COMPARATIVE SEQUENCE ANALYSIS OF TLR2 TLR4 AND TLR9 GENES AMONG SELECTED VERTEBRATES-A META-ANALYSIS

AVISHEK DAS[a], POKHRAJ GUHA[b], TAPAS KUMAR CHAUDHURI[b*]

[a,b]Cellular Immunology Laboratory, Department of Zoology, University of North Bengal, Raja Rammohunpur, Siliguri, West Bengal 734013, India
Email: dr_tkc_nbu@rediffmail.com

ABSTRACT

Objective: Toll-like receptors are the pattern recognition receptors that recognize a diverse set of conserved pathogens. The receptors are also constantly under selection pressure because of the host antigen modifications. The present study focuses on how selection and mutation have modified the TLRs throughout the evolution in selected groups.

Methods: We have selected the sequences of TLR2, 4 and 9 among Hominid group, *Homo sapiens, Bubalus bubalis* and *Danio rerio* in our analysis and analyzed different parameters like relative synonymous codon usage (RSCU), sequence divergence, amino acid composition and estimated evolutionary selection forces using Tajima's test.

Results: The phylogenetic assessment proved that positive selection influences TLR2 and TLR4, but neutral selection/balancing selection occurred in TLR9 which concluded from the Tajima's test. Synonymous codon usage described the selection of leucine and arginine in all the sequences which describe the structural similarities of TLRs. Values of nucleotide pairs and disparity index proved the close relationship of Hominid and Human between TLR2 and TLR4 and TLR9 where the distant relationship was found with Danio. It can be hypothesized that some of the codons may be best selected for binding with the antigens and it was selected in the genome and some were eliminated due to selection pressure.

Conclusion: The present study aimed to substantiate the closeness of TLR2 and TLR4 due to their functional similarity but distant with TLR9 because of the different antigens they recognized in the endosome.

Keywords: Pattern recognition receptors, RSCU, Tajima's test, Disparity index

INTRODUCTION

Toll-like receptors are one of the most interesting players of innate immune response in different vertebrate groups ranging from teleost to mammals. Ten different types of TLRs are found till date in human and other primate species. These receptors are present both on cellular surfaces and compartments [1] where they can recognize conserved molecular pattern molecules known as pathogen-associated molecular patterns (PAMPs). These genes are distributed throughout the genome in different chromosomes. Recent reports have shown that TLRs play a significant role in the pathogenesis of different infectious diseases including HIV which is highly prevalent in different populations of the World. [2-5].

Furthermore, the significant role of TLRs has also been reported elsewhere [6, 7]. It is already established that host antigen modification permits the innate immune related genes to modify themselves [8]. As TLR genes lie directly at the host-environment interface [9], therefore co-evolutionary forces always impose positive selection on the TLR genes, which results in sequence variability among the members of the family. Thus, sequence analysis becomes the primary tool to identify parts of the receptor molecule which are largely modified amongst different species due to evolutionary forces.

The rapid evolution of pathogens resulted in quick modification of the selection pressure giving an opportunity for adaptive evolution [10, 11]. Therefore, the immune-related genes have undergone adaptive radiation during the evolutionary process according to the antigens present in a particular environment. In this study, we have selected six different animal species for comparison of TLR sequences namely, *Homo sapiens* (Human), *Pan troglodytes* (chimpanzee), *Pongo abelii* (Sumatran orungotan), *Gorilla gorilla* (western gorilla), *Bubalus bubalis* (water buffalo) and *Danio rerio*

(zebrafish) respectively. The TLR genes selected for the analyses are TLR2, TLR4 and TLR9 respectively.

