HOSTED BY
ELSEVIER

Contents lists available at ScienceDirect

Asian Pacific Journal of Tropical Medicine

journal homepage: http://ees.elsevier.com/apjtm

Original research http://dx.doi.org/10.1016/j.apjtm.2016.03.016

Genetic diversity in Ebola virus: Phylogenetic and *in silico* structural studies of Ebola viral proteins

Alba Grifoni[1], Alessandra Lo Presti[2], Marta Giovanetti[2,3], Carla Montesano[3], Massimo Amicosante[1,4], Vittorio Colizzi[3], Alessia Lai[5], Gianguglielmo Zehender[5], Eleonora Cella[2,6], Silvia Angeletti[7], Massimo Ciccozzi[2,8*]

[1]*ProxAgen Ltd, Sofia, Bulgaria*

[2]*Department of Infectious Parasitic and Immunomediated Diseases, National Institute of Health, Rome, Italy*

[3]*Department of Biology, University of Rome 'Tor Vergata', Rome, Italy*

[4]*Department of Biomedicine and Prevention, University of Rome 'Tor Vergata', Rome, Italy*

[5]*Laboratory of Infectious Diseases and Tropical Medicine, University of Milan, Italy*

[6]*Department of Public Health and Infectious Diseases, Sapienza University of Rome, Rome, Italy*

[7]*Clinical Pathology and Microbiology Laboratory, University Hospital Campus Bio-Medico of Rome, Rome, Italy*

[8]*University Campus Bio-Medico, Rome, Italy*

ARTICLE INFO

Article history:
Received 15 Jan 2016
Received in revised form 20 Feb 2016
Accepted 1 Mar 2016
Available online 9 Mar 2016

Keywords:
Ebola virus
Proteins
Evolutionary analysis

ABSTRACT

Objective: To explore the genetic diversity and the modification of antibody response in the recent outbreak of Ebola Virus.

Methods: Sequences retrieved from public databases, the selective pressure analysis and the homology modeling based on the all protein (nucleoprotein, VP35, VP40, soluble glycoprotein, small soluble glycoprotein, VP30, VP24 and polymerase) were used.

Results: Structural proteins VP24, VP30, VP35 and VP40 showed relative conserved sequences making them suitable target candidates for antiviral treatment. On the contrary, nucleoprotein, polymerase and soluble glycoprotein have high mutation frequency.

Conclusions: Data from this study point out important aspects of Ebola virus sequence variability that for epitope and vaccine design should be considered for appropriate targeting of conserved protein regions.

1. Introduction

Zaire Ebola virus (EBOV), a member of the Filoviridae family, is a virulent Category A pathogen that causes considerable morbidity and mortality. The EBOV genome is a linear, non-segmented, single-stranded RNA approximately of 19 kb. The virus is filamentous and pleomorphic with a mean unit length of 1 200 nm [1]. The viral genome encodes for a nucleoprotein (NP), a glycoprotein (GP), a RNA dependent RNA polymerase (L), and four structural proteins termed VP24, VP30, VP35 and VP40 [2–4]. The structural proteins, VP40 and VP24, represent viral matrix proteins connecting the nucleocapsid to the viral envelope. NP, VP30 and L proteins are of fundamental importance in the replication and transcription of the Ebola genome [2,5]. The envelope GP is an integral membrane protein, which forms spike-like protrusions on the surface of the virion. Recently, surface GP level regulated by RNA editing mechanism has shown a fundamental role in EBOV pathogenicity and viral immune escape [6].

EBOV causes Ebola viral disease, characterized by fever, malaise, and other nonspecific symptoms such as myalgia, headache, vomiting, and diarrhea. About 30%–50% patients manifest hemorrhagic symptoms. Moreover, in some severe cases multi-organ dysfunction, including hepatic damage, renal failure, and central nervous system involvement occur, leading to shock and death [7]. 'Cytokine storm' with immune suppression of CD4 and CD8 lymphocytes is a candidate mechanism for production of the terminal hemorrhagic fever [8].

EBOV was first identified in 1976 during the epidemic of hemorrhagic fever in Zaire, now Democratic Republic of Congo, with the epicenter of the outbreak in Yambuku. Zaire EBOV appeared again in Democratic Republic of Congo in 1977 near Yambuku and subsequent outbreaks among humans have occurred in west-central Africa in distinct waves during 1994–1997 and 2001–2005 [9]. The recent and ongoing outbreak of EBOV Disease began in December 2013 in forested areas of Southeastern Guinea

*Corresponding author: Massimo Ciccozzi, Department of Infectious Parasitic and Immunomediated Diseases, National Institute of Health, Rome, Italy.
Tel: +39 0 649903187
E-mail: ciccozzi@iss.it
Peer review under responsibility of Hainan Medical College.

affecting additionally the West African countries of Liberia, Nigeria, and Sierra Leone. In Sierra Leone, a total of 8 698 confirmed cases with 3 587 confirmed deaths were reported in the Ebola Situation Report of 2 September 2015, of which 302 (221 deaths) among health care workers [10,11]. A significant decline in both Ebola cases and deaths was observed until April 2015, although, sporadic outbreaks and deaths continue to occur, including infection among health care workers [10]. In Africa, EBOV disease infection have been documented through the handling of infected chimpanzees, gorillas, fruit bats, monkeys, forest antelope and porcupines found ill or dead or in the rainforest. Ebola spreads to the community through person-to-person transmission, with infection resulting from direct contact with the blood, secretions, organs or other bodily fluids, and indirect contact with environments contaminated [12].

