

Identification of transition bias in oxidized low density lipoprotein receptor 1 gene in buffalo

N. Shabir¹, Chetan V. Jawale¹, Naveed A. Chikan², C. D. Bhong¹, D. N. Rank³ and C. G. Joshi¹

1. Department of Animal Biotechnology, College of Veterinary Science and Animal Husbandry, Anand Agricultural University, Anand, Gujarat - 388110, India; 2. School of Bioscience and Technology, Vellore Institute of Technology University, Vellore, Tamil Nadu, India; 3. Department of Animal Genetics and Breeding, College of Veterinary Sciences and Animal Husbandry, Anand Agricultural University, Anand, Gujarat - 388110, India

Corresponding author: N. Shabir, email: drnadurose@gmail.com

Received: 21-01-2014, Revised: 06-02-2014, Accepted: 12-02-2014, Published online: 08-03-2014

doi: 10.14202/vetworld.2014.135-140

How to cite this article: Shabir N, Jawale CV, Chikan NA, Bhong CD, Rank, DN and Joshi CG (2014) Identification of transition bias in oxidized low density lipoprotein receptor 1 gene in buffalo, *Veterinary World* 7(3): 135-140.

Abstract

Background and aim: Though transition bias has been previously demonstrated in cattle, however, there has not been any study that has explored transition bias in buffalo nuclear genome. The aim of the present study was to evaluate the nucleotide substitution pattern in the Intron I of *Oxidised Low Density Lipoprotein Receptor 1 (OLR1)* gene in four breeds of Indian buffalo using 24 different nucleotide substitution models and evaluate their association with DNA methylation.

Materials and Methods: Transition/transversion bias (R) was estimated by 24 different nucleotide substitution models available in MEGA 5.0. The transition/transversion bias (R) was estimated under the Kimura 2-parameter model. Substitution patterns and the transitions/transversions rates (r) were then estimated by Tamura-Nei-I and Tamura-Nei-II models. The CpG Island search was done by using CpG Plot Island online Software available at European Bioinformatics Institute (EBI) website.

Results: The frequency of transition was found to be 3.5 times higher than that of the transversion mutation frequency. Out of 9 nucleotide substitutions, 7 transitions and 2 transversions were found. Among all the nucleotide substitutions, thymine to cytosine substitutions was observed to be very high. CpG Island search tool revealed that IntronI of OLR1 genes is a CpG rich region, thus prone to methylation.

Conclusions: Higher transition frequency was found in the intronI of *OLR1* gene, however due to the richness of methylated CpGs in the evaluated stretch of genome, the higher T → C transitions could likely be a result of frequent deaminations of the methylated cytosines into thymines during the evolution of four buffalo breeds.

Keywords: buffalo, oxidized low density lipoprotein receptor 1, transitions.

Introduction

Since transitions (T → C, A → G) have been observed to occur at higher frequencies than transversions (T → A, T → G, C → A, C → G) in most of the DNA sequences [1], the phenomenon is commonly termed as transition bias. The nature of nucleotide substitutions not only aid in the comprehension of species evolution at DNA-sequence level but also help in calculating the evolutionary distance between the species. The transition/transversion mutation rate biases are predicted to differ significantly among the genes [2]. One of the other important features of such substitutions is their association with the type and intensity of natural selection, the sustenance of high allelic diversity in Major Histocompatibility Complexes (MHC's) being a prime archetype [3]. The evolutionary studies primarily entail the study of genetic distances in mitochondrial DNA [4], however, differential genomic nucleotide substitutions and SNPs in population have also been demonstrated to help in understanding their evolution. The alterations in Intronic regions, intron gain/loss

events in particular, across the generations of a species have recently been identified as potential targets for evolutionary studies [5]. Recently transition bias was observed in the bovine genome where in, they evaluated 15,110 SNPs and concluded that 69.8% of the total SNPs turned out to be transitions [6]. Shabir et al and Jawale [7,8] reported nine SNPs in Intron I of *Oxidized Low Density Lipoprotein Receptor 1 (OLR1)* gene in four breeds of Indian buffalo.

The present study was aimed at exploring nucleotide substitution pattern in intron-1 of *OLR1* gene of buffalo, using a comparative analysis of 24 different nucleotide substitution models.