MATERIALS AND METHODS

Statistical analyses were performed using Mega software ver. 6 (MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0), and R statistical software (ver-3.3.1) [12], Kyplot (ver-2.0) and MS-Excel. All concerned sequences were downloaded from the NCBI database (http://www. ncbi. nlm. nih. gov/) and aligned using Clustal Omega server (http://www. ebi. ac. uk/Tools/msa/clustalo/). Neighbor-joining tree was constructed among six animals for three different TLRs. The frequency of the 59 codons (excluding the single synonymous codons AUG [Met] and UGG [Trp] and the three termination codons) was calculated for TLR2, TLR4, and TLR9 genes and pertaining relative synonymous codon usage (RSCU) values were estimated using same Mega software (ver. 6). Nucleotide compositions, frequencies and the amino acid compositions for three different TLRs were also calculated. The distance data were generated for different animals for calculating distant relationships among the animals. Tajima's test of neutrality was also calculated for the three different TLRs to decipher how selection pressure acts on DNA sequences. Disparity indices were also calculated from the observed difference in substitution patterns for a pair of sequences. It works by comparing the nucleotide (or amino acid) frequencies in a given pair of sequences and using the number of observed differences between sequences. Heat-map for three different TLRs among six animals was constructed to see the relationship among the six different animals.

RESULTS

We have constructed the neighbor-joining tree for six different species and found that in the case of TLR2 and TLR4 the hominids and Homo sapiens have the common branching

pattern. Similarly, *Gorilla gorilla* and *Homo sapiens* originated from the common branch and *Pan troglodytes* and *Pongo abelii* shared the common stock. In the case of TLR2, it was seen that Bubalus bubalis also diverged from the primates group and *Danio rerio* became an outlier (fig. 1). In the case of TLR4, we found a different tree where *Homo sapiens* and *Pan troglodytes* cluster in a single line of descent and *Gorilla gorilla* diverged away (fig. 2). When we studied the TLR9 we found that *Pan troglodytes* and Homo sapiens cluster in a single line and *Gorilla gorilla* branched separately. On the other hand, Bubalus and Pongo segregate from the single line whereas *Danio rerio* branched out from the primates group (fig. 3).

Heat map generated for three different TLRs showed the shared genetic relationship among the animals and with the out-group members (fig. 4). The heat map color band ranges from red to dark green indicating lowest to the highest rate of evolutionary relationship.
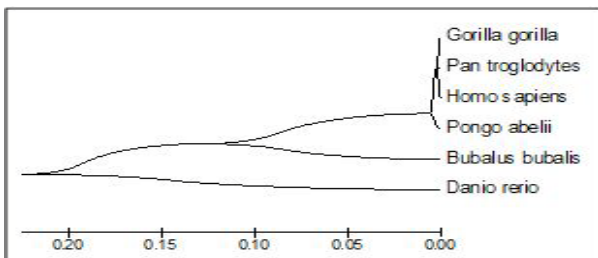

Fig. 2: Neighbor-joining tree was constructed to see the close relationship among six animals in case of TLR4
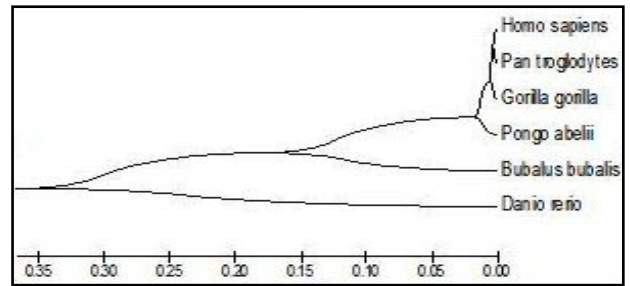

Fig. 1: Neighbor-joining tree was constructed to see the close relationship among six animals in case of TLR2
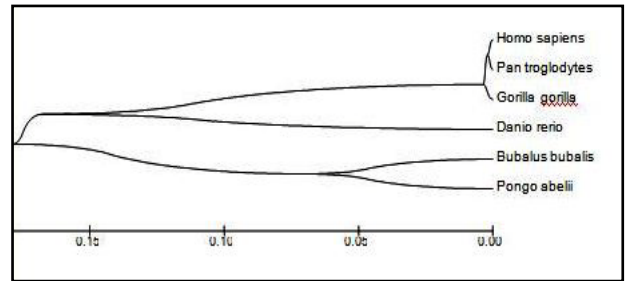

