DNA AND CELL BIOLOGY
Volume 34, Number 6, 2015

Pp. 1–11
DOI: 10.1089/dna.2014.2678

ORIGINAL RESEARCH ARTICLE

Ebolavirus Classification Based on Natural Vectors

Hui Zheng,[1] Changchuan Yin,[1] Tung Hoang,[1] Rong Lucy He,[2] Jie Yang,[1] and Stephen S.-T. Yau[3]

According to the WHO, ebolaviruses have resulted in 8818 human deaths in West Africa as of January 2015. To better understand the evolutionary relationship of the ebolaviruses and infer virulence from the relationship, we applied the alignment-free natural vector method to classify the newest ebolaviruses. The dataset includes three new Guinea viruses as well as 99 viruses from Sierra Leone. For the viruses of the family of *Filoviridae*, both genus label classification and species label classification achieve an accuracy rate of 100%. We represented the relationships among *Filoviridae* viruses by Unweighted Pair Group Method with Arithmetic Mean (UPGMA) phylogenetic trees and found that the filoviruses can be separated well by three genera. We performed the phylogenetic analysis on the relationship among different species of *Ebolavirus* by their coding-complete genomes and seven viral protein genes (glycoprotein [GP], nucleoprotein [NP], VP24, VP30, VP35, VP40, and RNA polymerase [L]). The topology of the phylogenetic tree by the viral protein VP24 shows consistency with the variations of virulence of ebolaviruses. The result suggests that VP24 be a pharmaceutical target for treating or preventing ebolaviruses.

Introduction

During the past 39 years, Ebola outbreaks have occurred 24 times in Africa. The most serious outbreak included 318 infected in 1976 with a case-fatality rate of 88% in the Democratic Republic of Congo (DRC). The latest 2014 Ebola outbreak began in Guinea in December 2013 and quickly spread to the neighboring countries of Liberia and Sierra Leone (Gatherer, 2014). By January 2015, 22,101 suspected cases had resulted in a total of 8818 deaths in these three countries. The infection and case-fatality rates have surpassed previous records (Table 1). The WHO designated this outbreak as a public health emergency of international concern. Early and accurate classification of the ebolaviruses could have prevented the spread of ebolaviruses within medical centers since the early symptoms of ebolavirus are similar to malaria, dysentery, typhoid fever, and other viral infections.

Ebolaviruses can attack a variety of tissue cells in mammals. After a human being is infected with this virus, the person may experience a number of different types of symptoms, including nausea, vomiting, diarrhea, internal or external bleeding, and fever. The disease often causes multiorgan dysfunction, including hepatic damage and renal failure, and possible death (Feldmann and Geisbert, 2011). There is neither a specific treatment nor a licensed vaccine for ebolaviruses available as yet.

Ebolaviruses belong to the *Filoviridae* family and *Ebolavirus* genus in the Baltimore V classification system. *Ebolavirus* has five species: *Zaire ebolavirus* with type virus Ebola virus (EBOV), *Bundibugyo ebolavirus* with type virus Bundibugyo virus (BDBV), *Reston ebolavirus* with type virus Reston virus (RESTV), *Sudan ebolavirus* with type virus Sudan virus (SUDV), and *Taï Forest ebolavirus* (also known as *Cote d'Ivoire ebolavirus*) with type virus Taï Forest virus (TAFV) (Kuhn *et al.*, 2014). Among them, EBOV has up to 90% case-fatality rate in some epidemics. The causative agent of the 2014 Guinea outbreak has been identified as a separated clade of *Zaire ebolavirus* (Baize *et al.*, 2014). SUDV has 41–65% lethality (McElroy *et al.*, 2014). BDBV resulted in two outbreaks with a case-fatality rate of 25% in 2008 and 51% in 2012 (Roddy *et al.*, 2012). TAFV infected two persons and it was nonfatal in both of them (Le Guenno, *et al.*, 1999). RESTV is nonpathological for humans, but highly virulent for nonhuman primates (Feldmann *et al.*, 2003). *Ebolavirus* is an obligate intracellular pathogen that enters a host cell to replicate. It has a characteristic thread-like structure and carries a negative-sense RNA genome of 19 kb length. Gire *et al.* (2014) presented a comprehensive genomic analysis of the 2014 ebolavirus epidemic and identified new mutations that generate genetically distinct sequence clades. Many of the mutations alter protein sequences and have biologically meaningful impacts. These mutations may provide insights on diagnostics, vaccines, and therapies (Gire *et al.*, 2014). The ebolavirus genome encodes seven proteins (glycoprotein [GP], nucleoprotein [NP], VP24, VP30, VP35, VP40, and RNA polymerase [L]) (Volchkov, 1999). GP is the only viral protein found on the surface of ebolavirus. It is responsible for mediating attachment and entry of the virus

[1]Department of Mathematics, Statistics, and Computer Science, University of Illinois at Chicago, Chicago, Illinois.
[2]Department of Biological Sciences, Chicago State University, Chicago, Illinois.
[3]Department of Mathematical Sciences, Tsinghua University, Beijing, People's Republic of China.

Table 1. Lethality of Previous Outbreaks Caused by *Zaire ebolavirus*

Time	*Country*	*Cases*	*Fatality*	*Rate*
1976	DRC	318	280	0.88
1994	Gabon	52	31	0.60
1995	DRC	315	254	0.81
1996	Gabon	92	67	0.73
2001–2002	Gabon	124	97	0.78
2003	Congo	178	157	0.88
2005	Congo	12	10	0.83
2007	DRC	264	187	0.71
2010	DRC	32	14	0.44

DRC, Democratic Republic of Congo.

into host cells (Francica, 2010). The NP of ebolavirus is necessary and sufficient for the formation of nucleocapsid-like structures in a mammalian expression system. It is also responsible for the NP–NP interaction and replication of the viral genome (Watanabe *et al.,* 2006). Xu (2014) showed that VP24 binds to the karyopherin alpha nuclear transporters and blocks the nuclear accumulation of STAT1, a very important component of the immune system (Xu *et al.*, 2014). An earlier study revealed that VP35 blocks production of interferon, a critical component for immune defense against viruses. L protein (LP) and other three virion structural proteins (NP, VP30, and VP35) have been shown to be necessary and sufficient for ebolavirus transcription and replication (Groseth *et al.,* 2005).

However, which protein critically contributes to virulence is still in debate. To classify new ebolaviruses and associate proteins with the virus virulence, we used the reliable alignment-free natural vector method (Yu *et al.*, 2013). Our phylogenetic analysis of the ebolaviruses is not only based on coding-complete genomes but also the seven protein genes in the genomes.

Materials and Methods

To accurately and quickly classify ebolavirus genomes and genes, we employed our alignment-free natural vector method (Yu *et al.*, 2013). The natural vector represents a virus nucleotide sequence by a 12-dimensional vector. The advantage of the method is that it does not rely on any model assumption. The natural vector method has been successfully used to classify and predict not only single-segmented viruses, such as HIV (Deng *et al.*, 2011) and West Nile viruses (Yu *et al.*, 2013), but also multiple-segmented viruses, such as influenza (Huang *et al.*, 2014). The viral genome space can be embedded into a 12-dimensional space using the natural vector method.

Let $S=(s_1, s_2, \ldots, s_n)$ be a nucleotide sequence with length n, $s_i \in \{A, C, G, T\}$, For $\alpha = A, C, G, T$, define $w_\alpha(\cdot):\{A, C, G, T\} \rightarrow \{0, 1\}$. Let $w_\alpha(s)=1$, *if* $s=\alpha$; $w_\alpha(s)=0$ otherwise.

