Viruses **2012**, *4*, 1425-1437; doi:10.3390/v4091425

OPEN ACCESS

viruses

ISSN 1999-4915

www.mdpi.com/journal/viruses

Article

Genetics-Based Classification of Filoviruses Calls for Expanded Sampling of Genomic Sequences

Chris Lauber [1] and Alexander E. Gorbalenya [1,2,*]

[1] Molecular Virology Laboratory, Department of Medical Microbiology, Leiden University Medical Center, 2333 ZA Leiden, The Netherlands; E-Mail: c.lauber@lumc.nl

[2] Faculty of Bioengineering and Bioinformatics, Lomonosov Moscow State University, 119899 Moscow, Russia

* Author to whom correspondence should be addressed; E-Mail: a.e.gorbalenya@lumc.nl; Tel.: +31-71-526-1436; Fax: +31-71-526-6761.

Received: 3 July 2012; in revised form: 23 August 2012 / Accepted: 24 August 2012 / Published: 31 August 2012

Abstract: We have recently developed a computational approach for hierarchical, genome-based classification of viruses of a family (DEmARC). In DEmARC, virus clusters are delimited objectively by devising a universal family-wide threshold on intra-cluster genetic divergence of viruses that is specific for each level of the classification. Here, we apply DEmARC to a set of 56 filoviruses with complete genome sequences and compare the resulting classification to the ICTV taxonomy of the family *Filoviridae*. We find in total six candidate taxon levels two of which correspond to the species and genus ranks of the family. At these two levels, the six filovirus species and two genera officially recognized by ICTV, as well as a seventh tentative species for Lloviu virus and prototyping a third genus, are reproduced. DEmARC lends the highest possible support for these two as well as the four other levels, implying that the actual number of valid taxon levels remains uncertain and the choice of levels for filovirus species and genera is arbitrary. Based on our experience with other virus families, we conclude that the current sampling of filovirus genomic sequences needs to be considerably expanded in order to resolve these uncertainties in the framework of genetics-based classification.

Keywords: genetics-based classification; taxonomy; DEmARC; filoviruses; ebolavirus; virus species; genetic diversity; pairwise evolutionary distances

1. Introduction

For a steadily growing number of viruses the genome sequence is the first and only information available. Because experimental characterization lags behind and is unlikely to be pursued for many viruses, comparative sequence analysis plays a central role in identifying commonalities and specifics between viruses. In this framework, researchers increasingly explore the usability of genetic sequences for virus classification. We have recently introduced a computational approach to hierarchically classify viruses of a family by relying solely on genetic data, coined DEmARC. Briefly, in DEmARC virus clusters are delimited by devising a threshold on the maximum intra-cluster (intra-taxon) divergence of viruses. This is done separately for each level of the hierarchical classification, all of which are selected using a cost function that measures the quality of virus clustering. The approach was extensively evaluated in a case study of picornaviruses [1]. Strikingly, the DEmARC-based picornavirus classification showed only few, but biologically notable, deviations from the ICTV taxonomy of the family *Picornaviridae* [2], the latter being developed by extensive efforts of expert picornavirologists who rely on various virus characteristics [3]. This analysis revealed important key parameters of DEmARC that distinguishes it from distance-based classification approaches and their applications of similar studies [4–17]. The DEmARC specifics include (i) the use of pairwise evolutionary distances (PEDs) instead of uncorrected p-distances, and (ii) a quantitative method to devise taxon levels and associated PED thresholds for virus clustering in a systematic and family-wide manner. We also reasoned that, in order to avoid a biased selection of genes/protein domains and to represent a virus genome as fully as possible, all family-wide conserved proteins must be used in a DEmARC-mediated analysis. As result, most incomplete genome sequences are excluded from the analysis. Since some of these differences are evolutionary-based, the DEmARC-based picornavirus classification enabled biological implications that are not (yet) available in taxonomy or through other approaches to virus classification. They include the prediction of known and currently unknown genetic diversity in the family and the proposed genetic separation of members of virus species [2]. So far, DEmARC was used to extensively revise the taxonomy of coronaviruses [18], and to propose the classification of two recently discovered insect nidoviruses [19,20] into the tentative new family "Mesoniviridae" [21]. In order to validate a general applicability of DEmARC in RNA virus taxonomy a systematic analysis of viruses from diverse families is most wanted.

In this study we sought to apply DEmARC to filoviruses, making it the first analysis of viruses with RNA genomes of negative polarity (ssRNA−). Filoviruses form the family *Filoviridae* of the order *Mononegavirales*, the latter combining all known ssRNA− viruses with non-segmented genomes [22]. Currently the filovirus genera *Marburgvirus* and *Ebolavirus*, which comprise one and five species, respectively, are recognized [23–25]. Additionally, a tentative genus "Cuevavirus" with a single species formed by Lloviu virus has been proposed [24]. Filovirus species are delimited using both phenotypic and genetic demarcation criteria including thresholds on percentage identity of full-length genomic sequences [24–27]. The filovirus genome of about 19 kb encodes seven structural proteins in separate open reading frames (ORFs). They include a nucleoprotein (NP), a spike glycoprotein ($GP_{1,2}$), two matrix proteins (VP40 and VP24), a multi-domain protein (L) with RNA-dependent RNA polymerase (RdRp) and methyltransferase function [28], an RdRp cofactor (VP35), and a transcriptional activator (VP30) [29]. Additionally, some filoviruses may encode lineage-specific proteins [30–33].

