny and molecular clock calibrations the divergence time of buffalo and cattle were estimated. From the present analysis of seven autosomal genes, buffalo and cattle were found to have diverged 10.4 million years ago.

Conclusion

As a high rate of amplification was achieved for buffalo from primers designed using cattle as reference despite divergence time of 10.4 MYA. A high rate of amplification between species should facilitate further population genetic studies in Bovini and possibly other ruminant subfamilies. The phylogenetic reconstruction of the Bovini tribe has resolved two distinct subtribes the Bubalina and Bovina. The divergence time of Bubalina (Buffalo) and Bovina (Cow) were estimated to be 10.4 MYA. The study can be extended by sequencing more number of genes in different members of Bovinae subfamily.

ACKNOWLEDGEMENTS

World bank funded National Agricultural Innovation Project is greatly acknowledged.

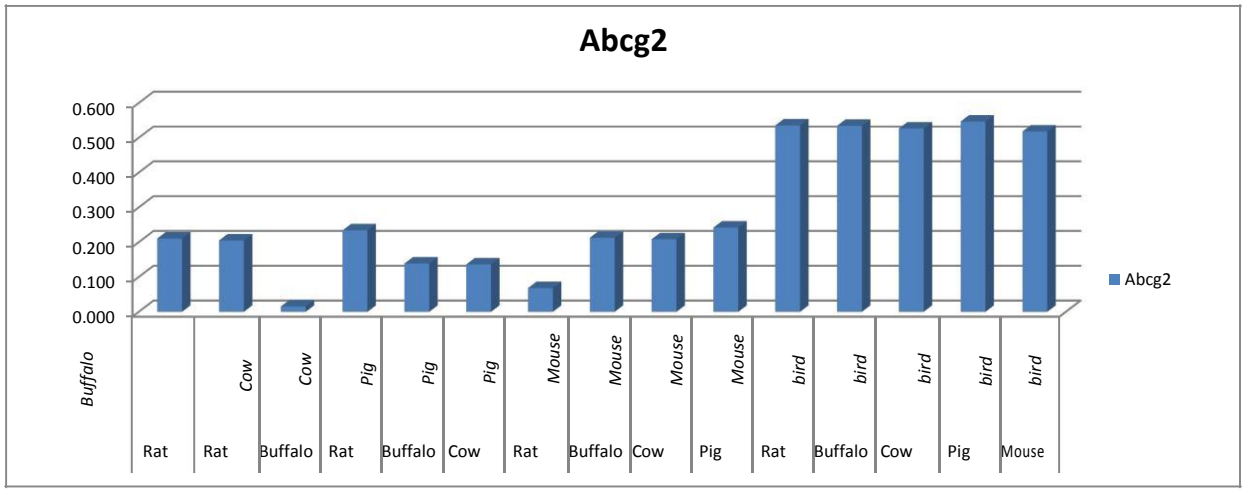

Figure 1(a). Abcg2 gene, estimates of evolutionary divergence between sequences.