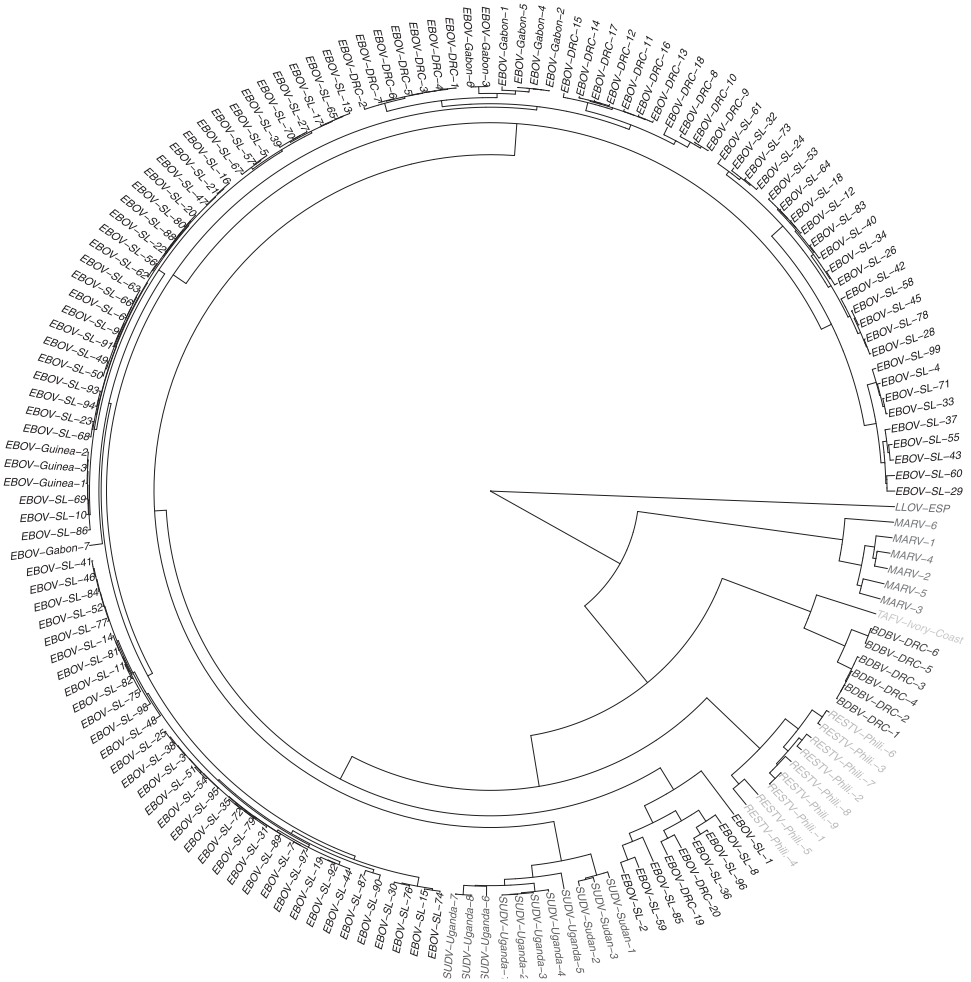

FIG. 1. Phylogenetic tree of 163 filoviruses based on the distance matrix derived by coding-complete genomes' natural vector through UPGMA. UPGMA, Unweighted Pair Group Method with Arithmetic Mean.

The natural vector of S is defined as $(n_A, n_C, n_G, n_T, \mu_A, \mu_C, \mu_G, \mu_T, D_A^2, D_C^2, D_G^2, D_T^2)$, where $n_\alpha = \sum_{i=1}^{n} w_\alpha(s_i)$ denotes the number of occurrences of letter k in S. $\mu_\alpha = \sum_{i=1}^{n} i \cdot \frac{w_\alpha(s_i)}{n_\alpha}$ is the mean position of letter k.

$$D_\alpha^2 = \sum_{i=1}^{n} \frac{(i-\mu_\alpha)^2 w_\alpha(s_i)}{n_\alpha n}.$$

After getting the natural vector of each virus sequence in the dataset, a distance matrix is constructed using the Euclidean distance between vectors for all viruses. For each virus, we find its nearest neighbor, which means the neighbor that has the smallest distance to the virus from the distance matrix. Then, we verify whether the taxonomy labels, including Family, Genus, and Species, are consistent with its neighbors. If it is not, we count it as one error of this classification.

To serve those users interested in virus classification and prediction based on natural vectors, we built the virus database and the online inquiry system: http://mathlab.math.uic.edu/VirusDB

The database stores all single-segmented and multiple-segmented virus reference sequences, including the six *Filoviridae* reference sequences (NC_014373.1, NC_014372.1, NC_001608.3, NC_004161.1, NC_006432.1, and NC_002549.1). Using this classification database, we classify the *Filoviridae* viruses by their genera and the *Ebolavirus* by their species with 100% accuracy in both cases.

We plotted the phylogenetic trees by UPGMA (Unweighted Pair Group Method with Arithmetic Mean). For a given virus sequence dataset, UPGMA forms each virus sequence into a cluster, then groups two smaller clusters of nodes to build up the phylogenetic tree until there is only one tree that contains all the virus sequences. It is the clustering method from a pairwise distance matrix and