Fig. 3: Neighbor-joining tree was constructed to see the close relationship among six animals for six different animals in case of TLR9
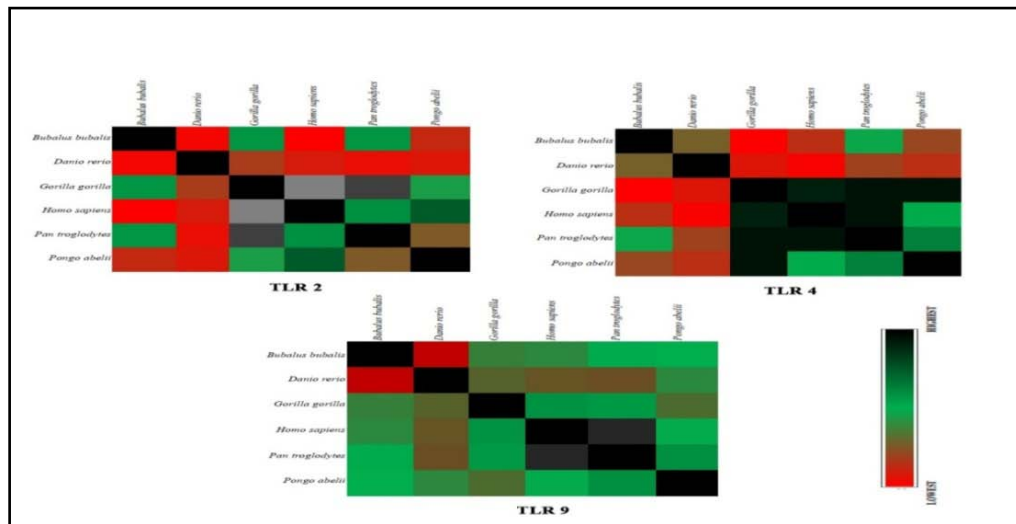

Fig. 4: Heatmap was constructed using R statistical software to show the genetic divergence among six animals

The most abundant codon was AAA (code for lysine amino acid) (count-265.7) for TLR2 and the RSCU value estimated 1.33 followed by the UUU (code for phenylalanine) (count-247.3) (RSCU = 1.28). The highest frequency of RSCU for TLR2 was found to be 2.91 and 1.88 that codes for arginine (AGA and AGG). The highest RSCU value for TLR4 was UUU (coding for phenylalanine) (count = 199.2) and followed by AAA (coding for lysine amino acid). The highest frequency of RSCU was calculated 3.17 which codes for AGA and 1.86 for AGG and codes for arginine. In the case of TLR9, the RSCU value was 2.24 (coding for arginine) (table 1).

Table 1: Relative synonymous codon usage is given in parentheses following the codon frequency of three TLRs

Codons	TLR2	TLR4	TLR9
UUU(F)	1.28	1.29	0.93
UUC(F)	0.72	0.71	1.07
UUA(L)	1.24	1.3	0.4
UUG(L)	1.23	1.12	0.85
CUU(L)	0.95	1.2	1.12
CUC(L)	0.78	0.79	1.27