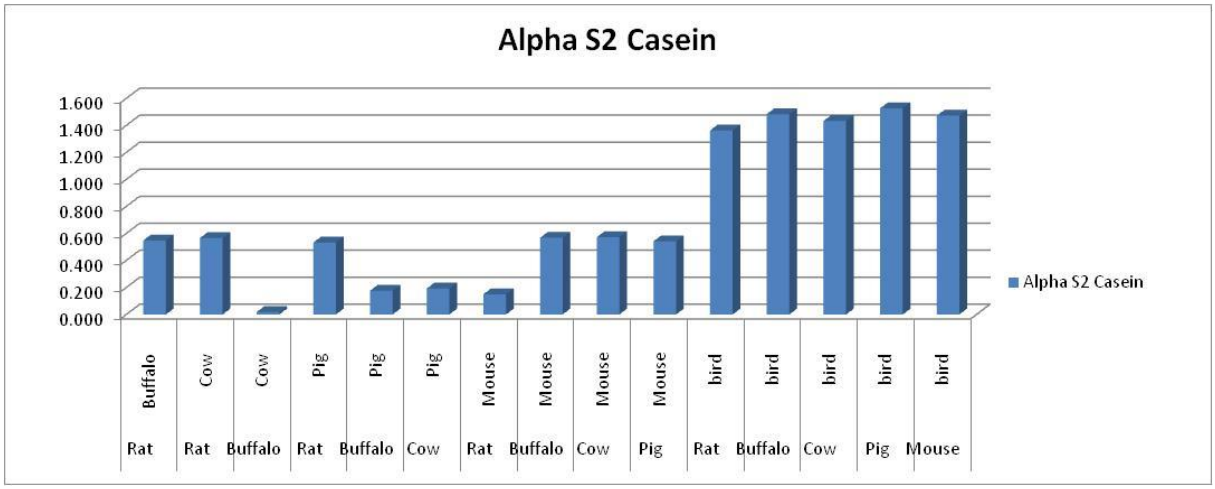

Figure 1(b). Alpha S2 Casein gene, estimates of evolutionary divergence between sequences.

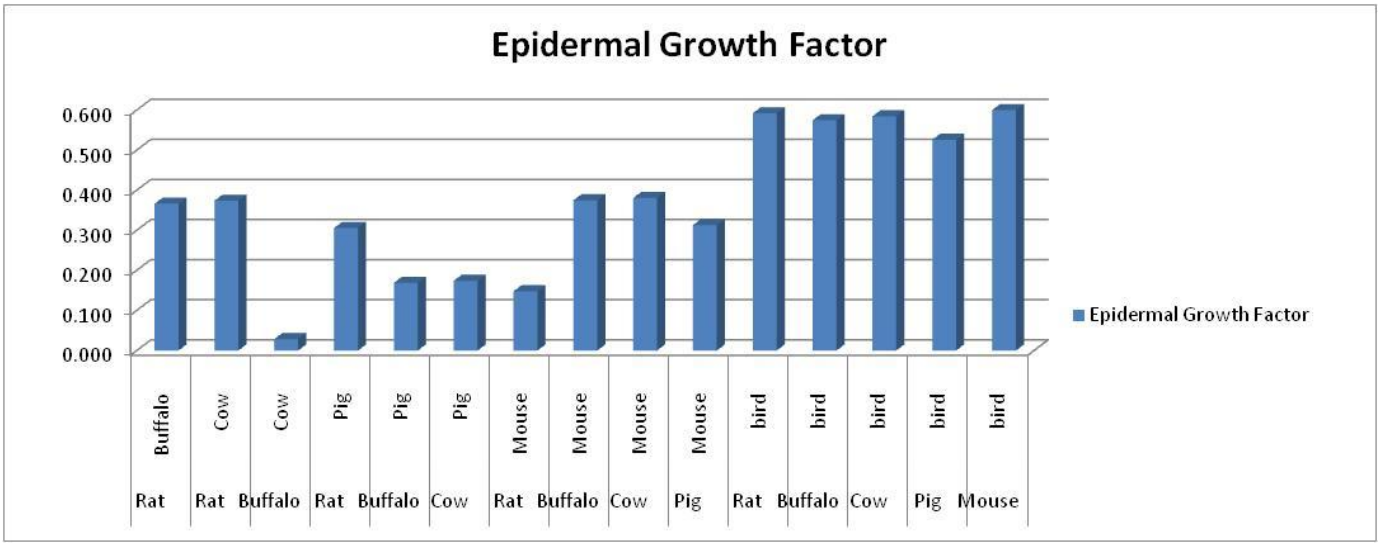

Figure1(c). Epidermal growth factor gene, estimates of evolutionary divergence between sequences.

