Viruses **2012**, *4*, 1318-1327; doi:10.3390/v4081318

OPEN ACCESS

viruses

ISSN 1999-4915

www.mdpi.com/journal/viruses

Article

PAirwise Sequence Comparison (PASC) and Its Application in the Classification of Filoviruses

Yiming Bao *, Vyacheslav Chetvernin and Tatiana Tatusova

National Center for Biotechnology Information, National Institutes of Health, Bethesda, MD 20892, USA; E-Mails: chetvern@ncbi.nlm.nih.gov (V.C.); tatiana@ncbi.nlm.nih.gov (T.T.)

* Author to whom correspondence should be addressed; E-Mail: bao@ncbi.nlm.nih.gov (Y.B.); Tel.: +1-301-435-5939; Fax: +1-301-402-9651.

Received: 3 July 2012; in revised form: 16 August 2012 / Accepted: 16 August 2012 / Published: 20 August 2012

Abstract: PAirwise Sequence Comparison (PASC) is a tool that uses genome sequence similarity to help with virus classification. The PASC tool at NCBI uses two methods: local alignment based on BLAST and global alignment based on Needleman-Wunsch algorithm. It works for complete genomes of viruses of several families/groups, and for the family of *Filoviridae*, it currently includes 52 complete genomes available in GenBank. It has been shown that BLAST-based alignment approach works better for filoviruses, and therefore is recommended for establishing taxon demarcation criteria. When more genome sequences with high divergence become available, these demarcations will most likely become more precise. The tool can compare new genome sequences of filoviruses with the ones already in the database, and propose their taxonomic classification.

Keywords: *Filoviridae*; filovirus; ICTV; International Committee on Taxonomy of Viruses; National Center for Biotechnology Information; NCBI; PAirwise Sequence Comparison; PASC; virus classification; virus taxonomy

1. Introduction

The family *Filoviridae* belongs to the order *Mononegavirales*. Filovirus genomes are single-stranded, negative sense RNAs of about 19 kb, which encode 7 structural proteins. There are currently two genera in this family (*Marburgvirus* and *Ebolavirus*), and a third genus ("Cuevavirus") was proposed recently to the International Committee on Taxonomy of Viruses (ICTV) [1,2]. The criteria used to separate the members of the two genera mainly include antigenic cross-reactivity, virion length, number of gene overlaps, number of protein products expressed by gene 4, and genome divergence. The genus *Ebolavirus* currently includes five species, all of which are represented by a single virus: *Bundibugyo ebolavirus* (member Bundibugyo virus), *Reston ebolavirus* (member Reston virus), *Sudan ebolavirus* (member Sudan virus), *Taï Forest ebolavirus* (member Taï Forest virus) and *Zaire ebolavirus* (member Ebola virus). The species demarcation criteria for ebolaviruses include whole genome differences, number of gene overlaps and geographic origin. The genus *Marburgvirus* and the tentative genus "Cuevavirus" each include one species, *Marburg marburgvirus* and the tentative "Lloviu cuevavirus", respectively, and there are currently no species demarcation criteria for the two. However, whereas only one "cuevavirus", Lloviu virus, is known, two viruses, Marburg virus and Ravn virus, have been designated members of the species *Marburg marburgvirus* [1–3].

PAirwise Sequence Comparison (PASC) is a tool that can help virus classification using genome sequences [4]. It calculates the percent of pairwise sequence identity for all published complete genomes within a virus family/group, and plots the frequency distribution. In many virus families/groups, there are clear peaks in percent identity that represent pairs of viruses belonging to the same species, to different species but to the same genus, or to different genera. When these peaks are well separated, the identity percentage at their boundaries can serve as one of the species/genera demarcation criteria.

In order to provide an online tool to use PASC for virus families to which this method already has been applied (e.g., families *Geminiviridae*, *Papillomoviridae*, *Picornaviridae* and *Potyviridae*), and to investigate the possibilities of applying PASC to a wider range of viral groups, the National Center for Biotechnology Information (NCBI) created a PASC resource [5] which currently covers over 50 viral families/groups. PASC can be used to evaluate whether sequence-based classification criteria can be established for these viral families/groups, and, if yes, PASC can suggest the taxonomic classification for viruses whose genome sequences are available.

Traditionally, genome identities were calculated based on pairwise global alignments (*i.e.*, sequences are aligned end-to-end) in PASC. During the development of the PASC tool at NCBI, we noticed that there were issues with the global alignment in cases where sequences have inconsistent starting position, are in different strand of the genomes, or have extremely low similarities in part of the genomes. A new method based on BLAST (*i.e.*, only regions with good alignment are used to calculate pairwise identities) was introduced to NCBI's PASC tool, which improved the result for many families.