CUA(L)	0.69	0.69	0.46
CUG(L)	1.11	0.9	1.9
AUU(I)	1.3	1.29	1.22
AUC(I)	0.64	0.65	1
AUA(I)	1.06	1.05	0.78
AUG(M)	1	1	1
GUU(V)	1.22	1.18	0.83
GUC(V)	0.71	0.65	1.19
GUA(V)	0.82	0.97	0.49
GUG(V)	1.26	1.2	1.49
UCU(S)	1.58	1.52	1.1
UCC(S)	1	0.96	1.42
UCA(S)	1.23	1.58	1
UCG(S)	0.17	0.14	0.2
CCU(P)	1.4	1.59	1.23
CCC(P)	0.94	0.87	1.05
CCA(P)	1.41	1.39	1.39
CCG(P)	0.25	0.15	0.33
ACU(T)	1.44	1.45	1.06
ACC(T)	0.77	0.79	1.21
ACA(T)	1.54	1.57	1.34
ACG(T)	0.25	0.18	0.38
GCU(A)	1.4	1.56	1.11
GCC(A)	0.9	0.85	1.27
GCA(A)	1.45	1.43	1.14
GCG(A)	0.24	0.17	0.49
UAU(Y)	1.34	1.29	1.05
UAC(Y)	0.66	0.71	0.95
UAA	1.19	1.26	0.63
UAG	0.67	0.64	0.6
CAU(H)	1.2	1.26	0.86
CAC(H)	0.8	0.74	1.14
CAA(Q)	0.97	1.02	0.67
CAG(Q)	1.03	0.98	1.33
AAU(N)	1.29	1.27	1.14
AAC(N)	0.71	0.73	0.86
AAA(K)	1.33	1.19	0.78
AAG(K)	0.67	0.81	1.22
GAU(D)	1.34	1.14	0.94
GAC(D)	0.66	0.86	1.06
GAA(E)	1.22	1.22	0.69
GAG(E)	0.78	0.78	1.31
UGU(C)	1.18	1.22	0.78
UGC(C)	0.82	0.78	1.22
UGA	1.14	1.1	**1.77**
UGG(W)	1	1	1
CGU(R)	0.25	0.3	0.52
CGC(R)	0.3	0.2	0.53
CGA(R)	0.27	0.25	0.48
CGG(R)	0.39	0.22	0.74
AGU(S)	1.09	1.05	0.86
AGC(S)	0.92	0.73	1.43
AGA(R)	2.91	3.17	1.5
AGG(R)	1.88	1.86	2.24
GGU(G)	0.89	0.84	0.74
GGC(G)	0.74	0.8	1.08
GGA(G)	1.25	1.47	0.96
GGG(G)	1.12	0.89	1.22

When we compared the amino acid compositions for the three different TLRs in six different animals, we observed that leucine was found to display the highest frequency in TLRs (fig 5) followed by Serine. The rarely occurring amino acids were tryptophan, tyrosine, and methionine in three different TLR codons (table 2).

Table 2: Average amino acid composition for three different TLRs (All frequencies are given in percentage)

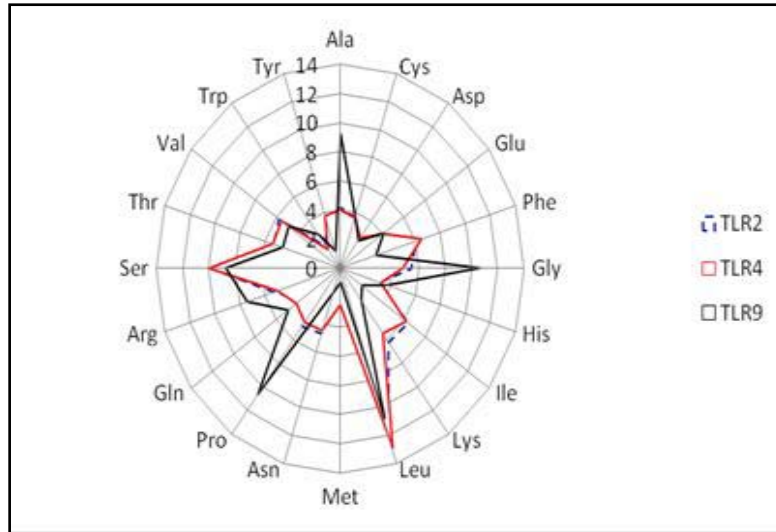|	Ala	Cys	Asp	Glu	Phe	Gly	His	Ile	Lys	Leu	Met	Asn	Pro	Gln	Arg	Ser	Thr	Val	Trp	Tyr
TLR 2	4.21	3.70	2.53	3.89	6.05	5.37	3.10	6.34	6.24	11.82	2.06	4.67	4.91	3.59	5.27	9.73	5.31	5.76	1.96	3.36
TLR 4	3.99	3.60	2.56	4.01	6.47	4.73	3.22	6.17	5.50	12.90	2.56	4.46	4.57	4.10	4.99	9.95	5.38	5.49	1.58	3.74
TLR 9	9.16	3.65	2.31	4.04	2.83	10.58	3.77	2.02	2.67	10.83	0.96	1.75	10.68	4.92	7.47	8.72	4.61	4.86	2.85	1.22