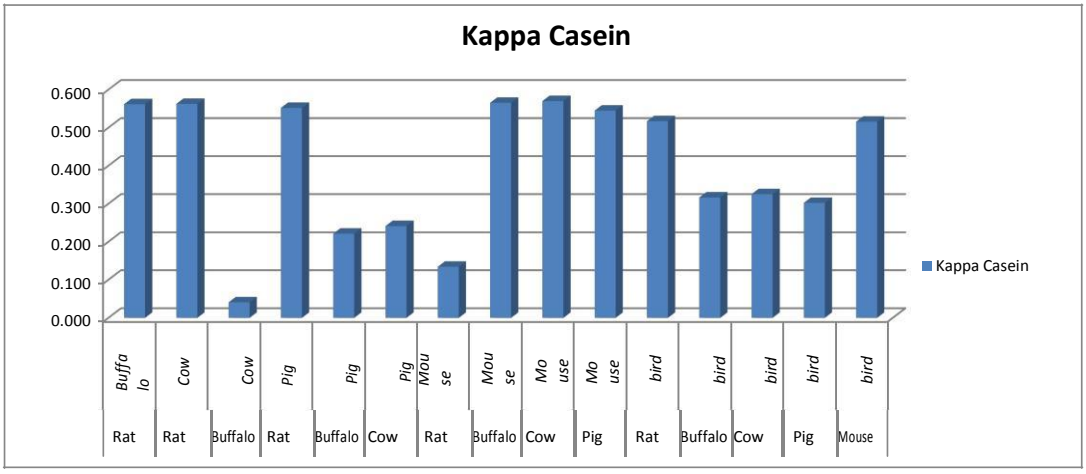

Figure 1(d). Kappa casein gene, estimates of evolutionary divergence between sequences.

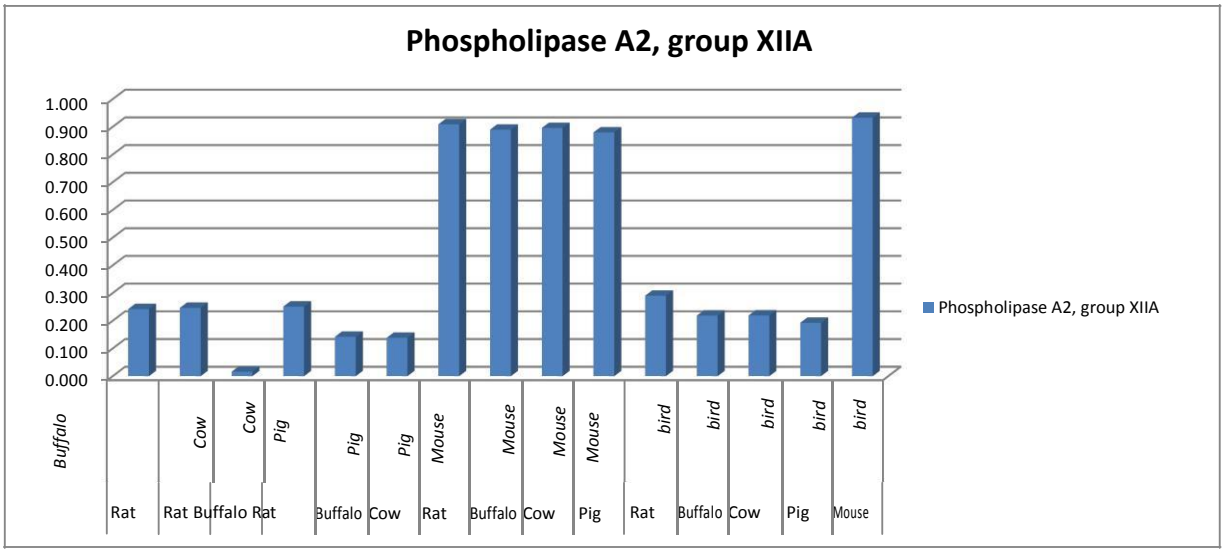

Figure 1(e). Phospholipase A2 group XIIA gene, estimates of evolutionary divergence between sequences.

