Evaluation of Amino Acids Change in Structural Protein of Filoviridae Family during Evolution Process

Maryam Mohammadlou*

Science Faculty, I. Azad University, Ahar branch, Ahar, Iran.

* Corresponding author. Tel.: +9891 4388 9462; e-mail: mohammadlou.maryam@gmail.com

Abstract: The likelihood ratios test results for VP24, VP30, VP35, and VP40 proteins of different species of Ebola viruses from Filoviridae family were determined using different models to obtain $2\Delta\ell$ values and to compare them with Chi-Square distribution 99% and 95%. Then, by using site models, BEB(Bayes Empirical Bayes) analysis gives positively selected sites of GP that are in the musin like domain. Results of likelihood ratio test using positive selection models lead to the conclusion that there are evidence for the relation between the change of amino acids in particular sites of structural proteins of this virus and increasing mortality rate of it.

Key words: Ebola virus, structural protein, likelihood ratios test, bayes empirical bayes.

1. Introduction

Nucleotide substitutions are divided into nonsynonymous substitutions (N), which result in changes of amino acid residues in the encoded proteins, and synonymous substitutions (S), which do not cause changes in amino acid residues. Positive Darwinian selection at the DNA sequence level is often tested by estimating the ratio (ω) of the rate of nonsynonymous (dN) nucleotide substitutions to that of synonymous (dS) substitutions between homologous protein-coding gene sequences [1]. An ω value significantly higher than 1 is interpreted as evidence for positive selection, $\omega <1$ suggests purifying selection (selective constraints), and $\omega = 1$ indicates neutral evolution. Detecting positive selection is generally difficult because positive selection often acts on a few sites and in a short period of evolutionary time and the signal may be swamped by the ubiquitous negative selection.

With determining nonsynonymous-synonymous substitution rate ratio (dN/dS or ω) a measure of selective pressure at the protein level will be provided [2], [3]. An ω greater than one indicates that nonsynonymous mutations offer fitness advantages and are fixed in the population at a higher rate than synonymous mutations. The positive selection can thus be detected by identifying cases where $\omega>1$. In a functional protein, many amino acids may be under strong structural and functional constraints and not free to vary. Thus, it is important to account for variation in selective pressure (and thus in the ω ratio) among sites if one hopes to detect positive selection affecting only a few amino acid residues [4], [5]. A number of such models were implemented previously [4], [6] based on the codon-substitution model [7], [8]. In the past few years, such site-specific models have been used to detect positive selection in a variety of genes and species [9]-[20]. Computer simulations also confirmed the power of those methods [21]-[23].

The Filoviridae family in the order Mononegavirales is divided into some genera including Ebolavirus and

Marburgvirus. Ebolavirus genus has Bundibugyo virus, Reston ebolavirus, Sudan ebolavirus, Tai Forest ebolavirus and Zaire ebolavirus species. Filoviruses are recognized as a significant threat to public health and conservation as they cause periodic human and non-human primate outbreaks with high mortality rates.

In the current research, some programming and software tools are used to determine how seven structural proteins of Filoviruses undergo some evolution changes and what results of these alternations in the amino acid framework are.

2. Method

The expressed proteins by the Filoviruses are: nucleoprotein (NP), glycoprotein (GP), RNA-dependent RNA polymerase (L), VP24, VP30, VP35, and VP40 [24], [25]. Gene-protein sequence data were downloaded from Virus Pathogen Database and Analysis Resource for each protein separately (ViPR, http://www.viprbrc.org). According to the classification of Ebolavirus by ViPR, five species of this genus are Bundibugyo virus, Reston ebolavirus, Sudan ebolavirus, Tai Forest ebolavirus, Zaire ebolavirus and Unclassified Ebolavirus. 50 sequences of VP24(protein length was 756 for all species), 58 sequences of VP30(protein length was 870 for Tai Forest ebolavirus and Bundibugyo virus species and 867 for other species) , 51 sequences of VP35(protein length was 1023 for Zaire ebolavirus, 990 for Reston ebolavirus and   Sudan ebolavirus and 1026 for Tai Forest ebolavirus and Bundibugyo virus species) , 52 sequences of VP40(protein length was 996 for Reston ebolavirus and 981 for other species) , 26 sequences of L(protein length was 6639 for Zaire ebolavirus and 6633 for other species), 39 sequences of GP (protein length was 2034 for Reston ebolavirus and 2031 for other species) and 42 sequences of NP(protein length was 2217 for Sudan ebolavirus and 2220 for other species)   are analyzed in this paper. We did not make any change in sequences. After these sequences were aligned poor sequences were completely removed and this resulted in different Taxa number. Several tools have been developed for alignment of sequences based on back- translation, but TranslatorX+T-Coffee was used according to the advantage of this tool [26]. EMBOSS Seqret from Sequence Format Conversion tools of EMBL-EBI was used to convert these sequences to Phylip interleaved and Nexus/Paup interleaved format. PAUP* tool (Phylogenetic Analysis Using Parsimony) version 4a.147 was utilized for inferring and interpreting phylogenetic trees but without branch length [27]. These trees and Phylip interleaved format text were implemented in the CODEML program of the PAML package [28], [29].