Marburg- and ebolaviruses are endemic to central Africa and the Philippines [26,27,34,35] while Lloviu virus was discovered in southern Europe [30]. Some filoviruses have been isolated from bats and domesticated pigs and can cause hemorrhagic fevers in primates with often fatal outcome. Since they comprise some of the most dangerous pathogens in the world, the taxonomy of filoviruses, especially at the species level, may have considerable practical implications.

2. Results and Discussion

The dataset of this study was formed by 56 complete filovirus genome sequences that we downloaded into the Viralis platform [36] in February 2012. For these filoviruses, a concatenated multiple alignment of the seven proteins that are conserved family-wide was constructed and submitted to PED calculation. The resulting PED values are distributed non-uniformly along the range of 0 to 1.2 substitutions per amino acid position (Figure 1A). Using DEmARC we identified six distance threshold candidates, each associated with the optimal clustering cost of zero (Figure 1B) (see Experimental section). With this highest possible support all intra-cluster PED values would fall below the respective distance threshold, indicating that all clusters in a respective level may not be improved further (see [1] for technical details). This result suggests the use of all six thresholds for building a classification. However, the number of ranks below the family level is limited to three in virus taxonomy (subfamily, genus, and species). To satisfy this limitation, three thresholds must be selected by using additional criteria, e.g., biological properties. Within the DEmARC framework we used those thresholds that are associated with the highest threshold support measure (TSM) values. This use of TSM values is non-canonical: in DEmARC they are commonly used for the selection of PED ranges in which thresholds are further identified by local cost optimization if that is attainable (in situations where the optimal clustering cost of zero cannot be achieved; see [1] for technical details). In the current analysis of filoviruses, however, using each of the PED threshold candidates within the six PED ranges would result in a clustering cost of zero (Figure 1B). Thus, for each range we arbitrarily selected the smallest observed PED value within the range as the threshold value (colored arrows in Figure 1B). We note that the produced classification accommodates laboratory-introduced genetic variation in some sequences due to virus propagation in tissue culture before sequencing. The observed continuous ranges with zero PED frequency imply that the scale of this variation is small compared to the natural genetic variation even at the lowest level of the derived classification.

The first selected threshold (PED of 0.120) results in seven clusters (Figure 1B) that match the official or tentative ICTV species of the family *Filoviridae*. These are species *Marburg marburgvirus* (comprising 34 virus sequences), species *Zaire ebolavirus* (8), species *Reston ebolavirus* (6), species *Sudan ebolavirus* (3), species *Taï Forest ebolavirus* (2), species *Bundibugyo ebolavirus* (2), and tentative species “Lloviu cuevavirus” (1). According to the second threshold (PED of 0.396), three clusters that match the official or tentative genera—*Marburgvirus*, *Ebolavirus*, and “Cuevavirus”—of the family are recognized (Figure 1B). The third threshold (PED of 0.806) joins viruses of the genus *Ebolavirus* and the tentative genus “Cuevavirus” into a single cluster while viruses of the genus *Marburgvirus* form the second cluster; these two clusters could be provisionally treated as tentative subfamilies. Hierarchical relationships of the 12 clusters according to the three applied distance thresholds are shown in Figure 2A. All delineated clusters form monophyletic lineages in the phylogeny of the 56 filoviruses (reciprocal monophyly) (Figure 3) when assuming the root to be closest to the branch leading to marburgviruses (which would correspond to midpoint rooting). The three threshold candidates not considered for classification (PED of 0.044, 0.204, and 0.274) would result in eight, six, and five clusters, respectively (Figure 1B). The clustering in eight clusters would split viruses of the species *Marburg marburgvirus* into two clusters formed by the RAVN and MARV lineage, respectively (Figure 2B). The clustering in six clusters would join viruses of the species *Taï Forest ebolavirus* and *Bundibugyo ebolavirus* into a single cluster (Figure 2C). The clustering in five clusters would join viruses of the species *Zaire ebolavirus*, *Taï Forest ebolavirus*, and *Bundibugyo ebolavirus* into a single cluster (Figure 2D). It would be reasonable to consider these eight-, six-, and five-cluster scenarios as alternative species groupings. On the other hand, it could indicate that the three taxonomic ranks below the family level may be insufficient to accurately classify the genetic diversity of filoviruses. In the DEmARC-mediated classification of coronaviruses we also observe numerous PED thresholds with zero clustering cost [37].

The PED threshold delineating the seven ICTV species is controlled by a single cluster, species *Marburg marburgvirus*, which shows by far the highest sampling (61% of all sequences) among all clusters of that level (Figures 2A and 3). This cluster contributes the largest intracluster PED value (0.120) of that level compared to other clusters for which values of at most 0.036 (between viruses of the species *Reston ebolavirus*) are observed (Figure 3). These considerable differences in the divergence of viruses from different filovirus species can be rationalized through two contrasting explanations that both exploit possible effects of biased virus sampling on the classification. First, the observed differences could be a result of the relatively poor sampling of virus sequences from the five ebolavirus and the “cuevavirus” species. Once sampling is improved, viruses from these species may show a genetic divergence comparable to marburgviruses. Alternatively, future improved virus sampling might show that the five ebolavirus species recognized by ICTV actually form a single species. The full PED range from 0 to around 0.4 (Figure 1A) might then be populated which would merge viruses of the five ebolavirus species into a single cluster. This would result in three species clusters in total corresponding to the three ICTV genera of the family *Filoviridae*. Consequently, ebolaviruses and “cuevaviruses” would form a single genus (in addition to the genus *Marburgvirus*) as suggested by the PED threshold of 0.806 (Figures 1 and 3). The above scenarios are just few from many possible (e.g., see the alternative species groupings in Figure 2B–D) and illustrate the current uncertainty about filovirus classification. The three-species scenario seems to be conceivable when comparing the filovirus PED distribution with that of the well-sampled family *Picornaviridae* with up to 260 available sequences per species (more than 1,200 sequences in total distributed among 38 species clusters) [1]. In the high-sampling case of picornaviruses, no PED values with zero frequency are observed which suggests that the current sampling of filovirus genome sequences may strongly underestimate the natural genetic diversity in the family. Expanded virus sampling might also lead to a better differentiation of filovirus proteins by evolutionary criteria. Currently, all seven proteins are conserved family-wide among known filoviruses, which may change for this family in the future when most diverged viruses are separated by (much) larger genetic distances, as we already observe for picornaviruses and many other families. This development would also affect the choice of proteins by DEmARC and, consequently, the resulting classification. Furthermore, the PED threshold values are likely to change in the future, even if the underlying virus clusters will be stable, given the large PED ranges they currently represent (Figure 1B). Thus, a definite decision about number and virus composition of filovirus species and genera as well as stable demarcation thresholds will only be possible if the sampling of filovirus genome sequences is expanded in the future.