Here, we describe the application of PASC to the classification of filoviruses.

2. Results and Discussion

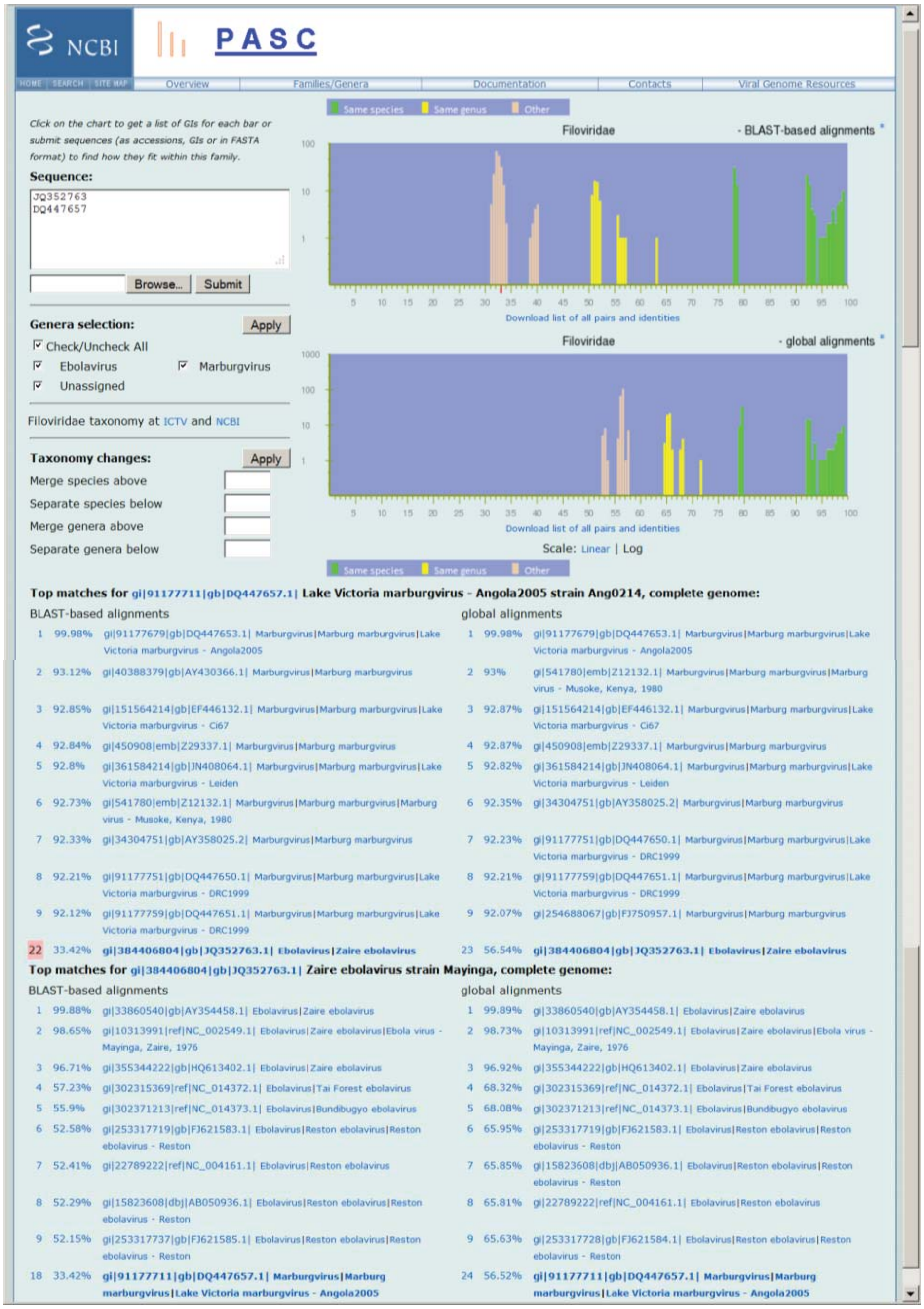
2.1. Establishing Demarcation Criteria

The PASC tool for the family of *Filoviridae* at NCBI can be accessed through [6] (a screen shot is shown in Figure 1). It currently contains 52 complete genome sequences available in GenBank. The upper right plot in Figure 1 displays the pairwise identity distribution of 28 non-redundant genomes calculated by the BLAST-based alignment, while the lower right plot shows the pairwise identity distribution of 27 non-redundant genomes by the global alignment (See Experimental Section for the alignment methods). The green, yellow and peach bars in the plots represent pairs of genomes that are assigned to the same species, different species but in the same genus, and different genera respectively in the current NCBI Taxonomy Database. PASC works best when the bars of different colors group together and are well separated, which is the case in both plots in Figure 1. The results clearly indicate that (a) sequence-based classification agrees with the current ICTV taxonomy for the family *Filoviridae*; (b) the taxonomic assignments of the filovirus genome sequences in GenBank are accurate; (c) species and genera demarcations can be set for the family.