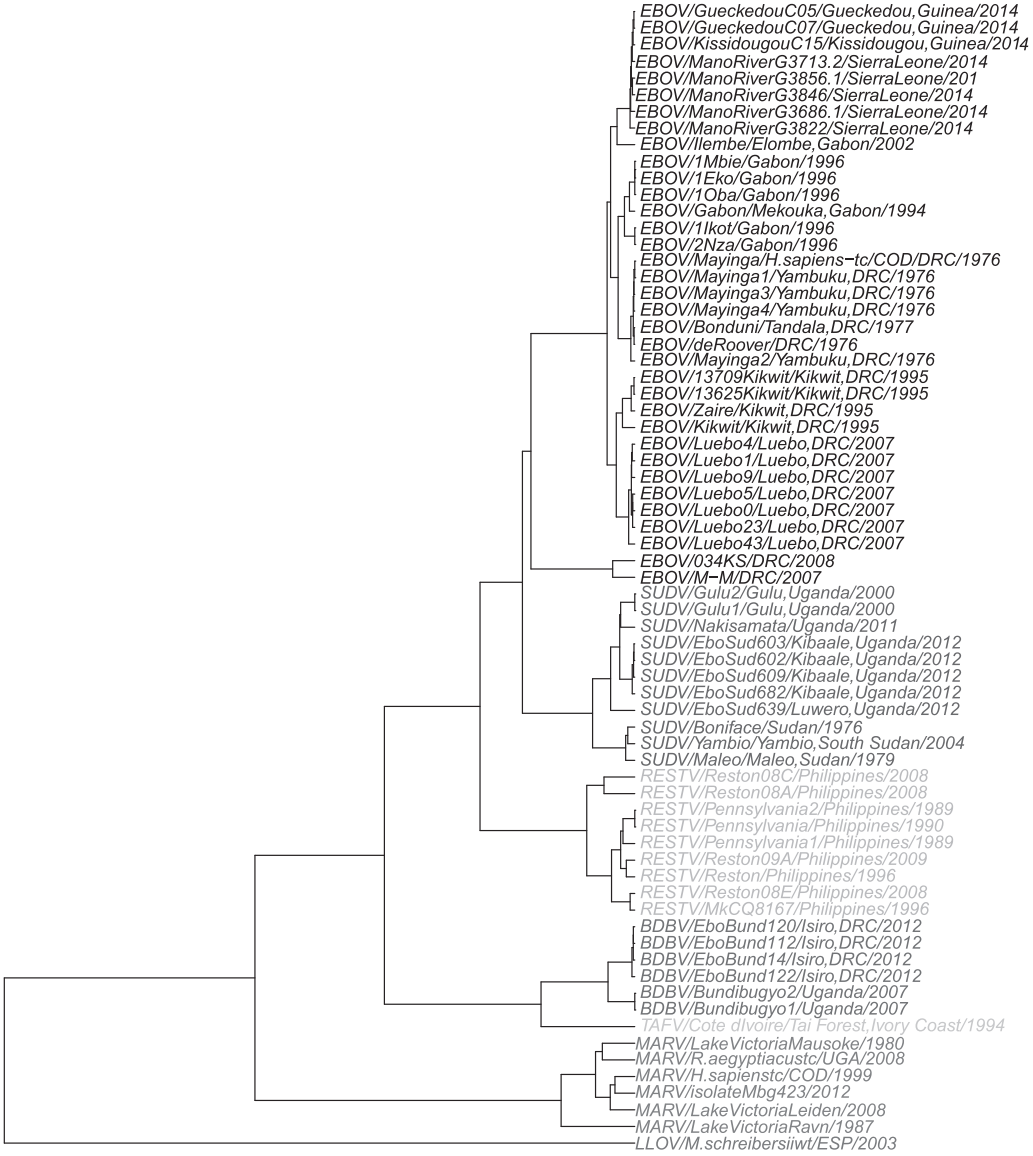

FIG. 2. Phylogenetic tree of 69 filoviruses based on the distance matrix derived by coding-complete genomes' natural vector through UPGMA.

widely used in sequence similarity analysis (Sourdis and Krimbas, 1987).

Results and Discussion

We performed phylogenetic analysis on genomes of ebolaviruses. The dataset includes 1 cuevavirus, 6 marburgvirus, all previous ebolaviruses, the 3 EBOVs from Guinea, and 99 EBOVs from Sierra Leone of the 2014 outbreak (Supplementary Table S1; Supplementary Data are available online at www.liebertpub.com/dna). Figure 1 shows the UPGMA phylogenetic tree based on the coding-complete genomes of the 163 viruses of the dataset. The tree was plotted based on the distance the matrix derived from the natural vectors of these filovirus genomes downloaded from the NCBI GenBank. Figure 1 displays clearly that the 99 Sierra Leone viruses belong to genus *Ebolavirus* and species *Zaire ebolavirus*, and it shows that the Sierra Leone viruses have very close relationship with the 3 Guinea viruses. This result is in agreement with the geographic distribution of the viruses and is consistent with the study of the newest Sierra Leone virus research (Gire *et al.*, 2014).

To get a more detailed relationship of filoviurses, we examined a subset of 163 viruses by keeping 5 Sierra Leone *ebolaviruses* instead of 99 for phylogenetic tree analysis (Supplementary Table S2). As illustrated in Figure 2, the 69 filoviruses are classified clearly by 3 genera (*Ebolavirus*, *Marburgvirus*, and *Cuevavirus*). Genus *Ebolavir*us includes 5 species: *Zaire ebolavirus*, *Sudan ebolavirus*, *Reston ebolavirus*, *Bundibugyo ebolavirus*, *Taï Forest ebolavirus*. *Marburgvirus* only has one species *Marburg marburgvirus* with type virus Marburg virus (MARV); *Cuevavirus* has the only species *Lloviu cuevavirus* with type virus Lloviu virus (LLOV). From the computation of the distances by the natural vector method, the smallest distance between *Marburgvirus* and *Ebolavirus* is 534.07 and the smallest distance between *Cuevavirus* and *Ebolavirus* is 810.52. This result indicates that the *Marburgvirus* and *Ebolavirus* genomes are closer than the *Cuevavirus* and *Ebolavirus* genomes. This observation is consistent with the virus virulence study. The human lethality rate reached about 70% in those infected with marburgvirus, and the patients have the same symptoms with ebolavirus. Marburgviruses have been studied together with ebolaviruses in previous research because of

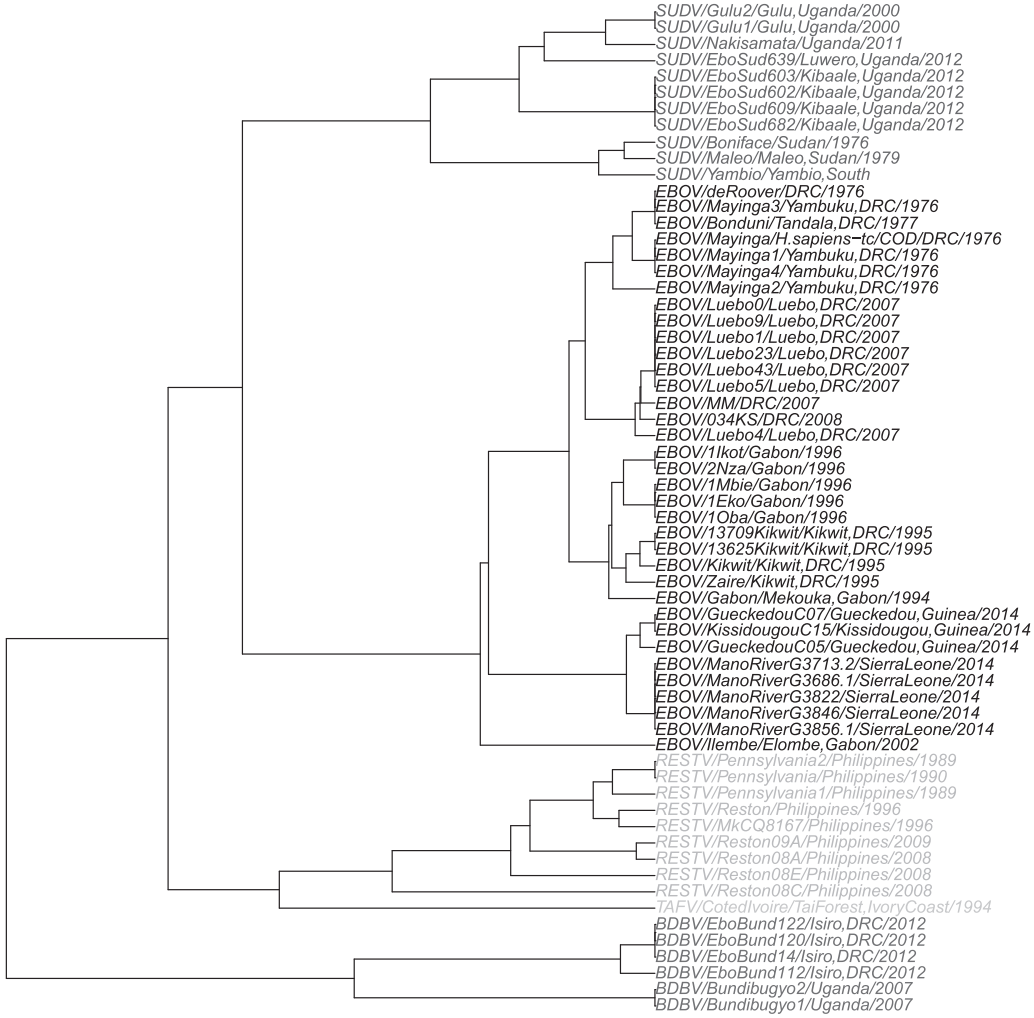

FIG. 3. Phylogenetic tree of ebolaviruses based on the distance matrix derived by glycoprotein (GP) sequences' natural vector through UPGMA.

their similar infection and replication mechanism (Monath, 1999; Bosio *et al.*, 2003). Therefore, the relative distance of the virus genome by the natural vector method could be used to infer the relationship of virus virulence.