We note that the use of pairwise sequence similarities is becoming increasingly popular for assisting decision-making in virus taxonomy (see the Introduction for literature). On the other hand, there is the current paradigm that the use of a single (e.g., genetic) criterion is insufficient for the demarcation of viral taxa [38]. In our opinion, a fully genetics-based classification provides a promising and meaningful foundation for virus taxonomy. Indeed, the genome is the principal carrier of genetic information and heredity; and it was already acknowledged that virus taxonomy should reflect the evolutionary history of viruses [39] which can be reconstructed from genetic data. However, analyzing the evolutionary record in genomes remains a challenging task with many parameters to define and is dependent on the amount of available genetic information (see above). Consequently, the technical implementation and associated choices made during an analysis (e.g., using a relatively simplistic measure of pairwise distances) may affect the quality of genetics-based classification. Different directions of future research efforts are conceivable, ranging from the development of improved evolutionary models for the calculation of genetic distances [40] to entirely different techniques of utilizing the genetic information for virus classification [41,42]. As the development and application of DEmARC illustrates, this line of research offers a possibility of tackling the virus classification problem in a systematic manner.

In the DEmARC framework, an upper limit on intra-species genetic divergence is imposed on loci encoding the family-wide conserved proteins. Consequently, when there is a strong support for species, viruses of the same species are genetically separated from viruses outside the species in these loci. For picornaviruses with their RNA genomes of positive polarity (ssRNA+), this separation may be promoted by mutation and limited through homologous recombination [2,43,44], as we argued [2]. Recombination among ssRNA− viruses was estimated to be generally rare [45] and this was explained by major differences in the replication cycle (due to the negative polarity of genomes) compared to ssRNA+ viruses, which limits the template-switching ability of the RdRp through rapid packaging of the genomic RNA with ribonucleoproteins [46]. However, homologous (intrasegmental) recombination can occur as was found for different ssRNA− viruses [47–49] including the prototype ebolavirus, Ebola virus [50]. The frequency with which both ssRNA+ and ssRNA− viruses recombine in nature remains to be shown but it may not be uncommon [51], and currently available recombination detection tools were shown to underestimate this frequency in certain situations [49]. Ultimately, both the rate of homologous recombination among RNA viruses in nature and the model of genetic separation of virus speciation should be probed experimentally.

Figure 1. Filovirus-wide pairwise evolutionary distance (PED) distribution and thresholds for virus clustering. (**A**) Shown is a frequency distribution (white bins) of around 1,500 PED values, which estimate the average number of amino acid substitutions per site between each pair of viruses. A mixture model (red curve) was fitted to the PED distribution and used to calculate a threshold support measure (TSM, green). For details on threshold delineation see [1]. A threshold candidate (peak in the TSM measure) may be used to group the viruses into clusters in which virus pairs with a PED not exceeding the threshold join the same cluster. The three candidates with highest TSM scores were used to hierarchically group the 56 filoviruses at three levels comprising seven, three, and two clusters, respectively. Each of the thresholds is located within a continuous PED range for which no PED values are observed with the current filovirus sampling (striped background shading). PED values in these ranges (if sampled in the future) could be associated with either the classification level delimited by the respective threshold (intra-cluster distances) or with the next higher level (inter-cluster distances). The actual PED values of the thresholds are thus uncertain (bright-colored horizontal bars and question marks); for simplicity we selected the values (0.120, 0.396, and 0.806) that correspond to the smallest value within the respective PED range. (**B**) The change in the number of derived clusters with respect to the value of the PED threshold candidate is shown. For each threshold candidate a continuous PED range with optimal clustering cost of zero (no intra-cluster PED values exceed the threshold) is highlighted in green. The six threshold candidates considered in this study show the following PED ranges (from left to right): 0.044–0.109, 0.120–0.203, 0.204–0.267, 0.274–0.369, 0.396–0.783, and 0.806–1.097. The three threshold candidates used in (A) are indicated by colored arrows; three alternative PED thresholds for the species level are shown by black arrows. Note that another clustering (12 clusters) with zero cost was not considered in this study because of the marginal PED range and low TSM support of the associated threshold (PED of 0.025).