Although the knowledge of clinical and pathogenic aspects of Ebola viral disease has recently improved the role of antibody response in viral clearance and protection against EBOV in humans is not fully understood. Fatal EBOV infection is characterized by a defective innate immune response, leading to uncontrolled release of inflammatory mediators and chemokines in the late stage of the disease, and correlates with the collapse of adaptive immunity with massive T and B lymphocyte apoptosis. Immune protection seems to be associated with the development of both cellular and humoral immunity [13–17].

Several amino acid differences have been characterized in the recent Ebola outbreak. A better knowledge of the viral protein structure modifications represent the key point for drug design and vaccination [18].

In this study, the selective pressure analysis was carried out to detect the presence of sites under positive selective pressure that could represent candidate ‘hot spot’ with a crucial rule in the viral escape and evolution. Homology modeling analysis has been performed to evaluate the virus evolution consequences in the protein recognition by host immune response. We previously performed these analyses considering only the contribution of GP protein [18].

In this paper, the study is extended on all EBOV genome transcripts to evaluate new targets for therapeutic and vaccine strategies. Bioinformatics and immune-informatic approaches can provide new insights into the pathogen's evolution, genetic diversity and heterogeneity and the related protective immune response against the virus to evaluate new targets for therapeutic and vaccine strategies.

2. Materials and methods

2.1. Sequence data set and phylogenetic analysis

Seven different data set were built, one for each protein [NP, VP35, VP40, soluble glycoprotein (sGP), VP30, VP24, L] downloading a number of sequences that ranged from 91 for NP and VP40 proteins to 101 to sGP protein. The small soluble glycoprotein (ssGP) was not investigated due to the fact that it is a portion with the same reading frame of the GP, already described [19]. All the sequences with known sampling date and geographical location were obtained from the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/).

The sampling dates for the sequences in the data set ranged from 1976 to 2014. All data sets were used to perform the selective pressure and the homology modeling analysis. All the sequences were aligned using ClustalX software and edited by using Bio-Edit software v. 7.0 [19], The best-fitting nucleotide substitution models were chosen with the hierarchical LRT strategy described by Swofford & Sullivan [20], as implemented in the MODELTEST v. 3.7 program [21].

2.2. Selective pressure analysis

Comparison of relative fixation rates of synonymous (silent) and non-synonymous (amino acid-altering) mutations provide a means for understanding the mechanisms of molecular sequence evolution. The non-synonymous/synonymous rate ratio ($\omega = d_N/d_S$) is an important indicator of selective pressure at the protein level, with $\omega = 1$ meaning neutral mutations, $\omega < 1$ purifying selection, and $\omega > 1$ diversifying positive selection.

The CODEML program implemented in the PAML 3.14 software package (http://abacus.gene.ucl.ac.uk/software/paml.html) [22] was used to investigate the adaptive evolution of the different data set of EBOV.

Six models of codon substitution: M_0 (one-ratio), M_{1a} (nearly neutral), M_{2a} (positive selection), M_3 (discrete), M_7 (beta), and M_8 (beta and omega) were used in this analysis [23]. Since these models are nested, we used codon-substitution models to fit the model to the data using the likelihood ratio test (LRT) [24]. The M_3 model, with three d_N/d_S (ω) classes, allows ω to vary among sites by defining a set number of discrete site categories, each with its own ω value. Through maximum-likelihood optimization, it is possible to estimate the ω and P values and the fraction of sites in the aligned data set that falls into a given category. Finally, the algorithm calculates the a posteriori probability of each codon belonging to a particular site category. Using the M_3 model, sites with a posterior probability exceeding 90% and a ω value > 1.0 were designated as being ‘positive selection sites’ [23]. The site rate variation was evaluated comparing M_0 with M_3, while positive selection was evaluated comparing M_1 with M_2. The Bayes empirical Bayes approach implemented in M_{2a} and M_8 was used instead to determine the positively selected sites by calculating the posterior probabilities of ω classes for each site [25]. It is worth noting that PAML LRTs have been reported to be conservative for short sequences (*e.g.* positive selection could be underestimated), although the Bayesian prediction of sites under positive selection is largely unaffected by sequence length [25,26]. The d_N/d_S rate (ω) was also estimated by the ML approach implemented in the program HyPhy to enforce the previous analysis [27]. Two different algorithms estimated site-specific positive and negative selection: the fixed effect likelihood and random effect likelihood. The fixed effect likelihood fits a ω rate to every site and uses the likelihood ratio to test if $d_N = d_S$. The random effect likelihood is a variant of the Nielsen–Yang approach which assumes that a discrete distribution of rates exists across sites and allows both d_S and d_N to vary independently site by site. The three methods have been described in more detail elsewhere [28]. In order to select sites under selective pressure and keep our test conservative, a P value of ≤ 0.1 or a posterior probability of ≥ 0.9 as relaxed critical values was assumed.