It can be seen from Figure 1 that the yellow and peach peaks are different in the two distribution plots. For the BLAST-based alignment plot, the yellow peaks are located between 50% and 64%, and the peach ones are between 30% and 41%; while in the global-alignment plot, they are between 64% and 72%, and between 52% and 58%, respectively. Originally, the Needleman–Wunsch global-alignment method [7] was used to calculate the genome identities in PASC [4]. It was later observed that this method was not optimized for some virus families [8]. For filoviruses, this is illustrated on Figure 2, where the dot matrix and text alignments between the genomes of Ebola virus (NC_002549) and Taï Forest virus (NC_014372) are shown (Similar alignments can be obtained by clicking the percentage of identities for the pairs of interest such as those listed in Figure 1). The identity between the two genomes calculated using the BLAST-based and global alignment is 57% and 68% respectively, neither of which matches the 63% reported by Towner *et al.* [9]. The BLAST-based method calculates percent identity only in the regions of alignment that in the case of filoviruses correspond to the 7 filoviral genes (see the dot matrix view at the upper section in Figure 2). This method essentially resembles the strategy used to classify bacteriophages based on protein sequence similarities [10,11]. The global alignment method forces alignment over the whole length of the sequence (see the dot matrix view at the lower section in Figure 2). Since, on average, any two random nucleotide sequences of the same size will have an identity of 25%, the identities obtained by the global alignment method are usually inflated, especially for genomes with high divergence. We therefore believe the BLAST-based alignment result represents the true relationship among genome sequences of filoviruses, and should be used to establish taxa demarcation criteria, which are between 64% and 77% for species, and between 41% and 50% for genera. These demarcation criteria better match those set up by the ICTV *Filoviridae* Study Group, which are currently set as 70% for species and 50% for genera [1]. The demarcation criteria suggested by PASC are in a range rather than precise percentages as in many other viral groups, as the filovirus genomes sequenced so far are not very diverse. When more genome sequences become available, these demarcations will most likely change to narrower ranges or even to a single percentage.

It is important to keep in mind that the demarcations obtained by the BLAST-based alignment in PASC are different from those determined by other algorithms using different datasets and/or different genome regions, and therefore can only be used in the PASC system at NCBI. In another word, one cannot apply the demarcations discussed here to sequence identities calculated by other algorithms such as ClustalX, or using a single gene rather than the whole genome.

Figure 1. Frequency distribution of pairwise identities for the complete genome sequences of filoviruses, and the application in classifying newly sequenced viruses. The two plots represent results obtained by the BLAST-based alignments and global alignments respectively. The green, yellow and peach bars in the plots represent pairs of genomes that are assigned to the same species, different species but in the same genus, and different genera respectively in the current National Center for Biotechnology Information (NCBI) Taxonomy Database. The top matches for each input genome (JQ352763 and DQ447657) to the existing genomes in GenBank and other input genome are listed, and their pairwise identities are shown. The small red bar on the X-axis of the top plot indicates the percentage of identity of the selected pair (#22 which is highlighted). Not all virus species names listed reflect the most recent International Committee on Taxonomy of Viruses (ICTV) species names.



Click on the chart to get a list of GIs for each bar or submit sequences (as accessions, GIs or in FASTA format) to find how they fit within this family.

Sequence:

```
JQ352763
DQ447657
```

Genera selection:

☑ Check/Uncheck All

☑ Ebolavirus ☑ Marburgvirus

☑ Unassigned

Filoviridae taxonomy at ICTV and NCBI

Taxonomy changes:

Merge species above □

Separate species below □

Merge genera above □

Separate genera below □

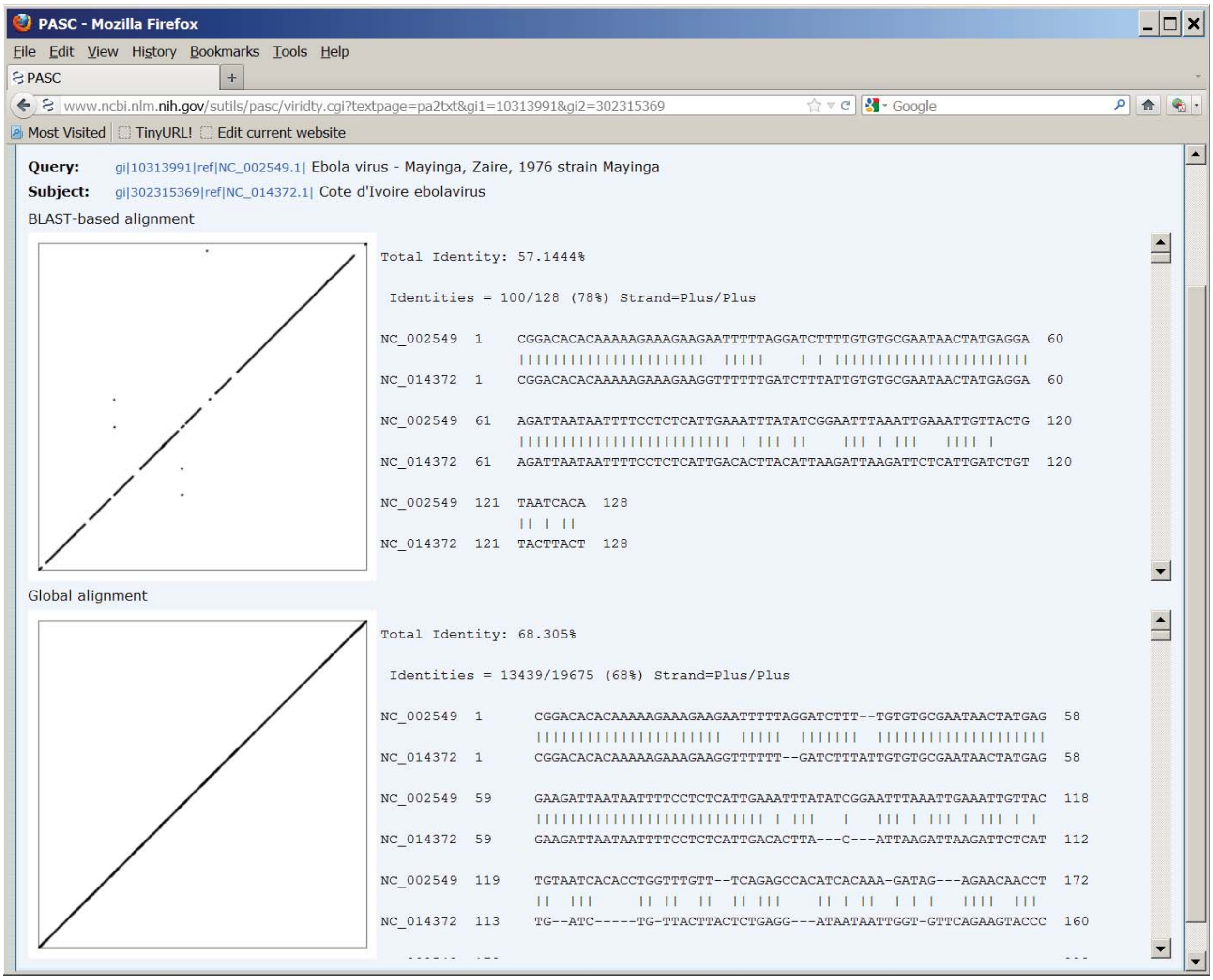
Top matches for gi|91177711|gb|DQ447657.1| Lake Victoria marburgvirus - Angola2005 strain Ang0214, complete genome:

BLAST-based alignments			global alignments		
1	99.98%	gi|91177679|gb|DQ447653.1| Marburgvirus|Marburg marburgvirus|Lake Victoria marburgvirus - Angola2005	1	99.98%	gi|91177679|gb|DQ447653.1| Marburgvirus|Marburg marburgvirus|Lake Victoria marburgvirus - Angola2005
2	93.12%	gi|40388379|gb|AY430366.1| Marburgvirus|Marburg marburgvirus	2	93%	gi|541780|emb|Z12132.1| Marburgvirus|Marburg marburgvirus|Marburg virus - Musoke, Kenya, 1980
3	92.85%	gi|151564214|gb|EF446132.1| Marburgvirus|Marburg marburgvirus|Lake Victoria marburgvirus - Ci67	3	92.87%	gi|151564214|gb|EF446132.1| Marburgvirus|Marburg marburgvirus|Lake Victoria marburgvirus - Ci67
4	92.84%	gi|450908|emb|Z29337.1| Marburgvirus|Marburg marburgvirus	4	92.87%	gi|450908|emb|Z29337.1| Marburgvirus|Marburg marburgvirus
5	92.8%	gi|361584214|gb|JN408064.1| Marburgvirus|Marburg marburgvirus|Lake Victoria marburgvirus - Leiden	5	92.82%	gi|361584214|gb|JN408064.1| Marburgvirus|Marburg marburgvirus|Lake Victoria marburgvirus - Leiden
6	92.73%	gi|541780|emb|Z12132.1| Marburgvirus|Marburg marburgvirus|Marburg virus - Musoke, Kenya, 1980	6	92.35%	gi|34304751|gb|AY358025.2| Marburgvirus|Marburg marburgvirus
7	92.33%	gi|34304751|gb|AY358025.2| Marburgvirus|Marburg marburgvirus	7	92.23%	gi|91177751|gb|DQ447650.1| Marburgvirus|Marburg marburgvirus|Lake Victoria marburgvirus - DRC1999
8	92.21%	gi|91177751|gb|DQ447650.1| Marburgvirus|Marburg marburgvirus|Lake Victoria marburgvirus - DRC1999	8	92.21%	gi|91177759|gb|DQ447651.1| Marburgvirus|Marburg marburgvirus|Lake Victoria marburgvirus - DRC1999
9	92.12%	gi|91177759|gb|DQ447651.1| Marburgvirus|Marburg marburgvirus|Lake Victoria marburgvirus - DRC1999	9	92.07%	gi|254688067|gb|FJ750957.1| Marburgvirus|Marburg marburgvirus
22	33.42%	**gi|384406804|gb|JQ352763.1| Ebolavirus|Zaire ebolavirus**	23	56.54%	**gi|384406804|gb|JQ352763.1| Ebolavirus|Zaire ebolavirus**