We further analyzed the relationships among ebolaviruses at the genus level. The five species of the *Ebolavirus* are separated well, as shown in Figure 2 in different colors. As we know, *Zaire ebolavirus* and *Sudan ebolavirus* are the two most fatal species; *Zaire ebolavirus* has about as high as 90% lethality, and *Sudan ebolavirus* has about 53% lethality. The tree shows that *Zaire ebolavirus* and *Sudan ebolavirus* are neighbors and the smallest distance between all tested EBOVs and SUDVs is 189.055. The smallest distance between *Zaire ebolavirus* and *Reston ebolavirus* is 285.177 and between *Sudan ebolavirus* and *Reston ebolavirus* is 194.203. The distance between *Zaire ebolavirus* and *Sudan ebolavirus* is smaller than the distance between *Zaire ebolavirus* and *Reston ebolavirus*. This distance difference is in agreement with the virulence difference of *Zaire ebolavirus*, *Sudan ebolavirus*, and *Reston ebolavirus*. Also, the viruses from the same country are classified together within each species. In the branch of *Zaire ebolavirus*, eight viruses from Guinea and Sierra Leone of the 2014 outbreak are separated from others and classified in the new strain. This result is consistent with the previous research (Baize *et al.*, 2014; Gire *et al.*, 2014). *Bundibugyo ebolavirus* and *Taï Forest ebolavirus* are in the same group despite the fact that *Bundibugyo ebolavirus* is a deadly species and *Taï Forest ebolavirus* is not so deadly. It is unusual that *Bundibugyo ebolavirus* always be classified with *Taï Forest ebolavirus*, but not with *Zaire ebolavirus* and *Sudan ebolavirus* in previous research (Baize *et al.*, 2014). We will perform more research on the *Bundibugyo ebolavirus* and *Taï Forest ebolavirus* relationship in the future.

In addition, we performed phylogenetic analysis using the natural vector method of seven proteins: NP, VP35, VP40, GP, VP30, VP24, and LP of ebolaviruses. We only focused on genus *Ebolavirus*.

The phylogenetic tree based on GP sequences is shown in Figure 3. All five species (*Zaire ebolavirus*, *Sudan ebolavirus*, *Reston ebolavirus*, *Bundibugyo ebolavirus*, and *Taï Forest ebolavirus*) are well separated, and the two fatal species *Zaire ebolavirus* and *Sudan ebolavirus* are in the same branch. The phylogenetic tree from GP proteins has deeper and richer branch structures than the corresponding trees of coding-complete genomes, LP and NP protein. This means that the GP sequences have more

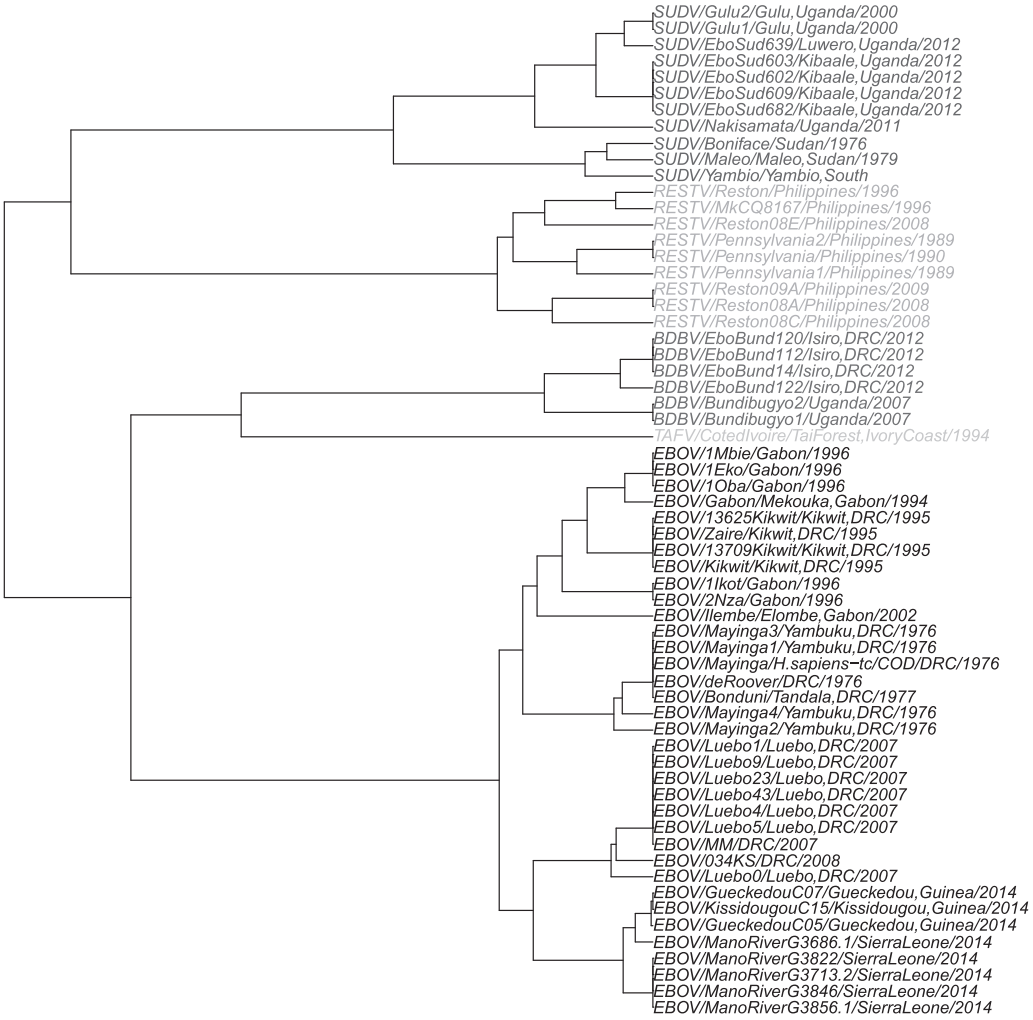

FIG. 4. Phylogenetic tree of ebolaviruses based on the distance matrix derived by nucleoprotein (NP) sequences' natural vector through UPGMA.

mutations, which help the virus survive in different hosts. Furthermore, the Euclidean distance between *Zaire ebolavirus* and *Sudan ebolavirus* is 35.94, while the distance between *Zaire ebolavirus* and *Reston ebolavirus* is 37.88. These distances are consistent with these viruses' virulence.

The phylogenetic tree of NP sequences is shown in Figure 4. The tree structure is relatively simple compared with coding-complete genomes and GP trees. There are only two main branches; one diverges to *Sudan ebolavirus* and *Reston ebolavirus* and the other goes to *Bundibugyo ebolavirus*, *Taï Forest ebolavirus*, and *Zaire ebolavirus*. For the Euclidean distance computations of NP, *Zaire ebolavirus* has a distance of 44.94 with *Bundibugyo ebolavirus*, a distance of 48.18 with *Reston ebolavirus*, and a distance of 49.83 with *Sudan ebolavirus*. Because these distances do not match the difference of the virulence of these species and many NP sequences are classified as one leaf in the tree, the result suggests that the NP sequences are stable and may not contribute significantly to the virus virulence.

It is noteworthy that the orders of the five *Ebolavirus* species from the GP tree and NP tree from top to bottom are SUDV-EBOV-RESTV-TAFV-BDBV and SUDV-BDBV-RESTV-TAFV-EBOV, respectively. The switched positions of the species in the two trees indicate that there is a recombination of GP and NP between species *Zaire ebolavirus* and *Bundibugyo ebolavirus*. The recombination between GP and NP was also studied previously (Wittmann *et al.*, 2007; Domazet-Lošo and Haubold, 2011).