For evolutionary analysis, the reference sequence Accession Number: NC_002549 was used to trace the exact position of the amino acids found under selection.

2.3. Amino acid mutation frequency analysis

Alignment between the Ebola protein reference sequences derived from NC_002549 (NP: P18272, GP: Q05320, VP24:

Q05322, VP30: Q05323, VP35: Q05127, VP40: Q05128, SGP: P60170, SSGP: Q9YMG2) and protein sequences of the new Ebola epidemic have been performed evaluating all the amino acid mutation frequencies and positions.

2.4. Homology modeling

The L protein has been excluded by the homology modeling study because it consists of more than 2 000 amino acid residue and it cannot be modeled. Amino acid mutation with a frequency lower than 0.1 has not been considered. A model of all the Ebola protein, considering the sequences of the new epidemic, has been generated using different homology modeling servers. The servers have been tested using VP24 because its structure has been already resolved. The obtained models have been compared with VP24 crystallography structure (PDB ID:4M0Q) and only the servers able to retrieve an RMSD $\leq$ 0.5 have been chosen for further analyses.

Phyre2 server [29] showed the best performance combining homology modeling and de novo modeling strategies in comparison with other homology modeling servers such as RaptorX, I-Tasser and SWISS-MODEL [30–32] GP protein could not be properly modeled using the different servers. Thus, loop modeling of the GP crystal structure in complex with a neutralizing antibody has been performed (PDB ID: 3CSY). Loops modeling of GP to further perform sGP and ssGP homology modeling experiments have been performed on YASARA (Yet Another Scientific Artificial Reality Application, http://www.yasara.com). All the resulting models obtained have been ‘repaired’ to obtain best protein quality and Alanine scanning has been performed using FoldX tools implemented in YASARA as previously reported [33].

3. Results

3.1. Evolutionary analysis

Genetic variability was determined by nucleotide sequencing of fragments ranging from 753 nt for the VP24 protein to 6 636 nt for the L protein. The Alfa parameter of the gamma distribution for all the protein analyzed was <1, showing as this distribution has a characteristic *L*-shape and suggesting a nucleotide substitution rate heterogeneity across sites.

Likelihood values and parameter estimates obtained from different data sets with site under selective pressure are listed in Table 1. Estimates of the transition/transversion rate ratio (ts/tv) are quite homogeneous among models in each data set and thus are not shown in Table 1.

The average non-synonymous/synonymous substitution rate ratio in VP40 protein ranges from 0.150 to 0.218 whereas in L protein from 0.089 to 0.097. These values suggests that in VP40 protein non-synonymous mutation has about 15–22% much chances as a synonymous mutation of being fixed in the population, whereas in L protein about from 9 to 10% for non-synonymous mutation to be fixed over time (Table 1).

Nevertheless having these proteins a ω ratio <1, the purifying selection dominates their evolution.

However, models that allow for positively selected sites (M_2, M_3 and M_8) significantly fit the data better that their counterpart models for estimating neutral and negatively selected sites, as suggested by the LRT (M_1, M_0 and M_7 respectively). The comparison between these models suggest the presence of positively selected sites with a proportion of 3% for VP40 protein and a small percentage proportion (<1%) for L protein.

Two statistically supported sites, under positive selection, each one with a probability >98%, respectively at amino acidic position 247 (F, L); 324 (I, L, V) were found for VP40 protein. Only one positive site, statistically supported, with a probability >99% at amino acidic position 1610 (F, L), for the L, have been found. Negative selected sites statistically supported, for all the protein of EBOV ranged from 1% to 5.1% for VP35 and L, were respectively found.

3.2. Homology modeling strategy

L amino acid sequence has been analyzed; however, no structure modeling has been performed because the protein has more than 2 000 amino acid residues. Of the overall 2 212 amino acid positions, 53 sites are mutated (2.4%) and, among them, 24 have a frequency higher than 0.1. Structural proteins NP, VP35, VP24, VP30 and glycoproteins sGP and ssGP have been therefore analyzed for mutation frequency higher than 0.1 and

Table 1

Likelihood values and parameters estimates for the selection analysis of the VP40 and L-polymerase gene proteins.

