Contents lists available at ScienceDirect

Asian Pacific Journal of Tropical Disease

journal homepage:www.elsevier.com/locate/apjtd

Document heading doi:10.1016/S2222-1808(11)60007-8

Molecular evolutionary studies of Lassa virus nucleoprotein 2

Okoror LE[1*], Eniolorunda TA[1], Okoror OI[2]

[1]Department of Microbiology and Biotechnology, College of Natural Sciences, Joseph Ayo Babalola University, PMB 5006, Ilesha, Osun State, Nigeria

[2]Department of Microbiology, Faculty of Natural Sciences, Ambrose Alli University, PMB 14, Ekpoma, Edo State, Nigeria

ARTICLE INFO

Article history:
Received 15 January 2011
Received in revised form 21 February 2011
Accepted 10 March 2011
Available online 28 March 2011

Keywords:
Lassa virus
Nucleoprotein
Arena viruses
Molecular evolutionary
MEGA
Lassa fever
Virulence diversity
Circulating antibodies
Nucleotide sequence
Viral glycoprotein
Phylogenetic tree
Synonymous substitution

ABSTRACT

Objective: To study the virulence diversity through molecular evolution, and to provide insight on circulating antibodies. **Methods:** The nucleotide sequences of 18 Lassa virus genomic RNA encoding Lassa virus nucleoprotein isolates collected from different parts of the world since the identification of the Josiah strain were obtained from the GenBank and nucleotide substitution among them studied using the computer program MEGA 4. The genetic distances among strains were predicted by pairwise nucleotide differences. **Results:** The rate of synonymous substitution was high 5.889 per nucleotide per year and nonsynonymous was higher at 49.664. The average predicted rate of synonymous and nonsynonymous using modified Nei-Gojobori (assuming transition/transversion bias=2) was 27.9 which was taken as the genetic distance between strains. The average number of synonymous sites is 150.741 while the average number of nonsynonymous sites is 392.259. The phylogenetic tree was inferred by unweighted pairwise grouping in MEGA4 and using neighbour-joining method. The time of emergence of Lassa virus was predicted to be around January 1920. However, the first human appearance of the virus was predicted to be around May (1 959±24) months. In synonymous substitution the rate of (G-T) rare was high. The nucleotide frequencies were 0.314 (A), 0.246 (T/U), 0.204 (C) and 0.235 (G). The transition/transversion ratio k_1=14.991 (purines) and k_2=69.916 (pyrimidines). The overall transition/transversion bias R=16.662 with a total of 620 position in the final data set. These figures are far higher than an earlier study using Lassa virus glycoprotein. The nucleotide diversity were also very high using the Taijima's model in MEGA 4. **Conclusions:** The divergence within strains always coincides with the period of epidemic which goes to confirm that the cause of epidemic outbreak should be the emergence of new strain and also why the infection remains endemic despite circulating antibodies. A comparison with a similar study with the viral glycoprotein concludes that the glycoprotein is more suited for vaccine development.

1. Introduction

Lassa virus belongs to a very large group of heamorrhagic fever viruses-arena viruses which are responsible for high rate of morbidity and mortality in areas of West Africa. Among the six arena viruses known so far only one-Lassa virus (LASV) is known to cause illness in humans. Lassa virus is the etiological agent of Lassa fever which is an acute and often fatal illness and is endemic to West Africa. There are an estimated 300 000 to 500 000 cases of Lassa fever each year[1-5] with mortality rate of 15% to 20% for hospitalized patients and as high as 50% during epidemics[2,6]. Presently, there is no licensed vaccine or immunotherapy available for preventing or treating this disease. Although the antiviral drug ribavirin is somewhat beneficial, it must be administered at an early stage of infection to successfully alter disease outcome, thereby limiting its utility[7]. Furthermore, there is no commercially available Lassa fever diagnostic assay, thus preventing early detection and rapid implementation of existing treatment regimens (*e.g.*ribavirin administration). The lack of adequate countermeasures and means of detection, coupled with the severity of disease,

*Corresponding author: Okoror LE, Department of Microbiology and Biotechnology, College of Natural Sciences, Joseph Ayo Babalola University, PMB 5006, Ilesha, Osun State, Nigeria.
E-mail: OLE LEOkoror@gmail.com; OOI Larison86@yahoo.com

contributed to the classification of LASV as a national institutes of allergy and infectious diseases (NIAID) category A pathogen and biosafety level-4 (BSL-4) agent. The LASV genome is comprised of two ambisense, single stranded RNA molecules, designated small (S) and large (L)[8]. Two genes on the S segment encode NP, GP1, and GP2; whereas, the L segment encodes the viral polymerase (L protein) and RING finger Z matrix protein. GP1 and GP2 subunits result from post-translational cleavage of a precursor glycoprotein (GPC) by the protease SKI-1/S1P[9]. GP1 serves a putative role in receptor binding, while the structure of GP2 is consistent with viral transmembrane fusion proteins[10]. Humoral immunity to LASV is commonly bipartite, displaying an initial IgM response after infection, with an ensuing mature IgG response[11].

