Sequence analysis of the GP, NP, VP40 and VP24 genes of Ebola virus isolated from deceased, surviving and asymptomatically infected individuals during the 1996 outbreak in Gabon: comparative studies and phylogenetic characterization

Eric M. Leroy,1 Sylvain Baize,1 Elie Mavoungou2 and Cristian Apetrei1

1Centre International de Recherches Médicales de Franceville, BP 769 Franceville, Gabon
2Institut für Tropenmedizin, Tübingen-Universität, Wilhelmstrasse 27, 72074 Tübingen, Germany

The aims of this study were to determine if the clinical outcome of Ebola virus (EBOV) infection is associated with virus genetic structure and to document the genetic changes in the Gabon strains of EBOV by sequencing the GP, NP, VP40 and VP24 genes from deceased and surviving symptomatic and asymptomatic individuals. GP and NP sequences were identical in the three groups of patients and only one silent substitution occurred in the VP40 and VP24 genes in asymptomatic individuals. A strain from an asymptomatic individual had a reverse substitution to the Gabon-94 sequence, indicating that minor virus variants may cocirculate during an outbreak. These results suggest that the different clinical outcomes of EBOV infection do not result from virus mutations. Phylogenetic analysis confirmed that Gabon-96 belonged to the Zaire subtype of EBOV and revealed that synonymous substitution rates were higher than nonsynonymous substitution rates in the GP, VP40 and VP24 genes. In contrast, nonsynonymous substitutions predominated over synonymous substitutions in the VP gene of the two Gabon strains, pointing to divergent evolution of these strains and to selective pressures on this gene.

Ebola virus (EBOV) is one of the most virulent human pathogens, killing up to 70–80% of patients within 5–7 days (Khan et al., 1999). EBOV is subdivided into four subtypes, namely Zaire, Sudan, Côte d’Ivoire and Reston. EBOV and Marburg virus (MBGV) are members of the family Filoviridae, order Mononegavirales, which groups together nonsegmented, negative-stranded RNA viruses, namely filoviruses, paramyxoviruses, rhabdoviruses and bornaviruses (Mayo & Pringle, 1998; Sanchez et al., 1993). The EBOV genome displays genes arranged linearly on a single negative-stranded RNA molecule that encodes the following seven structural proteins: nucleoprotein (NP), virion structural proteins (VP) VP35, VP40, glycoprotein (GP), VP30, VP24 and RNA-dependent RNA polymerase (L) (Feldmann et al., 1993). Viral proteins are thought to play a key role in host–virus interactions in EBOV infection and the corresponding genes may thus be subject to selective pressure. In humans, NP and VP40 elicit strong IgG responses (Baize et al., 1999; Leroy et al., 2000). EBOV GP is thought to induce endothelial cell disruption and cytotoxicity in blood vessels (Yang et al., 1998) and to mediate entry into the host cell (Watanabe et al., 2000). EBOV pathogenicity in guinea pigs is associated with mutations in VP24 (Volchkov et al., 2000). Moreover, several studies have shown that filovirus subtypes have different pathogenic potential, causing different mortality rates and disease severity and variable patterns of haemorrhagic effects. Because of a marked genetic divergence between subtypes, the lower pathogenic potential of the Sudan and Reston strains, relative to the Zaire strains, is thought to be due to mutations in the viral proteins (Bowen et al., 1980; Fisher-Hoch et al., 1992). Taken together, these data suggest that any mutation of the virus, particularly in GP, NP, VP40 or VP24, may account for the different clinical outcomes during natural human EBOV infection and that mutations in any of these proteins can influence virulence. We determined the complete open reading frame (ORF) sequences of these genes in EBOV isolates from deceased patients, surviving symptomatic patients and infected asymptomatic individuals during

Author for correspondence: Eric Leroy.
Fax +241 67 72 95, e-mail leroy@cirmf.sci.ga
Table 1. Synthetic oligonucleotides used for PCR and sequencing