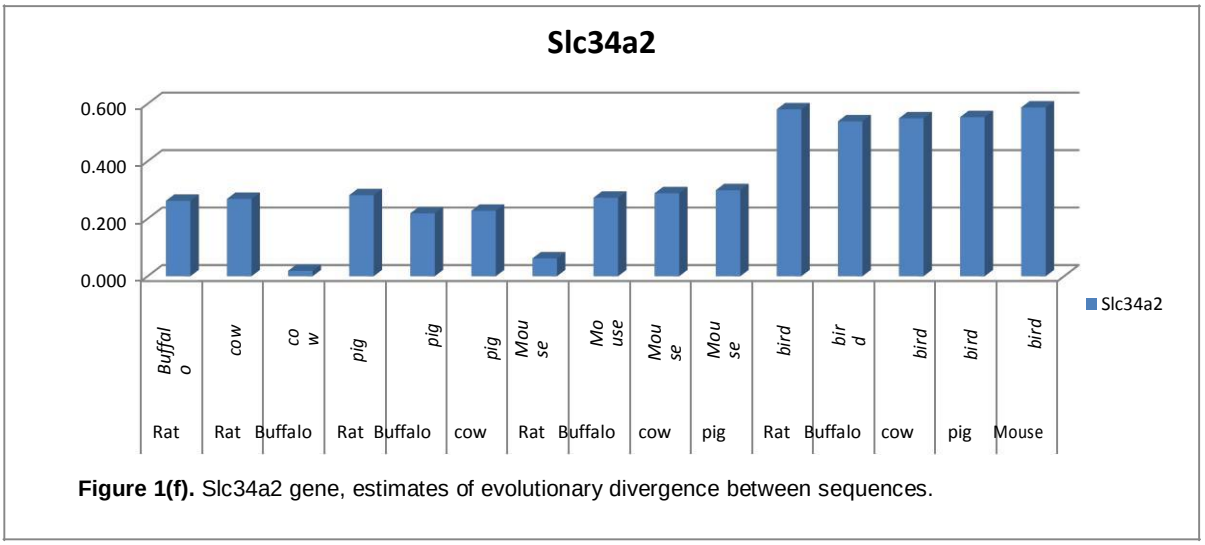

Figure 1(f). Slc34a2 gene, estimates of evolutionary divergence between sequences.