Over the years many aspects of Lassa fever have been defined including the clinical presentation, epidemiology, immunology, pathology, physiology and therapy with sparse reports on molecular evolution. Lassa fever has been reported to be seasonal and usually occur within the dry months of December to May, infections also occur in the months of February through April. However, the exact seasonality of Lassa virus is still not understood and therefore increase in availability of rodent host (*Mastomys natalensis*) may increase the risk of infection or transfer of the virus to humans. Transfer from humans to humans has also been reported[12]. Lassa virus infection has continued unabated despite the presence of circulating antibodies against the infection in the areas of epidemics with emergence of newer strains which might likely be the cause of new outbreaks, hence a detailed evolutionary study at molecular level becomes essential since serological studies over the years has not provided any information about the origin of the virus. Serological data do not also provide information on the origin of human to human infection. In 2004 Omilabu *et al*[13] identified two new strains whose gene sequences were deposited in the Genbank.

Lassa virus was first identified in Northern Nigeria, although reports have it that the virus co-evolved with the rodent host over a period of 9 million years. A study of the molecular evolution will provide better and concise information as to the origin of the virus in humans. Okoror *et al*[12] studied molecular evolution of Lassa virus glycoprotein though reported that the glycoprotein is well suited for evolutionary studies because it maintains the viral structure but a comparison of the glycoprotein with other sequences like the nucleoprotein is necessary to see which of the sequences is better adapted for evolutionary studies is imperative. The recent identification of two new strains point to the fact that the virus is in constant genetic change which has made this study necessary. Evolutionary study of the virus which is a determinant of the virulence diversity have been reported to give a knowledge of possible vaccine development for the virus and also give an in site into the host parasite relationship. A further study may eventually lead to the explanation as to why despite the circulating antibodies against Lassa virus there still occur yearly outbreak.

2. Materials and methods

The computer program MEGA 4 and CLC workbench 4 enabled us to separately estimate the rate of synonymous and nonsynonymous substitution as wells as study the entire evolution of Lassa virus nucleoprotein using 18 nucleotide sequences identified over the years and deposited in Genbank. The genetic distances between strains were estimated using pairwise nucleotide differences. The substitution pattern as well as transition and tranversional substitution and frequency of parallel independently occurring at the same site of different evolutionary lineages. Multiple sequence alignment of the sequences were also carried out and viewed in genedoc to observe for evolutionary conservation.

2.1. Genetic distances of isolates

Nucleotide differences between isolates were predicted using estimated genetic distances which includes the observed nucleotide differences, the number of synonymous sites (Ks), the number of nonsynonymous sites per synonymous sites (Ka) and the mean value Kt of Ks and Ka weighted by respective number of synonymous and nonsynonymous sites[14]. This was done by pairwise distance calculation with complete deletion of gaps/missing gaps using the model codon of modified Nei-Gojobori (distance). Codon position used were 1st, 2nd, 3rd and non-coding sites. Transitional and transversional changes between codons were also predicted.

2.2. Evolutionary rate

Evolutionary rates were done according to Takeda *et al*[15], and neighbour-joining method and all calculations were done in MEGA 4. It was calculated by regression analysis. Genetic distances between the earliest isolates and other isolates were plotted along the vertical axis and the isolation time along the horizontal axis.

2.3. Neighbour joining method

Neighbour joining method was done in MEGA 4 according to the method of Nei and Gojobori.

2.4. Phylogenetic tree

Phylogenetic tree was constructed using UPGMA in MEGA 4 as previously described. This was done using adjusted genetic distances ($d'y$) of both synonymous and nonsynonymous substitutions.

2.5. Substitution frequency and pattern

Relative substitution frequency was carried out using the methods of Taimura *et al*[16], and Taimura *et al*[17] with codon position 1st, 2nd, 3rd and non-coding sites. Maximum composited likelihood of the estimated nucleotide substitution was used. All calculations were done in MEGA 4. The substitution pattern corresponds to originally proposed pattern by Li *et al*[18], in which nucleotide site in the codon was classified according to its degeneracy (*i.e.*, nondegenerate, twofold degenerate, or fourfold degenerate), depending on how often nucleotide substitutions result in amino acid replacement. A site is fourfold degenerate if all possible changes at the site are synonymous; the third positions of 32 of the 61 sense codons (*e.g.*, the third position of valine) are of this type. A site is twofold degenerate if one of the three possible changes is synonymous; the third position of 24 of the 61 sense codons (*e.g.*, the third position of histidine) is of this type. A site is nondegenerate if all possible changes at this site are nonsynonymous or nonsense; the second positions of all sense codons belong to this type. Nucleotide changes are also classified as either transitional or transversional, and changes between codons are assumed to occur with different probabilities which are determined on the basis of the relative frequency of nucleotide changes in mammalian genes. Both the method of degeneracy and transition and transversion were used in this study but were done in MEGA 4.