Oligonucleotides are shown in the 3′ → 5′ direction and the length (bp) of the fragment amplified from each gene is also shown. To amplify the GP (GPa and GPb) and NP (NPa and NPb) genes, two overlapping segments representing almost the entire ORF were used.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Sequences 3′ → 5′</th>
<th>Length (bp)</th>
</tr>
</thead>
</table>
| GPa (first-round) | GATGAAGATTAAAGCGGACAGTGA  
TTGTCGTTCTCGGTTGATGTTG  
TGCAATGTTCAAGTACACAGTC  
AAGAGATTAAAGCGGACAGTGA  |
| GPb (first-round) | GTGAGCGTAATCTTCATCTCTCTTA  
TTGTTCACTTGAATTGCTCATGAG  
CAAGGAAGGCGAAGCTGCAGTGTCG  
GAATCACATTGGCATTGTTTAAAGC  |
| GPa (nested)  | GTGAGCGTAATCTTCATCTCTCTTA  
TTGTTCACTTGAATTGCTCATGAG  
CAAGGAAGGCGAAGCTGCAGTGTCG  
GAATCACATTGGCATTGTTTAAAGC  |
| GPb (nested)  | GTGAGCGTAATCTTCATCTCTCTTA  
TTGTTCACTTGAATTGCTCATGAG  
CAAGGAAGGCGAAGCTGCAGTGTCG  
GAATCACATTGGCATTGTTTAAAGC  |
| NPa (first-round) | GACAGTTTCCTTTCATATGC  
CCGGTTTCTTCCAGTACTCCT  |
| NPb (first-round) | CATCCCTGGTTCCAGTACTCCT  
ATTCTTAGAACTTCCATC  |
| NPa (nested)  | TTACAAACTTTTTCTGGAAA  
GCGAGTTACCGACCGATGTA  |
| NPb (nested)  | CATCCCTGGTTCCAGTACTCCT  
ATTCTTAGAACTTCCATC  |
| VP40 (first-round) | TACCCCTGAGTACAGAGTGC  
ATGCCGCCATTGGGAGATAAAG  |
| VP40 (nested) | TACCCCTGAGTACAGAGTGC  
ATGCCGCCATTGGGAGATAAAG  |
| VP24 (first-round) | ATTTAATTCATTGCCAGTTG  
ATTGAATTTAAAGGTGCTTACTC  |
| VP24 (nested) | ATTTAATTCATTGCCAGTTG  
ATTGAATTTAAAGGTGCTTACTC  |

The 1996 Ebola outbreaks in Gabon. The aims of this study were to investigate if genetic modifications in any of these four proteins could explain the different clinical outcomes in humans and to examine the phylogenetic features and profiles of the Gabon-94 and -96 strains.

Blood samples were obtained from nine individuals selected randomly during the EBOV outbreak in Booué, Gabon. These nine individuals comprised three symptomatic patients who died, three symptomatic patients who survived (Baize et al., 1999) and three asymptomatic infected individuals (Leroy et al., 2000). In each group, one viral gene sequence was determined in a sample obtained at the beginning of the epidemic, while the other two were determined from samples obtained later during the epidemic (Georges-Courbot et al., 1997). Samples were obtained with the patients’ verbal, informed consent and all specimens were drawn and manipulated according to the WHO guidelines on virus haemorrhagic fever agents in Africa (WHO, 1985). Blood samples, collected on EDTA, were transported on ice to CIRMF (Centre International de Recherches Médicales de Franceville, Gabon) for analysis. Peripheral blood mononuclear cells (PBMCs) were separated from whole blood by Ficoll–diatrizoate density-gradient centrifugation. Total RNA was extracted from PBMCs and first-strand cDNA was synthesized as described previously (Leroy et al., 2000). Amplifications of cDNA were carried out as described previously (Leroy et al., 2000) using the primers listed in Table 1. PCR products were then excised from the gel and extracted using the QIAquick Gel Extraction kit (Qiagen). VP24 fragments were sequenced on an ALF Express DNA sequencer (Pharmacia Biotech) using the Autocycle 200 Sequencing kit (Pharmacia Biotech). GP, NP and VP40 fragments were sequenced by ACTgene Laboratories (France) using the ABI 373A or 377 Automated Sequencer (Applied Biosystems). The CLUSTAL W (Thompson et al., 1994) profile alignment option was used to align the newly sequenced GP and NP genes of EBOV Gabon-96 to the filovirus nucleotide sequences of GP and NP that are available in GenBank. Gap-containing sites were removed prior to all analysis. Pairwise distances were calculated using the DNADIST program of the PHYLIP package (Felsenstein, 1993) using the Kimura two-parameter model of nucleotide substitution. Phylogenetic trees were constructed using both the neighbour-joining (NJ) and the maximum-likelihood (ML) methods. All phylogenetic trees were inferred using PAUP.
The sequences of the GP, NP, VP40 and VP24 genes of the EBOV strain isolated during the 1996 epidemic in Bouée, Gabon are available from GenBank under accession numbers AY058898, AY058895, AY058896 and AY058897, respectively.

