Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.

VIROLOGY 240, 138–146 (1998)
ARTICLE NO. VY978902

Variation in the Glycoprotein and VP35 Genes of Marburg Virus Strains

Anthony Sanchez,*[,1] Sam G. Trappier,* Ute Ströher,† Stuart T. Nichol,* Michael D. Bowen,* and Heinz Feldmann†

*Special Pathogens Branch, Division of Viral and Rickettsial Diseases, National Center for Infectious Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia; and †Institute of Virology, Philipps University, Marburg, Germany

Received August 18, 1997; returned to author for revision September 15, 1997; accepted October 16, 1997

Marburg virus, the prototype of the family *Filoviridae*, differs genetically, serologically, and morphologically from Ebola viruses. To better define the genetic variation within the species, VP35 and glycoprotein (GP) genes of representative human isolates from four known episodes of Marburg virus hemorrhagic fever were analyzed. The percentage nucleotide differences in the GP gene coding regions of Marburg viruses (0.1–21%) was nearly equal to the percentage amino acid changes (0–23%), while the percentage nucleotide differences in VP35 coding regions (0.3–20.9%) were higher than the percentage amino acid changes (0.9–6.1%), indicating a greater number of nonsynonymous changes occurring in the GP gene. The higher variation in the GP gene and the corresponding protein, especially those changes in the variable middle region of the GP, suggests that the variability may be the result of responses to natural host pressures. Analysis of the GP gene open reading frame shows a nonrandom distribution of nonsynonymous mutations that may indicate positive Darwinian selection is operating within the variable region. A heptad repeat region and an adjoining predicted fusion peptide are found in the C-terminal third of Marburg virus GPs, as has been previously shown for Ebola virus, and are similar to those found in transmembrane glycoproteins of retroviruses, paramyxoviruses, coronaviruses, and influenza viruses. Comparative analyses showed that there are two lineages within the Marburg virus species of filoviruses. The most recent isolate from Kenya (1987) represents a separate genetic lineage within the Marburg virus species (21–23% amino acid difference). However, this lineage likely does not represent a separate Marburg subtype, as the extent of divergence is less than that separating Ebola virus subtypes.

INTRODUCTION

Marburg (MBG) viruses are nonsegmented, negative-strand RNA viruses, and together with Ebola (EBO) viruses constitute the family *Filoviridae* (Jahrling *et al.*, 1995). Filoviruses can cause an extremely severe form of hemorrhagic fever in humans and/or nonhuman primates and are classified as Biosafety Level 4 agents (Centers for Disease Control and Prevention, 1993; Feldmann *et al.*, 1996; Peters *et al.*, 1996). The natural hosts have not been identified, and thus the factors that influence their evolution and ecology in the wild, and lead to human and nonhuman primate infections, are unknown.

Recent molecular studies have shown that these RNA viruses are similar in their genome organization and mechanisms of replication, but represent genetically distinct viruses (Feldmann *et al.*, 1992, 1996; Sanchez *et al.*, 1993, 1996; Peters *et al.*, 1996; Georges-Courbot *et al.*, 1997; Volchkov *et al.*, 1997). These differences are in agreement with studies that have shown a complete lack of serologic cross-reactivity, contrasting structural protein profiles, and differences in virion morphology between MBG and EBO viruses (Kiley *et al.*, 1988; Ksiazek *et al.*, 1992; Peters *et al.*, 1993, 1996; Feldmann *et al.*, 1994, 1996; Geisbert *et al.*, 1996). Isolates of MBG virus have tended to be fairly homogeneous, whereas EBO virus isolates appear more heterogeneous; four different EBO virus subtypes have been identified thus far (Sanchez *et al.*, 1996). The heterogeneity of EBO viruses has also been demonstrated at the genetic level through sequence analysis of their glycoprotein (GP) genes (Sanchez *et al.*, 1996). The nucleotide sequence differences between EBO subtypes varies from 37 to 41% (compared with a 55% difference between EBO and MBG viruses). For both MBG and EBO viruses, the GP genes are positioned fourth (3′ to 5′) in a series of seven linearly arranged genes and are the only genes to encode viral glycoproteins. The GP genes of EBO viruses were shown to have an unusual organization and to encode a nonstructural secreted glycoprotein (SGP) as the primary gene product in a single frame and the structural GP in two frames (expressed through transcriptional editing) (Volchkov *et al.*, 1995; Sanchez *et al.*, 1996). This unusual organization is absent in the GP genes of previously analyzed MBG viruses (Will *et al.*, 1993; Bukreyev *et al.*, 1995).

The VP35 genes of these viruses may also serve as a potential region of variability in MBG strains, since their gene products have been previously reported to differ in rates of migration in denaturing gels (Feldmann *et al.*,

[1] To whom correspondence should be addressed at 1600 Clifton Road, Building 15, Room SB611, Mail Stop G14, Atlanta, Georgia 30333. Fax: (404) 639-1118. E-mail: ans1@cdc.gov.

0042-6822/98

1994). The filovirus VP35 gene product is a structural protein that is not membrane associated and probably functions as a cofactor in the transcription complex. Its position (second) in the genome makes it analogous to the P proteins of paramyxoviruses and rhabdoviruses, but its function has not been conclusively identified. Analysis of this gene may supply additional information about the evolution of MBG virus, since environmental constraints on the biology of the VP35 protein are likely to be distinct from those affecting the GP.

To better define the genetic variation within MBG viruses, we have analyzed representative strains from the four known episodes of MBG virus hemorrhagic fever in humans. We have examined the GP and VP35 genes of these MBG virus isolates to expand on earlier phylogenetic analyses, which primarily focused on EBO viruses (Sanchez *et al.,* 1996; Georges-Courbot *et al.,* 1997; Volchkov *et al.,* 1997). We present here the results of these studies and a discussion of the evolution of filoviruses.