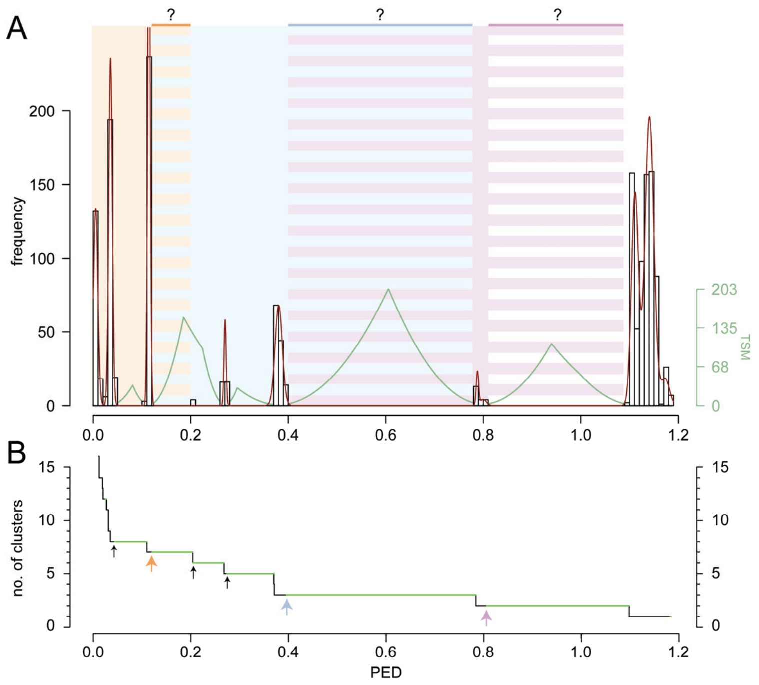


Figure 2. Genetics-based classification of filoviruses and alternative species groupings. A quadrant is used to visualize the classification of the 56 filoviruses with three hierarchical levels. The axes indicate intervirus genetic divergence (as PED) which increases linearly from the perimeter of the quadrant (zero PED) to its origin (maximum PED of 1.2). The three classification levels are highlighted using three basic colors (orange, blue, and purple). Each color exists in two shadings that highlight the limit on expected intragroup genetic divergence according to a distance threshold (soft shading) and the maximum observed intragroup genetic divergence (bright shading) of a cluster. (**A**) The genetics-based filovirus classification using the three top-ranked PED thresholds. It comprises seven, three, and two clusters, respectively, at the three hierarchical levels. The seven orange clusters of the lowest level correspond to the official or tentative ICTV species of the family *Filoviridae* [25] and are indicated by names. The three blue clusters correspond to the official or tentative ICTV genera of the family *Filoviridae* (from left to right: "Cuevavirus", *Ebolavirus*, and *Marburgvirus*); the two purple clusters correspond to supra-generic taxa currently not recognized in the ICTV filovirus classification. Outside the quadrant, the relative density of virus sampling per ICTV species is shown as gray shadings from low (light) to high (dark) sampling, which is in the range of 1 to 34. (**B**) An alternative classification with eight instead of seven clusters at the lowest level. It was derived by using the alternative species threshold at PED = 0.044. (**C**) An alternative classification with six instead of seven clusters at the lowest level. It was derived by using the alternative species threshold at PED = 0.204. (**D**) An alternative classification with five instead of seven clusters at the lowest level. It was derived by using the alternative species threshold at PED = 0.274.

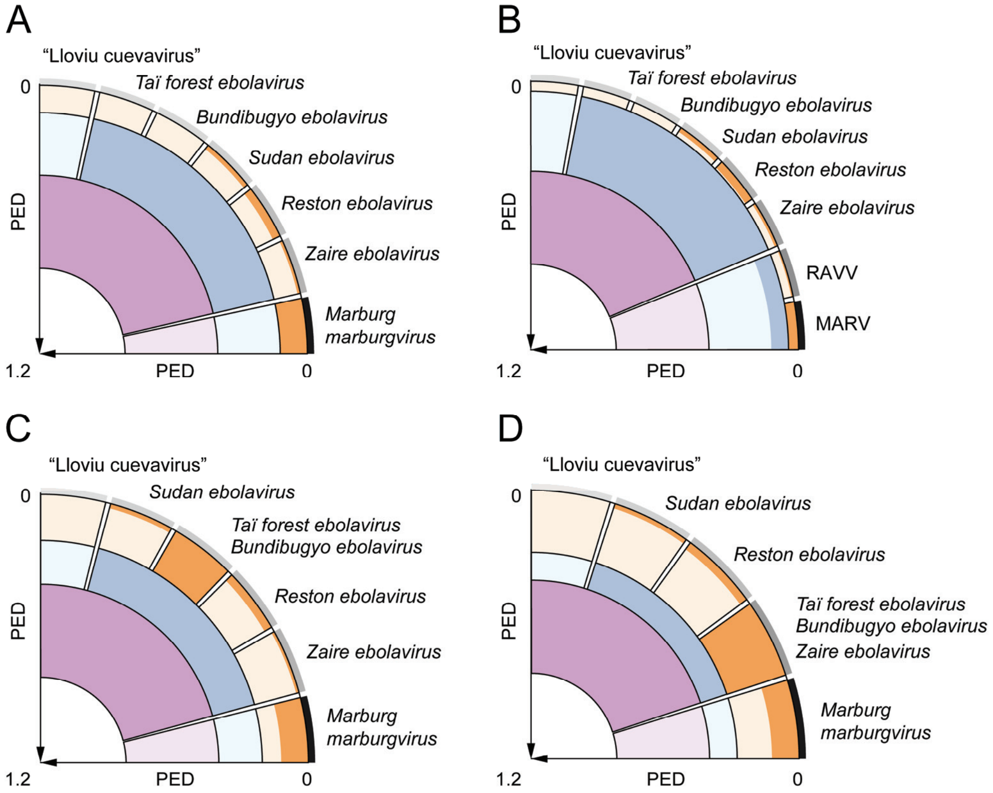
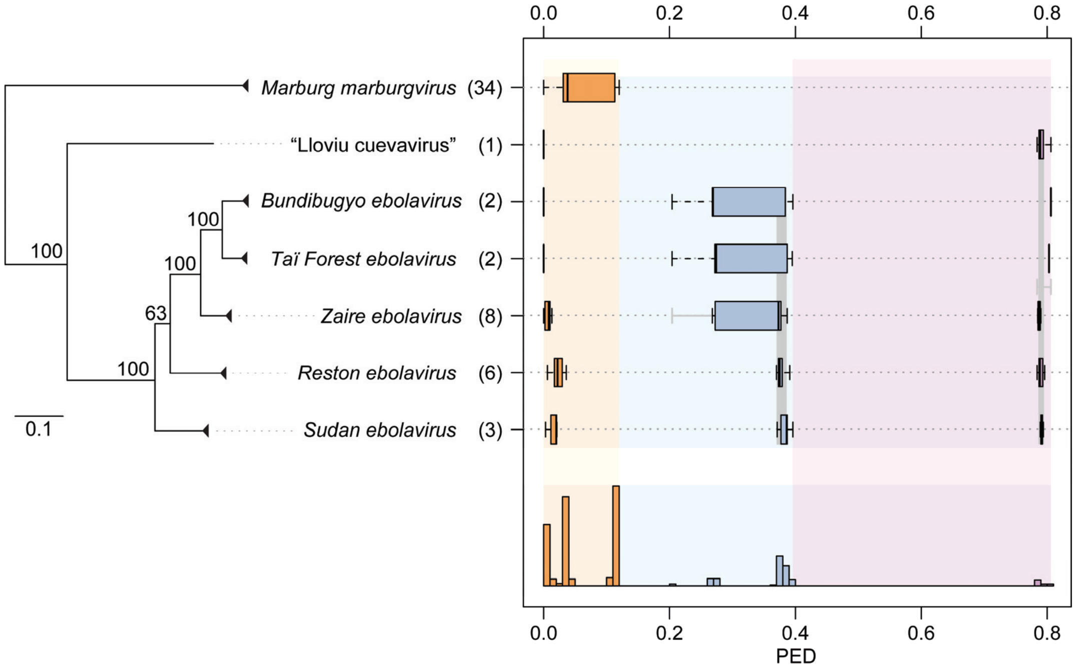