Materials and Methods

Sequence data: 52 *OLR1* (Intron I) sequences of four buffalo species viz. Surti, Jaffarabadi, Mehsana and Banni, previously deposited in NCBI (Accession numbers: GQ478023 to GQ478042 & GQ415420 to GQ415451) were aligned in MEGA 5.0. The pattern of nucleotide substitution, maximum likelihood estimate of transition/transversion bias and maximum likelihood fits using 24 different nucleotide substitutions models available in MEGA 5.0 were analyzed in *OLR1* (IntronI) in four breeds of buffalo viz. Jaffarabadi,

Copyright: The authors. This article is an open access article licensed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>) which permits unrestricted use, distribution and reproduction in any medium, provided the work is properly cited.

Table-1. SNPs in *OLR1* Intron I of Jaffarabadi, Surti, Mehsana and Banni buffalo

Nucleotide position	Base change*	Buffalo Breed	Nature of Mutation	NCBI SNP ID**
61	a>G	Banni	Transition	155458696
327	a>G	Mehsana	Transition	155458697
423	t>C	Surti	Transition	155458698
541	t>C	Mehsana	Transition	155458699
813	t>C	Banni	Transition	155458700
843	t>C	Jaffarabadi	Transition	155458701
866	t>A	Surti	Transversion	155458702
957	c>T	Banni	Transition	155458703
1128	c>G	Banni	Transversion	155458704

* Denotes Single Nucleotide Polymorphisms; Surti and Jaffarabadi [7], Mehsana and Banni [8]; ** National Centre for Biotechnology Information, USA Single Nucleotide Polymorphism ID available at NCBI SNP database.

Table-2. Maximum composite likelihood estimate of the pattern of nucleotide substitution using Tamura Nei models

	A	T	C	G
A	-	<i>2.59</i>	<i>1.1</i>	5.94
T	<i>2.3</i>	-	20.7	<i>1.39</i>
C	<i>2.3</i>	48.77	-	<i>1.39</i>
G	9.81	<i>2.59</i>	<i>1.1</i>	-

Each entry shows the probability of substitution from one base (row) to another base (column). Rates of different transitional substitutions are shown in bold and those of transversional substitutions are shown in italics.

Surti, Mehsana and Banni.

Nucleotide substitution pattern estimation: The transition/transversion bias (R) was primarily estimated under the Kimura 2-parameter model. Substitution patterns and the transitions/transversions rates (r) were then estimated by Tamura-Nei-I and Tamura-Nei-II models.

Further, the transition/transversion bias (R) was also estimated by 24 different nucleotide substitution models available in MEGA 5.0. Models with the lowest AICc scores (Akaike Information Criterion, corrected) were considered to describe the best substitution pattern. For each model, BIC value (Bayesian Information Criterion), Maximum Likelihood value (lnL), and the number of parameters (including 101 tree branch lengths) were also presented. The Transition/Transversion bias (R) was calculated by $R = [A * G * k1 + T * C * k2] / [(A + G) * (T + C)]$.

CpG island search: CpG islands were searched in the intron I of *OLR1* gene of buffalo to get an idea regarding the probability of CpG methylations which could lead higher mutations, cytosine to thymine, in particular due to the deamination of methyl cytosine. The CpG Island search was done by using CpG Plot Island online Software available at European Bioinformatics Institute (EBI) website (<http://www.ebi.ac.uk/Tools/emboss/cpgplot/>).

Results

In our earlier studies we reported three SNPs in Surti and Jaffarabadi Buffalo in *OLR1* gene of buffalo [7] and later six more SNPs (Table-1) were explored in the Intron I of *OLR1* gene in Mehsana and Banni buffalo [8]. Out of the total nine SNPs explored in four breeds of Indian buffalo in Intron I of *OLR1* gene, only two transversions were observed and the rest were transitions (Table-1). Four t>C, two a>G and one c>T transitions were present out of the seven transitions