PCR amplifications yielded DNA fragments of 2174 (GP), 1767 (NP), 1069 (VP40) and 853 (VP24) bp in size. Under the same conditions, no EBOV DNA was amplified from cells of uninfected humans. No differences in the nucleotide sequences of the GP and NP genes (2174 and 1767 nucleotides, respectively) were found among the three categories of patients. The VP40 gene showed only one synonymous substitution (C → T) in strains from the three asymptomatic individuals compared to those from the symptomatic individuals and the Zaire-76 and -95 strains. Analysis of the VP24 gene only revealed a G → A reverse mutation to the Gabon-94 strain in one asymptomatic individual. We then sequenced VP24 sequences from two more asymptomatic subjects (five in total) and found that these two individuals had identical sequences to those of the six symptomatic patients and the other two asymptomatic individuals.

The deduced GP amino acid sequences were compared to those of all known EBOV sequences. The mean genetic diversity between Bouée-96 and the other Zaire subtype strains was much lower (1–2%) than that between Bouée-96 and the other three subtypes (42–71%), at both the nucleotide and amino acid levels, suggesting that Bouée-96 belongs to the Zaire subtype (Fig. 2). Analysis of multialignment sequences in the sequenced region of the GP gene (2174 bp) showed only 36 nucleotide substitutions and 15 amino acid changes between Bouée-96 and Mayinga-76 (Fig. 1a). Most of these mutations were located in the middle of the GP gene, which encodes a variable region that seems to be subtype-specific. In this hypervariable region of 180 nucleotides (60 amino acids), we have observed 14 nucleotide substitutions (including 10 nonsynonymous) between Bouée-96 and Mayinga-76 and three nucleotide substitutions (all nonsynonymous) between the two Gabon strains. Almost 40% of substitutions are located in less than 9% of the GP gene sequence. The two Gabon strains form a subcluster within the Zaire subtype (Fig. 1). The nucleotide distance in the GP gene is only 0.39% (eight substitutions) between Bouée-96 and Gabon-94, while the nucleotide distance with Mayinga-76 and Kikwit-95 is 1.79% (36 substitutions) and 1.04% (21 substitutions), respectively. At the amino acid level, the genetic distance between Bouée-96 and Gabon-94 is only 0.99%, while the genetic distance with Mayinga-76 and Kikwit-95 is 2.14% and 0.99%, respectively. The sequenced region of the GP gene presented in Fig. 1(a) contains 18 N-glycosylation sites. Only one site, located in the hypervariable region, had changed from N–D–S to N–A–S between the Zaire and Gabon strains.