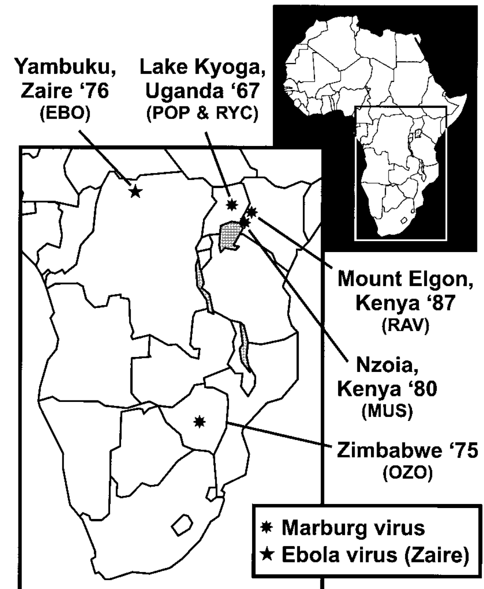

FIG. 1. African locations where isolated strains of MBG viruses and a Zaire species of EBO (used for phylogenetic analyses) are believed to have originated. The first identified outbreak of MBG virus disease in humans (1967) was traced to infected monkeys trapped near Lake Kyoga, Uganda, which were shipped to Germany (POP and RYC strains) and the former Yugoslavia (Siegert *et al.,* 1967). The second outbreak began when an Australian man became infected while traveling through what is now Zimbabwe, and who was later treated in Johannesburg, South Africa, where transmission led to two secondary cases (Gear *et al.,* 1975). The last two episodes occurred in Kenya in 1980 and 1987. The first of these is believed to have originated when a Frenchman working in Nzoia became infected and was transported to Nairobi for treatment, where he died and transmitted the disease to a Nigerian physician (MUS strain) (Smith *et al.,* 1982). The last episode (1987) involved a 15-year-old Danish boy (RAV strain) who was staying in Kisumu (located on the shore of Lake Victoria), but who visited and explored Kitum Cave in Mount Elgon 9 days prior to disease onset (Johnson *et al.,* 1997). Yambuku, Zaire is the site where the first identified outbreak of human disease caused by the Zaire species of Ebola virus occurred in 1976 (MAY strain).

MATERIAL AND METHODS

Viruses and passage histories

The MBG virus isolates used to generate sequence information for this study are as follows: (1) an isolate from the original 1967 outbreak from Marburg, Germany, strain M/Germany/Marburg/1967/Ratayczak (RYC strain) (Siegert *et al.,* 1967; Feldmann *et al.,* 1994), passaged three times in Vero E6 cells (ATCC CRL 1586); (2) a 1975 isolate from a South African case, strain M/S. Africa/Johannesburg/1975/Ozolin (OZO strain) (Gear *et al.,* 1975; Kiley *et al.,* 1988), passaged three times in Vero 76 cells and four times in Vero E6 cells; (3) a 1987 isolate from Kenya, strain M/Kenya/Kitum Cave/1987/Ravn (RAV strain) (Johnson *et al.,* 1997), passaged once in SW13 cells and three times in Vero E6 cells. Sequence information for the M/Germany/Marburg/1967/Popp (POP strain; passaged nine times in guinea pigs) and M/Kenya/Nairobi/1980/Musoke strains (MUS strains; isolated, plaque-purified three times, and passaged two to three times thereafter in Vero E6 cells) of MBG virus, and the E/Zaire/Yambuku/1976/057935/Mayinga strain (MAY strain; isolated, plaque-picked three times, and passaged two to four times thereafter in Vero E6 cells) of EBO virus (Zaire species) were previously determined or obtained from GenBank (Bukreyev *et al.,* 1995; Feldmann *et al.,* 1992; Sanchez *et al.,* 1993). Figure 1 shows the geographical locations where infections by the above viruses are believed to have originated.

Purification of viral RNA and RT-PCR

Virions were purified from clarified supernatant fluids by pelleting through a 20% sucrose cushion (Feldmann *et al.,* 1994), and genomic RNA was extracted as described elsewhere (Sanchez and Kiley, 1987) or with a commercial kit (RNaid Kit, BIO 101, Inc.). For examination of VP35 gene sequences, first-strand DNA synthesis was performed using the primer 5′-GCCTAGATGATAACAGATAT (plus-sense within NP gene and near polyadenylation site) and random primers (hexamer) as previously described (Sanchez *et al.,* 1989). DNA produced in reverse transcription (RT) reactions were diluted 1:50 in water and used as template in polymerase chain reaction (PCR) assays. For amplification of the entire VP35 genes, either the primer 5′-CTGTCTCAGCTAAGGAGCTTCA or 5′-ATGATTCATATTATAAGGTAC (NP gene, plus-sense) and primer 5′-GACGGTACCTTCCAGTAAAAGAACAC-TACGA were used with appropriate internal primers in PCR assays. Amplification was performed by mixing 10 μl of 10× reaction buffer (Boerhinger Mannheim), 8 μl of 2.5 mM dNTPs (Promega), 3.5 μl of each primer (100 ng/ml), 2 μl of DNA template, 1 μl *Taq* DNA polymerase (Boehringer Mannheim), and water to a final volume of

100 μl. Thermocycling was performed as follows: 40 cycles of denaturing at 94°C for 1 min (except for the first cycle, where the time was increased to 5 min), annealing at 50°C for 1 min, and then extension at 72°C for 1.5 to 3 min, followed by a final extension at 72°C for 5 min and cooling to 4°C.

Due to the length of the GP open reading frame (ORF) and the variation seen in the MBG strains, multiple primer sets were required to generate overlapping PCR products that spanned the coding region. As sequences were generated, specific primers were synthesized to amplify GP ORF regions. The primers 5′ TGAAGAACATTAATTGCTGGGTAA and 5′-CCTAATCATTACACATTTAACGTT correspond to plus-sense and minus-sense sequences flanking the MBG ORF (close to the N-terminal and C-terminal ends, respectively) and were used with internal primers to amplify the ends of the ORF. The primer 5′-CCGGTACC$(T)_{35}$, which binds to poly(A), was used with plus-sense primers to amplify sequences within the ORF to the 3′-end of the GP mRNA. First-strand cDNA synthesis reactions, using genomic or mRNA preparations as template, were performed either as previously described (Sanchez *et al.,* 1989), or using a commercial kit (SuperScript II, Life Technologies, GIBCO BRL). RT-PCR amplification assays used commercial kits and reagents (Boehringer Mannheim and Promega). Thermocycling was performed by denaturing at 94°C for 1 min (all cycles), and for the first 3 cycles annealing occurred at 37°C for 30 s with extension at 72°C for 2 min, followed by 30 cycles of annealing at 45, 50, or 55°C for 30 s and extension for 1 min at 72°C. Reactions were cooled to 4°C at the end of thermocycling.