Fig. 5: Radar plot showing the relative abundance of the amino acid present in three different TLR receptors

The identical average pair of nucleotide numbers for TLR2 was 9804, for TLR4 was 10684 and for TLR9 was 2888. Interestingly, transitional and trans version ratio (R= si/sv) for TLR2 was 0.84, for TLR4 was 1.09 and for TLR9 was found to be 0.89 respectively for three different positions in the TLR domain. TT (3147) was the most frequent nucleotide pair found in TLR2 followed by GG and CC pairs. The average identical nucleotide pair frequency was 10684 and was seen in the case of TLR4. The ratio of the R value was 1.09, and the common nucleotide pairs were TT (3532), AA (3267) and GG (1932) (fig 6). In the case of TLR9, the average identical pairs of nucleotide were 2888 whereas the frequency of R was 0.89 and the average nucleotide pairs found in the sequences were GG (901) and CC (849) (table 3).

Table 3: Nucleotide pair frequencies and average identical pairs number for three different TLRs

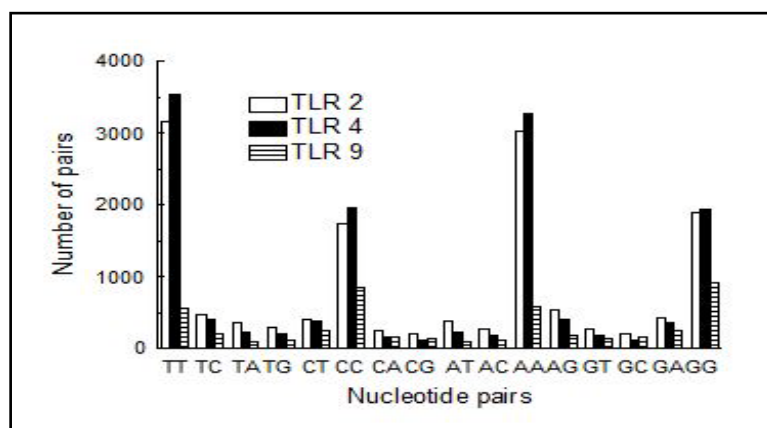|	TT	TC	TA	TG	CT	CC	CA	CG	AT	AC	AA	AG	GT	GC	GA	GG	Total	Average identical pairs
TLR2	3147	474	361	282	411	1739	250	202	374	269	3024	526	258	195	433	1894	13840	9804
TLR4	3532	391	219	205	374	1953	163	113	233	180	3267	393	175	107	364	1932	13602	10684
TLR9	556	201	93	110	244	849	146	144	88	108	582	182	142	154	248	901	4748.40	2888

Fig. 6: Graph showing nucleotide pairs frequencies in the TLR genes using Kyplot (ver-2.0)

Table 4: Tajima`s Neutrality test for three TLRs in six different selected sequences

	m	S	P_s	O_w	π	D
TLR2	6	3152	0.642872	0.281550	0.262669	-0.437153
TLR4	6	5628	0.567510	0.248545	0.232960	-0.408803
TLR9	6	2168	0.739679	0.323947	0.379939	+1.126589