	Model code	lnL	d_N/d_S	Estimates of parameters	Positively selected sites
VP40	M_0 one ratio	−1 862.86	0.190	ω = 0.190	247 (F,L)
	M_1 neutral	−1 853.89	0.150	P_0 = 0.849 79, (P_1 = 0.150 21)	324 (I,L,V)
	M_2 selection	−1 851.09	0.217	P_0 = 0.968 37, P_1 = 0.000 03 (P_2 = 0.031 63) ω_2 = 0.061 60	
	M_3 discrete	−1 851.11	0.217	P_0 = 0.822 54, P_1 = 0.145 82 (P_2 = 0.031 64) ω_0 = 0.061 58, ω_1 = 0.061 58, ω_2 = 4.996 09	
	M_7 beta	−1 854.43	0.200	P = 0.005 00 Q = 0.020 73	
	M_8 beta and ω	−1 851.10	0.218	P_0 = 0.968 76, (P_1 = 0.031 24), P = 6.652 79 Q = 99.000 00, ω = 99.000 00	
L-Pol	M_0 one ratio	−10 899.72	0.089	ω = 0.089 00	1 610 (F, L)
	M_1 neutral	−10 893.33	0.089	P_0 = 0.948 13, (P_1 = 0.051 87)	
	M_2 selection	−10 891.90	0.095	P_0 = 0.997 01, P_1 = 0.000 02 (P_2 = 0.002 99) ω_2 = 0.074 30	
	M_3 discrete	−10 891.97	0.095	P_0 = 0.000 25, P_1 = 0.996 76 (P_2 = 0.002 99) ω_0 = 0.000 03, ω_1 = 0.074 38, ω_2 = 7.121 06	
	M_7 beta	−10 893.69	0.091	P = 0.052 65 Q = 0.521 07	
	M_8 beta and ω	−10 891.97	0.097	P_0 = 0.997 18, (P_1 = 0.002 82), P = 8.224 05 Q = 99.000 00, ω = 7.420 18	

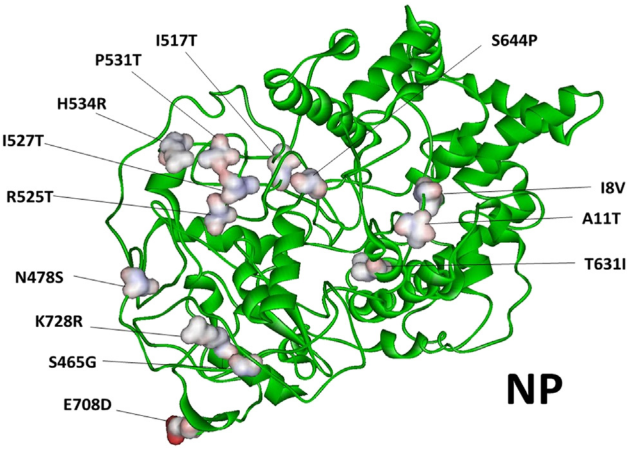

Figure 1. Ebola protein modeled structure of NP and amino acid positions with a mutation frequency higher than 0.1.

Six out of 340 amino acid positions are mutated in VP35 (1.8%). Among them amino acid positions 12, 41, 68, and 204 have a mutation frequency higher than 0.1 and alanine scanning results show that are all stabilizing mutations (Figure 2 and Table 2).

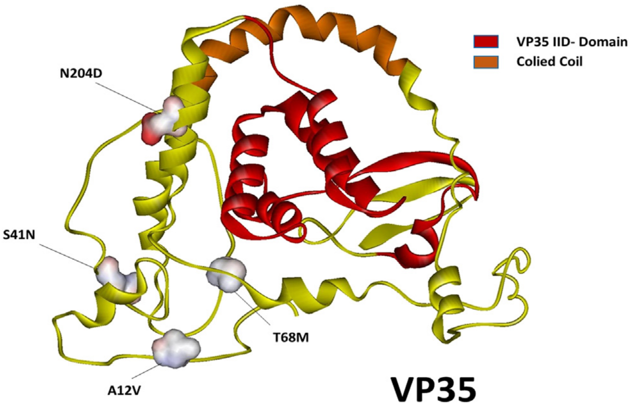

Figure 2. Ebola protein modeled structure of VP35 and amino acid positions with a mutation frequency higher than 0.1.

for the consequences that each amino acid mutated positions could induce on their structures (Figures 1–7 and Table 2).

NP mutation analysis shows that 32 out of 739 amino acid residues (4.33%) are mutated. Among them 13 out of 32 mutation have a frequency mutation higher than 0.1 (Figure 1). NP mutation in amino acid position 8 is a destabilizing mutation, while mutations in amino acid positions 11, 478, 517, 525, 527, 534, 708 are stabilizing mutations. Overall eight out thirteen frequent mutations induce a conformational change in NP structure (Table 2).