2.6. Nucleotide sequences

Sequences used in this study were from Genbank and are as follows:

L._virus_DGD112_NP, L._virus_DGD104_NP, L._virus_DGD87_NP, L._virus_DGD43_NP, L._virus_DGD35_NP, L._virus_DGD28_NP, L._virus_DGD13_NP, L._virus_DGD4_NP, L._virus_TA846_NP, L._virus_TA820_NP, L._virus_TA817_NP, L._virus_TA491_NP,L._virus_TA471_NP, L._virus_TA464_NP, L._virus_TA462_NP, L._virus_TA444_NP, L._virus_TA341_NP, L._virus_TA416_NP.

A graph of evolutionary differences representing their evolutionary distance was plotted against the time of identification of the new strain in order to determine the time of emergence of the virus using only human isolates since 1972. The date of isolation of all the human strains used were extracted from Bowen *et al*[30]. The graph was prepared using SPSS version 17.

3. Results

Alignment of the 18 nucleotide sequences of Lassa virus strains demonstrated insertion and deletions as well as substitutions among the strains collected over the years from different parts of the world. It also showed some conserved areas over the years especially regions of suspected high functionality. The number of observed nucleotide differences is shown in Table 1. Of the 628 sites, 488 were conserved. There were 399 nondegenerate sites, 101 two fold degenerate sites and 46 four fold degenerate sites.

Table 1
Number of observed nucleotide difference and the standard error of Lassa virus nucleoprotein sequences (synonymous).

Differences	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1		0.000	0.000	0.000	0.000	0.000	1.065	1.065	3.394	3.118	2.746	2.984	2.984	3.394	2.948	2.746	2.984	2.984
2	0.000		0.000	0.000	0.000	0.000	1.065	1.065	3.394	3.118	2.746	2.984	2.984	3.394	2.984	2.746	2.984	2.984
3	0.000	0.000		0.000	0.000	0.000	1.065	1.065	3.394	3.118	2.746	2.984	2.984	3.394	2.984	2.746	2.984	2.984
4	0.000	0.000	0.000		0.000	0.000	1.065	1.065	3.394	3.118	2.746	2.984	2.984	3.394	2.984	2.746	2.984	2.984
5	0.000	0.000	0.000	0.000		0.000	1.065	1.065	3.394	3.118	2.746	2.984	2.984	3.394	2.984	2.746	2.984	2.984
6	0.000	0.000	0.000	0.000	0.000		1.065	1.065	3.394	3.118	2.746	2.984	2.984	3.394	2.984	2.746	2.984	2.984
7	1.000	1.000	1.000	1.000	1.000	1.000		0.000	3.232	2.962	2.540	2.816	2.816	3.2.2	2.816	2.540	2.816	2.816
8	1.000	1.000	1.000	1.000	1.000	1.000	0.000		3.232	2.962	2.540	2.816	2.816	2.232	2.816	2.540	2.816	2.816
9	13.000	13.000	13.000	13.000	13.000	13.000	12.000	12.000		1.303	2.104	1.654	1.654	0.000	1.654	2.104	1.654	1.654
10	11.000	11.000	11.000	11.000	11.000	11.000	10.000	10.000	2.000		1.359	1.359	2.14	1.359	0.000	1.359	1.359	1.359
11	8.000	8.000	8.000	8.000	8.000	8.000	7.000	7.000	5.000	3.000		1.359	1.359	2.104	1.359	0.000	1.359	1.359
12	10.000	10.000	10.000	10.000	10.000	10.000	9.000	9.000	3.000	1.000	2.000		0.000	1.654	0.000	1.359	0.000	0.000
13	10.000	10.000	10.000	10.000	10.000	10.000	9.000	9.000	3.000	1.000	2.000	0.000		1.654	0.000	1.359	0.000	0.000
14	13.000	13.000	13.000	13.000	13.000	13.000	12.000	12.000	0.000	2.000	5.000	3.000	3.000		1.654	2.104	1.654	1.654
15	10.000	10.000	10.000	10.000	10.000	10.000	9.000	9.000	3.000	1.000	2.000	0.000	0.000	3.000		1.359	0.000	0.000
16	8.000	8.000	8.000	8.000	8.000	8.000	7.000	7.000	5.000	3.000	0.000	2.000	2.000	5.000	2.000		1.359	1.359
17	10.000	10.000	10.000	10.000	10.000	10.000	9.000	9.000	3.000	1.000	2.000	0.000	0.000	3.000	0.000	2.000		0.000
18	10.000	10.000	10.000	10.000	10.000	10.000	9.000	9.000	3.000	1.000	2.000	0.000	0.000	3.000	0.000	2.000	0.000	

Key: Strains:[1] #L._virus_DGD112_NP; [2] #L._virus_DGD104_NP; [3] #L._virus_DGD87_NP; [4] #L._virus_DGD43_NP; [5] #L._virus_DGD35_NP; [6] #L._virus_DGD28_NP; [7] #L._virus_DGD13_NP; [8] #L._virus_DGD4_NP; [9] #L._virus_TA846_NP; [10] #L._virus_TA820_NP; [11] #L._virus_TA817_NP; [12] #L._virus_TA491_NP; [13] #L._virus_TA471_NP; [14] #L._virus_TA464_NP; [15] #L._virus_TA462_NP; [16] #L._virus_TA444_NP; [17] #L._virus_TA341_NP; [18] #L._virus_TA416_NP
Number of the observed synonymous differences are shown on the right of the diagonal while the standard error are on the left of the diagonal as computed with MEGA 4.