The phylogenetic tree of VP24 sequences is shown in Figure 5. Among the seven protein trees, VP24 has the most similar structure to the tree of coding-complete genomes sequences. Since *Sudan ebolavirus*, *Reston ebolavirus*, *Bundibugyo ebolavirus*, and *Taï Forest ebolavirus* have low genetic diversity, we focus on the genetic diversity of *Zaire ebolavirus*. We regard to the viruses from the same outbreak (same year) as the subspecies in *Zaire ebolavirus*, six subspecies are well separated in six clear branches. The result shows that there are no more than three clades within each subspecies. This suggests that VP24 sequences are exactly the same for some viruses within the subspecies and it was confirmed after we checked the distance matrix. Initially, the viruses from the outbreak of 1976 and 1995 in the DRC have high (larger

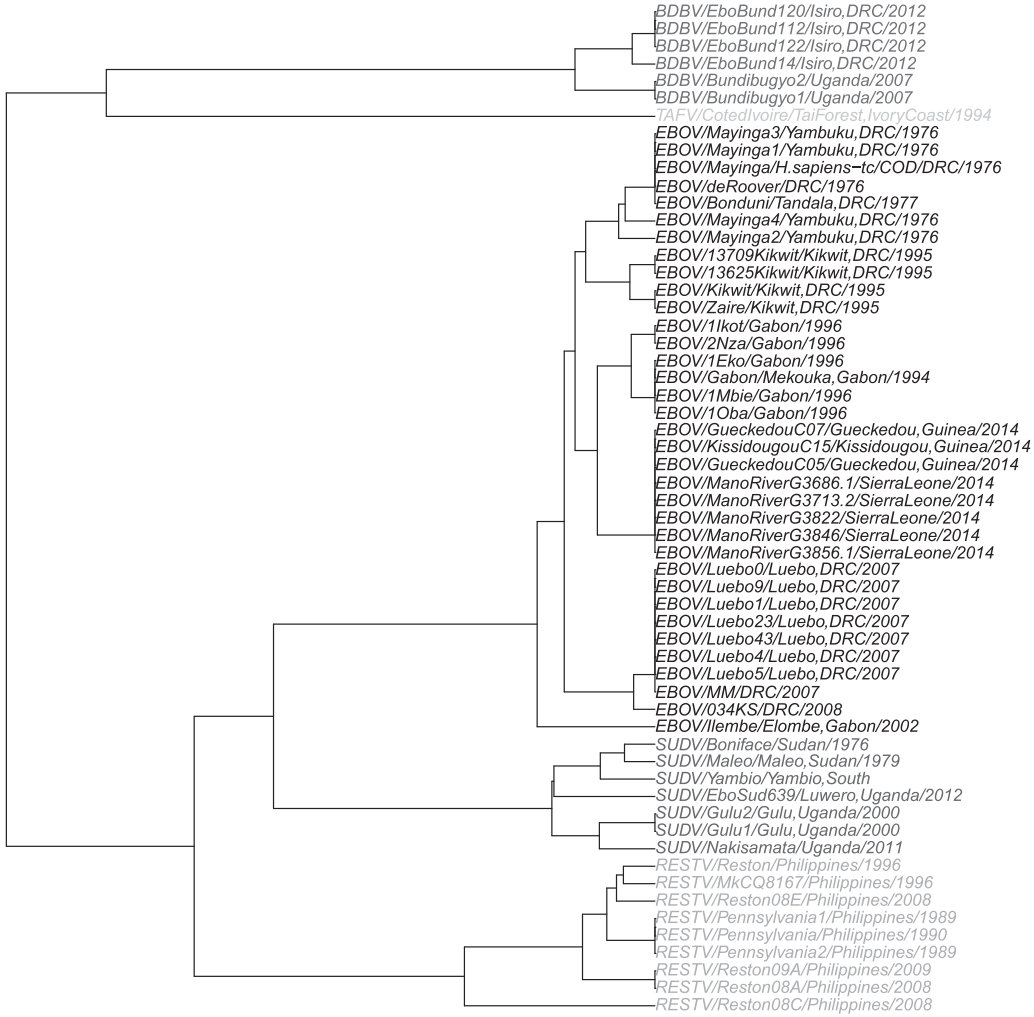

FIG. 5. Phylogenetic tree of ebolaviruses based on the distance matrix derived by VP24 sequences' natural vector through UPGMA.

than 80%) lethality as shown Table 1 (summarized from the WHO website). In addition, from the coding-complete genomes phylogenetic tree, these two subspecies are close to each other and are constructed together as a big branch on the top of the tree (Fig. 2). The viruses from the outbreak of 2007 in the DRC have a relatively lower (71%) lethality and are in a separate small branch from the other subspecies. The only virus from the 2002 outbreak itself as a single clade locates at the bottom of the tree. In the middle are the viruses from the outbreak in 1996 in Gabon and the viruses from the 2014 outbreak. These two subspecies are combined together as a big branch; also, these two subspecies are the only two subspecies not from DRC in this data set. The lethality in 1996 is 73% and we know that the current outbreak case-fatality rate is about 40% as of January 2015. This result suggests that VP24 plays a key role in the different virulence in different outbreaks according to the classification and the different lethality of each subspecies. This discovery is consistent with other medical research that VP24 is the molecular determinant of ebolavirus virulence (Xu, 2014). We suggest that VP24 should be the most important target to block when we develop drugs or vaccines.

The phylogenetic tree of ebolaviruses is shown in Figure 6. First of all, in the VP35 phylogenetic tree, *Reston ebolavirus* and *Sudan ebolavirus* are classified together.