Figure 3. Intragroup genetic divergence of filoviruses. Box-and-whisker graphs are used to plot level-specific PED distributions for the seven clusters of the lowest classification level (bright-shaded orange, blue, and purple). The seven orange clusters correspond to the official or tentative ICTV species of the family *Filoviridae* [25] and are indicated by names; virus sampling per ICTV species is shown in brackets. The combined PED distributions of clusters of the second and third level are shown as gray box-and-whisker graphs. The expected range of level-specific PED values, bordered by two distance thresholds, is indicated by light-shaded background colors. The respective part of the PED distribution covering the full range of intragroup PED values of the three levels is show at the bottom. The ICTV species are grouped vertically according to a maximum likelihood phylogeny shown at the left. Internal nodes of the tree that correspond to ICTV species are collapsed (triangles); all of them have a bootstrap support value of 100.



3. Experimental Section

The seven filovirus proteins that are conserved in all known members of the family (NP, VP35, VP40, $GP_{1,2}$, VP30, VP24, and L) were aligned separately at the amino acid level using the program Muscle version 3.52 [52] followed by manual correction. The seven protein alignments were concatenated to form a single alignment of 5,143 positions with a gap content of 5.5%. To estimate the genetic similarity between virus pairs, PED values were calculated on this concatenated alignment using the program Tree-Puzzle version 5.2 [53]; the WAG amino acid substitution matrix was applied [54]. Proteins that are not conserved family-wide, like sGP and ssGP of ebolaviruses and "cuevaviruses" or certain hypothetical proteins encoded by the anti-sense genomic RNA, were not included in the calculation of PED values. For these proteins, an accurate estimation of genetic divergence may be approached only for the selected filoviruses that encode these proteins. The distribution of all PED values was partitioned into intra-rank and inter-rank ranges using a systematic approach implemented in DEmARC [1]. This partitioning is achieved through the inference of PED thresholds below which two viruses are grouped together. We refer to the resulting virus groups as "clusters" in the context of genetic classification by DEmARC. A cluster might correspond to a viral taxon officially recognized by ICTV.

A derivative of the multiple alignment used for PED calculation, from which strongly conserved blocks [55] (in total 3,522 alignment positions, 68.5%) have been extracted by BAGG [56], formed the dataset for unrooted tree reconstruction by PhyML version 3.0 [57]; the WAG amino acid substitution matrix was applied [54]; support for internal nodes was obtained through a non-parametric bootstrap analysis with 100 replicates.

4. Conclusions

DEmARC offers a systematic and quantitative framework for virus classification that utilizes genome sequences, the only information available for a growing number of viruses. The striking agreement on species and genus taxa between the DEmARC-mediated filovirus classification of this study and the taxonomy of the family *Filoviridae* could be considered a cross-validation for both. However, we note that this classification is one of many equally strongly supported classifications, as DEmARC identified in total six potential taxon levels. Each of these taxon levels is associated with a large continuous range of PED values that are not sampled (yet). Consequently, we conclude that the current coverage of the natural genetic diversity of filoviruses is limited and needs to be considerably expanded, also concerning hosts not sampled so far, in order to gain certainty about both filovirus taxa and levels.

Acknowledgments

We thank Igor Sidorov, Alexander Kravchenko, and Dmitry Samborskiy for expert administration of the Viralis platform. This research was partially supported through the European Union Seventh Framework Programme (FP7/2007–2013) under the program SILVER (grant agreement no. 260644), the Collaborative Agreement in Bioinformatics between Leiden University Medical Center and Moscow State University (MoBiLe), and Leiden University Fund.