Table 2
Number of observed nucleotide difference and the standard error of lassa virus nucleoprotein sequences (nonsynonymous).

Strain NO	Differences																	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1		0.940	2.105	0.940	0.940	1.402	1.866	2.105	7.320	7.352	7.524	7.132	7.287	7.459	7.287	7.416	7.321	7.578
2	1.000		1.866	0.000	0.000	1.004	2.105	1.866	7.300	7.322	7.488	7.107	7.260	7.320	7.260	7.390	7.288	7.548
3	5.000	4.000		1.866	1.866	2.146	0.940	0.000	7.169	7.177	7.273	6.976	7.140	7.180	7.140	7.091	7.157	7.438
4	1.000	0.000	4.000		0.000	1.004	2.105	1.866	7.300	7.322	7.448	7.107	7.260	7.320	7.260	7.390	7.288	7.548
5	1.000	0.000	4.000	0.000		1.004	2.105	1.866	7.300	7.322	7.448	7.107	7.260	7.320	7.260	7.390	7.288	7.548
6	2.000	1.000	5.000	1.000	1.000		2.373	2.146	7.211	7.251	7.417	7.014	7.167	7.255	7.167	7.324	7.190	7.452
7	4.000	5.000	1.000	5.000	5.000	6.000		0.940	7.180	7.199	7.303	6.992	7.160	7.314	7.160	7.110	7.182	7.460
8	5.000	4.000	0.000	4.000	4.000	5.000	1.000		7..169	7.177	7.273	6.976	7.140	7.180	7.140	7.091	7.157	7.438
9	88.000	87.000	85.000	87.000	87.000	86.000	86.000	85.000		2.014	4.447	3.701	3.363	0.940	3.363	4.317	3.473	3.916
10	88.000	87.000	85.000	87.000	87.000	86.000	86.000	85.000	4.000		4.455	3.599	3.254	2.271	3.254	4.254	3.356	3.939
11	90.000	89.000	87.000	89.000	89.000	88.000	88.000	87.000	21.000	21.000		3.933	3.880	4.554	3.880	2.611	3.939	4.692
12	83.000	82.000	80.000	82.000	82.000	81.000	81.000	80.000	13.000	13.000	18.000		0.954	3.834	0.954	3.693	1.376	3.087
13	84.000	83.000	81.000	83.000	83.000	82.000	82.000	81.000	12.000	12.000	17.000	1.000		3.501	0.000	3.632	0.999	2.918
14	89.000	88.000	86.000	88.000	88.000	87.000	87.000	86.000	1.000	5.000	22.000	14.000	13.000		3.501	4.411	3.596	4.031
15	84.000	83.000	81.000	83.000	83.000	82.000	82.000	81.000	12.000	12.000	17.000	1.000	0.000	13.000		3.632	0.999	2.918
16	90.000	89.000	87.000	89.000	89.000	88.000	88.000	87.000	21.000	21.000	6.000	18.000	17.000	22.000	17.000		3.726	4.595
17	83.000	82.000	80.000	82.000	82.000	81.000	81.000	80.000	13.000	13.000	18.000	2.000	1.000	14.000	1.000	18.000		3.083
18	87.000	86.000	84.000	86.000	86.000	85.000	85.000	84.000	17.000	17.000	20.000	10.000	9.000	18.000	9.000	20.000	10.000	

Key: Strains: [1] #L._virus_DGD112_NP; [2] #L._virus_DGD104_NP; [3] #L._virus_DGD87_NP; [4] #L._virus_DGD43_NP; [5] #L._virus_DGD35_NP; [6] #L._virus_DGD28_NP; [7] #L._virus_DGD13_NP; [8] #L._virus_DGD4_NP; [9] #L._virus_TA846_NP; [10] #L._virus_TA820_NP; [11] #L._virus_TA817_NP; [12] #L._virus_TA491_NP; [13] #L._virus_TA471_NP; [14] #L._virus_TA464_NP; [15] #L._virus_TA462_NP; [16] #L._virus_TA444_NP; [17] #L._virus_TA341_NP; [18] #L._virus_TA416_NP
Number of the observed nonsynonymous differences is shown on the right of the diagonal and the standard error on the left of the diagonal as computed with MEGA 4.

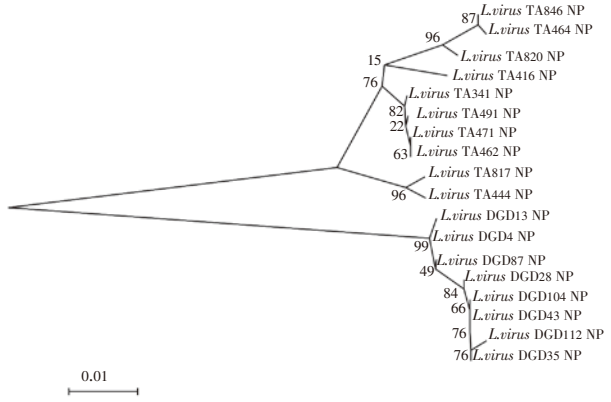

Figure 1. Phylogenetic tree inferred by neighbour-joining method of Lassa virus.