Amino acid alignment (Fig. 1b), phylogenetic tree construction (Fig. 2) and genetic distance calculations for the NP gene and its product were consistent with those of GP, showing that the Bouée-96 strain belongs to the Zaire subtype. However, the genetic diversity observed between EBOV NP subtypes was lower than that found in the analysis of GP. The mean nucleotide distances in the NP gene between the Zaire and Sudan subtypes are around 30%, compared to 70% in the GP gene. The mean amino acid distances in NP between the same subtypes are around 45%, compared to 65–66% in GP. The variability between subtypes is located mainly in the N-terminal part of the NP gene (Fig. 1b). Substitutions were not located in a particular region but were observed throughout the NP gene and its product (Fig. 1b). NP amino acid distances between the Zaire and Sudan subtype strains are much higher than nucleotide distances (45 versus 30%), in contrast with those in the GP gene (66 versus 72%). In contrast with GP, genetic distances in the NP gene between Bouée-96 and Gabon-94 are similar to, or higher than, those between Bouée-96 and the other Zaire subtype strains. This nucleotide distance is 0.91% (18 substitutions) between Bouée-96 and Gabon-94, 0.74% between Bouée-96 and Mayinga-76 and only 0.51% between Bouée-96 and Kikwit-95. Genetic variability in NP between the two Gabon strains is high (127%), relative to that between Bouée-96 and Mayinga-76 or Kikwit-95 (0.95 and 0.63%, respectively). This higher variability in NP between the Gabon strains compared with other viral proteins seems to be a feature of the ‘Gabon cluster’.

One goal of this study was to determine whether the clinical outcome of EBOV infection is associated with a particular virus genetic structure. No difference was found in the nucleotide sequences of the GP and NP genes among strains from fatalities, survivors and asymptomatic individuals. Moreover, we identified only one synonymous substitution in the VP40 gene and another in the VP24 gene between symptomatic patients and asymptically infected individuals. Taken together, these results show that GP, NP, VP40 and VP24 from the three categories of patients are identical and suggest that fatalities, survivors and asymptomatic individuals were infected with the same virus. Thus, the different clinical outcomes of EBOV infection in humans does not seem to be due to virus mutations. These data are consistent with those obtained during the Kikwit outbreak in 1995, when no substitutions were identified in the most divergent region of the GP gene (249 nucleotides) between isolates from fatalities and isolates from survivors. Moreover, no substitutions were identified in the same GP gene region of isolates from convalescing patients (Rodriguez et al., 1999). The strong conservation of virus genetic structure observed in the three groups of EBOV-infected individuals in our study suggests that differences in clinical outcome are related to host immune reactivity. Despite this high genetic stability, sequencing of the genes encoding L, VP30 and VP35 will be necessary to confirm that the three groups of individuals were infected by the same virus. Also, these proteins could also be involved in EBOV pathogenesis, as VP35 has been described...
Fig. 1. For legend see facing page.
Evolution of Ebola virus in Gabon

Fig. 1. (a) Predicted amino acid sequences of the GP gene of all known EBOV strains. The reference sequence of EBOV (Mayinga-76) and that of the Yambuku outbreak strain in the Democratic Republic of Congo (DRC) in 1976 are shown at the top. The N-glycosylation sites, mapping as NXT (where X can be any amino acid residue except proline or tryptophan), are boxed. Amino acid changes between the reference sequence and Boué-96 are shaded in grey. (b) Predicted amino acid sequences in the NP gene of all known EBOV strains. The reference sequence of EBOV (Mayinga-76) and that of the Yambuku outbreak strain in DRC are shown at the top. Amino acid changes between the reference sequence and Boué-96 are shaded in grey. Dots indicate residues that are identical to the reference sequence and dashes indicate gaps introduced for optimal alignment. Amino acids are numbered according to their position in the sequence.

as a type I interferon antagonist (Basler et al., 2000) and VP30 and L are part of the replicase complex (Mühlberger et al., 1999).

Phylogenetic analysis of the VP24 gene showed only one and eight substitutions in the Boué-96 sequences from the three deceased patients, the three survivors and the four asymptomatic individuals relative to Gabon-94 and Zaire-76, respectively, indicating a low genetic variability in this gene. Interestingly, another asymptomatic individual had exactly the same VP24 sequence as the Gabon-94 strain. This indicates that several virus variants can cocirculate during the same Ebola outbreak and that the variant corresponding to this asymptomatic individual is minor. The low genetic variability of EBOV was shown further by analysis of the NP and GP genes. The high mutation rate between Boué-96 and the Côte d’Ivoire, Sudan and Reston subtypes, and the low mutation rate between Boué-96 and the other Zaire subtype strains demonstrates that Boué-96 belongs to the Zaire subtype.