Sequence analysis

Virus sequences were determined by direct sequencing of RT-PCR products, using an automated nonisotopic method (dye-terminator cycle sequencing; Perkin–Elmer). PCR products were isolated from TAE–agarose gels, using a commercial extraction kit (QIAEX II, Qiagen Inc.), eluted into 50 μl of water, and used in sequencing reactions. ABI 373 and 377 automated sequencers were used to derive sequence information and generate sequence files. Computer-assisted nucleic acid and predicted amino acid sequence analyses were performed as previously described (Sanchez *et al.,* 1996). Phylogenetic analyses of nucleotide alignments were performed using the PAUP software, Version 3.1.1 (developed by D. L. Swofford), run on a Power Macintosh 8100/110 (Apple).

RESULTS

The entire VP35 gene and the coding region of the GP gene sequences of the RYC, OZO, and RAV strains of MBG virus were successfully amplified and sequenced. Phylogenetic analyses were performed using the coding regions of these genes, and the corresponding ORFs of the MAY strain of EBO virus were included to provide an outgroup. Figure 2 shows phylogenetic trees based on these nucleotide sequences. Both trees show a similar topology in that the EBO taxon is well separated from the MBG taxa and that the most recent MBG virus isolate (RAV) appears distinct from the other four MBG strains and is extremely well supported by bootstrap values (Felsenstein, 1993). Relationships within the POP, RYC, MUS, and OZO clade in the VP35 tree are not fully resolved, as evidenced by poor bootstrap support for the node connecting MUS and OZO. A clearer relationship of these viruses is seen when GP ORF sequences are analyzed, with the 1980 MUS strain showing a greater similarity to the 1967 viruses than does the 1975 OZO strain. The same phylogenetic profile (as described above) is seen when amino acid sequences are used in maximum parsimony analysis (trees not shown), except that branch lengths are shorter.

To the right of the trees in Fig. 2 are shown matrices of percentage identity at the nucleotide (above diagonal) and amino acid level (below diagonal) for each gene ORF. For the VP35 gene, the nucleotide differences between the RAV and the other MBG virus strains is ~20–21%, while the differences within the POP/RYC/MUS/OZO grouping are at most only 6%. However, corresponding comparisons using amino acid sequences show a decrease to ~6 and 3%, respectively. Compared with EBO, both the nucleotide and the amino acid sequences of the MBG strains differed by ~52–53% and ~65%, respectively. Comparisons of GP ORF sequences showed a similar percentage nucleotide difference between RAV and the other MBG viruses (~21–22%), but the differences in amino acid sequences did not decrease and instead stayed at the same approximate level as nucleotide differences (~22–23%). Within the POP/RYC/MUS/OZO group, a difference of ~6–9% is seen at the amino acid level. In comparison with EBO, the MBG strains differed in the GP ORF nucleotide and predicted amino acid sequences by ~54–55% (comparable to VP35 ORF) and ~69% (4% greater than the VP35 amino acid sequence), respectively. These data indicate that within the MBG strains and between MBG and EBO, there is a greater variation in amino acid sequence for the GP than for the VP35.

Alignments of the predicted VP35 and GP amino acid sequences of the MBG and EBO viruses (used in phylogenetic analyses) were performed (Fig. 3). The VP35 alignment shows only three sites where there is variation in two or more MBG strains, and amino acid changes are primarily found from residues 27 to 184 (alignment numbers 46–204). In contrast, the middle of the GP alignment shows the greatest variation and is flanked by long conserved regions; the N-terminal region (signal sequence), however, does show a fair amount of divergence. The central variable region shows two marked