Top matches for gi|384406804|gb|JQ352763.1| Zaire ebolavirus strain Mayinga, complete genome:

BLAST-based alignments			global alignments		
1	99.88%	gi|33860540|gb|AY354458.1| Ebolavirus|Zaire ebolavirus	1	99.89%	gi|33860540|gb|AY354458.1| Ebolavirus|Zaire ebolavirus
2	98.65%	gi|10313991|ref|NC_002549.1| Ebolavirus|Zaire ebolavirus|Ebola virus - Mayinga, Zaire, 1976	2	98.73%	gi|10313991|ref|NC_002549.1| Ebolavirus|Zaire ebolavirus|Ebola virus - Mayinga, Zaire, 1976
3	96.71%	gi|355344222|gb|HQ613402.1| Ebolavirus|Zaire ebolavirus	3	96.92%	gi|355344222|gb|HQ613402.1| Ebolavirus|Zaire ebolavirus
4	57.23%	gi|302315369|ref|NC_014372.1| Ebolavirus|Tai Forest ebolavirus	4	68.32%	gi|302315369|ref|NC_014372.1| Ebolavirus|Tai Forest ebolavirus
5	55.9%	gi|302371213|ref|NC_014373.1| Ebolavirus|Bundibugyo ebolavirus	5	68.08%	gi|302371213|ref|NC_014373.1| Ebolavirus|Bundibugyo ebolavirus
6	52.58%	gi|253317719|gb|FJ621583.1| Ebolavirus|Reston ebolavirus|Reston ebolavirus - Reston	6	65.95%	gi|253317719|gb|FJ621583.1| Ebolavirus|Reston ebolavirus|Reston ebolavirus - Reston
7	52.41%	gi|22789222|ref|NC_004161.1| Ebolavirus|Reston ebolavirus	7	65.85%	gi|15823608|dbj|AB050936.1| Ebolavirus|Reston ebolavirus|Reston ebolavirus - Reston
8	52.29%	gi|15823608|dbj|AB050936.1| Ebolavirus|Reston ebolavirus|Reston ebolavirus - Reston	8	65.81%	gi|22789222|ref|NC_004161.1| Ebolavirus|Reston ebolavirus
9	52.15%	gi|253317737|gb|FJ621585.1| Ebolavirus|Reston ebolavirus|Reston ebolavirus - Reston	9	65.63%	gi|253317728|gb|FJ621584.1| Ebolavirus|Reston ebolavirus|Reston ebolavirus - Reston
18	33.42%	**gi|91177711|gb|DQ447657.1| Marburgvirus|Marburg marburgvirus|Lake Victoria marburgvirus - Angola2005**	24	56.52%	**gi|91177711|gb|DQ447657.1| Marburgvirus|Marburg marburgvirus|Lake Victoria marburgvirus - Angola2005**

Figure 2. Dot matrix and text views of pairwise alignment between genome sequences of Ebola virus (NC_002549) and Taï Forest virus (NC_014372), using the BLAST-based and global alignment methods. Not all virus species names listed reflect the most recent ICTV species names.