The rate of synonymous substitution was high 5.889 per nucleotide per year and nonsynonymous was higher at 49.664. The average predicted rate of synonymous and nonsynonymous using modified Nei-Gojobori (Table 1and 2) (assuming transition/transversion bias=2) was 27.9 which was taken as the genetic distance between strains. The average number of synonymous sites is 150.741. The average number of nonsynonymous sites is 392.259. The phylogenetic tree was inferred by unweighted pairwise grouping in MEGA4 and using neighbour-joining method (Figure 1 and 2). The time of emergence of Lassa virus was predicted to be around January [(1 920±24) months]. And the time of emergence of the virus in humans was predicted to be May 1959, about ten years before the first Lassa virus report in Nigeria. In synonymous substitution the rate of (G-T) rare was high. The nucleotide frequencies were 0.314 (A), 0.246 (T/U), 0.204 (C) and 0.235 (G). The transition/transversion ratio k_1=14.991.

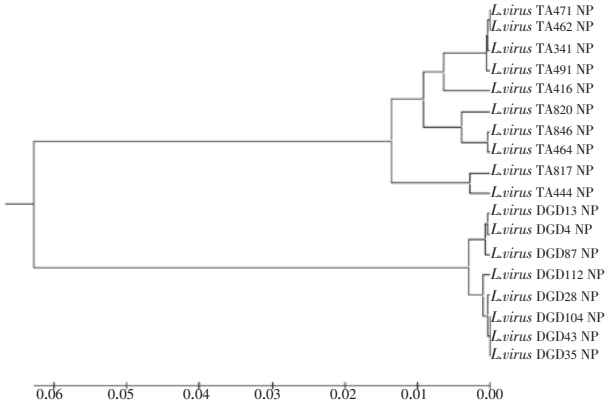

Figure 2. Phylogenetic tree inferred by UPGMA showing the evolutionary history of Lassa virus.
The branch length is the same as those shown in evolutionary distances in table 1a and b. The evolutionary distances were maximum composite likelihood method and are in the units of number of base substitution per site. Codon position included are 1st, 2nd, 3rd positions + non coding sites. There were 254 positions in the final data set. The phylogenetic relationship was done using the program MEGA 4.

(purines) and k_2=69.916 (pyrimidines). The overall transition/transversion bias R=16.662 with a total of 620 position in

the final data set (Table 3). These figures are far higher than an earlier study using Lassa virus glycoprotein. The nucleotide diversity was also very high using the Taijima's model in MEGA 4 (Table 4). Of the 628 sites, 399 were non degenerate while 101 were two fold degenerate and 46 four fold degenerate. The neighbour joining method was used to explore the evolutionary relationship of isolates using Kt as the genetic distance (Figure 1). This is based on the principle that the sum of the branch length (shown by numbers on the branch) at each cluster stage is to be minimised. Numbers represent estimated branch lengths in nucleotide. Notice that all strains of Lassa virus evolved from one focal point. The branch length in Figure 2 is the same as those shown in evolutionary distances in Table 1a and b. The evolutionary distances were maximum composite likelihood method and are in the units of number of base substitution per site. Codon position included is the 1st, 2nd, 3rd positions and non coding sites. There were 254 positions in the final data set. The phylogenetic relationship was done using the program MEGA 4 (Figure 2).

Thus this method is suitable for estimating genetic relationship among isolates without taking into account evolutionary rate or isolation time. The nucleotide diversity was high at 0.099 905 (Table 4). Amino acid substitution was high where only 99 of the total 209 sites were conserved. The graph of the nucleotide distance against their time of identification of new human strains (Figure 3) predicted the first human infection to be around May 1959 using the 1972 Josiah strain as a reference strain.

Table 3
The nucleotide substitution pattern and frequency of Lassa virus nucleoprotein gene sequences.

	A	T	C	G
A	–	0.59	0.49	8.46
T	0.75	–	34.19	0.56
C	0.75	41.26	–	0.56
G	11.29	0.59	0.49	–

Each entry shows the probability of substitution from one base (row) to another base (column) instantaneously. Only entries within a row should be compared. Rates of different transitional substitutions are shown in bold and those of transversionsal substitutions are shown in italics. The nucleotide frequencies are 0.314 (A), 0.246 (T/U), 0.204 (C), and 0.235 (G). The transition/transversion rate ratios are k_1=14.991 (purines) and k_2=69.916 (pyrimidines). The overall transition/transversion bias is R=16.662, where R=[A×G ×k_1 + T×C×k_2]/[(A+G)×(T+C)]. Codon positions included were 1st+2nd+3rd+Noncoding. All positions containing gaps and missing data were eliminated from the dataset (Complete deletion option). There were a total of 620 positions in the final dataset. All calculations were conducted in MEGA4.