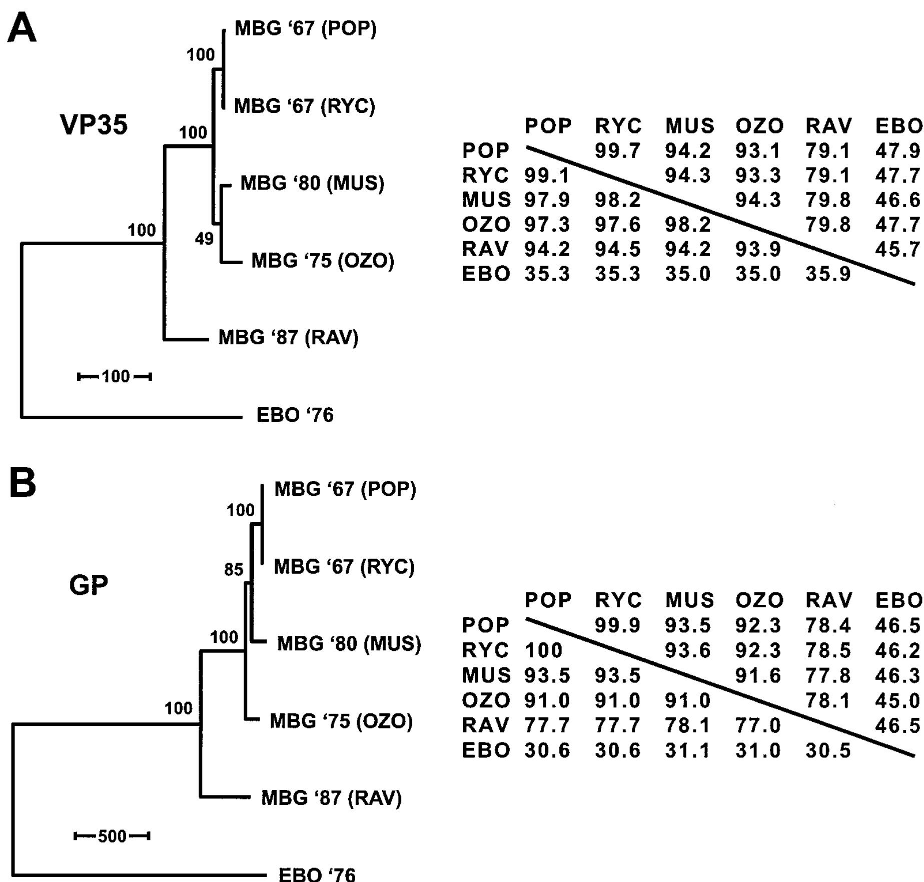

FIG. 2. Phylogeny of MBG viruses based on coding region nucleotide sequences. Phylogenetic analyses were performed using the VP35 and GP gene ORFs of MBG virus strains (990 and 2046 bases in length, respectively), together with those of an EBO virus (MAY strain; 1023 and 2030 bases) included as an outgroup. (A and B) Phylogenetic trees generated from VP35 and GP ORF sequences, respectively. Maximum parsimony (PAUP) analyses using a weighting of 4:1 (transversion:transition) were used to generate trees (shown are the most parsimonious trees). Bars indicate numbers of steps per unit length, and bootstrap values (percentage confidence values derived from 1000 replicates) are shown at branch points. Next to the trees are matrices showing the percentage identity when nucleotide (above diagonal) and predicted amino acid (below diagonal) sequences are aligned. VP35 and GP sequences for the RYC, OZO, and RAV strains of MBG have been deposited in GenBank (Accession Nos. AF005732, AF005730, AF005731, AF005735, AF005733, and AF005734, respectively). The sequences for the POP and MUS strains of MBG and those of EBO were derived from GenBank submissions Z29337, Z12132, and L11365, respectively.

boxes (totaling ~200 residues) where most of the divergence in MBG sequences is located. The N-terminal-most box corresponds to a similar variable region found in the GPs of EBO virus species (EBO residues ~300–500) (Sanchez *et al.*, 1996). Cysteine residues conserved in all the filoviruses are present in both proteins. Three are found in the VP35, while the GP alignment contains 5 in the N-terminal third and 7 in the C-terminal third. The highly conserved C-terminal ~180 residues contain a sequence very similar to an immunosuppressive motif that has been described for the P15 glycoprotein of oncogenic retroviruses (Volchkov *et al.*, 1992; Will *et al.*, 1993). This conserved sequence also partially overlaps an alpha-helical heptad repeat region (4-3 hydrophobic amino acids) (Chambers *et al.*, 1990; Gallaher, 1996). Immediately N-terminal to this helical region in all these viruses are predicted fusion peptides (Gallaher, 1996). The GP alignment also shows that potential glycosylation sites are partly conserved. There are 14 predicted N-linked glycosylation sites that are found in all MBG strains, over half of which are located in the two variable boxes, and six sites are found in all MBG strains except for RAV, again located in the variable boxes. Nine sites are either unique to RAV or found in RAV and some of the other MBG strains, with all but one located in the variable boxes. These findings show that the central variable region is the area where N-linked glycosylation seems to be focused.

To more closely examine the evolutionary processes that are driving MBG GP gene evolution, the synonymous and nonsynonymous nucleotide substitutions in the GP alignment were mapped at each codon (Fig. 4). Stop codons were excluded from the analysis. Synonymous substitutions were found to be distributed throughout the GP gene though this pattern deviated from a random binomial (Poisson) distribution due to the presence of seven codons (numbers 107, 282, 358, 424, 448, 590, 611) exhibiting three or more synonymous substitutions

A VP35

```
POP   ...................                             s   h                                .
RYC   ...................                                                              .
MUS   ...................                                                              .   r
OZO   ...................                                                l                 .     h
RAV   ...................                          t              p         i  s           .
EBO   mttrtkgrghtaattqndr pgpelsgw   q    r   sd  cd..i nnpglcyasqmqqtkpn ktrnsq q  p cnhsfee vq   s atv q
      -------------------MWDSSYMQQVSEGLMTGKVPIDQVFGANPLEKLYKRRKPKGTVGLQCSPCLMSKATSTDDIVWDQL-IVKKTLADLLIPIN 100
                                                                       *  *

POP     i       n
RYC     i y     n
MUS     i       s
OZO     tl      s
RAV     m       s m  k         d    v   k                                                           y qk
EBO   q tiase ....leq  tsl ...nglk  yd ak isslnrvca  v    l  mt   a  t   te  wa   q   g sl eesa rgkie rdet
      RQ-SDIQSTL-EVTTRVHEIERQLHEITPVLKMGRTLEAISKGMSEMLAKYDHLVISTGRTTAPAAAFDAYLNEHGVPPPQPAIFKDLGVAQQACSKGTM 200

POP      e
RYC      e
MUS      a
OZO      a
RAV      q                                                                                              a
EBO    pqsvre fnnlnstts t  n g  di     rnim d    fg a  q v  ic lgkd nsldiihae qas a  dsp c  iqitkrvpi qdaa
      VKN-TTDAADKMSKVLELSEETFSKPNLSAKDLALLLFTHLPGNNTPFHILAQVLSKIAYKSGKSGAFLDAFHQILSEGENAQAALTRLSRTFDAFLGVV 300

POP
RYC
MUS                  s
OZO
RAV       k
EBO       hirsrgdi  a      p   s k  r     fqlqd k lg
      PPVIRVKNFQTVPRPCQKSLRAVPPNPTIDKGWVCVYSSEQGETRALKI 349
                                        *
```

FIG. 3. Alignment of predicted (A) VP35 and (B) GP amino acid sequences. Dots indicate gaps inserted in the alignment. A consensus sequence (plurality = 4 or more) is shown in the bottom lanes, with dashes indicating no consensus reached. Cysteine residues conserved in all sequences are identified by asterisks under the consensus sequence, and predicted N-linked glycosylation sites (N-X-T/S) for all of the MBG virus strains are double underlined. N-linked sites found in all MBG virus strains except for RAV are italicized and in boldface font, while those found in RAV and not in all the other MBG strains are shown underlined. Two regions of the GP alignment that show a higher degree of variability are enclosed in boxes. Near the C-terminus are a set of two heptad repeats (for EBO there is one uninterrupted repeat), marked by dashes and flanked by brackets, that overlap a highly conserved immunosuppressive motif. Immediately N-terminal to this region is a predicted fusion peptide sequence (boldface text in consensus with "+" underneath).