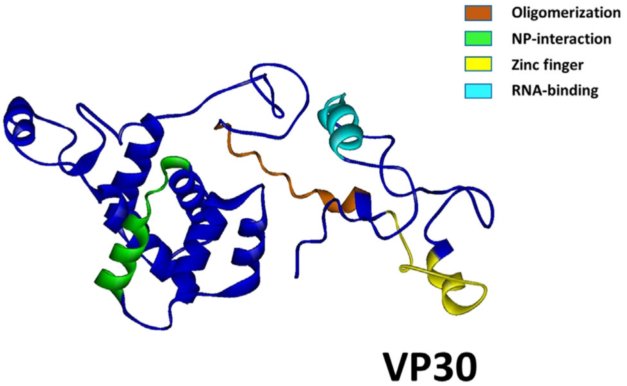

Figure 3. Ebola protein modeled structure of VP30 and amino acid positions with a mutation frequency higher than 0.1.

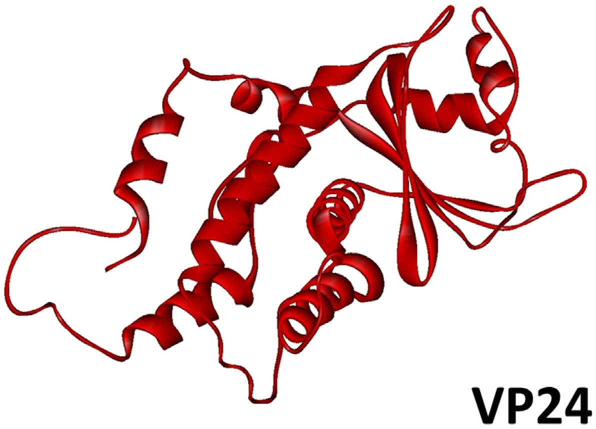

Figure 4. Ebola protein modeled structure of VP24 and amino acid positions with a mutation frequency higher than 0.1.

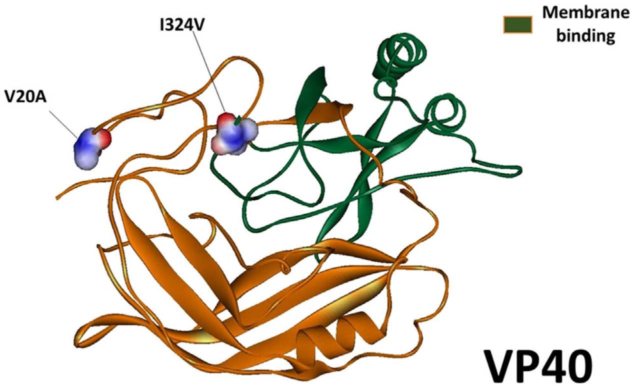

Figure 5. Ebola protein modeled structure of VP40 and amino acid positions with a mutation frequency higher than 0.1.

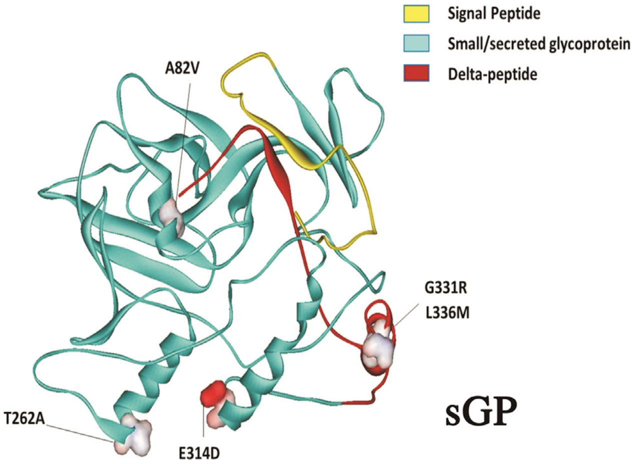

Figure 6. Ebola protein modeled structure of sGP and amino acid positions with a mutation frequency higher than 0.1.

Structural proteins VP24 and VP30 have structures highly conserved (Figures 3 and 4) with a low number of amino acid mutated positions (VP24 = 0.8% and VP30 = 1.5%). None of the amino acid mutations for each position have a frequency higher than 0.1.

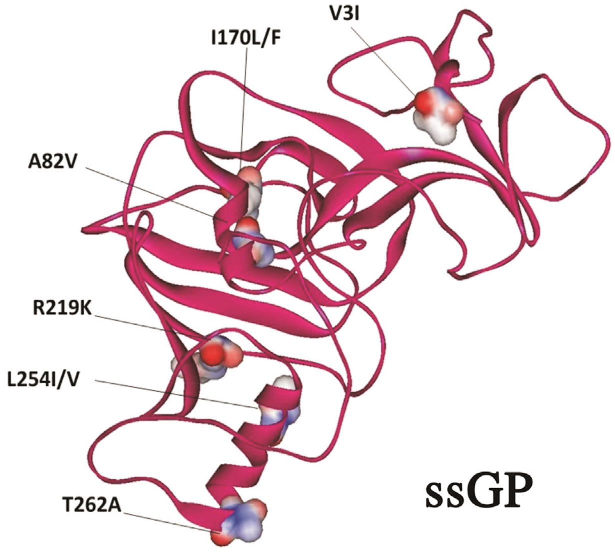

Figure 7. Ebola protein modeled structure of ssGP and amino acid positions with a mutation frequency higher than 0.1.