Table 4
The probable nucleotide diversity of Lassa virus nucleoprotein.

m	S	ps	Θ	π	D
18	137	0.220 968	0.064 243	0.099 905	2.344 264

Results from Tajima's Neutrality Test for 18 sequences
The Tajima test statistic [1] was estimated using MEGA4 [2]. All positions containing gaps and missing data were eliminated from the dataset (Complete deletion option).
m= number of sites, S=Number of segregating sites, ps = S/m, Θ = ps/a1, and π = nucleotide diversity. D= the Tajima test statistic.

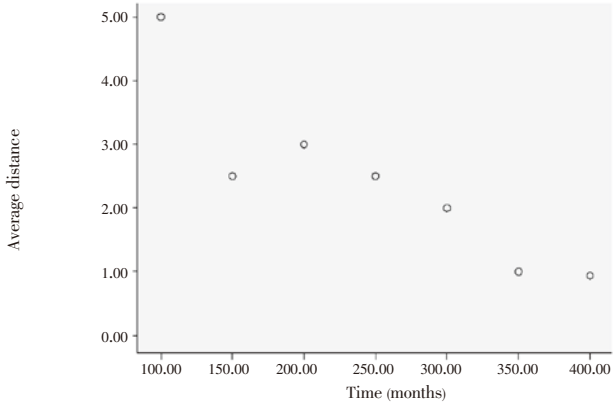

Figure 3. Evolution rate of Lassa virus and the first human infection. The 1972 strain was taken as the stand point for first case. Using 4 strains and assuming that they all evolved at a constant rate, the evolutionary distance was plotted against time of emergence. A dotted line from the point of joins at the dot plot, backwards will terminate at the month of first emergence. Or the slope of the plot in months minus the proposed first emergence date will give the date of first emergence of human straisn. The dot plot was done in SPSS version 17.

4. Discussion

The phylogenetic tree was inferred with 18 Lassa virus nucleoprotein nucleotide sequences using MEGA 4 and this was consistence with results from other studies[12,15]. Okoror *et al*[12] studied the Lassa virus glycoprotein nucleotide sequences using MEGA 4 in all calculations while Takeda compared 18 nucleotide of capsid protein VP1 in 18 isolates of EV70 in two separate studies with earlier study using oligonucleotide mapping of the entire genome[19]. The emergence of Lassa virus was predicted to be (1920±24) months which opposes earlier studies that suggested that the virus probably co-evolve millions of years ago with the rodent host. However, the rodent strain might be different from the human strain since all the strains compared in this study were all human strains. It is suggested that as soon as the strains crosses the species barrier to infect human there will be a change in the nucleotide sequences. Hence the prediction of the first human infection using only the human strains shows that Lassa virus may have first infected man around the May of 1959. This might as well be the year the virus crossed the species barrier to infect humans. This study also revealed that the divergence times are usually during the period of high epidemics which confirms that the emergent of newer strains were responsible for the yearly epidemic outbreaks. The concentration of strains at the period of epidemics also inferred that the variations in strains is usually location dependent rather time dependent. This is more so since only fewer strains were isolated during the non epidemic period. And occurrence of epidemics almost at the same time in all the endemic areas. These observations indicate that numerous genetically different strains were produced just before the start of any epidemic relative to the location of the virus and the rodent reservoir. The fact that epidemics occur at the same time in different

areas by different strains of the virus confirms earlier report by Banseh[20] who suggested that the virus vary with location. In this study therefore we suggest that the virus vary according to both location and time as epidemic outbreak in a location is usually caused by different strains. The high rate of transition/transversion at the same time shows that there is co-circulation of genetically and antigenically different strains in the population. This could be further confirmed by the isolation of two different strains by Omilabu *et al*[13]. in patients admitted to Irrua Specialist Hospital who live in the same locality and at the same time. This goes to point that the nucleoprotein is suitable for evolutionary studies but the glycoprotein is more suited for vaccine development if the study by Okoror *et al*[12] is anything to compare with.

The nucleotide substitution rates of the Lassa virus nucleoprotein gene obtained here (27.6) weighted average for both synonymous and nonsynonymous per nucleotide per year were higher than those of an earlier study by Okoror *et al*[12]. using the Lassa virus glycoprotein which was suggested to be responsible for maintaining viral structure. This goes to confirm that the Lassa virus nucleoprotein is best suited for evolutionary studies than the glycoprotein since it gives a better insight of the periodic nucleotide changes especially with a nucleotide diversity as high as 0.099 905. This is in opposition to earlier studies in non-structural protein gene[15], nucleoprotein gene[21,22] and PB2 of influenza[23] A virus and for haemagglutinin[24] and NS genes[25] of influenza B virus; they were in the range of 1.1×10^{-3} to 2.3×10^{-3} substitutions per nucleotide per year for the observed nucleotide differences. In addition, a frequency as low as 1.4×10^{-4} per nucleotide per year was reported for the 26S structural protein gene of the equine encephalitis virus. The high rate of the Lassa virus glycoprotein gene was accounted for by the high rate of nonsynonymous substitution as well as the synonymous substitution though the nonsynonymous substitution was higher per nonsynonymous and synonymous sites per nucleotide per year respectively. The nonsynonymous substitution rate was almost four times higher than that of influenza HA 14.1×10^{-3} and almost 3 times higher than that of EV70 VP1 gene put at 3.8×10^{-3} per nucleotide per year. A lower synonymous rate does not give rise to parallel substitution, hence it could be said that there occur newer strains at each epidemic and there even confirms the occurrence of different strains at each epidemic. This could also explain why the virus escapes circulatory antibodies to Lassa virus despite the persistence nature of the virus in the population. The lack of parallel substitution encourages the strains to be completely different both genetically and antigenically to previously existing strains. When two strains have a parallel substitution, this substitution may not be counted as a change between the strains. Thus, the accumulation of parallel substitutions for a longer period must erase actually occurring substitutions detectable between isolates. This might be the reason for the lower synonymous substitution rate of lassa virus nucleoprotein gene virus than of EV70. This is more so when lassa virus nucleoprotein gene was evaluated for about forty years and EV70 for only 10 years. Even the lassa virus glycoprotein[12] which was evaluated for a lower number of years also had lower synonymous substitution rate.