We compared the Tajima`s neutrality test for selection and mutation rate where we took a number of sequences (m = 6) and compared with three TLRs separately. In TLR2, the numbers of segregating sites (S) were 3152 and the nucleotide variation frequency was

observed to be 0.262669. The D value for TLR2 was calculated-0.437153. The p/s (the number of segregating sites/number of sequences) value for TLR2 was estimated to be 0.642872. In the case of TLR4 number of segregating sites was calculated 5628, the nucleotide variation frequency was found to be 0.232960, and D value was-0.408803. The p/s (number of segregating sites/number of sequences) value for TLR4 was 0.567510. When we compare the sequences for TLR9 we observed that the numbers of segregating sites (S) were estimated to be 2168, the nucleotide variation was much higher for TLR9 (0.379939). The D value was also positive

(+1.126589) which quite interesting (table 4) was. We also calculated the disparity index where we observed that in the case of TLR2 the pairwise matching of different taxa was much closer. When we compared the observed values of TLR4 for disparity index the concerned differences were high for Bubalus and Danio, but the similarities were closer with the hominids, but the least difference was found between Homo and Pongo. In the case of TLR9, the observed differences were highest with Bubalus and Danio among all the TLRs. The least difference was found with Bubalus–Pongo 0.869 (table 5).

Table 5: Disparity index of TLR2 for six different animal sequences

		1	2	3	4	5	6
1	*Bubalus bubalis*		4.252	5.238	5.125	5.466	5.234
2	*Danio rerio*	0.000		0.000	0.000	0.000	0.000
3	*Gorilla gorilla*	0.000	1.000		0.000	0.001	0.000
4	*Homo sapiens*	0.000	1.000	1.000		0.002	0.000
5	*Pan troglodytes*	0.000	1.000	0.364	0.260		0.000
6	*Pongo abelii*	0.000	1.000	1.000	1.000	1.000	

Disparity index of TLR4 for six different animal sequences

		1	2	3	4	5	6
1	*Bubalus bubalis*		20.000	4.828	4.963	5.014	4.830
2	*Danio rerio*	0.000		4.933	4.763	4.643	5.232
3	*Gorilla gorilla*	0.000	0.000		0.000	0.000	0.000
4	*Homo sapiens*	0.000	0.000	1.000		0.000	0.009
5	*Pan troglodytes*	0.000	0.000	1.000	1.000		0.039
6	*Pongo abelii*	0.000	0.000	1.000	0.296	0.072	

Disparity index of TLR9 for six different animal sequences

		1	2	3	4	5	6
1	*Bubalus bubalis*		48.438	3.784	3.991	4.147	0.869
2	*Danio rerio*	0.000		46.314	47.012	46.332	37.255
3	*Gorilla gorilla*	0.000	0.000		0.000	0.000	2.777
4	*Homo sapiens*	0.000	0.000	1.000		0.000	2.906
5	*Pan troglodytes*	0.000	0.000	1.000	1.000		2.976
6	*Pongo abelii*	0.010	0.000	0.000	0.000	0.000	

DISCUSSION

A common inference in comparative sequence analysis is that genome sequences have evolved with the same pattern of nucleotide substitution in DNA or homogeneity of the evolutionary process which changed the DNA sequences as well as the protein. Violation of this assumption is known to adversely affect the accuracy of phylogenetic assumptions and tests of evolutionary hypotheses [13]. Selection and mutations are two factors that modify the genomes constantly throughout the evolutionary period. An emerging goal of evolutionary biology is to understand the forces that govern how populations and species evolve. In terms of molecular evolution, this problem has often been framed in explaining the relative contributions of genetic drift and natural selection to extant patterns of genetic variation [14, 15]. The sequences for the TLRs are present in various chromosomes in the animal world. It recognizes a diverse set of conserved antigens like pathogen-associated molecular patterns (PAMPs). Our investigations produced a diverse set of data for providing a significant change in the genes of TLRs due to their functional properties. It also confers a statistical model of how it changed according to the pathogens they recognized.