PASC - Mozilla Firefox
File Edit View History Bookmarks Tools Help
PASC
www.ncbi.nlm.nih.gov/sutils/pasc/viridty.cgi?textpage=pa2txt&gi1=10313991&gi2=302315369
Google
Most Visited
TinyURL!
Edit current website

Query: gi|10313991|ref|NC_002549.1| Ebola virus - Mayinga, Zaire, 1976 strain Mayinga
Subject: gi|302315369|ref|NC_014372.1| Cote d'Ivoire ebolavirus

BLAST-based alignment

```
Total Identity: 57.1444%

 Identities = 100/128 (78%) Strand=Plus/Plus

NC_002549  1    CGGACACACAAAAAGAAAGAAGAATTTTTAGGATCTTTTGTGTGCGAATAACTATGAGGA  60
                ||||||||||||||||||||||  |||||    | | |||||||||||||||||||||||
NC_014372  1    CGGACACACAAAAAGAAAGAAGGTTTTTTGATCTTTATTGTGTGCGAATAACTATGAGGA  60

NC_002549  61   AGATTAATAATTTTCCTCTCATTGAAATTTATATCGGAATTTAAATTGAAATTGTTACTG  120
                ||||||||||||||||||||||||| | ||| ||    ||| | |||   |||| |
NC_014372  61   AGATTAATAATTTTCCTCTCATTGACACTTACATTAAGATTAAGATTCTCATTGATCTGT  120

NC_002549  121  TAATCACA  128
                || | ||
NC_014372  121  TACTTACT  128
```

Global alignment

```
Total Identity: 68.305%

 Identities = 13439/19675 (68%) Strand=Plus/Plus

NC_002549  1      CGGACACACAAAAAGAAAGAAGAATTTTTAGGATCTTT--TGTGTGCGAATAACTATGAG  58
                  ||||||||||||||||||||||  |||||  |||||||  ||||||||||||||||||||
NC_014372  1      CGGACACACAAAAAGAAAGAAGGTTTTTT--GATCTTTATTGTGTGCGAATAACTATGAG  58

NC_002549  59     GAAGATTAATAATTTTCCTCTCATTGAAATTTATATCGGAATTTAAATTGAAATTGTTAC  118
                  ||||||||||||||||||||||||||| | |||   |   ||| | ||| | ||| | |
NC_014372  59     GAAGATTAATAATTTTCCTCTCATTGACACTTA---C---ATTAAGATTAAGATTCTCAT  112

NC_002549  119    TGTAATCACACCTGGTTTGTT--TCAGAGCCACATCACAAA-GATAG---AGAACAACCT  172
                  ||  |||     || ||  ||  || |||    || | ||  | | |   ||||  |||
NC_014372  113    TG--ATC-----TG-TTACTTACTCTGAGG---ATAATAATTGGT-GTTCAGAAGTACCC  160
```

2.2. Classification of Lloviu Virus

Lloviu virus was recently discovered and proposed to represent a new species in a new genus, based on the differences in its genome sequence and gene products compared to those of existing genera [12]. Its genome shows higher similarity to genomes of ebolaviruses than to marburgvirus genomes. In the PASC result, the peach color peak between 38% and 40% in the BLAST-based alignment method (upper right plot in Figure 1) represents pairs of Lloviu virus and ebolavirus genomes. If the classification was solely determined by PASC, the genera demarcation could also be set to between 35% and 38%, which would put Lloviu virus in the genus *Ebolavirus*. This demonstrates that the determination of demarcation by PASC can be arbitrary, and should not be used as the only criterion for virus classification. Other biological properties have to be taken into consideration.

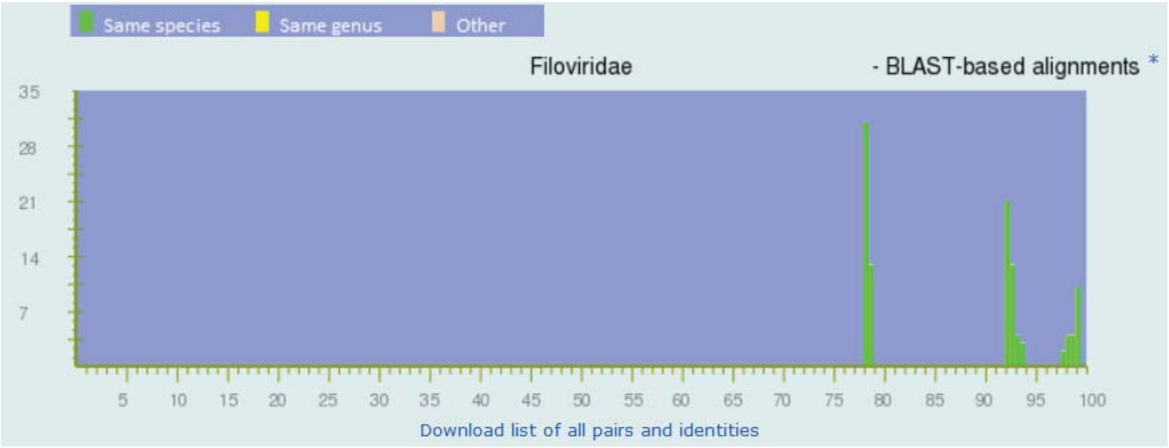
2.3. Sub-Species Grouping of Marburgvirus

Although only one species is currently recognized in the genus *Marburgvirus*, five major lineages can be found in the phylogenetic analysis of all genomes under the species [1]. The four more closely related lineages constitute Marburg virus (MARV), and the distant lineage constitutes Ravn virus (RAVV).