3. Results

3.1. Lineage-Specific/Branch Models

These models allow ω vary among branches of phylogeny and are used to detect positive selection acting on particular lineages. One-ratio model or null model assumes an equal ω-ratio for all branches in the phylogeny and free-ratios model assumes an independent ω-ratio for each branch. Then Likelihood Ratio Test (LRT) of these models can be compared and tests if they are significantly different. Where free-ratios model significantly favored we can conclude that there is variable selective pressure in the phylogeny and branches with ω>1 show us weak evidence of positive selection.

Results for comparison of One-ratio and Free-ratio models are in Table 1. Suppose the log-likelihood values under the two models are ℓ_0 and ℓ_1, respectively. Then, $2\Delta\ell=2(\ell_1-\ell_0)$ can be compared with χ^2 distribution (chi-square table, probability 95%, and 99%) and regard degree of freedom (df: difference in a number of parameters between One-ratio and Free-ratio model) to test whether these two models are significantly different. [30], [31]. Fig.1 was drawn for more illustration. Taxa names were written as GenBank Accession, species name, and collection date respectively. Numbers on branches are dN/dS of

Free-ratio model. If dN/dS for a branch was 0/0 or 0/number, it was not written. Red color branches have dN/dS>1.

In next step, Two-ratio model, a more specific model respect to Free-ratio model, was applied to alignment sequences of GP, NP and L proteins. The Free-ratio model is parameter-rich and is unlikely to produce accurate estimates for all ω ratios. Based on data in Table 1, in the case of VP24, VP30, VP35 and VP40, there is not a meaningful difference between these two models and investigation with Two-ratio model on these proteins was not performed.

Two-ratio model allows two or more ω ratios. In Table 2 it was specified which branches have different rates in tree files and also results of the comparison of LnL values of this model with One-ratio model was reported. In all cases, there are only two different ω ratios and then df equals one. Where Two-ratio model significantly favored ($2\Delta\ell= (\ell_1-\ell_0) > \chi^2_{5\%}$ where ℓ_1 and ℓ_0 are -LnL values of Two-ratio model and One-ratio model respectively) and branches tested have ω>1, we can conclude that there is evidence of positive selection on those lineages. In Fig.1 Specified branches for Two-ratio model were shown with capital letter labels in accordance with Table 2.

3.2. Site Models

In this section test of evidence for models of Site-Specific Positive Selection has been done. Site models unlike Branch models that require ω>1 over whole sequence and positive selection tends to act only on specific amino acids or domains, allow the ω ratio to vary among sites (among codons or amino acids in the protein) [31]. In other word, models of variable ω ratios among sites were used to test for the presence of sites under diversifying selection (with ω > 1) and to identify them. In this paper, a site refers to an amino acid or codon rather than a nucleotide. Following five models have been used for the ω distribution and implemented in the CODEML program of the PAML package [28], [29]. Model M1 (neutral) assumes two classes of sites in the protein: the conserved sites at which ω = 0 and the neutral sites at which ω =1. Model M2 (selection) adds a third class of sites with ω as a free parameter, thus allowing for sites with ω >1. Model M3 (discrete) uses a general discrete distribution with three site classes, with the proportions (p_0, p_1, and p_2) and the ω ratios (ω_0, ω_1, and ω_2) estimated from the data. Model M7 (beta) uses a beta distribution $B(p, q)$, which, depending on parameters p and q, can take various shapes in the interval (0, 1). Model M8 (beta and ω) adds an extra class of sites to the beta (M7) model, with the proportion and the ω ratio estimated from the data, thus allowing for sites with ω > 1. Data for these models have been given in Table 3 and Table 4 for GP and L proteins respectively. From these models, we construct three LRTs (Tables 5, 6, 7), which compare M0 (one ratio) with M3 (discrete), M1 (neutral) with M2 (selection), and M7 (beta) with M8 (beta & ω). For VP24, VP30, VP35, and VP40 proteins, there were not the considerable difference between LnL (log-Likelihood) value of model M1 (neutral) and M2 (selection) and between LnL value of model M7 (beta) and M8 (beta & ω) then data for these proteins were not reported. When the alternative models (M2, M3, and M8) suggest the presence of sites with ω > 1, all three tests can be considered tests of positive selection [30], [31]. However, the comparison of M0 with M3 may also be considered a test of variable ω values among sites. After ML estimates of parameters are obtained, the Bayes theorem is used to calculate the posterior probabilities of site classes for each site [31]. If the ω ratios for some site classes are>1, sites with high posterior probabilities for those classes are likely to be under diversifying selection.