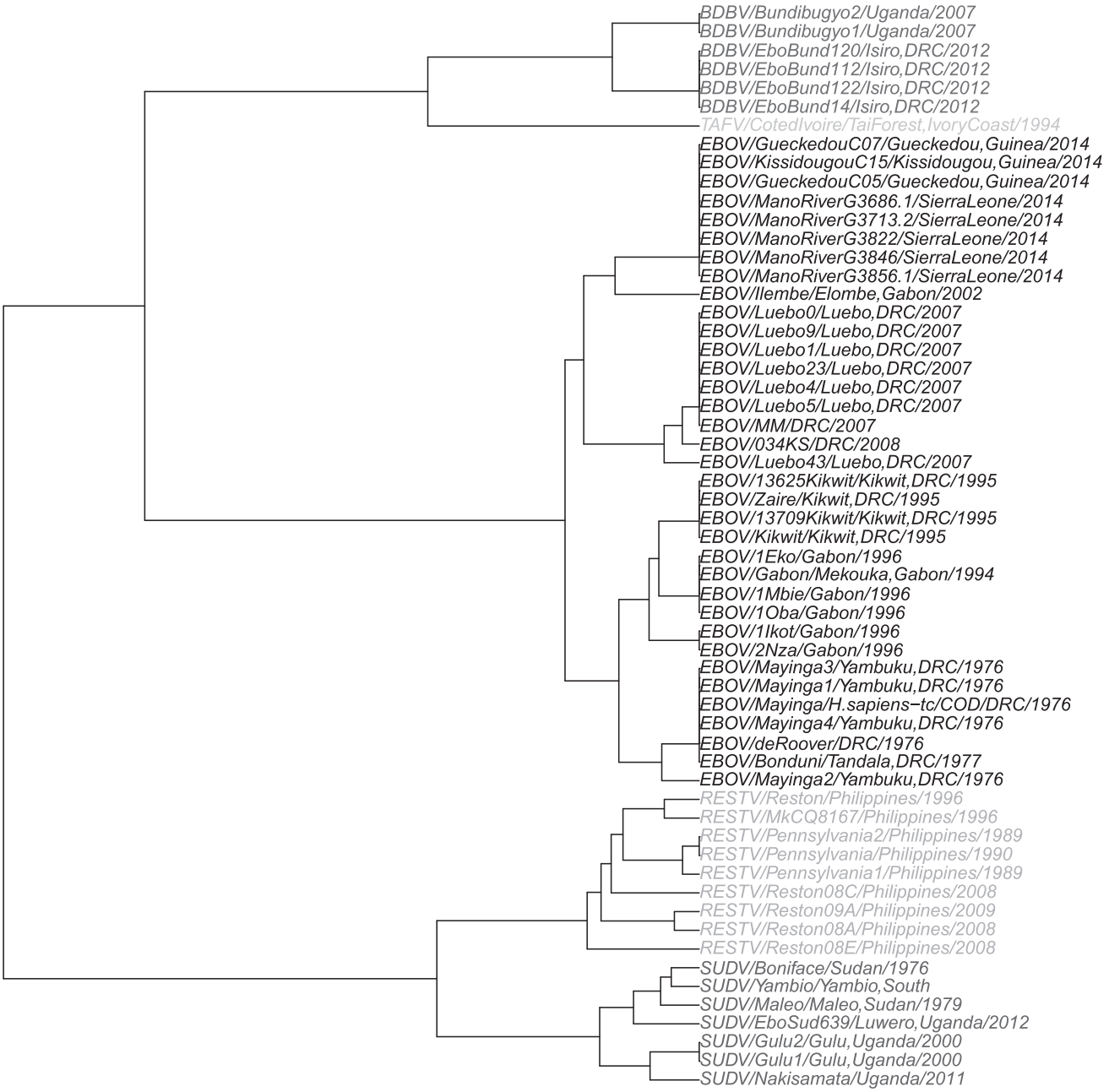

FIG. 6. Phylogenetic tree of ebolaviruses based on the distance matrix derived by VP35 sequences' natural vector through UPGMA.

As we know, *Reston ebolavirus* has high lethality only among monkeys and will not infect human beings. *Sudan ebolavirus* has at least 50% lethality in people of the previous outbreaks. *Sudan ebolavirus* should have more similarity with *Zaire ebolavirus* than *Reston ebolavirus*. Furthermore, regarding the genetic diversity among the *Zaire ebolavirus* species, this phylogenetic tree classifies the viruses in Gabon in 1996 and the viruses of DRC in 1995 together. This suggests that VP35 may not be as important in determining virus virulence.

The phylogenetic tree of VP30 is shown in Figure 7. We find that the viruses in Gabon in 1996 are classified with the Gabon virus of 2002, instead of the other three viruses in Gabon in 1996. Also, one Sierra Leone virus of 2014 is classified with the three Guinea viruses, instead of the other four Sierra Leone viruses. This result indicates that there are some gene mutations in VP30 from viruses in different geographic locations.

The phylogenetic tree based on VP40 sequences is presented in Figure 8. Two viruses from Gabon in 1996 are classified with the Gabon virus in 2002, instead of with the other three viruses from Gabon in 1996. Three of the five viruses from Sierra Leone in 2014 are in the same clade with the three viruses from Guinea. This result indicates that

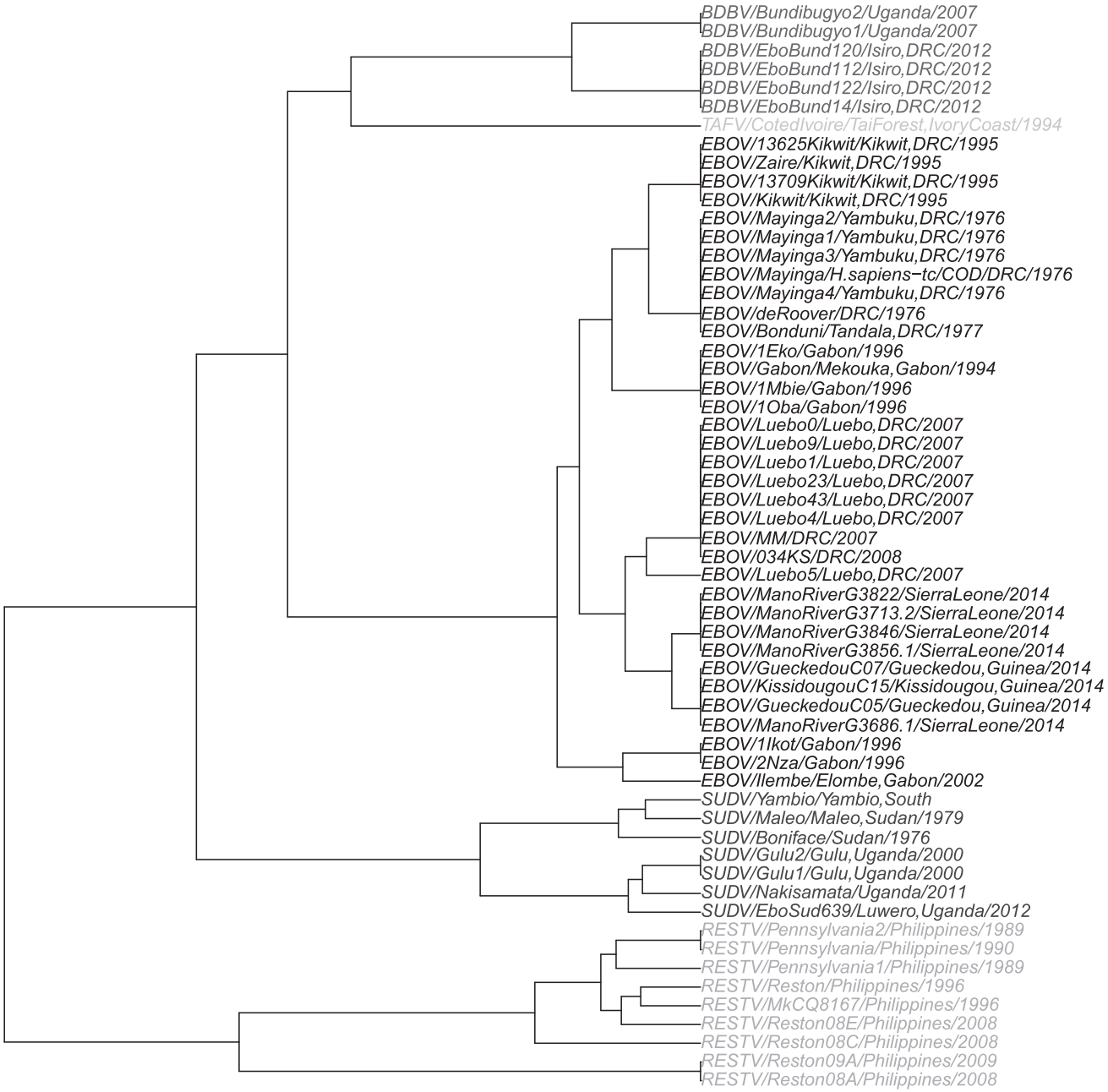

FIG. 7. Phylogenetic tree of ebolaviruses based on the distance matrix derived by VP30 sequences' natural vector through UPGMA.

there have been some gene mutations in VP40 from viruses in different geographic locations.

The phylogenetic tree based on LP sequences is presented in Figure 9. Although the five species are well separated, *Zaire ebolavirus* is classified with *Reston ebolavirus* but not *Sudan ebolavirus*. The distance between *Zaire ebolavirus* and *Reston ebolavirus* is 57.28, and the distance between *Zaire ebolavirus* and *Sudan ebolavirus* is 63.83. These distances do not match the different virulence. Furthermore, the virus locations are a little bit mixed within each species. The result suggests that the L-gene is not important for the classification of ebolaviruses.