The PASC result for the species *Marburg marburgvirus* (Figure 3) supports this division. In Figure 3, the peak above 97% represents the pairs of viruses within each of the five major lineages; the peak between 91% and 94% represents viruses between the four lineages of MARV; and the peak below 79% represents viruses between RAVV and MARV. Since these peaks are well-separated, PASC can be used to suggest sub-species classifications of newly sequenced marburgviruses.

Similar sub-species patterns are not observed in PASC for ebolaviruses, and the identities of members within each ebolavirus species are above 94% (data not shown), suggesting that intra-species variations of ebolaviruses are less than those of marburgviruses, at least when based on currently available genomes.

Figure 3. Frequency distribution of pairwise identities from the complete genome sequences of marburgviruses. The three peaks from right to left represent the pairs of viruses within each of the five major lineages of marburgviruses, between the four lineages of Marburg virus (MARV), and between Ravn virus (RAVV) and MARV.



2.4. Classifying Newly Sequenced Viruses

The use of PASC for classification of viruses with newly sequenced genomes is demonstrated in Figure 1. The accession numbers of two filovirus genomes, DQ447657 (Marburg virus—Angola2005 isolate Ang0214) and JQ352763 (Ebola virus isolate Mayinga), were entered into the "Sequence" box of the PASC page for the *Filoviridae* family, and compared with existing filovirus genomes and with each other.

The nine sequences most similar to DQ447657 are listed in Figure 1, together with their pairwise identities. The most similar was DQ447653 (Marburg virus—Angola2005 isolate Ang1379c), with a sequence identity of 99.98% by the BLAST-based alignment method. Since this is in the region consisting of green bars, it suggests that DQ447657 and DQ447653 belong to the same species (*i.e.*, that DQ447657 represents a variant of marburgviruses).

Similarly, a sequence identity of 99.88% between JQ352763 and its closest neighbor, AY354458 (Ebola virus variant Kikwit) suggests that JQ352763 represents a variant of ebolaviruses.

The identity of 33.42% between the two input sequences DQ447657 and JQ352763 is also reported by PASC (see the pair number 22 highlighted in pink background when the number is clicked). Such a report is very useful in determining the relationships among genomes of novel species—if the new genomes are similar enough to each other (*i.e.*, their identity is in the green peaks), they represent the same new species; otherwise, they represent different new species.

3. Experimental Section

3.1. Source of Genome Sequences and Taxonomy Information

Complete genome sequences of viruses in the family *Filoviridae* were retrieved from the NCBI viral genome collection [13], which contains both reference sequences and other genomes for all species. There are currently 52 complete genome sequences in the database (as of July, 2012). The data set is updated daily to reflect sequence and/or taxonomy changes and to include new genomes released in GenBank.

3.2. Pairwise Genome Alignment and Identity Calculation

Pairwise global alignment of the genomes were performed and their identities were calculated as described before [4].

For BLAST-based alignment, two sets of BLAST [14] searches were performed on each pair of genome sequences. In the first set, the nucleotide sequence of one genome was translated to protein sequence in all six frames and searched against the nucleotide sequence of the other genome using tblastn (tblastn searches translated nucleotide database using a protein query). The amino acid alignments were then converted back to nucleotide alignments. In the second set, two genome sequences were compared using blastn (nucleotide to nucleotide alignment), followed by the creation of a consistent and non-redundant set by trimming the overlapping hits from both sets. Pairwise identity was calculated as the total number of identical bases in the local hits divided by the longest sequence length.

3.3. Removal of Redundant Sequences

For optimization purposes, sequences with identities higher than 99.5% were represented by one sequence in the dataset, excluding redundant sets. As a result, 27 and 28 non-redundant genomes in the family of *Filoviridae* were used in the global and BLAST-based alignment method, respectively. The numbers of non-redundant genomes are different for the two methods, because the global alignment method results in higher identities (see discussion in Section 2.1) therefore more redundant sequences.