observed. The frequency of t>C transition was high in the population studied and was present in all the four breeds of buffalo, however, the two transversions viz. t>A and c>G were respectively present in Banni and Surti breed of buffalo (Table-1). The frequency of transitions was 3.5 times higher than the frequency of transversions as evinced by the transition/transversion bias (R) estimated under Kimura 2 parameter model [8] which equaled to 3.50. The maximum Log likelihood for this computation was -1689.545. Substitution patterns and the Transitions/Transversions rates (r) estimated by Tamura-Nei-I [9] and Tamura-Nei-II model [10] are enlisted in Table-2 which demonstrates the significantly higher values of transitions. The AICc score, as calculated using 24 models, for HKY model was the lowest (3459.1) among all the 24 models used and thus displayed the best Nucleotide Substitution pattern. The transition/transversion bias calculated by the HKY model was 3.5021 (Table-3) which is nearly equivalent to the values calculated by the Kimura 2 parameter model. However, the transition bias computed by Jukes Cantor (JC) and related models displayed the lowest value of transition/transversion bias (0.5) and a frequency of 0.25 for each nucleotide. Further, the nucleotide frequencies of the 52 *OLR1* (IntronI) nucleotide sequences ranged between 0.25-0.3313 (A), 0.25-0.351 (T/U), 0.1687-0.25 (C), and 0.25-0.1885 (G) as calculated by 25 different nucleotide substitution models (Table-4). The nucleotide substitution rates calculated by 24 different nucleotide substitution models are displayed in Table-5 which shows that the Thymine to Cytosine substitutions occurred at the highest frequency among all the other nucleotide substitutions. The CpG island search did not reveal any putative CpG islands in Intron I of *OLR1* gene of buffalo (Figure-1). Further, the Observed/Expected ratio of CpG as calculated by the EBI software turned out to be greater than 0.60 and the Percent C + Percent G was observed to be greater than 50.

Table-3. AICc scores (Akaike Information Criterion, corrected), BIC value (Bayesian Information Criterion), Maximum Likelihood value (*lnL*) and Transition/ Transversion bias (*R*) using 24 different nucleotide substitution models

Model*	Parameters	AICc	BIC	lnL	R**
HKY	105	3459.1	4404.7	1624.4	3.5021
TN93	106	3459.4	4414	1623.5	3.5029
HKY+I	106	3461.1	4415.7	1624.4	3.5021
HKY+G	106	3461.1	4415.7	1624.4	3.5021
TN93+I	107	3461.4	4425	1623.5	3.5029
TN93+G	107	3461.4	4425	1623.5	3.5029
GTR	109	3462.2	4443.8	1621.9	3.5027
T92	103	3463	4390.6	1628.3	3.5021
HKY+G+I	107	3463.1	4426.7	1624.4	3.5021
TN93+G+I	108	3463.4	4436	1623.5	3.5029
GTR+I	110	3464.2	4454.8	1621.9	3.5026
GTR+G	110	3464.2	4454.8	1621.9	3.5026
T92+I	104	3465	4401.6	1628.3	3.5021
T92+G	104	3465	4401.6	1628.3	3.5022
GTR+G+I	111	3466.2	4465.8	1621.9	3.5026
T92+G+I	105	3467	4412.6	1628.3	3.5022
K2	102	3583.4	4502	1689.5	3.5023
K2+I	103	3585.4	4513	1689.5	3.5023
K2+G	103	3585.4	4513	1689.5	3.5023
K2+G+I	104	3587.5	4524	1689.5	3.5023
JC	105	3596.9	4542.5	1693.3	0.5
JC+I	106	3598.9	4553.5	1693.3	0.5
JC+G	106	3598.9	4553.5	1693.3	0.5
JC+G+I	107	3600.9	4564.5	1693.3	0.5

** GTR: General Time Reversible; HKY: Hasegawa-Kishino-Yano; TN93: Tamura-Nei; T92: Tamura 3-parameter; K2: Kimura 2-parameter; JC: Jukes-Cantor. * Assumed or estimated values of transition/transversion bias (*R*) are shown for each model

Table-4. The frequency of four different nucleotides in Intron I of *OLR1* gene in Buffalo.

Model*	Freq* A	Freq T	Freq C	Freq G
HKY	0.3115	0.351	0.149	0.1885
TN93	0.3115	0.351	0.149	0.1885
HKY+I	0.3115	0.351	0.149	0.1885
HKY+G	0.3115	0.351	0.149	0.1885
TN93+I	0.3115	0.351	0.149	0.1885
TN93+G	0.3115	0.351	0.149	0.1885
GTR	0.3115	0.351	0.149	0.1885
T92	0.3313	0.3313	0.1687	0.1687
HKY+G+I	0.3115	0.351	0.149	0.1885
TN93+G+I	0.3115	0.351	0.149	0.1885
GTR+I	0.3115	0.351	0.149	0.1885
GTR+G	0.3115	0.351	0.149	0.1885
T92+I	0.3313	0.3313	0.1687	0.1687
T92+G	0.3313	0.3313	0.1687	0.1687
GTR+G+I	0.3115	0.351	0.149	0.1885
T92+G+I	0.3313	0.3313	0.1687	0.1687
K2	0.25	0.25	0.25	0.25
K2+I	0.25	0.25	0.25	0.25
K2+G	0.25	0.25	0.25	0.25
K2+G+I	0.25	0.25	0.25	0.25
JC	0.25	0.25	0.25	0.25
JC+I	0.25	0.25	0.25	0.25
JC+G	0.25	0.25	0.25	0.25
JC+G+I	0.25	0.25	0.25	0.25