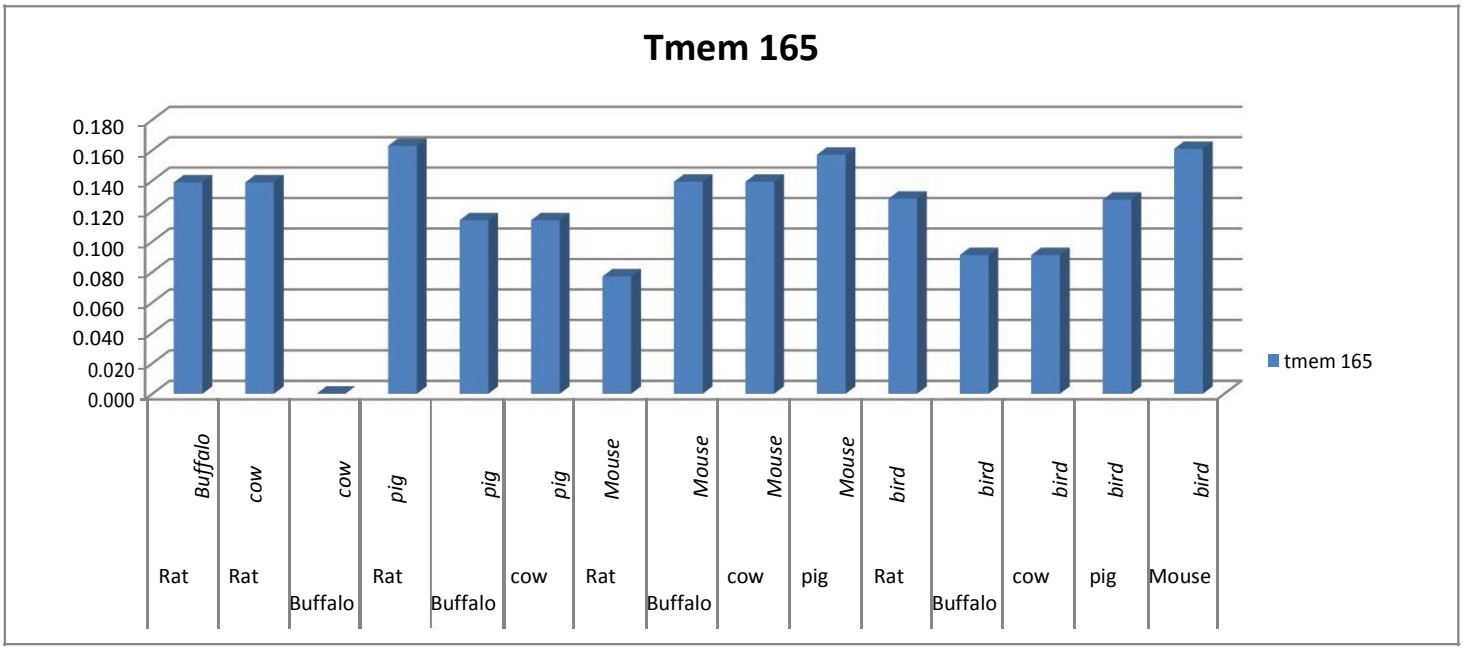

Figure 1(g). Tmem 165 gene, estimates of evolutionary divergence between sequences.

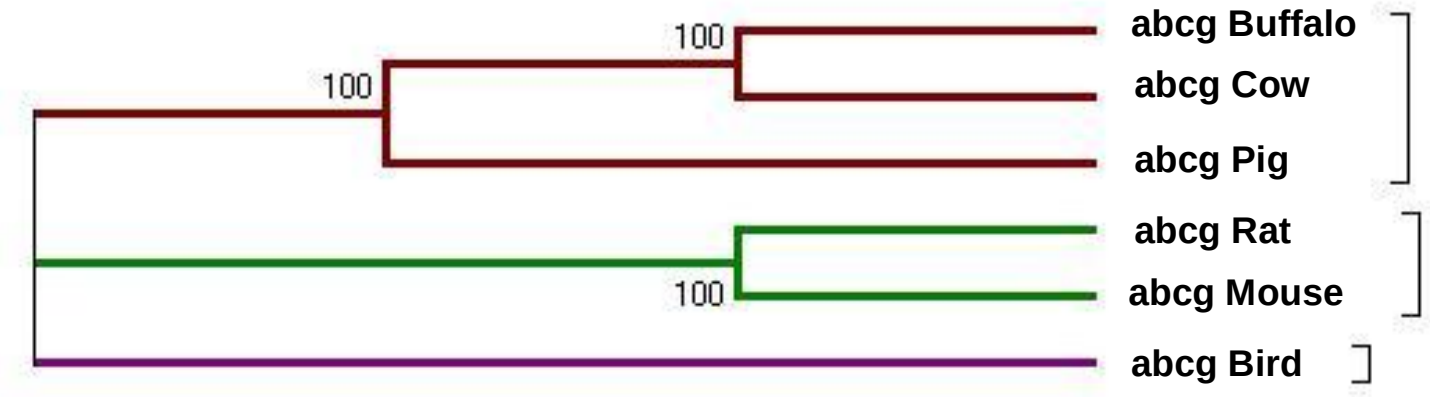

Figure 2(a). Molecular phylogenetic anaylsis of abcg2 gene in different organisms by maximum likelihood method.

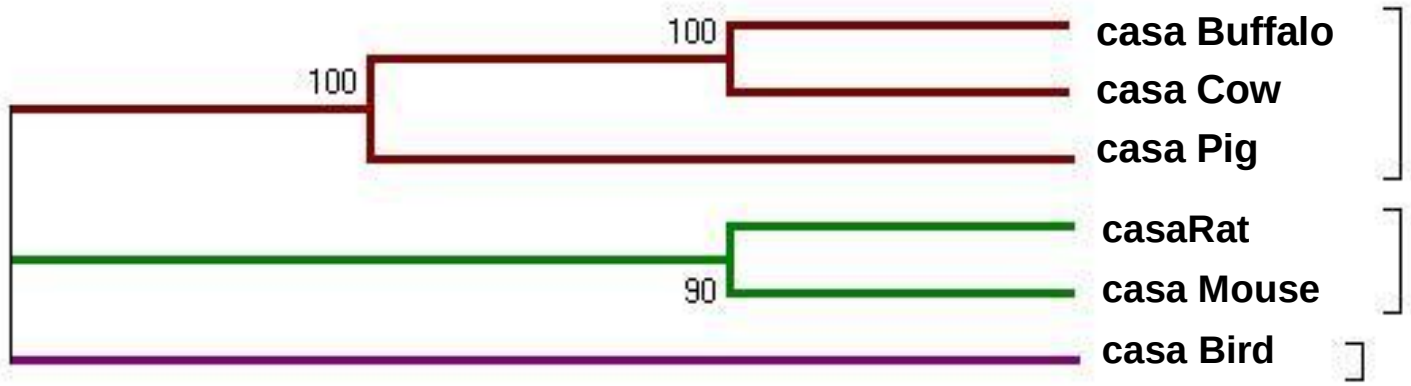

Figure 2(b). Molecular phylogenetic anaylsis of casa2 gene in different organisms by maximum likelihood method.