($0.01 < P < 0.025$, $\chi^2 = 7.47$, $df = 2$). The distribution of nonsynonymous substitutions exhibited a much different pattern with a strong deviation from a random binomial distribution ($P < 0.001$, $\chi^2 = 79.52$, $df = 2$). The majority of nonsynonymous distributions clustered at the 5′ end of the GP gene ORF (codons 2 through 32) and in the middle of the GP gene between codons 201 and 501. The ratio of nonsynonymous to synonymous substitutions at 16 codons in this middle region exhibited a ratio ≥3, with four exhibiting a ratio ≥4 (codons 267, 364, 370, 414). All of these 16 codons were found to reside in the two most divergent regions of the GP identified by the amino acid alignment. Thus, the GP gene of MBG viruses exhibits a strongly nonrandom distribution of nonsynonymous changes and a number of sites are showing evidence of selection for amino acid changes over silent substitutions.

DISCUSSION

A phylogeny profile of the family Filoviridae has been derived from GP gene coding sequences (Sanchez *et al.*, 1996). This phylogeny showed that EBO and MBG viruses represent very divergent lineages of filoviruses. Within Ebola viruses there are four separate monophyletic lineages that differ from one another by approximately 37–40% at the nucleotide level. However, within individual lineages, the variation in nucleotide sequence has been shown to be less than 2%, indicating a remarkable degree of stability over a 20-year period (Sanchez *et al.*, 1996; Georges-Courbot *et al.*, 1997; Volchkov *et al.*, 1997; A. Sanchez, unpublished data). Our analyses of MBG virus VP35 and GP gene sequences indicate a somewhat different pattern for these filoviruses. The most recent isolate, the RAV strain (1987), represents a separate lineage of MBG virus, while the strains that preceded it constitute another lineage. Phylogenetic analyses using either VP35 or GP sequences clearly demonstrate this distinction, but the phylogenetic relationship of the 1967 (POP & RYC), 1975 (OZO), and 1980 (MUS) isolates was resolved for GP sequences only. It is interesting to note that the Zimbabwe isolate (OZO) is genetically closer to the Uganda (RYC & POP) and one Kenya strain (MUS) than is the most recent Kenyan isolate (RAV). This geographical discordance (~600–700 vs 100 km) may be explained if the natural host is a migratory species with a large range. Alternatively, this geographical inconsistency may merely reflect a divergence of the RAV strain following coevolution with a distinct natural host.