* Frequency; ** GTR: General Time Reversible; HKY: Hasegawa-Kishino-Yano; TN93: Tamura-Nei; T92: Tamura 3-parameter; K2: Kimura 2-parameter; JC: Jukes-Cantor.

Discussion

In our previous studies we have demonstrated the phylogenetic analysis and thus analyzed the genetic distance between the two Jaffarabadi and Surti buffalo based on intronic region of *OLR1* gene [11]. However, this study focuses on the nucleotide substitution pattern in intronI of *OLR1* gene in four breeds of buffalo. In previous studies in the nuclear DNA of mammals, the frequency of transition mutations appeared to be twice higher than transversions which is well evinced by the nucleotide substitution patterns in synonymous and non-coding SNPs in humans [12], and in SNPs in mice [13] however, the frequency of transitions was three

times higher than that of transversions in our study entailing the Intron I of *OLR1* gene in Buffalo. It is also in agreement with a previous study in bovines where they evaluated 15,110 SNPs and found that 69.8% of the total SNPs were transitions [6]. The lower values of transition/transversion bias by Jukes cantor and related models is on account of the fact that Jukes cantor model assumes the equality of nucleotide substitution among sites and equal nucleotide frequencies [14].

The nucleotide substitution rates calculated by 24 different nucleotide substitution models show that t>C nucleotide substitutions were present at the highest frequency. The earlier studies on CpG methylation

Table-5. Nucleotide Substitution rates calculated by various models in Intron I of *OLR1* of Buffalo

Model*	A=>T	A=>C	A=>G	T=>A	T=>C	T=>G	C=>A	C=>T	C=>G	G=>A	G=>T	G=>C
HKY	0.04	0.02	0.15	0.03	0.12	0.02	0.03	0.28	0.02	0.25	0.04	0.02
TN93	0.03	0.01	0.08	0.03	0.18	0.02	0.03	0.42	0.02	0.13	0.03	0.01
HKY+I	0.04	0.02	0.15	0.03	0.12	0.02	0.03	0.28	0.02	0.25	0.04	0.02
HKY+G	0.04	0.02	0.15	0.03	0.12	0.02	0.03	0.28	0.02	0.25	0.04	0.02
TN93+I	0.03	0.01	0.08	0.03	0.18	0.02	0.03	0.42	0.02	0.13	0.03	0.01
TN93+G	0.03	0.01	0.08	0.03	0.18	0.02	0.03	0.42	0.02	0.13	0.03	0.01
GTR	0.04	0	0.08	0.03	0.17	0	0	0.4	0.08	0.13	0	0.06
T92	0.03	0.02	0.13	0.03	0.13	0.02	0.03	0.26	0.02	0.26	0.03	0.02
HKY+G+I	0.04	0.02	0.15	0.03	0.12	0.02	0.03	0.28	0.02	0.25	0.04	0.02
TN93+G+I	0.03	0.01	0.08	0.03	0.18	0.02	0.03	0.42	0.02	0.13	0.03	0.01
GTR+I	0.04	0	0.08	0.03	0.17	0	0	0.4	0.08	0.13	0	0.06
GTR+G	0.04	0	0.08	0.03	0.17	0	0	0.4	0.08	0.13	0	0.06
T92+I	0.03	0.02	0.13	0.03	0.13	0.02	0.03	0.26	0.02	0.26	0.03	0.02
T92+G	0.03	0.02	0.13	0.03	0.13	0.02	0.03	0.26	0.02	0.26	0.03	0.02
GTR+G+I	0.04	0	0.08	0.03	0.17	0	0	0.4	0.08	0.13	0	0.06
T92+G+I	0.03	0.02	0.13	0.03	0.13	0.02	0.03	0.26	0.02	0.26	0.03	0.02
K2	0.03	0.03	0.19	0.03	0.19	0.03	0.03	0.19	0.03	0.19	0.03	0.03
K2+I	0.03	0.03	0.19	0.03	0.19	0.03	0.03	0.19	0.03	0.19	0.03	0.03
K2+G	0.03	0.03	0.19	0.03	0.19	0.03	0.03	0.19	0.03	0.19	0.03	0.03
K2+G+I	0.03	0.03	0.19	0.03	0.19	0.03	0.03	0.19	0.03	0.19	0.03	0.03
JC	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08
JC+I	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08
JC+G	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08
JC+G+I	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08