4. Conclusion

According to likelihood ratios test results in Table 1, for VP24, VP30, VP35, and VP40 proteins 2Δℓ are smaller than Chi-Square distribution 99% and even 95% (considering the degree of freedom). This means free-ratios model does not prefer to one-ratio model and it is pointless to have any conclusion about

variable selective pressure in the phylogeny of these proteins and discussion for even weak evidence of positive selection in these cases is worthless. But for GP, NP, and L proteins, free-ratios model significantly favored and branches with ω>1 have been shown with red lines in Fig.1.   We can presume there is the positive selection in these proteins. Then for a better understanding of positive selection, data in Table 2 and results of Likelihood Ratio Test have been considered. In this table with comparing $2\Delta\ell$ with Chi-Square distribution 95% and ω value in the last column for each protein, we can conclude that for GP protein in branches with label D and J and for L proteins in branches with label A, C, and D there is evidence of positive selection.

In the next step, when we use site models, BEB(Bayes Empirical Bayes)analysis [32] gives positively selected sites for GP that are in the musin like domain, that it is crystallographically disordered and internal fusion loop regions [33], [34]. As it was determined GP is the only viral protein found on the surface of the Ebola virion that is responsible for mediating attachment, fusion, and entry of the virus. The first step of entry is an attachment. Because the mucin domain is the prominent feature on the virion surface, it seems likely that initial attachment steps occur through interactions of this domain with host cell surface factors [35]-[37]. For viruses, the mucin domain was determined to enforce intracellular signaling in dendritic cells, with aiding in replication steps or downstream entry [38]. BEB determined positively selected amino acids sites in musin like domain have changed in terms of their hydropathy, polarity or charge between different species that show different mortality rate. For example amino acid at site 413 in Zaire species is Glutamine but in Bundibugyo and Tai Forest species at the same site, there is Isoleucine and phenylalanine respectively.

Most of BEB positively selected sites (83%) in L protein are in region 1608-1806 (9% of all amino acid sites). In these sites almost always Bundibugyo, Tai Forest, and Sudan species have different amino acids. As mentioned previously these species as well as in branch model have shown evidence of positive selection.

Results of likelihood ratio test using positive selection models give evidence for the relation between the change of amino acids in special sites of proteins structure of Ebola different species and change of mortality rate in these species but it not improved yet.

5. Figures and Tables

Table 1. Likelihood Ratio Test Results for Comparison of One-Ratio and Free-Ratios Models. Chi-Square Distribution 95% and 99% for Specified Degree of Freedom Given in the Last Two Columns Are Used to Test If These Two Models Are Significantly Different

| Protein name | -lnL | | $2\Delta\ell$ | df | $\chi^2_{1\%}$ | $\chi^2_{5\%}$ |
	One-ratio model	Free-ratio model				
VP24	3506.816	3480.876	51.880	74	105.202	95.0815
VP30	4328.535	4283.889	89.292	90	124.116	113.145
VP35	4942.251	4918.138	48.226	72	102.816	92.803
VP40	4757.992	4723.315	69.354	76	107.582	97.351
NP	12127.284	12060.974	132.620	70	100.425	90.531
GP	9515.068	9467.859	94.418	56	83.513	74.468
L	25153.706	25078.825	149.762	46	71.201	62.829

Table 2. Comparison of One-Ratio and Two-Ratios Models by Likelihood Ratio Test (LRT). LnL Value of One-Ratio Model for GP, NP and L Proteins Are 9515.068, -12127.284 and -25153.706 Respectively. In All Cases Degree of Freedom=1. For Each Protein, Labels of Specified Branches Are in First and ω for These Branches Are in Last Column. These Labels Have Been Shown in Fig. 1. ω for All Other Branches Is Same.

GP				NP				L			
label	-LnL two-ratios model	$2\Delta\ell$	ω	label	-LnL two-ratios model	$2\Delta\ell$	ω	label	-LnL two-ratios model	$2\Delta\ell$	ω
A	9511.93	6.266	0.635	A	12127.28	0	0.073	A	25149.38	8.640	∞
B	9515.01	0.100	0.111	B	12127.26	0.044	0.076	B	25151.85	3.700	0.056
C	9514.99	0.152	0.147	C	12126.67	1.214	0.178	C	25144.66	18.09	∞
D	9509.34	11.44	∞	D	12126.02	2.51	7.891	D	25145.24	16.91	∞
E	9514.98	0.164	0.103	E	12127.13	0.302	0.090	E	25148.63	10.14	0.072
F	9514.41	1.314	0.242	F	12126.68	1.194	0.368	F	25151.71	3.986	0.139
G	9513.75	2.632	∞	G	12125.48	3.608	∞	G	25130.77	45.85	0.003
H	9513.75	2.630	∞	H	12125.51	3.544	∞	H	25153.21	0.992	0.166
I	9515.06	0.000	0.138								
J	9513.05	4.034	1.122								
K	9514.88	0.358	0.074								

Table 3. Parameter Estimates and Log-Likelihood Values under Models of Variable ω Ratios among Sites for GP Protein

*- p = number of parameters in the ω distribution

** - d_N/d_S = average over sites

*** - Bayes Empirical Bayes (BEB) analysis [32] Positively Selected Sites: 310 V, 381 L, 403 -, 413 Q, 435 D, 494 T, 496 F, 509 H, 513 A, 564 A. Amino Acids Refer to 1st sequence (GenBank Accession KM655246)