Differences in the nucleotide sequence of the GP gene

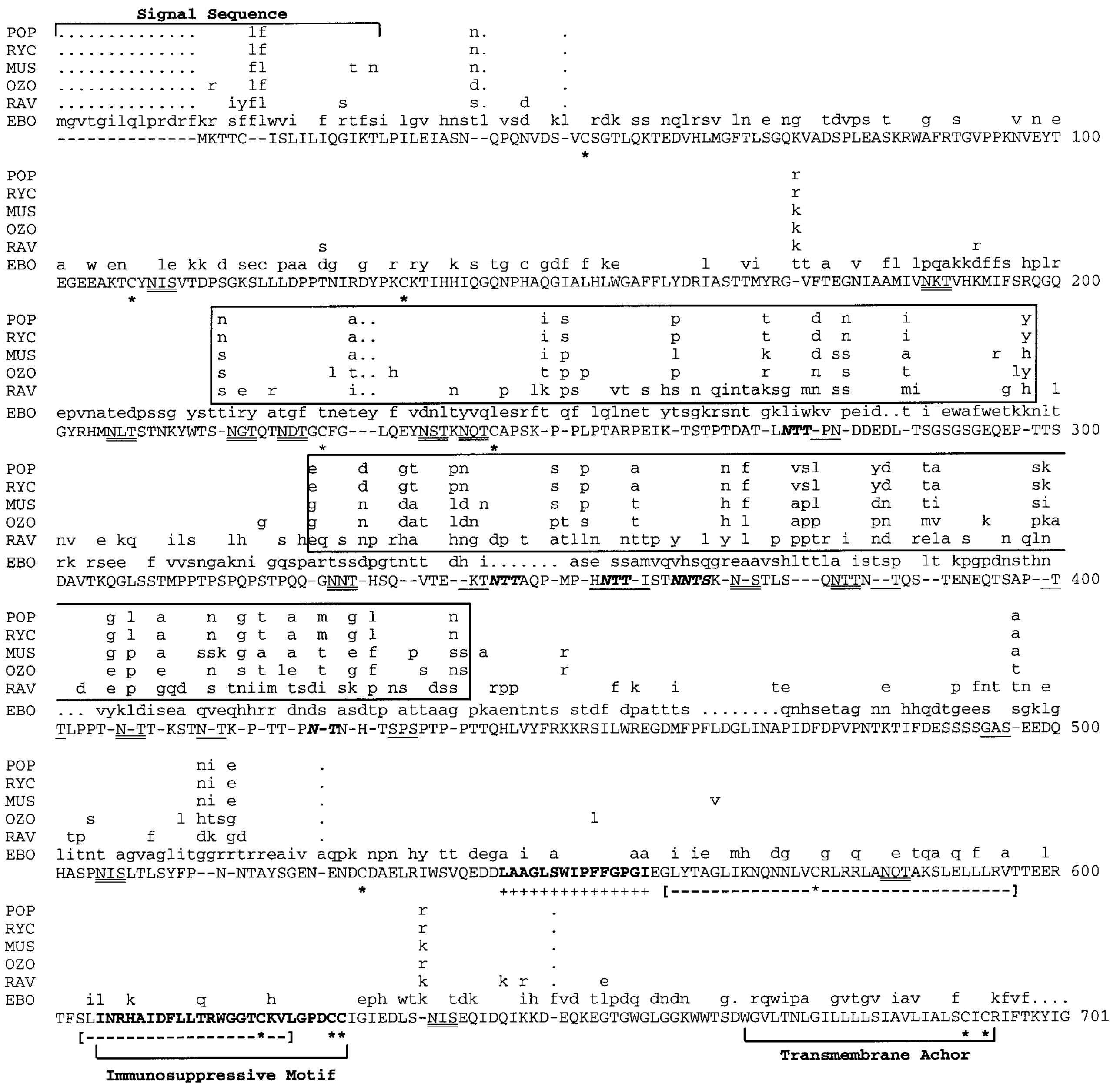

FIG. 3—*Continued*

ORFs of MBG viruses (0.1–21%) were comparable to the percentage amino acid changes (0–23%). In contrast, the nucleotide differences in VP35 ORFs (0.3–20.9%) were higher than amino acid changes (0.9–6.1%), which indicates that the number of nonsynonymous changes is greater in the GP gene than in the VP35 gene. This finding may indicate that the variation in the GP genes of MBG viruses are changes that have been selected for and have not occurred as purely random events. The greater variation in the MBG GP gene versus the VP35 gene is not inconsistent with the variation seen between glycoprotein and P genes of other nonsegmented negative-strand RNA viruses. It was previously shown that there is more variation in the GP gene of the POP and MUS strains of MBG virus compared with other genes (Bukreyev *et al.*, 1995). In addition, alignments of MBG or EBO GP amino acid sequences show that the middle portion of this molecule can be very divergent and is the primary site for N- and O-linked glycosylation (Sanchez *et al.*, 1993). It is possible that this variable region reflects the adaptation of filoviruses as they responded to natural host selective pressures. During this time, GP may have been under greater selective pressure than other virus proteins, possibly from phenotypic changes in cell receptors or the immune system of the natural host.

According to the theory of positive Darwinian selec-

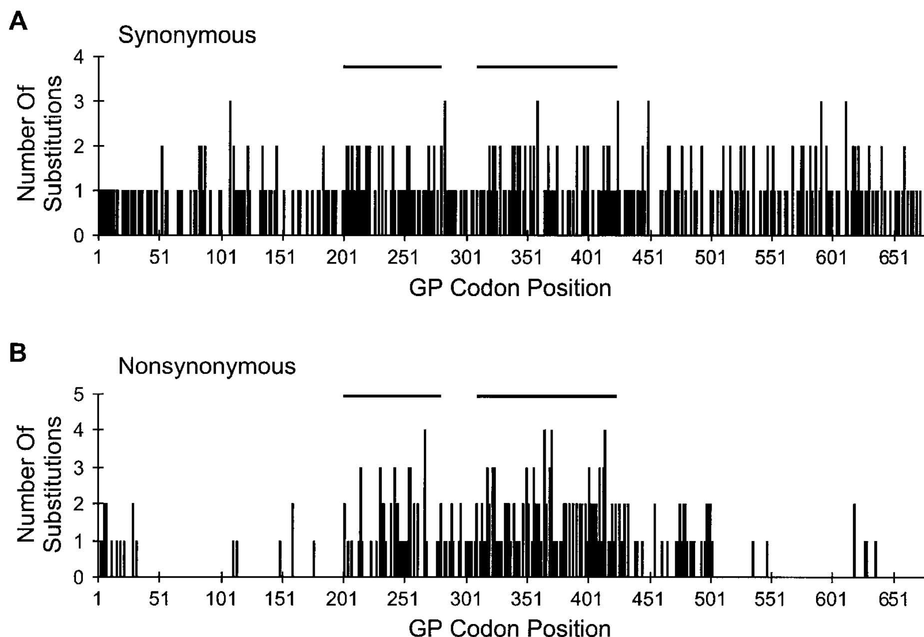

FIG. 4. Distribution of synonymous (A) and nonsynonymous (B) substitutions in the codons of MBG GP genes (minus stop codons). Bars above plots indicate those sequences encoding the variable regions identified in the GP amino acid alignment (Fig. 3B).

tion, nonsynonymous substitutions that cause adaptive amino acid changes are preferred over silent synonymous substitutions (Endo *et al.*, 1996). In the immunoexposed proteins of viral pathogens, the pressures of the host immune system are thought to drive positive selection of epitopes. The resulting antigenic polymorphism provides a mechanism for evading the immune response. Viral proteins thought to be evolving by positive Darwinian selection under immune pressure include the hemagglutinin gene of influenza and the *env* gene of human immunodeficiency virus (Fitch *et al.,* 1991; Yamaguchi and Gojobori, 1997). The strongly nonrandom distribution of nonsynonymous changes in the MBG GP gene and presence of codons exhibiting high ratios of nonsynonymous/synonymous substitutions strongly suggest that the GP is undergoing positive Darwinian selection. Mapping of epitopes within the MBG GP that are recognized by the vertebrate immune response will be necessary to confirm or refute this hypothesis.

Processes in tissue culture or experimentally infected animals that lead to changes in virus phenotypes, whether through pressures from immune responses or some other selective mechanism, may be different from those acting in the natural host. There are differences between the POP strain of MBG virus, which was derived from a virus stock that had been passaged nine times in guinea pigs (Bukreyev *et al.*, 1995) and the RYC strain that we sequenced (both derived from the 1967 outbreak in Marburg, Germany). It is likely that passaging in guinea pigs has led to the selection of a modified strain, which is phenotypically altered in the VP35 protein (three amino acid changes) from the initial virus, but has not led to a change in the GP amino acid sequence.