As an alignment-free sequence analysis method, the natural vector method not only takes less computation time compared with other alignment-based methods but also works with high accuracy on classification. The natural vector method has advantages compared to other methods because it does not rely on any parameters. For example, the k-mer method depends on the choice of k's, and the criteria of how to determine the parameters in the k-mer method are controversial because different datasets need different parameters. This parameter-free feature allows the natural vector method to produce accurate, reliable, and consistent results.

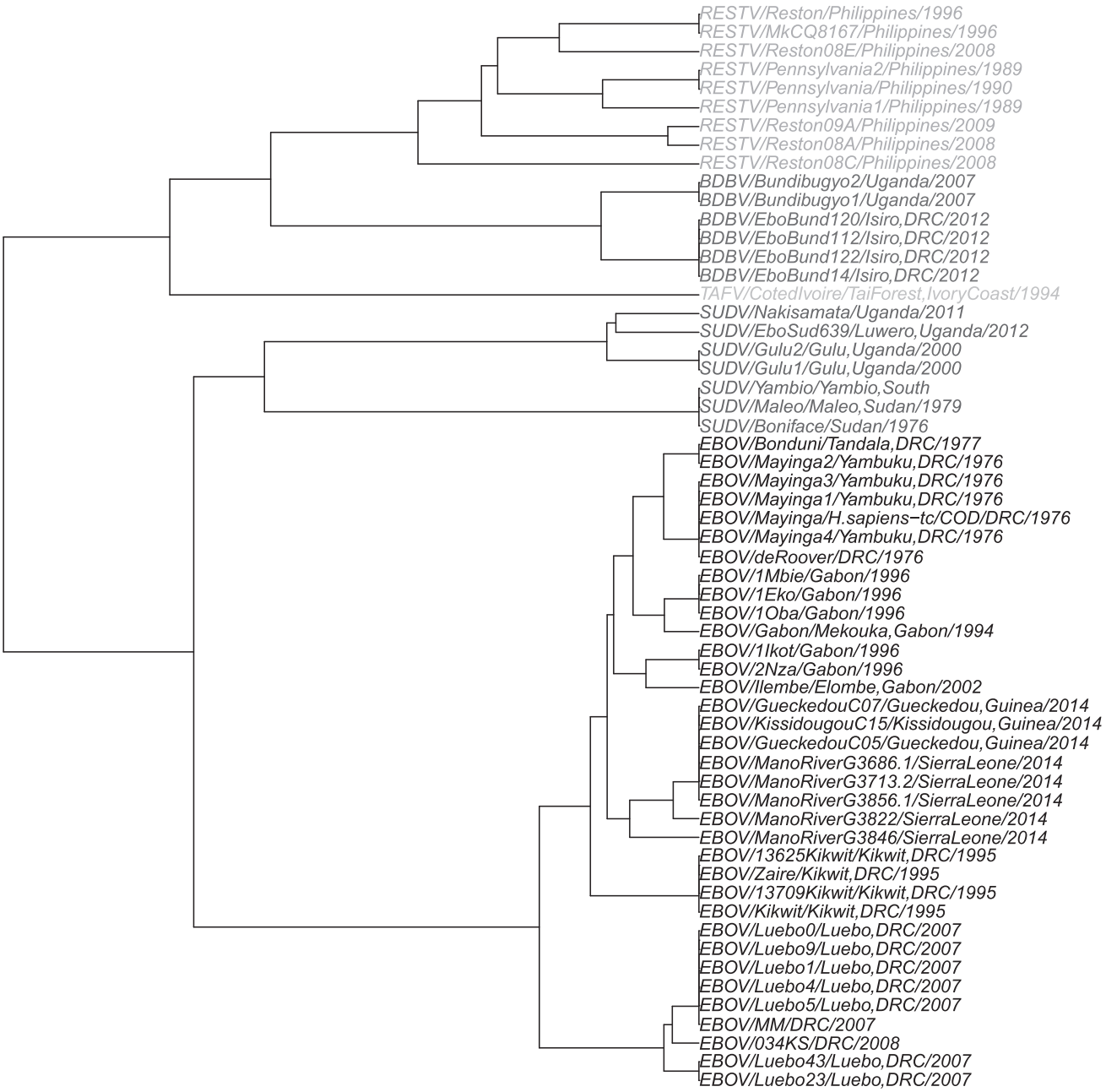

FIG. 8. Phylogenetic tree of ebolaviruses based on the distance matrix derived by VP40 sequences' natural vector through UPGMA.

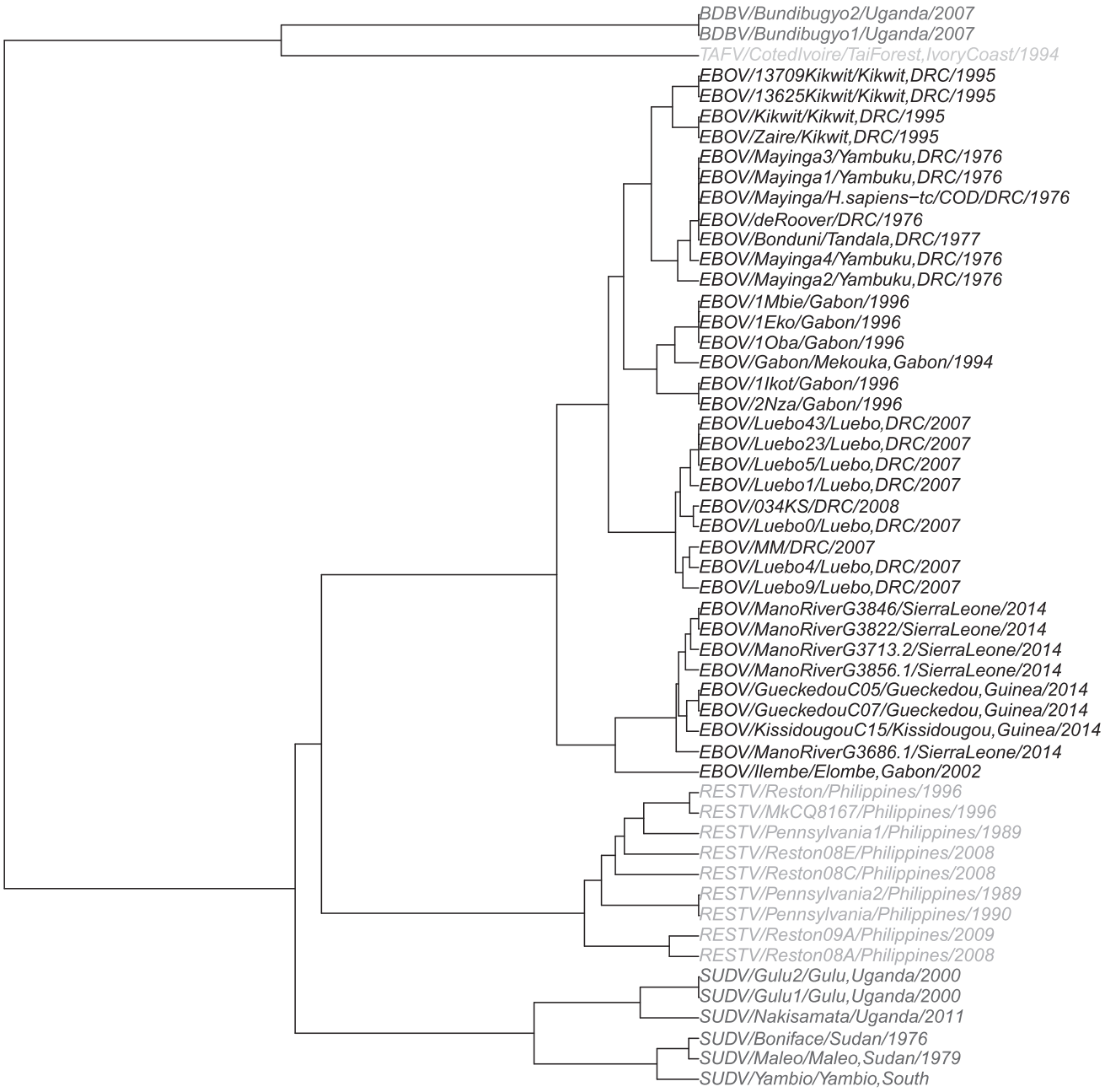

FIG. 9. Phylogenetic tree of ebolaviruses based on the distance matrix derived by L protein (LP) sequences' natural vector through UPGMA.