Model	*p	parameters	ℓ	d_N/d_S**	Positively Selected Sites
M0: one ratio	1	$\omega = 0.137$	-9515.068	$=\omega$	None
M1: neutral	1	$p_0 = 0.737$, $\omega_0 = 0.041$ $p_1 = 0.262$, $\omega_1 = 1.000$	-9283.642	0.292	Not allowed
M2: selection	3	$p_0 = 0.737$, $\omega_0 = 0.041$ $p_1 = 0.143$, $\omega_1 = 1.000$ $p_2 = 0.118$, $\omega_2 = 1.000$	-9283.642	0.292	-
M3: discrete	5	$p_0 = 0.307$, $\omega_0 = 0.000$ $p_1 = 0.454$, $\omega_1 = 0.078$ $p_2 = 0.238$, $\omega_2 = 0.774$	-9266.513	0.2201	-
M7: beta	2	$p = 0.268$, $q = 1.1015$	-9274.738	0.1943	Not allowed
M8: beta & ω	4	$p_0 = 0.820$, $p = 0.478$, $q = 6.042$ $P_1 = 0.179$, $\omega = 1.000$	-9266.974	0.2375	10 sites***

Table 4. Parameter Estimates and Log-Likelihood Values under Models of Variable ω Ratios among Sites for L Protein

Model	*p	parameters	ℓ	d_N/d_S**	Positively Selected Sites
M0: one ratio	1	$\omega = 0.043$	-25153.706	$=\omega$	None
M1: neutral	1	$p_0 = 0.874$, $\omega_0 = 0.018$ $p_1 = 0.125$, $\omega_1 = 1.000$	-24615.010	0.142	Not allowed
M2: selection	3	$p_0 = 0.874$, $\omega_0 = 0.018$	-24615.010	0.142	40 sites***

		p_1 = 0.138 , ω_1 = 1.000 p_2 = 0.111 , ω_2 = 1.000			
M3: discrete	5	p_0 = 0.685 , ω_0 = 0.003 p_1 = 0.250 , ω_1 = 0.098 p_2 = 0.064 , ω_2 = 0.704	-24436.128	0.072	-
M7: beta	2	p = 0.160 , q= 2.160	-24448.311	0.064	Not allowed
M8: beta & ω	4	p_0 = 0.965 , p = 0.195 , q= 4.234 P_1 = 0.034 , ω = 1.038	-24435.650	0.074	(NEB) 54 sites**** (BEB) 37 sites*****

*- p = number of parameters in the ω distribution, ** - d_N/d_S = average over sites, *** - Bayes Empirical Bayes (BEB) analysis [32] Positively Selected Sites : 138 L, 338 M, 347 Q, 534 S, 945 E, 1608 Q, 1611 F, 1612 K, 1615 I, 1616 S, 1652 N, 1655 Y, 1659 D, 1664 G, 1666 S, 1675 E, 1676 R, 1678 Q, 1679 E, 1686 H, 1687 D, 1692 N, 1693 L, 17722 -, 1727 H, 1760 F, 1761 Q, 1763 -, 1764 -, 1767 H, 1769 S, 1777 Q, 1794 T, 1797 L, 1800 S, 1805 T, 1806 Q, 2080 R, 2231 Y, 2232 R.Amino Acids Refer to 1st sequence (GenBank Accession KR063671)

**** - Naive Empirical Bayes(NEB) analysis Positively Selected Sites: 138 L, 139 S, 270 N, *338 M*, 347 Q, 534 S, 945 E, 1608 Q, *1611 F*, 1612 K, 1613 T, 1615 I, 1616 S, 1621 P, 1652 N, 1655 Y, 1659 D, **1664 G**, 1665 S, 1666 S, 1667 T, *1675 E*, 1676 R, 1678 Q, 1679 E, 1684 D, 1686 H, *1687 D*, 1692 N, *1693 L*, 1720 Q, 1722 -, 1726 T, 1727 H, 1738 D, *1760 F*, 1761 Q, 1763 -, 1764 -, 1767 H, 1769 S, 1776 H, 1777 Q, *1794 T*, 1796 Q, 1797 L, 1800 S, 1805 T, 1806 Q, 1810 S, 2072 C, 2080 R, 2231 Y, 2232 R. Sites inferred under selection at the 99% level are listed in bold, and those at the 95% level are in italic. Amino Acids Refer to 1st sequence (GenBank Accession KR063671)

*****- Bayes Empirical Bayes (BEB) analysis [32] Positively Selected Sites : 138 L, 338 M, 347 Q, 534 S, 945 E, 1608 Q, 1611 F, 1612 K, 1615 I, 1616 S, 1655 Y, 1659 D, *1664 G*, 1666 S, 1675 E, 1676 R, 1678 Q, 1679 E, 1686 H, 1687 D, 1692 N, 1693 L, 1722 -, 1727 H, 1738 D, 1760 F, 1761 Q, 1763 -, 1764 -, 1767 H, 1769 S, 1777 Q, 1794 T, 1800 S, 1805 T, 1806 Q, 2232 R. Sites inferred under selection at the 95% level are in italic. Amino Acids Refer to 1st sequence (GenBank Accession KR063671)\