Conflict of Interest

The authors declare no conflict of interest.

References and Notes

1. Lauber, C.; Gorbalenya, A.E. Partitioning the genetic diversity of a virus family: Approach and evaluation through a case study of picornaviruses. *J. Virol.* **2012**, *86*, 3890–3904.
2. Lauber, C.; Gorbalenya, A.E. Toward genetics-based virus taxonomy: Comparative analysis of a genetics-based classification and the taxonomy of picornaviruses. *J. Virol.* **2012**, *86*, 3905–3915.

3. Knowles, N.J.; Hovi, T.; Hyypia, T.; King, A.M.Q.; Lindberg, A.M.; Pallansch, M.A.; Palmenberg, A.C.; Simmonds, P.; Skern, T.; Stanway, G.; *et al*. Family *Picornaviridae*. In *Virus Taxonomy, Ninth Report of the International Committee on Taxonomy of Viruses*; King, A.M.Q., Adams, M.J., Carstens, E.B., Lefkowitz, E.J., Eds.; Elsevier Academic Press: Amsterdam, The Netherlands, 2012; pp. 855–880.
4. Adams, M.J.; Antoniw, J.F.; Bar-Joseph, M.; Brunt, A.A.; Candresse, T.; Foster, G.D.; Martelli, G.P.; Milne, R.G.; Fauquet, C.M. The new plant virus family Flexiviridae and assessment of molecular criteria for species demarcation. *Arch. Virol.* **2004**, *149*, 1045–1060.
5. Adams, M.J.; Antoniw, J.F.; Fauquet, C.M. Molecular criteria for genus and species discrimination within the family Potyviridae. *Arch. Virol.* **2005**, *150*, 459–479.
6. Bao, Y.; Kapustin, Y.; Tatusova, T. Virus classification by Pairwise Sequence Comparison (PASC). In *Encyclopedia of Virology*; Mahy, B.W.J., van Regenmortel, M.H.V., Eds.; Elsevier: Oxford, UK, 2008; Volume 5: pp. 342–348.
7. Bernard, H.U.; Burk, R.D.; Chen, Z.G.; van Doorslaer, K.; zur Hausen, H.; de Villiers, E.M. Classification of papillomaviruses (PVs) based on 189 PV types and proposal of taxonomic amendments. *Virology* **2010**, *401*, 70–79.
8. Chan, Y.F.; Sam, I.C.; Abubakar, S. Phylogenetic designation of enterovirus 71 genotypes and subgenotypes using complete genome sequences. *Inf. Genet. Evol.* **2010**, *10*, 404–412.
9. Fauquet, C.M.; Bisaro, D.M.; Briddon, R.W.; Brown, J.K.; Harrison, B.D.; Rybicki, E.P.; Stenger, D.C.; Stanley, J. Revision of taxonomic criteria for species demarcation in the family Geminiviridae, and an updated list of begomovirus species. *Arch. Virol.* **2003**, *148*, 405–421.
10. Gonzaalez, J.M.; Gomez-Puertas, P.; Cavanagh, D.; Gorbalenya, A.E.; Enjuanes, L. A comparative sequence analysis to revise the current taxonomy of the family Coronaviridae. *Arch. Virol.* **2003**, *148*, 2207–2235.
11. Lefkowitz, E.J.; Wang, C.; Upton, C. Poxviruses: Past, present and future. *Virus Res.* **2006**, *117*, 105–118.
12. Maes, P.; Klempa, B.; Clement, J.; Matthijnssens, J.; Gajdusek, D.C.; Kruger, D.H.; van Ranst, M. A proposal for new criteria for the classification of hantaviruses, based on S and M segment protein sequences. *Inf. Genet. Evol.* **2009**, *9*, 813–820.
13. Matthijnssens, J.; Ciarlet, M.; Heiman, E.; Arijs, I.; Delbeke, T.; McDonald, S.M.; Palombo, E.A.; Iturriza-Gomara, M.; Maes, P.; Patton, J.T.; *et al.* Full genome-based classification of rotaviruses reveals a common origin between human Wa-like and porcine rotavirus strains and human DS-1-like and bovine rotavirus strains. *J. Virol.* **2008**, *82*, 3204–3219.
14. Oberste, M.S.; Maher, K.; Kilpatrick, D.R.; Pallansch, M.A. Molecular evolution of the human enteroviruses: Correlation of serotype with VP1 sequence and application to picornavirus classification. *J. Virol.* **1999**, *73*, 1941–1948.
15. Schuffenecker, I.; Ando, T.; Thouvenot, D.; Lina, B.; Aymard, M. Genetic classification of "Sapporo-like viruses". *Arch. Virol.* **2001**, *146*, 2115–2132.
16. Shukla, D.D.; Ward, C.W. Amino-acid sequence homology of coat proteins as a basis for identification and classification of the potyvirus group. *J. Gen. Virol.* **1988**, *69*, 2703–2710.
17. Zheng, D.P.; Ando, T.; Fankhauser, R.L.; Beard, R.S.; Glass, R.I.; Monroe, S.S. Norovirus classification and proposed strain nomenclature. *Virology* **2006**, *346*, 312–323.