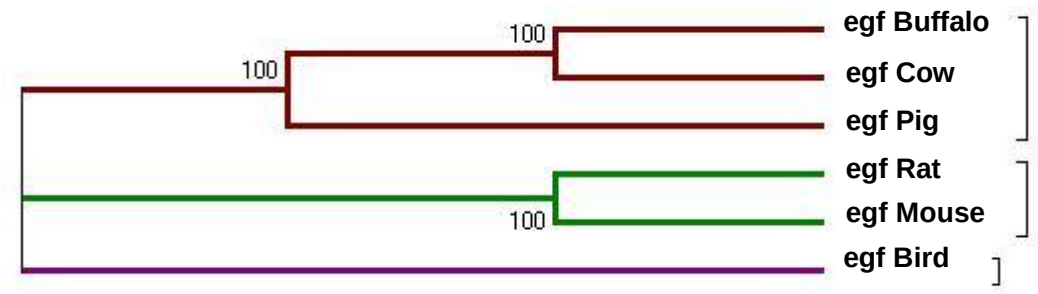

Figure 2(c). Molecular phylogenetic anaylsis of egf gene in different organisms by maximum likelihood method.

Figure 2(d). Molecular phylogenetic anaylsis of kappa casein gene in different organisms maximum likelihood method.

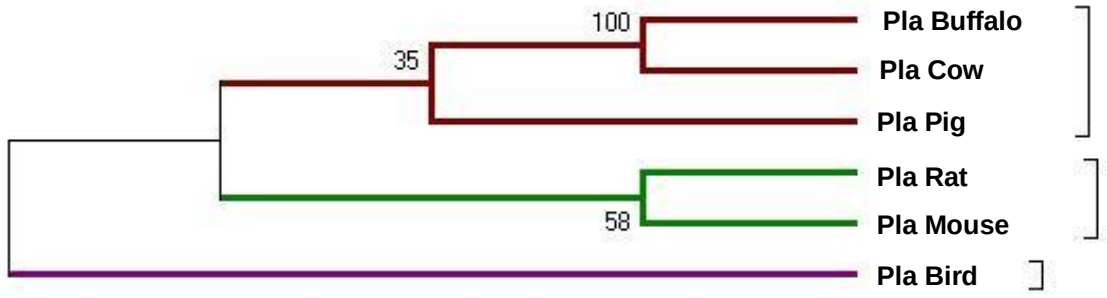

Figure 2(e). Molecular phylogenetic anaylsis of pla2g XIIA gene in different organisms by maximum likelihood method.

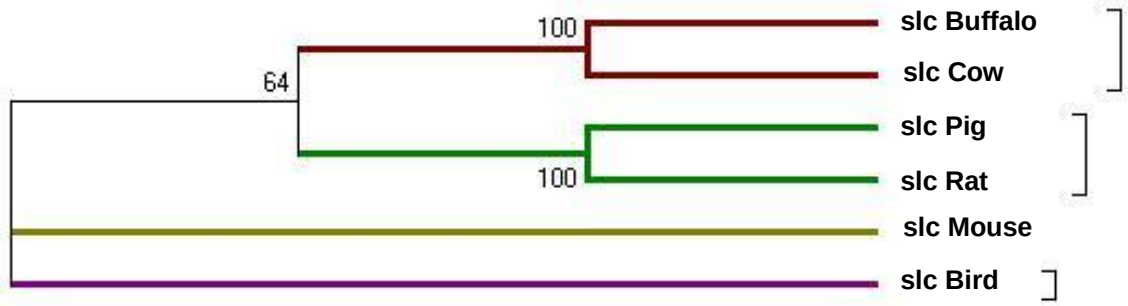

Figure 2(f). Molecular phylogenetic anaylsis of Slc gene in different organisms by maximum likelihood method.