There was a high remarkable feature of the nucleotide substitution patterns of lassa virus transitions, *i.e*., low transversions. The relative frequency of transitions (C*-U and A*->G) was almost 70% (R=16.662 in the 620 sites in the final data sets in all three estimations, *i.e*., in total substitutions, at the third position of codons, and at fourfold-degenerate sites. This is an indication that transitions were found in almost all the sites in the data set. Although transitions in naturally occurring mutations are known to be much more frequent than expected from random substitutions[18,26], the frequency of transitions in lassa virus glycoprotein gene was much higher than that reported for the pseudogene DNA of human origin (55.6% on average)[26], for the third codon position of the Mahoney and Sabin strains of type 1 poliovirus (77.1%), for the third codon position of influenza virus (77.7%)[27], and for fourfold-degenerate codon position of three coding regions (gag, pol, and env) of retroviruses (66.4%)[28]. This study is in opposition with earlier report by Takeda *et al*[18], which suggested that the short period for the analysis of the EV70 VP1 gene was responsible for its high transition since this study analysed sequences that were isolated close to 40 years ago. Such high frequency was also observed for Lassa virus glycoprotein which was analysed for a shorter period of time. Hence the time of isolation is not a parameter to judge is transition. This study also suggests that virulence diversity in arenaviruses as studied is essential for the development of vaccine against the virus as earlier reported by Jahrling and Peters[29]. It is also suggested that the reasons for a yearly epidemic outbreak is due to the fact that each time the rodents strain crosses the species barrier there is always a re-modification of genetic or molecular make-up of the virus, hence not been recognised by already circulating antibodies. It is also likely that several outbreaks in different locations at the same time with newer strains could also be responsible for the reasons why the virus surpasses the circulating antibodies. The high rate of non-synonymous mutation reported in Lassa virus glycoprotein gives them an advantage in vaccine production as against Lassa virus nucleoprotein which has high synonymous mutation. This is probably due to the fact that the glycoprotein is more exposed to the environment and therefore more prone to external changes.

Conflict of interest statement

We declare that we have no conflict of interest.

Acknowledgements

OLE conceived the study and did all the computational study as well as wrote the paper. OOI was actively involved in interpretation of the results and the proof reading of the final manuscript up to the point of submission. She also supervised most of the evolutionary analysis.

References

[1] Schlie K, Maisa A, Lennartz A, Ströher U, Garten W, Strecker T. Characterization of Lassa virus glycoprotein oligomerization and influence of cholesterol on virus replication. *J Virol* 2010: **84**(2): 983-992.

[2] Schlie K, Maisa A, Lennartz F, Strecker T. Inhibition of Lassa virus glycoprotein cleavage and multicycle replication by site 1 protease-adapted alpha(1)-antitrypsin variants. *PLoS Negl Trop Dis* 2009; **4**(2): 923-933.

[3] Kernéis S, Koivogui L, Magassouba N, Koulemu K, Lewis R, Aplogan A, et al. Prevalence and risk factors of Lassa virus seropositivity in inhabitants of the forest region of Guinea: a cross sectional study. *PLOs Negl Dis* 2009; **3**(11): e548.

[4] Branco LM, Grove JN, Geske FJ, Boisen ML, Muncyn IJ, Magliato SA, et al. Lassa virus-like particles displaying all major immunological determinants as a vaccine candidate for lassa heamorrhagic fever. *Virol J* 2010; **7**: 296.

[5] Geisbert TW, Jones S, Fritz EA, Shurtleff AC, Geisbert JB, Leibscher R, et al. Development of a new vaccine for the prevention of Lassa fever. *PLOs Med* 2005; **2**(6): e183.

[6] Fichet-Calvet E, Rogers DJ. Risk maps of Lassa fever in West Africa. *PLoS Negl Trop Dis* 2009; **3**(3): e388.

[7] Bredenbeek PJ, Molenkamp R, Spaan WJM, Deubel V, Marianneau P, Salvato MS, et al. A recombinant yellow fever 17D vaccine expressing lassa virus glycoprotein. *Virol* 2006; **345**(2): 299-304.