We have taken the sequences of TLR2, TLR4, and TLR9 among which TLR2 and TLR4 are cell surface receptors whereas, TLR9 resides in the endosome where there is a constant pressure of infectious pathogens like HIV [2, 3] and others.

When we assessed the phylogenetic tree, we found that different species of Hominid group and Humans emerged from the same lineage of common ancestors. Bubalus, took the separate line although it has a

common line of primate lineage. Danio, a fish species has been found to be an out-group member among them. But in the case of TLR9 Homo-Pan–Gorilla clustered in a single group, but Bubalus and Pongo diverge from another line of common descent. The average sequence divergence between the human-chimpanzee pair is 1.24%, 1.62% for the human-gorilla pair, and 1.63% for the chimpanzee-gorilla pair [16]. The average sequence divergences between orangutans and humans, chimpanzees, and gorillas were 3.08%, 3.12%, and 3.09% respectively as calculated from the GenBank data and other sequence data [16]. We have also found similar results in our data set. It might hint towards the convergent evolution of the TLRs in Hominid group and Human and also in Bubalus. It was found that *Danio rerio* have some common TLR variants orthologs to some mammalian TLRs. It indicates the functional similarity of the receptors during evolution and common ancestry of them [17]. Pairwise genetic distance maps w showed the close association similar to phylogenetic tree among the Hominids and Humans and distant relationship with *Danio rerio*. It was observed that evolutionary divergence was lower among the Hominids and Human than the other two (Bubalus and Danio). The intermediate color inferred the intermediated divergence among the animals. Lowest similarities were found among *Danio rerio* with other species [18].

The high frequencies of amino acids are lysine (AAA) and phenylalanine (UUU) that has been found in the case of TLR2 and TLR4. In TLR9 proline (CCA) is the most abundant amino acid that is present in the above-mentioned genes. It indicates the effect of positive selection operative on the above-mentioned codons. But interestingly the frequencies of RSCU are highest for arginine in all TLRs.

When we compared the amino acid composition for the different TLRs we found that the most abundant amino acid is leucine which is represented in a repeated fashion in the TLRs domain. The base composition and R value for TLR4 were highest among all. It might be possible that TLR4 are frequently used against bacterial invasion during evolution. When we compare the D values, we found that TLR2 and TLR4 have negative values. Negative values of Tajima's D indicate an excess of low-frequency alleles and can result from population expansions or positive selection [11]. Disparity index assessed the relationship among the groups and values greater than zero (0) indicates larger differences in the animal taxa. Interestingly, a close relationship was found among Bubalus-Pongo in the case of TLR9 established that they have diverged from common ancestors. The distant relationship was found for Bubalus–Danio in the case of TLR4 and TLR9 but quiet close in case of TLR2. The close relationship among the Hominid–Human-Bubalus was found in the case of all three TLRs.

We have found positive values for TLR9 which might be an upshot of excess intermediate frequency alleles and can result from population bottlenecks, structure and/or balancing selection for six sequences. The highest segregating sites were observed in the case of TLR4 which might be a consequence of mixing of the ancestral genes. Thus, it could be concluded from our result that TLR2 and TLR4 have constantly been under positive selection, but neutral mutation has been assessed in the case of TLR9.

CONCLUSION

After a comprehensive analysis of sequence data, we might arrive at the conclusion that TLR2 and TLR4 have some common functional similarities and common ancestry among different animals. It might be due to the similar surface antigens that they recognize and result of co-evolution or convergent evolution occurring among them. Similar amino acids present in TLR2 and TLR4 indicated the preference of the synonymous codons in their DNA sequences. The negative value of Tajima's test for both TLR2 andTLR4 indicates the selection pressure and convergent evolution among different animals. This might draw some attention that both of them are cell surface receptor and convergent evolution might have occurred due to the functional similarities and recognition of the antigens. But in the case of TLR9, there are lots of differences compared to the other two TLRs. It indicates the neutral selection that pressurized the TLR9 in the evolutionary process. The selection and mutation are the two evolutionary processes that might be responsible for the diversification of the receptors in the evolution of the TLRs in the ancestry of different animals. Similarities were also observed among the *Homo sapiens* and other Hominid group. The distant relationship was found with *Danio rerio* because of the speciation. Finally, we have inferred that Hominid and Human have co-evolved from the common ancestors of Danio sp. due to selection and speciation.