### Table 5. Likelihood Ratio Statistics (2Δℓ) for GP Protein

$\chi^2_{5\%}$	$\chi^2_{1\%}$ Comparison			2Δℓ
	df			
M0 (one ratio) vs. M3 (discrete) . . .	497.110	4		13.28
	9.49			
M1 (neutral) vs. M2 (selection) . . .	0.000	2		9.21
	5.99			
M7 (beta) vs. M8 (beta & v)	15.528	2		9.21
	5.99			

Table 6. Likelihood Ratio Statistics (2Δℓ) for L Protein

$\chi^2_{5\%}$	$\chi^2_{1\%}$ Comparison			2Δℓ
	df			
M0 (one ratio) vs. M3 (discrete) . . .	1435.156	4		13.28
	9.49			
M1 (neutral) vs. M2 (selection) . . .	0.000	2		9.21
	5.99			
M7 (beta) vs. M8 (beta & v) . . .	25.322	2		9.21
	5.99			

Table 7. Likelihood Ratio Statistics (2Δℓ) for NP Protein

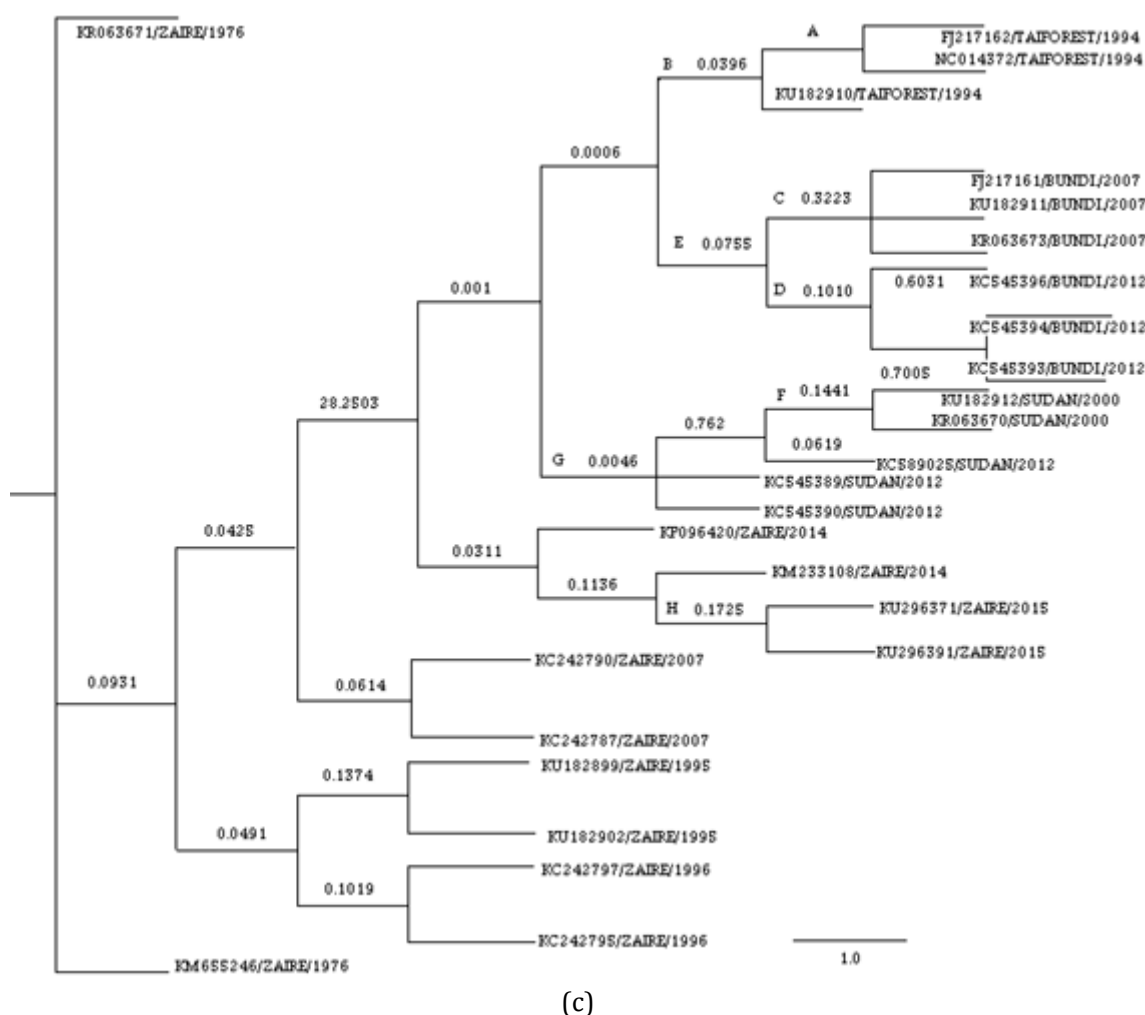| $\chi^2_{5\%}$	$\chi^2_{1\%}$ Comparison			2Δℓ
	df			
M0 (one ratio) vs. M3 (discrete) . . .	1095.654	4		13.28
	9.49			
M1 (neutral) vs. M2 (selection) . . .	0.000	2		9.21
	5.99			
2	9.21 5.99		1.3558	M7
(beta) vs. M8 (beta & v). . .				