[8] Carrion Jr R, Patterson JL, Johnson C, Gonzales M, Moreira CR, Ticer A, et al. A ML29 reassortant virus protects guinea pigs against a distantly-related Nigeria strain of lassa virus and can provide sterilizing immunity. *Vaccine* 2007; **25**(20): 4093-4102.

[9] Flatz L, Rieger T, Merkler D, Bergthaler A, Regen T, Schedensack M, et al. T-dependent lassa fever pathogenesis. *PLoS Patho* 2010; **6**(3): e1000836.

[10] Gallaher WR, DiSimone C, Buchmeier MJ. The viral transmembrane superfamily: Possible divergence of arenavirus and filovirus glycoproteins from a common RNA virus ancestor. *BMC Microbiol* 2001; **38**: 2670-2677.

[11] Bausch DG, Rollin PE, Demby AH, Coulibaly M, Kanu J, Conteh AS, et al. Diagnosis and clinical virology of Lassa fever as evaluated by enzyme-linked immunosorbent assay, indirect fluorescentantibody test, and virus isolation. *J Clin Microbiol* 2000; **21**: 234-239.

[12] Okoror LE, Alaiya HT, Imarenezor KP, Aliga P, Erimafa JT. Molecular evolution of Lassa virus glycoprotein. *Schriftenreihe Informatik* 2007; **26**: 21-29.

[13] Omilabu SA, Badaru SO, Okokhere P, Asogun D, Drosten C, Emmerich P, et al. Lassa fever, Nigeria, 2003 and 2004. *Emerg Infect Dis* 2005; **11**(10): 1642-1644.

[14] Botten J, Alexander J, Pasquetto V, Sidney J, Barrowman P, Ting J, et al. Identification of protective Lassa virus epitopes that are restricted to by HLA-A2. *J Virol* 2006; **80**(17): 8351-8361.

[15] Takeda N, Tanimura M, Tiyamura K. Molecular evolution of major capsid protein VP1 of enterovirus 70. *J Virol* 1994; 844-862.

[16] Tamura K, Nei M, Kumar S. Prospects for inferring very large phylogenies by using the neighbor-joining method. *PNAS* 2004; **101**:11030-11035.

[17] Tamura K, Dudley J, Nei M, Kumar S. MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. *Molecular Biol Evolution* 2007. DOI: 10.1093/molbev/msm092.

[18] Li WH, Wu CI, Luo CC. A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. *Mol Biol Evol* 1985; **2**: 150-174.

[19] Miyamura KM, Tanimura N, Kono TR, Yamazaki S. Evolution of enterovirus 70 in nature: all isolates were recently derived from a common ancestor. *Arch Virol* 1986; **89**: 1-14.

[20] Banseh DG, Rollin PE, Demby AH, Coulibally M, Karu J, Conteh AS. Diagnosis and clinical virology of lassa fever as evaluated by enzyme-linked immunosorbent assay. Indirect fluorescence antibody test and virus isolation. *Clin Microbial* 2000; **38**: 2670-2677.

[21] Gorman OT, Bean WJ, Kawaoka Y, Webster RG. Evolution of the nucleoprotein gene of influenza A virus. *J Virol* 1990; **64**: 1487-1497.

[22] Shu LL, Bean WJ, Webster RG. Analysis of the evolution and variation of the human influenza A virus nucleoprotein gene from 1933 to 1990. *J Virol* 1993; **67**: 2723-2729.

[23] Gorman OT, Donis RO, Kawaoka Y, Webster RG. Evolution of influenza A virus PB2 genes: implications for evolution of the ribonucleoprotein complex and origin of human influenza A virus. *J Virol* 1990; **64**: 4893-4902.

[24] Kamagae Y, Sugita S, Endo A, Ishida M, Senya S, Osako K, et al. Evolutionary pattern of the hemoglutinin gene of influenza B viruses isolated in Japan: co-circulating lineages in the same epidemic season. *J Virol* 1990; **64**(6): 2860-2865.

[25] Yamashita M, Krystal M, Fitch WM, Palese P. Influenza B virus evolution: co-circulating lineages and comparison of evolutionary pattern with those of influenza A and C viruses. *Virology* 1998; **163**: 112-122.

[26] Gojobori T, Li WH, Graur D. Patterns of nucleotide substitution in pseudogenes and functional genes. *J Mol Evol* 1982; **18**: 360-369.

[27] Saitou N. Patterns of nucleotide substitutions in influenza A virus genes. *Jpn J Genet* 1987; **62**: 439-443.

[28] Ghafoor A, Burney ML, Iqbal J, Khan Z. Acute haemorrhagic conjunctivitis (AHC) epidemic of 1981. *J Pak Med Assoc* 1984; **34**: 245-246.

[29] Hass M, Westerkofsky M, Müller S, Becker-Ziaja B, Busch C, Günther S. Mutational analysis of lassa virus promoter. *J Virol* 2006; **80**(24): 12414-12419.

[30] Cosset F, Marianneau P, Verney G, Gallais F, Tordo N, Percheur E et al. Characterization of Lassa virus cell entry and neutralization with lassa virus pseudoparticles. *J Virol* 2009; 3228-3237.