Seven out of 326 amino acid positions are mutated in VP40 structural protein (2.1%). Among them, amino acid positions 20 and 324 have a mutation frequency higher than 0.1. Alanine scanning analysis shows a stabilizing mutation only in amino acid position 324, which is located in the VP40 membrane-binding region (Figure 5 and Table 2).

5.8% of mutated amino acid positions (21 out of 364), five of them with a frequency higher than 0.1 as shown in Figure 6. Specifically, amino acid positions 82, 262 and 314 are located in the small-secreted GP portion, while amino acid positions 331 and 336 are located in the Δ-peptide. In addition, in amino acid position 314 the mutation is stabilizing while only amino acid positions 331 and 336 have destabilizing mutations (Table 2). Due to sequence overlapping with $GP_{1,2}$, ssGP mutation analysis has been extracted by our previous results on GP [8]. Among them, 6 out of 14 mutations have a frequency mutation higher than 0.1 (Figure 7).

Table 2

Ebola proteins mutated amino acid positions.

Ebola protein	Amino acid position	Reference	Mutation	Mutation frequency	Alanine scan ΔΔG (kCal/mol)
NP	8	*I*	*V*	0.945054945	0.631 007
NP	11	*A*	*T*	0.945054945	−1.946 6
NP	465	*S*	*G*	0.736263736	
NP	478	*N*	*S*	0.736263736	−0.577 056
NP	517	*I*	*T*	0.945054945	−0.862 761
NP	525	*R*	*T*	0.736263736	−2.598 44
NP	527	*I*	*T*	0.945054945	−0.658 525
NP	531	*P*	*T*	0.736263736	
NP	534	*H*	*R*	0.846153846	−0.847 835
NP	631	*T*	*I*	0.736263736	
NP	644	*S*	*P*	0.945054945	
NP	708	*E*	*D*	0.736263736	−0.922 518
NP	728	*K*	*R*	0.736263736	
VP35	12	*A*	*V*	0.858695652	−4.061 62
VP35	41	*S*	*N*	0.858695652	−2.001 55
VP35	68	*T*	*M*	0.858695652	−1.650 1
VP35	204	*N*	*D*	0.945652174	−2.273 44
VP40	20	*V*	*A*	0.769230769	
VP40	324	*I*	*V*	0.823529412	−1.048 45
sGP	82	*A*	*V*	0.455445545	
sGP	262	*T*	*A*	0.653465347	
sGP	314	*E*	*D*	0.514851485	−2.036 51
sGP	331	*G*	*R*	0.653465347	3.873 55
sGP	336	*L*	*M*	0.514851485	1.511 98

Amino acid positions with a mutation frequency higher than 0.1 and relative Alanine scan values are shown. Destabilizing mutation are in bold, stabilizing mutation are in italic. The ΔΔG error margin is approximately 0.5 kCal/mol, so changes in that range are insignificant and have not been shown.

4. Discussion

Many factors may influence the circulation and the genetic evolution of EBOV strains, including Virus infectivity, pre-existing immunity in the population, and antigenic variability of the virus. Antigenic variation, in particular, may play an important role in the ability of these viruses to escape the human immune response.

In this context, the analysis of the EBOV proteins gene evolution is important not only because some of the protein are an important target for the immune response during EBOV infection, but also to better understand the evolutionary dynamic of this virus.

This paper studied the evolution of EBOV proteins using ML techniques. In this analysis only two of the analyzed proteins, showed a high average ratio of non-synonymous to synonymous nucleotide substitution (VP40 and L proteins) ranging from 0.08 to 0.22. As this value is not higher than the threshold of $\omega > 1$, is therefore not indicative of positive selection [19]. Moreover, the alpha parameter value was below 1, meaning that, most of the sites along the gene may be invariable because they are under strong purifying selection. A high proportion of amino acids can be largely invariable, probably because amino acid substitutions are not tolerated or selected for (*i.e.* strong purifying selection). Because the average ω seems usually not sensitive enough to detect Darwinian selection at the molecular level, we used in this case the codon substitution models to detect sites under positive selection. By using these methods, we have identified three sites under positive selection in two different gene proteins (VP40 and L). These results can indicates that although in these proteins a high proportion of amino acids can be largely invariable probably due at structural and functional constraints, adaptive evolution may occur at certain sites of the genome. On the contrary, several amino acid differences have been characterized in the recent Ebola outbreak in EBOV G glycoprotein [18]. A better knowledge of the viral protein structure modifications is the key point for drug design and vaccination. Beside the fact that different studies have been focusing on Ebola GP, also VP24 VP30 VP35 and VP40 structural proteins have been recently investigated as potential drug targets [34–37]. Therefore, the evaluation of the genetic heterogeneity in these proteins assumes an increased importance not only in the viral immune escape but also in the drug design. In this study, structural proteins VP24, VP30, VP35 and VP40 show relatively conserved sequences making them suitable target candidates for antiviral treatment as previously suggested [36,37]. On the contrary, NP, L and soluble GP have high mutation frequency. In this context, NP and GP have been chosen as vaccine candidates and for antibody epitopes design as they are the most abundant protein in the infected cells [38–41], but with the frequency of mutation that characterized them attention should be used in the target region.