(a)

(b)

(c)
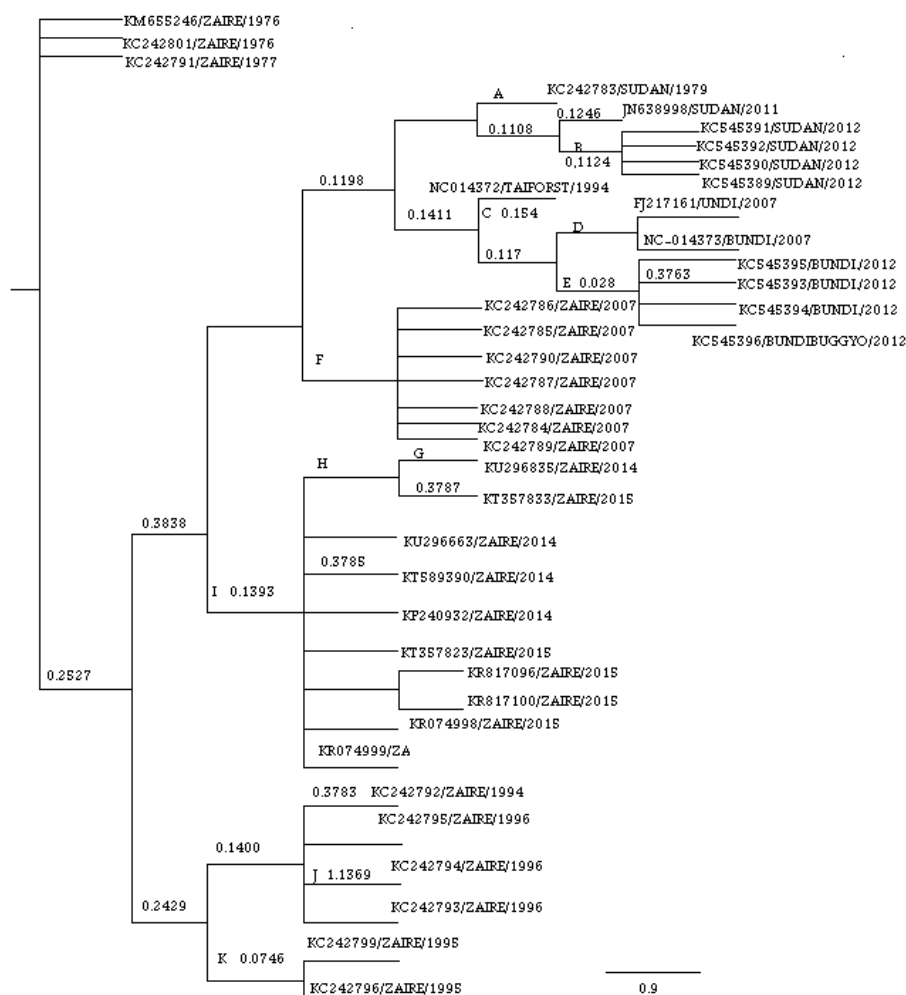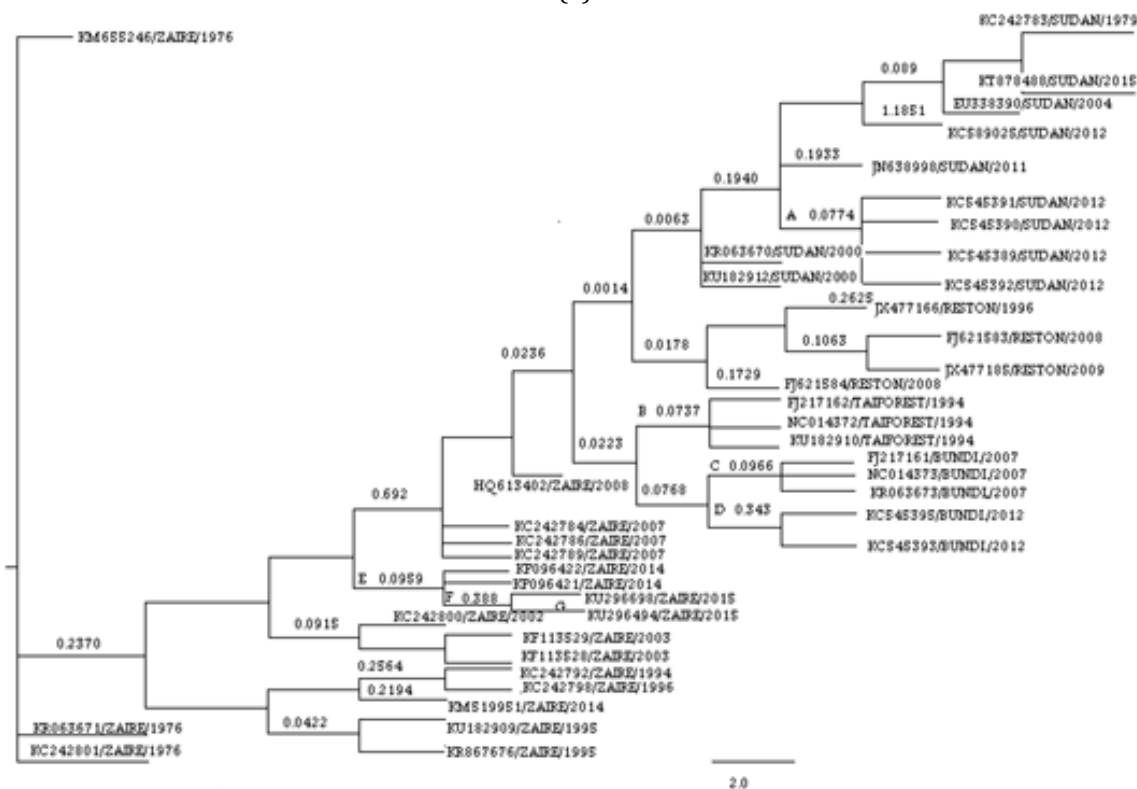
Fig. 1. Phylogeny for three proteins of Ebolavirus genus: a) GP b) NP and c) L.