Conclusion

In this work, we employed the natural vector method to classify the 163 filoviruses by the three genera: *Ebolavirus*, *Marburgvirus*, and *Cuevavirus*. The phylogenetic tree indicates that the *Marburgvirus* is similar to *Ebolavirus*. The new 99 Serra Leone viruses belong to *Zaire ebolavirus* and they are very close to the three Guinea viruses. Within *Ebolavirus*, we plot phylogenetic trees based on the whole genetic sequences and their seven proteins. We conclude that VP24 is more important compared with other six proteins when the virus infects and replicates. This study will benefit the medical and scientific communities to make it possible in the future to track the source of viral disease outbreaks, prevent their spread, and design pharmaceutical products for disease therapy.

Acknowledgments

This research is supported by the USA Natural Science Foundation (DMS-1120824 to S.S.-T.Y., 1119612 to R.L.H.), National Institutes of Health (5 SC3 GM098180-04 to R.L.H.), National Natural Sciences Foundation of China (31271408 to S.S.-T.Y.,), Tsinghua University startup fund and Tsinghua University independent research project grant.

Disclosure Statement

No competing financial interests exist.

References

Baize, S., Pannetier, D., Oestereich, L., Rieger, T., Koivogui, L., Magassouba, N.F., and Günther, S. (2014). Emergence of Zaire Zaire ebolavirus disease in Guinea—preliminary report. N Engl J Med **371,** 1418–1425.

Bosio, C.M., Aman, M.J., Grogan, C., Hogan, R., Ruthel, G., Negley, D., Mohamadzadeh., M., Bavari, S., and Schmaljohn, A. (2003). Ebola and Marburg viruses replicate in monocyte-derived dendritic cells without inducing the production of cytokines and full maturation. J Infect Dis **188,** 1630–1638.

Deng, M., Yu, C., Liang, Q., He, R.L., and Yau, S.S.T. (2011). A novel method of characterizing genetic sequences: genome space with biological distance and applications. PLoS One **6,** e17293.

Domazet-Lošo, M., and Haubold, B. (2011). Alignment-free detection of horizontal gene transfer between closely related bacterial genomes. Mob Genet Elements **1,** 230.

Feldmann, H., and Geisbert, T.W. (2011). Ebola haemorrhagic fever. Lancet **377,** 849–862.

Feldmann, H., Jones, S., Klenk, H.D., and Schnittler, H.J. (2003). Ebola virus: from discovery to vaccine. Nat Rev Immunol **3,** 677–685.

Francica, J.R. (2010). A Study of the Zaire Ebolavirus Glycoprotein: Disruption of Host Surface Protein Function and Evasion of Immune Responses (Doctoral Dissertation, University of Pennsylvania).

Gatherer, D. (2014). The 2014 Zaire ebolavirus disease outbreak in west Africa. J of Gen Virol **95(Pt 8),** 1619–1624.

Gire, S.K., Goba, A., Andersen, K.G., Sealfon, R.S., Park, D.J., Kanneh, L., *et al.* (2014). Genomic surveillance elucidates Zaire ebolavirus origin and transmission during the 2014 outbreak. Science **345,** 1369–1372.

Groseth, A., Feldmann, H., Theriault, S., Mehmetoglu, G., and Flick, R. (2005). RNA polymerase I-driven minigenome system for Zaire ebolaviruses. J Virol **79,** 4425–4433.

Huang, H.H., Yu, C., Zheng, H., Hernandez, T., Yau, S.C., He, R.L., Yang, J., and Yau, S.S.T. (2014). Global comparison of multiple-segmented viruses in 12-dimensional genome space. Mol Phylogenet Evol **81,** 29–36.

Kuhn, J.H., Andersen, K.G., Bào, Y., Bavari, S., Becker, S., Bennett, R.S., *et al.* (2014). Filovirus RefSeq entries: evaluation and selection of filovirus type variants, type sequences, and names. Viruses **6,** 3663–3682.

Le Guenno, B., Formenty, P., and Boesch, C. (1999). Ebola virus outbreaks in the Ivory Coast and Liberia, 1994–1995. Curr Top Microbiol Immunol **235,** 77–84.

McElroy, A.K., Erickson, B.R., Flietstra, T.D., Rollin, P.E., Nichol, S.T., Towner, J.S., and Spiropoulou, C.F. (2014). Ebola hemorrhagic fever: novel biomarker correlates of clinical outcome. J Infect Dis **210,** 558–566.

Monath, T.P. (1999). Ecology of Marburg and Zaire ebolaviruses: speculations and directions for future research. J Infect Dis **179(Suppl 1),** S127–S138.

Roddy, P., Howard, N., Van Kerkhove, M.D., Lutwama, J., Wamala, J., Yoti, Z., *et al.* (2012). Clinical manifestations and case management of Ebola haemorrhagic fever caused by a newly identified virus strain, Bundibugyo, Uganda, 2007–2008. PLoS One **7,** e52986.

Sourdis, J., and Krimbas, C. (1987). Accuracy of phylogenetic trees estimated from DNA sequence data. Mol Biol Evol **4,** 159–166.

Volchkov, V.E., Volchkova, V.A., Chepurnov, A.A., Blinov, V.M., Dolnik, O., Netesov, S.V., and Feldmann, H. (1999). Characterization of the L gene and 5′trailer region of Zaire ebolavirus. J Gen Virol **80,** 355–362.

Watanabe, S., Noda, T., and Kawaoka, Y. (2006). Functional mapping of the nucleoprotein of Zaire ebolavirus. J Virol **80,** 3743–3751.

Wittmann, T.J., Biek, R., Hassanin, A., Rouquet, P., Reed, P., Yaba, P., Pourrut, X., Real, L.A., Gonzalez, J.-P., and Leroy, E.M. (2007). Isolates of Zaire ebolavirus from wild apes reveal genetic lineage and recombinants. Proc Natl Acad Sci U S A **104,** 17123–17127.

Xu, W., Edwards, M.R., Borek, D.M., Feagins, A.R., Mittal, A., Alinger, J.B., *et al.* (2014). Zaire ebolavirus VP24 targets a unique NLS binding site on karyopherin alpha 5 to selectively compete with nuclear import of phosphorylated STAT1. Cell Host Microbe **16,** 187–200.

Yu, C., Hernandez, T., Zheng, H., Yau, S.C., Huang, H.H., He, R.L., and Yau, S.S.T. (2013). Real time classification of viruses in 12 dimensions. PLoS One **8,** e64328.

Address correspondence to:
Stephen S.-T. Yau, PhD
Department of Mathematical Statistics
Tsinghua University
Beijing 100084
People's Republic of China

E-mail: yau@uic.edu

Received for publication September 12, 2014; received in revised form February 12, 2015; accepted February 20, 2015.