Depending on the area of the predicted amino acid sequences of the GPs of different filoviruses compared, they can appear either very conserved or extremely variable. The C- and N-terminal regions tend to be generally hydrophobic and contain long regions of conserved residues, whose functions include a signal sequence for directing the protein into the endoplasmic reticulum, and a transmembrane sequence to help anchor the molecule in the virion envelope. These regions are also likely to function in the formation of virion spike structures, since they contain most of the highly conserved cysteine residues. The central, hydrophilic, and variable region of the GP is a platform for the addition of glycans and is presumed to project outward from the virion. The predicted amount of N- and O-linked carbohydrate is considerable, accounting for a third to half of the molecular weight, and likely has a role in the resistance of the virion to neutralization by specific antibody. The positioning of a predicted fusion peptide adjoining heptad repeats N-terminal to the transmembrane anchor in filovirus GPs, initially shown for Ebola virus by Gallaher (1996), is similar to motifs seen in the transmembrane glycoproteins of retroviruses, paramyxoviruses, coronaviruses, and influenza viruses (Chambers *et al.*, 1990; Lambert *et al.*, 1996; Chan *et al.*, 1997). The structural similarity of MBG and EBO GPs argues for a common mechanism by which filovirus GPs form trimers (Feldmann *et al.*, 1991) and mediate virus entry into host cells. This shared mechanism of glycoprotein structure and function, especially with respect to filoviruses and oncogenic retroviruses (Volchkov *et al.*, 1992; Sanchez *et al.*, 1996; Gallaher, 1996), points to either a common ancestral gene sequence or convergent evolution.

The current taxonomy of the family *Filoviridae* places MBG and EBO viruses into a single genus, *Filovirus*. However, it is apparent from our phylogenetic analyses, together with prior studies that described important differences in the biology of these viruses (Kiley *et al.*, 1988; Feldmann *et al.*, 1992, 1994; Sanchez *et al.*, 1993, 1996; Volchkov *et al.*, 1995), that EBO and MBG viruses can readily be classified into separate genera. In our opinion, the four EBO subtypes should be considered as separate species, and MBG strains should all be placed in a single species. An earlier study of the RAV strain of MBG virus suggested that this virus represented a new subtype, comparable to that seen with EBO subtypes (Johnson *et al.*, 1997). Our GP gene finding that RAV differed ~21–23% from that of the other MBG strains (at both nucleotide and amino acid level) is slightly more than half that found between EBO species. We believe that this level of difference is insufficient to consider RAV as a separate MBG species, but is sufficient to be regarded as a subspecies or variant. To better characterize the MBG lineage, it will be necessary to analyze additional novel isolates. However, as with EBO, identification of such viruses has relied on rare outbreaks in human and nonhuman primates, although this obstacle should be greatly reduced when the natural reservoir is eventually determined.

ACKNOWLEDGMENTS

The authors thank Carina Eckel and Steffi Lindow for their excellent technical assistance. We also acknowledge contribution of John O'Connor for editing this paper. This work was partially supported by grants of the Deutsche Forschungsgemeinschaft (Fe 286/4-1; SFB 535) and the Kempkes-Stiftung (21/95).

REFERENCES

Bukreyev, A. A., Volchkov, V. E., Blinov, V. M., Dryga, S. A., and Netesov, S. V. (1995). The complete nucleotide sequence of the Popp (1967) strain of Marburg virus: A comparison with the Musoke (1980) strain. *Arch. Virol.* **140**, 1589–1600.

Centers for Disease Control and Prevention and National Institutes of Health. (1993). "Biosafety in Microbiological and Biomedical Laboratories." U. S. Department of Health and Human Services Publication No. (CDC) 93-8395.

Chambers, P., Pringle, C. R., and Easton, A. J. (1990). Heptad repeat sequences are located adjacent to hydrophobic regions in several types of virus fusion glycoproteins. *J. Gen. Virol.* **71**, 3075–3080.

Chan, D. C., Fass, D., Berger, J. M., and Kim, P. S. (1997). Core structure of gp41 from the HIV envelope glycoprotein. *Cell* **89**, 263–273.

Endo, T, Ikeo, K., and Gojobori, T. (1996). Large-scale search for genes on which positive selection may operate. *Mol. Biol. Evol.* **13**, 685–690.

Feldmann, H., Will, C., Schikore, M., Slenczka, W., and Klenk, H.-D. (1991). Glycosylation and oligomerization of the spike protein of Marburg virus. *Virology* **182**, 353–356.

Feldmann, H., Mühlberger, E., Randolf, A., Will, C., Kiley, M. P., Sanchez, A., and Klenk, H.-D. (1992). Marburg virus, a filovirus: Messenger RNAs, gene order, and regulatory elements of the replication cycle. *Virus Res.* **24**, 1–19.

Feldmann, H., Nichol, S. T., Klenk, H.-D., Peters, C. J., and Sanchez, A. (1994). Characterization of filoviruses based on differences in structure and antigenicity of the virion glycoprotein. *Virology* **199**, 469–473.

Feldmann, H., and Klenk, H.-D. (1996). Marburg and Ebola viruses. *Adv. Virus Res.* **47**, 1–52.

Felsenstein, J. (1993). PHYLIP: Phylogeny Inference Package, Version 3.57 [Computer program]. University of Washington, Seattle, WA.

Fitch, W. M., Leiter, J. M. EJ. M., Li, X., and Palese, P. (1991). Positive Darwinian evolution in human influenza A viruses. *Proc. Natl. Acad. Sci. USA* **88**, 4270–4274.

Gallaher, W. R. (1996). Similar structural models of the transmembrane proteins of Ebola and avian sarcoma viruses. *Cell* **85**, 477–478.

Gear, J. S. S., Cassel, G. A., Gear, A. J., Trappler, B., Clausen, L., Meyers, A. M., Kew, M. C., Bothwell, T. H., Sher, R., Miller, G. B., Schneider, J., Koornhof, H. J., and Gomperts, E. D., Isaäcson, M., and Gear, J. H. S. (1975). Outbreak of Marburg virus disease in Johannesburg. *Br. Med. J.* **4**, 489–493.