Unrooted topology is used here. Branches are drawn in proportion to their lengths, defined as the expected number of nucleotide substitutions per codon. Maximum-Likelihood estimates of branch lengths were obtained under the "free-ratios" model, which assumes an independent ω ratio (d_N/d_S) for each branch in the tree. Estimates of the ω ratios under that model are shown along branches. Taxa names were written as GenBank Accession, species name and collection date respectively. Capital letters on branches indicate which branches are specified in Two-ratio model.

References

[1] Nei, M., & Kumar, S. (2000). *Molecular Evolution and Phylogenetics*. Oxford: Oxford University Press.

[2] Miyata, T., Miyazawa, S., & Yasunaga, T. (1979). Two types of amino acid substitutions in protein evolution. *J. Mol. Evol.*, *12*, (219–23).

[3] Li, W. H., Wu, C. I., & Luo, C. C. (1985). A new method for estimating synonymous and nonsynonymous rates of nucleotide substitutions considering the relative likelihood of nucleotide and codon changes. *Mol. Biol. Evol. 2*, 150–174.

[4] Nielsen, R., & Yang, Z. (1998). Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics, 148*, 929–936.

[5] Suzuki, Y., & Gojobori, T. (1999). A method for detecting positive selection at single amino acid sites. *Mol. Biol. Evol., 16*, 1315–1328.

[6] Yang, Z. (2000). Maximum likelihood estimation on large phylogenies and analysis of adaptive evolution in human influenza virus A. *J. Mol. Evol., 51,*423–432.

[7] Goldman, N., & Yang, Z. (1994). A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Mol. Biol. Evol., 11*, 725–736.

[8] Muse, S.V., & Gaut, B. S. (1994). A likelihood approach for comparing synonymous and nonsynonymous nucleotide substitution rates, with application to the chloroplast genome. *Mol. Biol. Evol., 11*, 715–724.

[9] Zanotto, P. M., Kallas, E. G., Souza, R. F., & Holmes, E. C. (1999). Genealogical evidence for positive selection in the nef gene of HIV-1. *Genetics, 153*, 1077–1089.

[10] Bishop, J. G., Dean, A. M., & Mitchell-Olds, T. (2000). Rapid evolution in plant chitinases: Molecular targets of selection in plant-pathogen coevolution. *Proc. Natl. Acad. Sci. USA, 97*, 5322–5327.

[11] Bielawski, J. P., & Yang, Z. (2001). Positive and negative selection in the DAZ gene family. *Mol. Biol. Evol., 18*, 523–529.

[12] Ford, M. J. (2001). Molecular evolution of transferrin: evidence for positive selection in salmonids. *Mol. Biol. Evol., 18*, 639–647.

[13] Haydon, D. T., Bastos, A. D., Knowles, N. J., & Samuel, A. R. (2001). Evidence for positive selection in foot-and-mouth disease virus capsid genes from field isolates. *Genetics, 157*, 7–15.

[14] Swanson, W. J., Yang, Z., Wolfner, M. F., & Aquadro, C. F. (2001). Positive Darwinian selection in the evolution of mammalian female reproductive proteins. *Proc. Natl. Acad. Sci., USA 98*, 2509–2514.

[15] Mondragon-Palomino, M., Meyers, B. C., Michelmore, R. W., & Gaut, B. S. (2002). Patterns of positive selection in the complete NBS-LRR gene family of Arabidopsis thaliana. *Genome Res., 12*, 1305–1315.