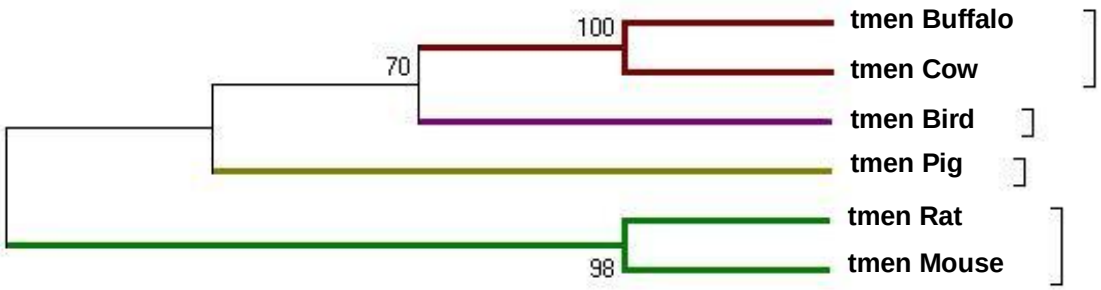

Figure 2(g). Molecular phylogenetic anaylsis of tmem gene in different organisms by maximum likelihood method.

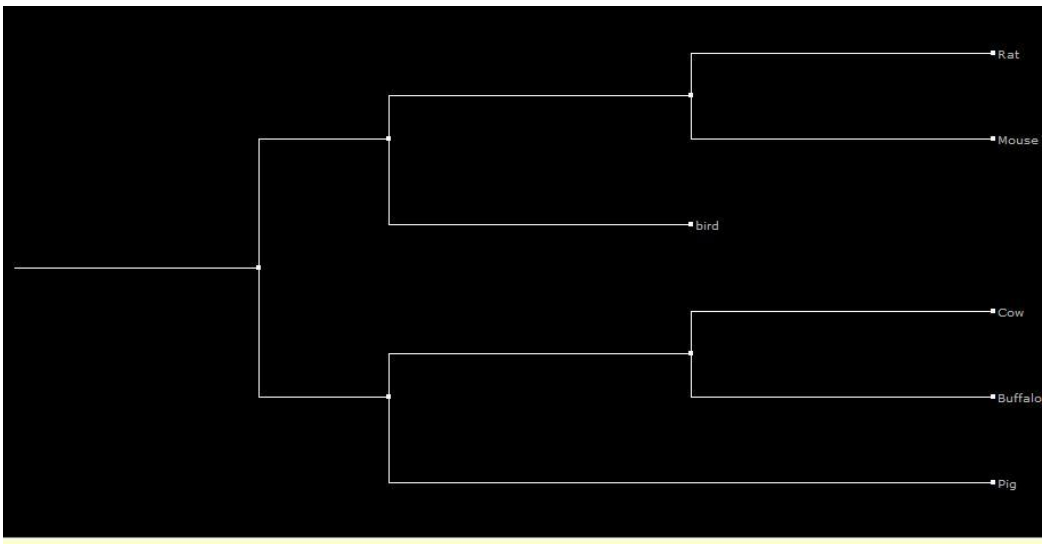

Figure 3. Species tree reconstructed with maximum likelihood method.