** GTR: General Time Reversible; HKY: Hasegawa-Kishino-Yano; TN93: Tamura-Nei; T92: Tamura 3-parameter; K2: Kimura 2-parameter; JC: Jukes-Cantor.

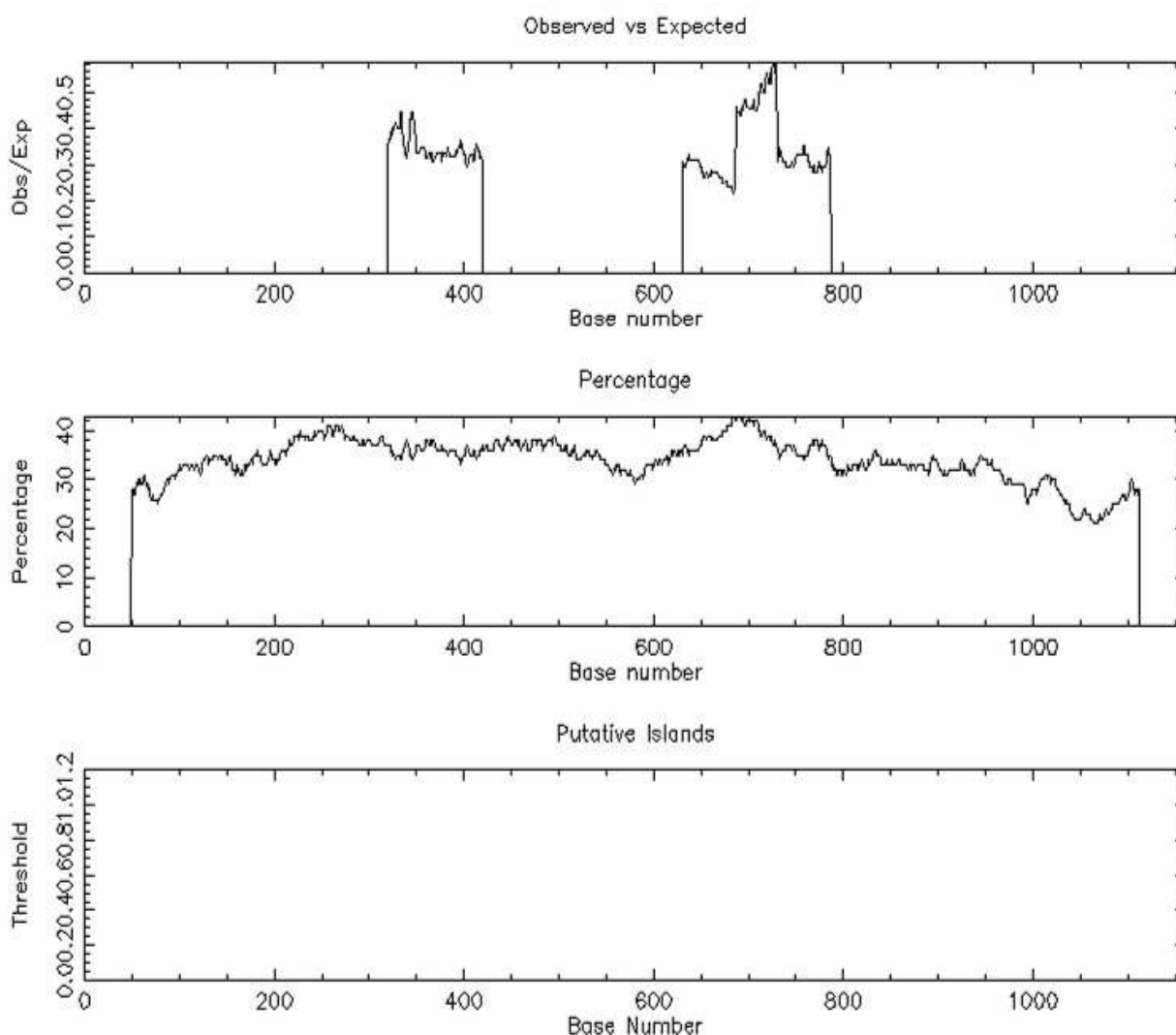


Figure-1. Shows the observed/expected ratio of CpG, the percentage of CpG throughout the length of Intron I of *OLR1*, and the absence of putative CpG islands in Intron I of *OLR1*.