18. De Groot, R.J.; Baker, S.C.; Baric, R.; Enjuanes, L.; Gorbalenya, A.E.; Holmes, K.V.; Perlman, S.; Poon, L.L.; Rottier, P.J.M.; Talbot, P.J.; *et al*. Family *Coronaviridae*. In *Virus Taxonomy, Ninth Report of the International Committee on Taxonomy of Viruses*; King, A.M.Q., Adams, M.J., Carstens, E.B., Lefkowitz, E.J., Eds.; Elsevier Academic Press: Amsterdam, The Netherlands, 2012; pp. 806–828.
19. Nga, P.T.; Parquet, M.D.; Lauber, C.; Parida, M.; Nabeshima, T.; Yu, F.X.; Thuy, N.T.; Inoue, S.; Ito, T.; Okamoto, K.; *et al.* Discovery of the first insect nidovirus, a missing evolutionary link in the emergence of the largest RNA virus genomes. *PLoS Pathog.* **2011**, *7*, e1002215.
20. Zirkel, F.; Kurth, A.; Quan, P.L.; Briese, T.; Ellerbrok, H.; Pauli, G.; Leendertz, F.H.; Lipkin, W.I.; Ziebuhr, J.; Drosten, C.; *et al.* An insect nidovirus emerging from a primary tropical rainforest. *mBio* **2011**, *2*, e00077-11.
21. Lauber, C.; Ziebuhr, J.; Junglen, S.; Drosten, C.; Zirkel, F.; Nga, P.T.; Morita, K.; Snijder, E.J.; Gorbalenya, A.E. Mesoniviridae: A proposed new family in the order Nidovirales formed by a single species of mosquito-borne viruses. *Arch. Virol.* **2012**, doi:10.1007/s00705-012-1295-x.
22. Easton, A.J.; Pringle, C.R. Order *Mononegavirales*. In *Virus Taxonomy, Ninth Report of the International Committee on Taxonomy of Viruses*; King, A.M.Q., Adams, M.J., Carstens, E.B., Lefkowitz, E.J., Eds.; Elsevier Academic Press: Amsterdam, The Netherlands, 2012; pp. 653–657.
23. Adams, M.J.; Carstens, E.B. Ratification vote on taxonomic proposals to the International Committee on Taxonomy of Viruses (2012). *Arch. Virol.* **2012**, *157*, 1411–1422.
24. Kuhn, J.H.; Becker, S.; Ebihara, H.; Geisbert, T.W.; Johnson, K.M.; Kawaoka, Y.; Lipkin, W.I.; Negredo, A.I.; Netesov, S.V.; Nichol, S.T.; *et al.* Proposal for a revised taxonomy of the family Filoviridae: Classification, names of taxa and viruses, and virus abbreviations. *Arch. Virol.* **2010**, *155*, 2083–2103.
25. Kuhn, J.K.; Becker, S.; Ebihara, H.; Geisbert, T.W.; Jahrling, P.B.; Kawaoka, Y.; Netesov, S.V.; Nichol, S.T.; Peters, C.J.; Volchkov, V.E.; *et al*. Familiy *Filoviridae*. In *Virus Taxonomy, Ninth Report of the International Committee on Taxonomy of Viruses*; King, A.M.Q., Adams, M.J., Carstens, E.B., Lefkowitz, E.J., Eds.; Elsevier Academic Press: Amsterdam, The Netherlands, 2012; pp. 665–671.
26. Towner, J.S.; Khristova, M.L.; Sealy, T.K.; Vincent, M.J.; Erickson, B.R.; Bawiec, D.A.; Hartman, A.L.; Comer, J.A.; Zaki, S.R.; Stroher, U.; *et al.* Marburgvirus Genomics and association with a large hemorrhagic fever outbreak in Angola. *J. Virol.* **2006**, *80*, 6497–6516.
27. Towner, J.S.; Sealy, T.K.; Khristova, M.L.; Albarino, C.G.; Conlan, S.; Reeder, S.A.; Quan, P.L.; Lipkin, W.I.; Downing, R.; Tappero, J.W.; *et al.* Newly Discovered Ebola Virus Associated with Hemorrhagic Fever Outbreak in Uganda. *PLoS Pathog.* **2008**, *4*, e1000212.
28. Ferron, F.; Longhi, S.; Henrissat, B.; Canard, B. Viral RNA-polymerases—A predicted 2 '-O-ribose methyltransferase domain shared by all Mononegavirales. *Trends Biochem. Sci.* **2002**, *27*, 222–224.
29. Feldmann, H.; Muhlberger, E.; Randolf, A.; Will, C.; Kiley, M.P.; Sanchez, A.; Klenk, H.D. Marburg Virus, A Filovirus—Messenger-RNAs, gene order, and regulatory elements of the replication cycle. *Virus Res.* **1992**, *24*, 1–19.