In conclusion, this study pointed out important aspects of EBOV sequence variability that for epitope and vaccine design should be taken in consideration for appropriate targeting of conserved protein regions.

Conflict of interest statement

We declare that we have no conflict of interest.

Acknowledgements

This work was supported by National Institute of Health and Proxagent Ltd in collaboration with University of Rome 'Tor Vergata'.

References

[1] Kuhn JH, Andersen KG, Baize S, Bao Y, Bavari S, Berthet N, et al. Nomenclature- and database-compatible names for the two Ebola virus variants that emerged in Guinea and the Democratic Republic of the Congo in 2014. *Viruses* 2014; **6**(11): 4760-4799.

[2] Feldmann H, Kiley MP. Classification, structure, and replication of filoviruses. *Curr Top Microbiol Immunol* 1999; **235**: 1-21.

[3] Feldmann H, Volchkov VE, Volchkova VA, Stroher U, Klenk HD. Biosynthesis and role of filoviral glycoproteins. *J Gen Virol* 2001; **82**(Pt 12): 2839-2848.

[4] Stahelin RV. Membrane binding and bending in Ebola VP40 assembly and egress. *Front Microbiol* 2014; **5**: 300.

[5] Rougeron V, Feldmann H, Grard G, Becker S, Leroy EM. Ebola and marburg haemorrhagic fever. *J Clin Virol* 2015; **64**: 111-119.

[6] Volchkova VA, Dolnik O, Martinez MJ, Reynard O, Volchkov VE. RNA editing of the GP eene of Ebola virus is an important pathogenicity factor. *J Infect Dis* 2015; **212**(Suppl 2): S226-S233.

[7] Azarian T, Lo Presti A, Giovanetti M, Cella E, Rife B, Lai A, et al. Impact of spatial dispersion, evolution, and selection on Ebola Zaire virus epidemic waves. *Sci Rep* 2015; **5**: 10170.

[8] Wauquier N, Becquart P, Padilla C, Baize S, Leroy EM. Human fatal Zaire Ebola virus infection is associated with an aberrant innate immunity and with massive lymphocyte apoptosis. *PLoS Negl Trop Dis* 2010; **4**(10): e837.

[9] Pourrut X, Kumulungui B, Wittmann T, Moussavou G, Delicat A, Yaba P, et al. The natural history of Ebola virus in Africa. *Microbes Infect* 2005; **7**(7–8): 1005-1014.

[10] World Health Organization. *Weekly epidemiological record. Seasonal influenza vaccine composition for tropical and subtropical countries: WHO Expert group meeting*. April 2015, p. 23-24 [Online] Available at: http://www.who.int/wer/2015/wer9036.pdf?ua=1.

[11] World Health Organization. *Ebola situation Report*. 2 September 2015 [Online] Available at: http://apps.who.int/ebola/current-situation/ebola-situation-report-2-september-2015.

[12] Gire SK, Goba A, Andersen KG, Sealfon RS, Park DJ, Kanneh L, et al. Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak. *Science* 2014; **345**(6202): 1369-1372.

[13] Audet J, Kobinger GP. Immune evasion in ebolavirus infections. *Viral Immunol* 2015; **28**(1): 10-18.

[14] Gupta M, Mahanty S, Ahmed R, Rollin PE. Monocyte-derived human macrophages and peripheral blood mononuclear cells infected with Ebola virus secrete MIP-1alpha and TNF-alpha and inhibit poly-IC-induced IFN-alpha in vitro. *Virology* 2001; **284**(1): 20-25.

[15] Takada A, Ebihara H, Jones S, Feldmann H, Kawaoka Y. Protective efficacy of neutralizing antibodies against Ebola virus infection. *Vaccine* 2007; **25**(6): 993-999.

[16] Wilson JA, Hevey M, Bakken R, Guest S, Bray M, Schmaljohn AL, et al. Epitopes involved in antibody-mediated protection from Ebola virus. *Science* 2000; **287**(5458): 1664-1666.

[17] Wong G, Kobinger GP, Qiu X. Characterization of host immune responses in Ebola virus infections. *Expert Rev Clin Immunol* 2014; **10**(6): 781-790.

[18] Giovanetti M, Grifoni A, Lo Presti A, Cella E, Montesano C, Zehender G, et al. Amino acid mutations in Ebola virus glycoprotein of the 2014 epidemic. *J Med Virol* 2015; **87**(6): 893-898.