Geisbert, T. W., and Jahrling, P. B. (1996). Differentiation of filoviruses by electron microscopy. *Virus Res.* **39**, 129–150.

Georges-Courbot, M.-C., Sanchez, A., Lu, C.-Y., Baize, S., Leroy, E., Lansout-Soukate, J., T'nissan, Georges, A. J., Trappier, S. G., Zaki, S. R., Swanepoel, R., Leman, P. A., Rollin, P. E., Peters, C. J., Nichol, S. T., and Ksiazek, T. G. (1997). Isolation and phylogenetic characterization of Ebola viruses causing different outbreaks in Gabon. *Emerg. Inf. Dis.* **3**, 59–62.

Jarhling, P. B., Kiley, M. P., Klenk, H.-D., Peters, C. J., Sanchez, A., and Swanepoel, R. (1995). Family Filoviridae. *In* "Virus Taxonomy, Sixth Report of the International Committee on Taxonomy of Viruses" (F. A. Murphy, C. M. Fauquett, D. H. L. Bishop, S. A. Ghabrial, A. W. Jarvis, G. P. Martelli, M. A. Mayo, and M. D. Summers, Eds.). *Arch. Virol.* **10**(Suppl.), 289–292.

Johnson, E. D., Kohnson, B. K., Silverstein, D., Tukei, P., Geisbert, T. W., Sanchez, A., and Jahrling, P. B. (1997). Characterization of a new Marburg virus isolated from a 1987 fatal case in Kenya. *Arch. Virol.* **11**(Suppl), 101–114.

Kiley, M. P., Cox, N. J., Elliott, L. H., Sanchez, A., DeFries, R., Buchmeier, M. J., Richman, D. D., and McCormick, J. B. (1988). Physicochemical properties of Marburg virus: Evidence for three distinct virus strains and their relationship to Ebola virus. *J. Gen. Virol.* **69**, 1957–1967.

Ksiazek, T. G., Rollin, P. E., Jahrling, P. B., Johnson, E., Dalgard, D. W., and Peters, C. J. (1992). Enzyme immunosorbent assay for Ebola virus antigens in tissues of infected primates. *J. Clin. Microbiol.* **30**, 947–950.

Lambert, D. M., Barney, S., Lambert, A. L., Guthrie, K., Medinas, R., Davis, D. E., Bucy, T., Erickson, J., Merutka, G., and Petteway, S. R., Jr. (1996). Peptides from conserved regions of paramyxovirus fusion (F) proteins are potent inhibitors of viral fusion. *Proc. Natl. Acad. Sci. USA* **93**, 2186–2191.

Peters, C. J., Johnson, E. D., Jahrling, P. B., Ksiazek, T. G., Rollin, P. E., White, J., Hall, W., Trotter, R., and Jaax, N. (1993). Filoviruses. *In* "Emerging Viruses" (S. Morse, Ed.), pp. 159–175. Oxford Univ. Press, New York.

Peters, C. J., Sanchez, A., Rollin P. E., Ksiazek, T. G., and Murphy, F. A. (1996). Filoviridae: Marburg and Ebola viruses. *In* "Fields Virology" (B. N. Fields, D. M. Knipe, P. M. Howley, *et al.*, Eds.), Third ed., Vol. 1, pp. 1161–1176. Lippincott-Raven Press, Philadelphia.

Sanchez, A., and Kiley, M. P. (1987). Identification and analysis of Ebola virus messenger RNA. *Virology* **157**, 414–420.

Sanchez, A., Kiley, M. P., Holloway, B. P., McCormick, J. B., and Auperin, D. D. (1989). The nucleoprotein gene of Ebola virus: Cloning, sequencing and in vitro expression. *Virology* **170**, 81–91.

Sanchez, A., Kiley, M. P., Holloway, B. P., and Auperin, D. D. (1993). Sequence analysis of the Ebola virus genome: organization, genetic elements, and comparison with the genome of Marburg virus. *Virus Res.* **29**, 215–240.

Sanchez, A., Trappier, S. G., Mahy, B. W. J., Peters, C. J., and Nichol, S. T. (1996). The virion glycoproteins of Ebola viruses are encoded in two reading frames and are expressed through transcriptional editing. *Proc. Natl. Acad. Sci. USA* **93,** 3602–3607.

Siegert, R., Shu, H.-L., Slenczka, W., Peters, D., and Müller G. (1967). Zur Äthiologie einer unbekannten von Affen ausgegangenen Infektionskrankheit. *Dtsch. Med. Wochenschr.* **92,** 2341–2343.

Smith, D. H., Johnson, B. K., Isaäcson, M., Swanapoel, R., Johnson, K. M., Bagshawe, A., Siongok, T., and Keruga, W. K. (1982). Marburgvirus disease in Kenya. *Lancet* **1,** 816–820.

Volchkov, V. E., Blinov, V. M., and Netesov, S. V. (1992). The envelope glycoprotein of Ebola virus contains an immunosuppressive-like domain similar to oncogenic retroviruses. *FEBS Lett.* **305,** 181–184.

Volchkov, V. E., Becker, S., Volchkova, V. A., Ternovoj, V. A., Kotov, A. N., Netesov, S. V., and Klenk, H.-D. (1995). GP mRNA of Ebola virus is edited by the Ebola virus polymerase and by T7 and vaccinia virus polymerases. *Virology* **214,** 421–430.

Volchkov, V. E., Volchkova, V., Eckel, C., Klenk, H.-D., Bouloy, M., LeGuenno, B., and Feldmann, H. (1997). Emergence of subtype Zaire Ebola virus in Gabon. *Virology* **232,** 139–144.

Will, C., Mühlberger, E., Linder, D., Slenczka, W., Klenk, H.-D., and Feldmann, H. (1993). Marburg virus gene 4 encodes the virion membrane protein, a type I transmembrane glycoprotein. *J. Virol.* **67,** 1203–1210.

Yamaguchi, Y., and Gojobori, T. (1997). Evolutionary mechanisms and population dynamics of the third variable envelope region of HIV within single hosts. *Proc. Natl. Acad. Sci. USA* **94,** 1264–1269.