30. Negredo, A.; Palacios, G.; Vazquez-Moron, S.; Gonzalez, F.; Dopazo, H.; Molero, F.; Juste, J.; Quetglas, J.; Savji, N.; Martinez, M.D.; *et al.* Discovery of an ebolavirus-like filovirus in Europe. *PLoS Pathog.* **2011**, *7*, e1002304.
31. Sanchez, A.; Trappier, S.G.; Mahy, B.W.J.; Peters, C.J.; Nichol, S.T. The virion glycoproteins of Ebola viruses are encoded in two reading frames and are expressed through transcriptional editing. *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93*, 3602–3607.
32. Volchkov, V.E.; Becker, S.; Volchkova, V.A.; Ternovoj, V.A.; Kotov, A.N.; Netesov, S.V.; Klenk, H.D. GP mRNA of Ebola virus is edited by the Ebola virus polymerase and by T7 and vaccinia virus polymerases. *Virology* **1995**, *214*, 421–430.
33. Volchkova, V.A.; Klenk, H.D.; Volchkov, V.E. Delta-peptide is the carboxy-terminal cleavage fragment of the nonstructural small glycoprotein sGP of Ebola virus. *Virology* **1999**, *265*, 164–171.
34. Leroy, E.M.; Kumulungui, B.; Pourrut, X.; Rouquet, P.; Hassanin, A.; Yaba, P.; Delicat, A.; Paweska, J.T.; Gonzalez, J.P.; Swanepoel, R. Fruit bats as reservoirs of Ebola virus. *Nature* **2005**, *438*, 575–576.
35. Miranda, M.E.; Ksiazek, T.G.; Retuya, T.J.; Khan, A.S.; Sanchez, A.; Fulhorst, C.F.; Rollin, P.E.; Calaor, A.B.; Manalo, D.L.; Roces, M.C.; *et al.* Epidemiology of Ebola (subtype Reston) virus in the Philippines, 1996. *J. Infect. Dis.* **1999**, *179*, S115–S119.
36. Gorbalenya, A.E.; Lieutaud, P.; Harris, M.R.; Coutard, B.; Canard, B.; Kleywegt, G.J.; Kravchenko, A.A.; Samborskiy, D.V.; Sidorov, I.A.; Leontovich, A.M.; *et al.* Practical application of bioinformatics by the multidisciplinary VIZIER consortium. *Antivir. Res.* **2010**, *87*, 95–110.
37. Lauber, C.; Gorbalenya, A.E. Molecular Virology Laboratory, Department of Medical Microbiology, Leiden University Medical Center, 2333 ZA Leiden, The Netherlands. Genetics-based classification of coronaviruses. Unpublished work, to be submitted for publication, 2012.
38. Van Regenmortel, M.H.V. Virus species and virus identification: Past and current controversies. *Inf. Genet. Evol.* **2007**, *7*, 133–144.
39. Mayo, M.A.; Pringle, C.R. Virus taxonomy—1997. *J. Gen. Virol.* **1998**, *79*, 649–657.
40. Dimmic, M.W.; Rest, J.S.; Mindell, D.P.; Goldstein, R.A. rtREV: An amino acid substitution matrix for inference of retrovirus and reverse transcriptase phylogeny. *J. Mol. Evol.* **2002**, *55*, 65–73.
41. Hraber, P.; Kuiken, C.; Waugh, M.; Geer, S.; Bruno, W.J.; Leitner, T. Classification of hepatitis C virus and human immunodeficiency virus-1 sequences with the branching index. *J. Gen. Virol.* **2008**, *89*, 2098–2107.
42. Pons, J.; Barraclough, T.G.; Gomez-Zurita, J.; Cardoso, A.; Duran, D.P.; Hazell, S.; Kamoun, S.; Sumlin, W.D.; Vogler, A.P. Sequence-based species delimitation for the DNA taxonomy of undescribed insects. *Syst. Biol.* **2006**, *55*, 595–609.
43. Jiang, P.; Faase, J.A.J.; Toyoda, H.; Paul, A.; Wimmer, E.; Gorbalenya, A.E. Evidence for emergence of diverse polioviruses from C-cluster coxsackie A viruses and implications for global poliovirus eradication. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 9457–9462.
44. Lukashev, A.N. Recombination among picornaviruses. *Rev. Med. Virol.* **2010**, *20*, 327–337.

45. Chare, E.R.; Gould, E.A.; Holmes, E.C. Phylogenetic analysis reveals a low rate of homologous recombination in negative-sense RNA viruses. *J. Gen. Virol.* **2003**, *84*, 2691–2703.
46. Holmes, E.C. The evolutionary genetics of emerging viruses. *Annu. Rev. Ecol. Evol. Systemat.* **2009**, *40*, 353–372.
47. Archer, A.M.; Rico-Hesse, R. High genetic divergence and recombination in arenaviruses from the Americas. *Virology* **2002**, *304*, 274–281.
48. Charrel, R.N.; Feldmann, H.; Fulhorst, C.F.; Khelifa, R.; de Chesse, R.; de Lamballerie, X. Phylogeny of New World arenaviruses based on the complete coding sequences of the small genomic segment identified an evolutionary lineage produced by intrasegmental recombination. *Biochem. Biophys. Res. Comm.* **2002**, *296*, 1118–1124.
49. Hao, W.L. Evidence of intra-segmental homologous recombination in influenza A virus. *Gene* **2011**, *481*, 57–64.
50. Wittmann, T.J.; Biek, R.; Hassanin, A.; Rouquet, P.; Reed, P.; Yaba, P.; Pourrut, X.; Real, L.A.; Gonzalez, J.P.; Leroy, E.M. Isolates of Zaire ebolavirus from wild apes reveal genetic lineage and recombinants. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 17123–17127.
51. Chare, E.R.; Holmes, E.C. A phylogenetic survey of recombination frequency in plant RNA viruses. *Arch. Virol.* **2006**, *151*, 933–946.
52. Edgar, R.C. MUSCLE: A multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* **2004**, *5*, 113.
53. Schmidt, H.A.; Strimmer, K.; Vingron, M.; von Haeseler, A. TREE-PUZZLE: Maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics* **2002**, *18*, 502–504.
54. Whelan, S.; Goldman, N. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol. Biol. Evol.* **2001**, *18*, 691–699.
55. Castresana, J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* **2000**, *17*, 540–552.
56. Antonov, I.V.; Leontovich, A.M.; Gorbalenya, A.E. BAGG (Blocks Accepting Gaps Generator). 2008. Available online: http://www.genebee.msu.su/~antonov/bagg/cgi/bagg.cgi (accessed on 16 February 2012).
57. Guindon, S.; Gascuel, O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* **2003**, *52*, 696–704.