One feature of members of the family Filoviridae is the contrast between the high genetic diversity between subtypes and the low intrasubtype variability. Indeed, the Boué strain diverges from other strains of the Zaire subtype by only 1–2%, despite the fact that Boué-96 and Zaire-76/95 were isolated 20 years apart and more than 3000 km apart. This is consistent with another study of the most variable 249 nucleotide region of the GP gene (Rodriguez et al., 1999). In contrast to most RNA viruses, EBOV is characterized by high genetic stability, which may be due to four main factors: low error rate of RNA polymerase, slow replication in the natural host, small number of natural hosts and weak immunological pressure. In contrast
Fig. 2. Phylogenetic trees of EBOV strains for the (a) GP and (b) NP gene sequences presented in Fig. 1 (a) (GP) and (b) (NP). Analysis was based on amino acid sequences and the trees were constructed by the ML method. Distance matrix-based trees were constructed with the NJ method (Saitou & Nei, 1987) using the HKY85 model of nucleotide substitution. Confidence estimates were assessed using 1000 bootstrap replicates (implemented using SEQUEN, NEIGHBOR and CONSENS from the PHYLIP package), with JTT distance matrices generated using PUZZLEBOOT (http://www.tree-puzzle.de). ML analysis was carried out with the PROTML program of the MOLPHY package using the JTT empirical matrix model of amino acid substitution (Jones et al., 1992); the trees presented are those with the highest likelihood score. ML trees were also constructed with the NUCML program of the MOLPHY package, using the HKY85 model of nucleotide substitution incorporating a transition/transversion ratio of four and an \( \alpha \) shape parameter (using eight discrete categories of rate variation) for a gamma distribution of rate heterogeneity among sites, both estimated on the NJ tree. The NJ tree topology was used as the starting tree in a heuristic search using tree bisection–reconnection branch swapping to search for the ML tree. MBGV sequences (Musoke and Popp strains) were used as the outgroup strains. Vertical branches are for clarity only; the lengths of the horizontal branches are proportional to the single-base changes. Sequences used in tree construction are available from GenBank. MBGV strains are indicated as MAR. EBOV strains are indicated as EBO-Cl (Côte d’Ivoire), -Z (Zaire), -S (Sudan) and -R (Reston).

with other genes, the rate of nonsynonymous substitutions between the two Gabon strains in the NP gene is significantly higher than that between Bouvé-96 and Mayinga-76 and between Bouvé-96 and Kikwit-95 (1.27 versus 0.95 and 0.63, respectively). This comparatively high level of nonsynonymous substitutions between the two Gabon strains in the NP gene relative to other genes points to selective constraints on NP. It has been shown that when nonsynonymous substitutions predominate over synonymous substitutions in a viral gene, this gene is subjected to positive selection (Hughes & Nei, 1988). This positive Darwinian selection is driven generally by host immune pressure (Fitch et al., 1991; Hughes & Nei, 1989). This particular NP pattern observed in the two Gabon strains of EBOV differs from the evolutionary patterns of most RNA virus genes, in which synonymous substitutions predominate strongly over nonsynonymous substitutions (the neutral theory of molecular evolution) (Gojobori et al., 1990). Two main hypotheses may be proposed to explain this particular NP pattern: (i) the two Gabon strains may have diverged much earlier than suggested by GP analysis (Suzuki & Gojobori, 1997) and (ii) the two Gabon strains diverged recently, but the separation was accompanied by a change in the natural host (species or genus), leading to different immunological pressures on the NP gene. Nevertheless, the coexistence of two different strains, without apparent genetic recombination, in similar geographical and temporal conditions may be possible if the viruses have their own ecological niche. These features of EBOV genetic structure may have implications for the development of treatments and vaccines against EBOV infection. The NP gene pattern of nonsynonymous nucleotide substitutions suggest that NP is subject to functional constraints that induce relatively rapid molecular evolution under immunological pressure. For this reason, the NP gene is a potential candidate target for antiviral
Evolution of Ebola virus in Gabon


References


Received 5 July 2001; Accepted 31 August 2001