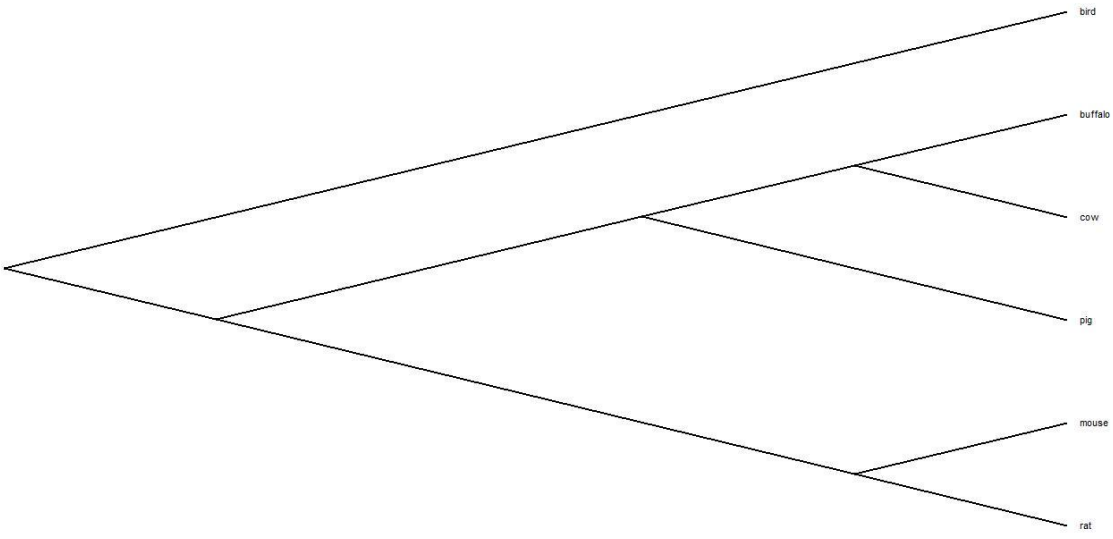

Figure 4. Species tree reconstructed with Bayesian method, A and B are the calibration points.

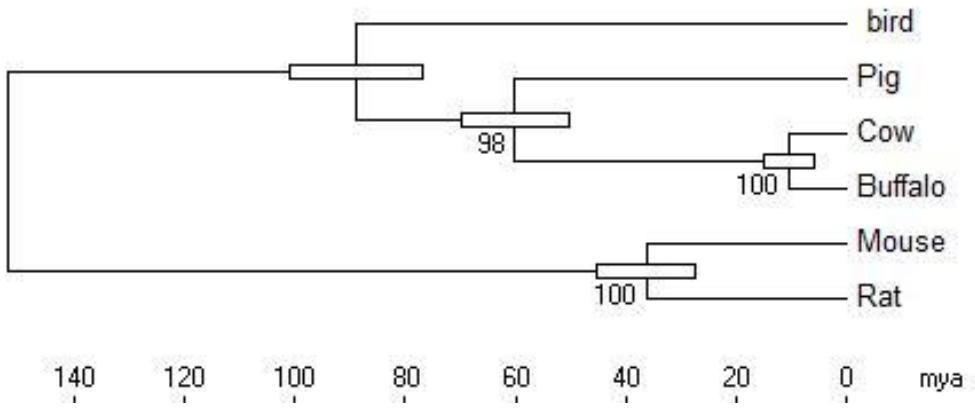

Figure 5. Species tree: Scale shows the divergence time of the species.

REFERENCES

Chenhong Li, Guillermo Ortí, Gong Zhang, Guoqing Lu (2007). A Practical Approach to Phylogenomics: The Phylogeny of Ray-Finned Fish (Actinopterygii) as a Case Study. BMC Evol. Biol., 7: 44 doi:10.1186/1471-2148-7-44.

Drummond AJ, Rambaut A (2007).BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol. 7:214doi:10.1186/1471-2148-7-214.

Gatesy J (2009). Whales and even –toed ungulates (Cetattiodactyla). The Time tree of Life. Oxford university press, pp. 511-515.

Goldammer T, Weikard R, Miziara MN, Brunner RM, Agarwala R, Schäffer AA, Womack JE, Amaral MEJ (2007). A radiation hybrid map of river buffalo (Bubalusbubalis) chromosome 7 and comparativemapping to the cattle and human genomes. Cytogenetic Genome research, 119: 235–241.

Guindon S, Gascuel O (2003). PhyML: A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Systematic Biol., 52(5): 696-704.

Kimura M (1980). A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol., 16: 111-120.

Murphy WJ, Eizirik E (2009).Placental mammals (Eutheria). The time tree of Life.Oxford University press, pp. 471-474.

Rozen S, Skaletsky H (2000). Primer3 on the WWW for general users and for biologist programmers. Methods Mol. Biol., 132: 365-86.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011). MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. Mol. Biol. Evol. (submitted).

Thompson JD, GHiggins D, Gibson TJ (1994).CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res., 22 (22): 4673-4680.