[19] Ciccozzi M, Babakir-Mina M, Lo Presti A, Farchi F, Zehender G, Ebranati E, et al. Genetic variability of the small t antigen of the novel KI, WU and MC polyomaviruses. *Arch Virol* 2010; **155**(9): 1433-1438.

[20] Wilgenbusch JC, Swofford D. Inferring evolutionary trees with PAUP*. *Curr Protoc Bioinforma* 2003; http://dx.doi.org/10.1002/0471250953.bi0604s00.

[21] Posada D, Crandall KA. MODELTEST: testing the model of DNA substitution. *Bioinformatics* 1998; **14**(9): 817-818.

[22] Yang Z. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci* 1997; **13**(5): 555-556.

[23] Yang Z, Nielsen R, Goldman N, Pedersen AM. Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* 2000; **155**(1): 431-449.

[24] Nielsen R, Yang Z. Likelihood models for detecting positively selected amino acid sites and applications to the *HIV-1 envelope* gene. *Genetics* 1998; **148**(3): 929-936.

[25] Yang Z, Wong WS, Nielsen R. Bayes empirical bayes inference of amino acid sites under positive selection. *Mol Biol Evol* 2005; **22**(4): 1107-1118.

[26] Anisimova M, Bielawski JP, Yang Z. Accuracy and power of bayes prediction of amino acid sites under positive selection. *Mol Biol Evol* 2002; **19**(6): 950-958.

[27] Pond SL, Frost SD, Muse SV. HyPhy: hypothesis testing using phylogenies. *Bioinformatics* 2005; **21**(5): 676-679.

[28] Kosakovsky Pond SL, Frost SD. Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Mol Biol Evol* 2005; **22**(5): 1208-1222.

[29] Kelley LA, Sternberg MJ. Protein structure prediction on the Web: a case study using the Phyre server. *Nat Protoc* 2009; **4**(3): 363-371.

[30] Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, et al. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res* 2014; **42**(W1): W252-W258.

[31] Kallberg M, Wang H, Wang S, Peng J, Wang Z, Lu H, et al. Template-based protein structure modeling using the RaptorX web server. *Nat Protoc* 2012; **7**(8): 1511-1522.

[32] Zhang Y. I-TASSER server for protein 3D structure prediction. *BMC Bioinforma* 2008; **9**(1): 40.

[33] Van Durme J, Delgado J, Stricher F, Serrano L, Schymkowitz J, Rousseau F. A graphical interface for the FoldX forcefield. *Bioinformatics* 2011; **27**(12): 1711-1712.

[34] Audet J, Wong G, Wang H, Lu G, Gao GF, Kobinger G, et al. Molecular characterization of the monoclonal antibodies composing ZMAb: a protective cocktail against Ebola virus. *Sci Rep* 2014; **4**: 6881.

[35] Clinton TR, Weinstock MT, Jacobsen MT, Szabo-Fresnais N, Pandya MJ, Whitby FG, et al. Design and characterization of ebolavirus GP prehairpin intermediate mimics as drug targets. *Protein Sci* 2015; **24**(4): 446-463.

[36] Hernandez H, Marceau C, Halliday H, Callison J, Borisevich V, Escaffre O, et al. Development and characterization of broadly cross-reactive monoclonal antibodies against all known Ebola virus species. *J Infect Dis* 2015; **212**(Suppl 2): S410-S413.

[37] Raj U, Varadwaj PK. Flavonoids as multi-target inhibitors for proteins associated with Ebola virus: in-silico discovery using virtual screening and molecular docking studies. *Interdiscip Sci Com Life Sci* 2015; http://dx.doi.org/10.1007/s12539-014-0246-5.

[38] Dziubanska PJ, Derewenda U, Ellena JF, Engel DA, Derewenda ZS. The structure of the C-terminal domain of the Zaire Ebola virus nucleoprotein. *Acta Crystallogr D Biol Crystallogr* 2014; **70**(9): 2420-2429.

[39] Matassov D, Marzi A, Latham T, Xu R, Ota-Setlik A, Feldmann F, et al. Vaccination with a highly attenuated recombinant vesicular stomatitis virus vector protects against challenge with a lethal dose of Ebola virus. *J Infect Dis* 2015; **212**(Suppl 2): S443-S451.

[40] Tsuda Y, Caposio P, Parkins CJ, Botto S, Messaoudi I, Cicin-Sain L, et al. A replicating cytomegalovirus-based vaccine encoding a single Ebola virus nucleoprotein CTL epitope confers protection against Ebola virus. *PLoS Negl Trop Dis* 2011; **5**(8): e1275.

[41] Wong G, Qiu X, Ebihara H, Feldmann H, Kobinger GP. Characterization of a bivalent vaccine capable of inducing protection against both Ebola and cross-clade H5N1 influenza in mice. *J Infect Dis* 2015; **212**(Suppl 2): S435-S442.