have demonstrated that methylations usually occur in the CpG rich regions of the genome, usually promoters/Introns, and the methylated cytosines after deamination get mutated into thymines [15]. No putative CpG islands were identified in the Intron I of *OLRI* gene of buffalo (Figure-1) which means that this region of *OLRI* is a CpG rich region and has higher possibilities of being methylated.

Previous study [16] have demonstrated that cytosine of most CpGs are methylated which leads to frequent deaminations of cytosine leading to mismatches. Recently, a study revealed high mutation rates in low to moderately methylated CpG sites at different scales in human embryonic stem cells [17]. Though t>C transition were highly frequent nucleotide substitution in our previous studies, but was observed to occur at a frequency of 0.25 in the studied population which demonstrated that 75% of the population possessed the “Thymine” nucleotide at those loci. Since, it is a CpG rich region and prone to methylation, so the methylated cytosines might have likely deaminated over time and mutated to thymine which might have later got fixed in the population, however, the unmethylated cytosines at these loci must have remained un-mutated, though at lesser frequencies, as evinced by their percentage in the population in our previous studies. The observations in this study suggest that the CpG rich regions in intron I of *OLRI* gene are more prone to mutations which is in agreement with a study where in it was revealed a significant correlation between the extent of germ line methylation and the substitution rate at human CpG sites [18].

However, it is worth to note that apart from the hyper mutability of the methylated CpGs, a study have revealed that the physical characteristics of DNA is a significant factor resulting in local sequence bias in human mutation [19]. Even though this study focuses on the single nucleotide polymorphisms in the intronic region of *OLRI* gene, it would be interesting to analyze the synonymous and non-synonymous changes in the coding regions of candidate genes involved in milk production, since the milk candidate genes are believed to follow polygenic inheritance. However, recently a study indicated that the human nervous system is more likely to be affected by monogenic diseases than the immune system [20] which should encourage more research at the molecular level to explore the monogenic/polygenic mutation inheritance of candidate genes involved in milk production.

Conclusions

To summarize the findings, the transition mutation frequency in Intron I of *OLRI* gene in four breeds of Indian buffaloes was found to be 3.5 times higher than that of the transversion mutation frequency. The higher T → C transitions might be a result of frequent deaminations of the methylated cytosines into thymines during the evolution of four buffalo breeds. Though, this study gives an insight into the probable

nucleotide substitution patterns in four breeds of buffalo, however, larger stretches of DNA, introns and exons, need to be evaluated to derive a generalized conclusion on the nucleotide substitution patterns in buffalo genome.

Authors' contributions

The present study is a review and analysis of master's research work of NS and CVJ. NS and CVJ did the review and drafted the manuscript. NAC and CDB assisted in data analysis. DNR and CGJ supervised the work and critically analyzed the data. All authors read and approved the final manuscript.

Acknowledgements

The authors are thankful to Indian Council of Agricultural Research (ICAR) for providing the necessary funding. Authors of thankful to Head, Dept. of Animal Genetics and Breeding, Head, Dept. of Animal Biotechnology and Director of Research, Anand Agricultural University, Anand for providing the necessary facilities for this work.

Competing interests

The authors declare that they have no competing interests.