[16] Twiddy, S. S., Woelk, C. H., & Holmes, E. C. (2002). Phylogenetic evidence for adaptive evolution of dengue viruses in nature. *J. Gen. Virol., 83*, 1679–1689.

[17] Takebayashi, N., Brewer, P. B., Newbigin, E., & Uyenoyama, M. K. (2003). Patterns of variation within selfincompatibility loci. *Mol. Biol. Evol., 20,* 1778–1794.

[18] Filip, L. C., & Mundy, N. I. (2004). Rapid evolution by positive Darwinian selection in the extracellular domain of the abundant lymphocyte protein CD45 in primates. *Mol. Biol. Evol., 21,*1504–1511.

[19] Lane, R. P., Young, J., Newman, T., & Trask, B. J. (2004). Species specificity in rodent pheromone receptor repertoires. *Genome Res., 14*, 603–608.

[20] Moury, B. (2004). Differential selection of genes of cucumber mosaic virus subgroups. *Mol. Biol. Evol., 21*, 1602–1611.

[21] Anisimova, M., Bielawski, J. P., & Yang, Z. (2001). The accuracy and power of likelihood ratio tests to detect positive selection at amino acid sites. *Mol. Biol. Evol., 18*, 1585–1592.

[22] Anisimova, M., Bielawski, J. P., &Yang, Z. (2002). Accuracy and power of Bayes prediction of amino acid sites under positive selection. *Mol. Biol. Evol., 19*, 950–958.

[23] Wong, W. S. W., Yang, Z., Goldman, N., & Nielsen, R. (2004). Accuracy and power of statistical methods for detecting adaptive evolution in protein coding sequences and for identifying positively selected sites. *Genetics, 168,*1041–1051.

[24] Sanchez, A., Geisbert, T. W., Feldmann, H., (2007).Filoviridae: Marburg and Ebola viruses. *Lippincott Williams and Wilkins*. Philadelphia, PA, USA, 1279–1304.

[25] Feldmann, H., & Kiley, M. P. (1999). Classification, structure, and replication of filoviruses. *Curr. Top. Microbiol. Immunol.*, *235*, 1–21.

[26] Abascal, F., Zardoya, R., Telford, M. J. (2010). TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. *Nucleic Acids Res.*, *38*, 7-13.

[27] Swofford, D. L. (2002). PAUP*. Phylogenetic analysis using parsimony (*and other methods). Sinauer Associates, Sunderland, Massachusetts.

[28] Yang, Z. (1997). PAML: A program package for phylogenetic analysis by maximum likelihood computer applications in biosciences.

[29] Yang, Z. (2000). Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Mol. Biol. Evol., 17,* 32–43.

[30] Ziheng, Y. (1998). Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Mol. Biol. Evol., 15(5),* 568–573.

[31] Nielsen, R., & Yang. Z. (1998). Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics, 148,* 929–936.

[32] Yang, W., & Nielsen (2005). *Mol. Biol. Evol.*, *22*, 1107-1118.

[33] Jeffrey, E. L., & Saphire, O. E., (2009). Ebolavirus glycoprotein structure and mechanism of entry. *Future Virol., 4(6),* 621–635.

[34] Lee, J. E., Fusco, M. H., Hessell, A. J., Oswald, W. B., Burton, D. R., & Saphire, E. O. (2008). Structure of the Ebola virus glycoprotein bound to an antibody from a human survivor. *Nature, 454(7201),* 177–183.

[35] Alvarez, C. P., Lasala, F., Carrillo, J., Muniz, O., Corbi, AL., *et al.* (2002). C-type lectins DCSIGN and L-SIGN mediate cellular entry by Ebola virus in cis and in trans. *J. Virol., 76,* 6841-6844.

[36] Takada, A., Fujioka, K., Tsuiji, M., Morikawa, A., Higashi, N., *et al.* (2004). Human macrophage C-type lectin specific for galactose and N-acetylgalactosamine promotes filovirus entry*. J. Virol., 78,* 2943-2947.

[37] Simmons, G., Reeves, J. D., Grogan, C. C., Vandenberghe, L. H., Baribaud, F., *et al.* (2003). DC-SIGN and DC-SIGNR bind ebola glycoproteins and enhance infection of macrophages and endothelial cells. *Virology*, *305*, 115-123.

[38] Martinez, O., Valmas, C., & Basler, C. F. (2007).Ebola virus-like particle-induced activation of NF-kappaB and Erk signaling in human dendritic cells requires the glycoprotein mucin domain. *Virology, 364,* 342-354.

Maryam Mohammadlou was born in Tabriz in 1983. She received her B. Sc. degree in applied chemistry from Tehran University in 2005 and M. Sc. in physical chemistry from Urmia University in 2008.

She has worked as technical director of a food factory and university lecturer in Payamnoor University of Urmia and Khoy branches and I. Azad University. Researches in the field of polymer and surfactant aquatic equilibrium as super molecules in living cells led to publishing articles during this time.

Ms. Mohammadlou in present time is working as research assistant in I. Azad University in the field of structural change detection of biomolecules