3.4. Identity Distribution Plot

The identity distribution chart was plotted based on pairwise alignments computed between all non-redundant genomes. The pair is represented in green if both genomes belong to the same species according to their assignment in NCBI's taxonomy database; in yellow if the two genomes belong to different species within the same genus; and in peach if they belong to different genera. Both linear and log scales are available for the Y-axis (number of pairs).

3.5. Compare External Genomes against Existing Ones

To compare external (or new) genomes against existing sequences in the database, a user may specify the query genomes in the "Sequence:" box using either their GenBank Accession/GI numbers, entering the raw data in FASTA format, or uploading a sequence file after clicking the "Browse" button. Up to 25 sequences can be added in one submission. After sequences are submitted, PASC will compute pairwise identities between user-provided genomes and the existing genome sequences within the family. PASC produces a list of pairwise identities, from the highest to the lowest, between this input genome and (1) the rest of input genomes (if there are more than one), and (2) 5 to 10 closest matches to existing genomes within the family. The identity distribution chart will depict the currently selected genome with a different color. The user can click on each genome's number to make it current, or can click the identity to see details of the alignment.

4. Conclusions

The PASC tool was created at NCBI using two pairwise alignment methods on all available complete genome sequences in public databases, and is available for the family *Filoviridae*. The BLAST-based alignment result better represents the true similarities among genome sequences of filoviruses, and is recommended for establishing the demarcation criteria, which are between 64% and 77% for species, and between 41% and 50% for genera. When more genome sequences with high divergence become available, these demarcations will most likely change to narrower ranges or even to a precise percentage. These demarcations are specifically associated with the method used here therefore should not be used as references for identities calculated using other algorithms. This tool can compare new genome sequences of filoviruses with existing ones in the database, and propose their taxonomic classification.

Acknowledgments

We thank Michael Baxter and Stacy Ciufo for comments on the manuscript. This research was supported by the Intramural Research Program of the NIH, National Library of Medicine.

Conflict of Interest

The authors declare no conflict of interest.

References and Notes

1. Kuhn, J.H.; 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.* 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: London, UK, 2011; pp. 665–671.
2. 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.
3. Adams, M.J.; Carstens, E.B. Ratification vote on taxonomic proposals to the International Committee on Taxonomy of Viruses. *Arch. Virol.* **2012**, *157*, 1411–1422.
4. Bao, Y.; Kapustin, Y.; Tatusova, T. Virus Classification by Pairwise Sequence Comparison (PASC). In *Encyclopedia of Virology*, 3rd ed.; Mahy, B.W.J., van Regenmortel, M.H.V., Eds.; Elsevier: Oxford, UK, 2008; Volume 5, pp. 342–348.
5. PASC. Available online: http://www.ncbi.nlm.nih.gov/sutils/pasc (accessed on 17 August 2012).
6. PASC for the family of *Filoviridae*. Available online: http://www.ncbi.nlm.nih.gov/sutils/pasc/viridty.cgi?cmdresult=main&id=333&scale=Log (accessed on 17 August 2012).
7. Needleman, S.B.; Wunsch, C.D. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.* **1970**, *48*, 443–453.
8. Bao, Y.; Chetvernin, V.; Kapustin, Y.; Tatusova, T. National Center for Biotechnology Information, National Institutes of Health, Bethesda, MD, USA. Unpublished work, 2012.
9. Towner, J.S.; Sealy, T.K.; Khristova, M.L.; Albariño, 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.
10. Lavigne, R.; Seto, D.; Mahadevan, P.; Ackermann, H.W.; Kropinski, A.M. Unifying classical and molecular taxonomic classification: Analysis of the Podoviridae using BLASTP-based tools. *Res. Microbiol.* **2008**, *159*, 406–414.
11. Lavigne, R.; Darius, P.; Summer, E.J.; Seto, D.; Mahadevan, P.; Nilsson, A.S.; Ackermann, H.W.; Kropinski, A.M. Classification of Myoviridae bacteriophages using protein sequence similarity. *BMC Microbiol.* **2009**, *9*, 224.
12. Negredo, A.; Palacios, G.; Vázquez-Morón, S.; González, F.; Dopazo, H.; Molero, F.; Juste, J.; Quetglas, J.; Savji, N.; de la Cruz Martínez, M.; *et al.* Discovery of an ebolavirus-like filovirus in europe. *PLoS Pathog.* **2011**, *7*, e1002304.
13. Bao, Y.; Federhen, S.; Leipe, D.; Pham, V.; Resenchuk, S.; Rozanov, M.; Tatusov, R.; Tatusova, T. National center for biotechnology information viral genomes project. *J. Virol.* **2004**, *78*, 7291–7298.

14. Altschul, S.F.; Madden, T.L.; Schäffer, A.A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D.J. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **1997**, *25*, 3389–3402.