References

1. Brown, W.M., Prager, E.M., Wang, A. and Wilson, A.C. (1982) Mitochondrial DNA sequences of primates, tempo and mode of evolution. *J. Mol. Evol.*, 18: 225–239.
2. Yang, Z. and Yoder, A.D. (1999) Estimation of the transition/transversion rate bias and species sampling. *J. Mol. Evol.*, 48: 274–283.
3. Hughes, A.L. (2002) Natural selection and the diversification of vertebrate immune effectors. *Immunol. Rev.*, 190: 161–168.
4. Pakendorf, B. and Stoneking, M. (2005) Mitochondrial DNA and Human Evolution. *Annu. Rev. Genomics. Hum. Genet.*, 6: 165–83.
5. Garcia-Espana, A., Mares, R., Sun, T.T. and Desalle, R. (2009) Intron evolution: testing hypotheses of intron evolution using the phylogenomics of tetraspanins. *PLoS One*, 4:e4680.
6. Jiang, Z., Wu, X.L., Zhang, M., Michal, J.J. and Wright, R.W. Jr. (2008) The complementary neighborhood patterns and methylation-to-mutation likelihood structures of 15,110 single-nucleotide polymorphisms in the bovine genome. *Genetics* 180: 639–47.
7. Shabir, N., Jawale, C.V., Bhong, C.D., Kurkute, A.S., Patel, T.B., Rank, D.N. and Joshi, C.G. (2011a) SNP exploration in the Oxidised Low Density Lipoprotein Receptor 1 (OLR1) gene in Bubalus Bubalis. *Buffalo Bulletin.*, 30: 267–271.
8. Jawale, C.V. (2009) Identification of SNPs in OLR1 gene of Mehsana and Banni buffalo by cloning and sequencing. An MVSc dissertation submitted to AAU, Anand, India.
8. 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.
9. Tamura, K., Nei, M., Kumar, S. (2004) Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proc. Natl. Acad. Sci. USA*, 101: 11030–11035.
10. Tamura, K., Dudley, J., Nei, M. & Kumar, S. (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol. Biol. Evol.*, 24:1596–1599.
11. Shabir, N., Jawale, C.V., Bhong, C.D., Naikoo, N., Rank,

- D.N., Joshi, C.G. (2011b). Haplotype and phylogenetic analysis of OLR1 (Intron I) gene in Jaffarabadi and Surti buffalo. *Vet. World.*, 4:396-398.
12. Cargill, M., Altshuler, D., Ireland, J., Sklar, P., Ardlie, K., Patil, N., Shaw, N., Lane, C.R., Lim, E.P., Kalyanaraman, N., Nemesh, J., Ziaugra, L., Friedland, L., Rolfe, A., Warrington, J., Lipshutz, R., Daley, G.Q., Lander, E.S. (1999) Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nat. Genet.*, 22: 231-238.
 13. Lindblad-Toh, K., Winchester, E., Daly, M.J., Wang, D.G., Hirschhorn, J.N., Lavoie, J.P., Ardlie, K., Reich, D.E., Robinson, E., Sklar, P., Shah, N., Thomas, D., Fan, J.B., Gingeras, T., Warrington, J., Patil, N., Hudson, T.J., Lander, E.S. (2000) Large-scale discovery and genotyping of single-nucleotide polymorphisms in the mouse. *Nat. Genet.*, 24: 381-386.
 14. Jukes, T.H. and Cantor, C.R. (1969) Evolution of Protein Molecules. New York, NY, USA: Academic Press.
 15. Zemojtel, T., Kielbasa, S.M., Arndt, P.F., Behrens, S., Bourque, G. and Vingron, M. (2011) CpG deamination creates transcription factor-binding sites with high efficiency. *Genome. Biol. Evol.*, 3:1304-11.
 16. Miranda, T. B. and Jones, P. A. (2007) DNA methylation: The nuts and bolts of repression. *J. Cell. Physiol.* 213: 384-390.
 17. Xia, J., Han, L. and Zhao, Z. (2012) Investigating the relationship of DNA methylation with mutation rate and allele frequency in the human genome. *BMC Genom.* 13 (Suppl) 8:S7. doi: 10.1186/1471-2164-13-S8-S7.
 18. Mugal, C.F. and Ellegren, H. (2011) Substitution rate variation at human CpG sites correlates with non-CpG divergence, methylation level and GC content. *Genome Biol.* 12: R58. doi: 10.1186/gb-2011-12-6-r58.
 19. Nevarez, P.A., DeBoever, C.M., Freeland, B.J., Quitt, M.A. and Bush, E.C. (2010) Context dependent substitution biases vary within the human genome. *BMC Bioinformatics.* 11: 462.
 20. Freudenberg, J., Gregersen, P.K. and Freudenberg-Hua, Y. (2012) A simple method for analyzing exome sequencing data shows distinct levels of nonsynonymous variation for human immune and nervous system genes. *PLoS One.* 7: e38087.