ACKNOWLEDGMENT

CONFLICT OF INTERESTS

The authors declared no conflict of interest

REFERENCES

1. Akira S, Takeda K, Kaisho T. Toll-like receptors: critical proteins linking innate and acquired immunity. Nat Immunol 2001;2:675–80.
2. Kumar SG, Manjula J, Rajasekaran C, Manoj MN, Subbarao PV, Chandrasekhar BN. Trends in the diagnosis of human immunodeficiency virus. Asian J Pharm Clin Res 2013;6:1-4.
3. Doke S, Shabri M, Shaikh G, Patil SM. AIDS in pregnancy–a review. Asian J Pharm Clin Res 2010;3:151-8.
4. Kalla PK, Lalnunthari K, Arjunan S. Cloning and molecular characterization of GAG gene from HIV-1into E. coli DH5A host. Int J Curr Pharm Res 2014;6:41-4.
5. Schwartz DA, Cook DN. Polymorphisms of the toll-like receptors and human disease. CID 2005;41:403–7.
6. Saxena R, Khan F, Masood M, Qureshi Z, Rathore M. Review on organ transplantation: a social, medical need. J Crit Rev 2016;3:23-9.
7. Wu H, Chadban SJ. Roles of toll-like receptors in transplantation. Curr Opin Organ Transplant 2014;19:1-7.
8. Netea MG, Wijmenga C, O'Neill LA. Genetic variation in toll-like receptors and disease susceptibility. Nat Immunol 2012;13:535-42.
9. Wlasiuk G, Nachman MW. Adaptation and constraint at toll-like receptors in primates. Mol Biol Evol 2010;27:2172–86.
10. Nielsen R, Bustamante C, Clark AG, Glanoski S, Sackton TB, Hubisz MJ, *et al.* A scan for positively selected genes in the genomes of humans and chimpanzees. PLoS Biol 2005;3:976–85.
11. Biswas S, Joshua AM. Genomic insights into positive selection. Trends Genet 2006;22:437-46.
12. Ihaka R, Genleman RR. A language for data analysis and graphics. J Computational Graphical Statistics 1996;5:299-314.
13. Kumar S, Gadagkar SR. Disparity index: a simple statistic to measure and test the homogeneity of substitution patterns between molecular sequences. Genetics 2001;158:1321–7.
14. Kimura M. The Neutral Theory of Molecular Evolution. Cambridge University Press; 1983.
15. Gillespie JH. The Causes of Molecular Evolution, Oxford University Press; 1991.
16. Feng-Chi Chen, Wen-Hsiung L. Genomic divergences between Humans and other Hominoids and the effective population size of the common ancestor of humans and chimpanzees. Am J Hum Genet 2001;68:444–56.
17. Jault C, Pichon L, Chluba J. Toll-like receptor gene family and TIR-domain adapters in Danio rerio. Mol Immunol 2004;40:759-71.
18. Singh J, Mukhopadhyay CS, Arora JS, Kaur S. Biocomputational characterization and evolutionary analysis of bubaline dicer1 enzyme. Asian Australas J Anim Sci 2015;28:876-87.

How to cite this article

- Avishek Das, Pokhraj Guha, Tapas Kumar Chaudhuri. Comparative sequence analysis of TLR2 TLR4 and TLR9 genes among selected vertebrates-a meta-analysis. Int J Pharm Pharm Sci 2016;8(11):180-185.