This form should be used for all taxonomic proposals. Please complete all those modules that are applicable (and then delete the unwanted sections). For guidance, see the notes written in blue and the separate document "Help with completing a taxonomic proposal"

Please try to keep related proposals within a single document; you can copy the modules to create more than one genus within a new family, for example.

MODULE 1: **TITLE, AUTHORS, etc**

Code assigned:	***2016.013aS***	(to be completed by ICTV officers)

Short title: Create 9 new species in the genus *Pegivirus* (family *Flaviviridae*)
(e.g. 6 new species in the genus *Zetavirus*)

Modules attached (modules 1 and 11 are required)	**2** ☒ **3** ☐ **4** ☐ **5** ☐ **6** ☐ **7** ☐ **8** ☐ **9** ☐ **10** ☐

Author(s):

Donald B. Smith, Paul Becher, Jens Bukh, Ernest A. Gould, Gregor Meyers, Thomas Monath, A. Scott Muerhoff, Alexander Pletnev, Rebecca Rico-Hesse, Jack T. Stapleton, Peter Simmonds

Corresponding author with e-mail address:

Donald Smith, Donald.smith.mail@gmail.com

List the ICTV study group(s) that have seen this proposal:

A list of study groups and contacts is provided at http://www.ictvonline.org/subcommittees.asp . If in doubt, contact the appropriate subcommittee chair (fungal, invertebrate, plant, prokaryote or vertebrate viruses)	*Flaviviridae*

ICTV Study Group comments (if any) and response of the proposer:

The proposal is from the Flaviviridae Study Group

Date first submitted to ICTV: 23rd June 2016
Date of this revision (if different to above):

ICTV-EC comments and response of the proposer:

MODULE 2: **NEW SPECIES**

creating and naming one or more new species.

If more than one, they should be a group of related species belonging to the same genus. All new species must be placed in a higher taxon. This is usually a genus although it is also permissible for species to be "unassigned" within a subfamily or family. Wherever possible, provide sequence accession number(s) for **one** isolate of each new species proposed.

Code	***2016.013aS***	(assigned by ICTV officers)

To create 9 new species within:

Genus:	***Pegivirus***
Subfamily:	
Family:	***Flaviviridae***
Order:	

Fill in all that apply.

- If the higher taxon has yet to be created (in a later module, below) write "**(new)**" after its proposed name.
- If no genus is specified, enter "**unassigned**" in the genus box.

Name of new species:	**Representative isolate: (only 1 per species please)**	**GenBank sequence accession number(s)**
Pegivirus C	PNF2161 (GBV-C, primate)	U44402
Pegivirus D	Horse_A1 (equine)	KC145265
Pegivirus E	C0035 (equine)	KC410872
Pegivirus F	PDB-1698 (bat)	KC796080
Pegivirus G	PDB-620 (bat)	KC796076
Pegivirus H	AK-790 (human PgV2)	KT439329
Pegivirus I	PDB-1715 (bat)	KC796088
Pegivirus J	CC61 (rodent)	KC815311
Pegivirus K	PPgV_903 (pig)	KU351669

Reasons to justify the creation and assignment of the new species:

- Explain how the proposed species differ(s) from all existing species.
  - If species demarcation criteria (see module 3) have previously been defined for the genus, **explain how the new species meet these criteria**.
  - If criteria for demarcating species need to be defined (because there will now be more than one species in the genus), please state the proposed criteria.
- Further material in support of this proposal may be presented in the Appendix, Module 11

A set of 26 *Pegivirus* sequences that differed from each other by > 0.11 of amino acid positions over their complete coding sequence was used to assess amino acid sequence diversity across the genome. There were two regions where mean amino acid diversity was consistently < 0.6; 888-1635 and 2398-2916 (numbered relative to U22303, Figure 1). Phylogenetic analysis of *Pegivirus* sequences in these two regions produced congruent trees, providing independent evidence that these sequences are phylogenetically distinct (Figure 2A, B). For both regions, the distribution of amino acid distances between these sequences, whether calculated using SSE v1.2 as p-distances, Kimura distances or using a matrix of similarity, were distributed in a series of peaks (Figure 2C, D) with discontinuities at 0.28-0.34 (positions 888 -1635) and 0.35-0.37 (2398-2916). Using an amino acid p-distance of > 0.31 for positions 888-1635 to demarcate *Pegivirus* species, the sequences currently described would represent 11 different species (Table 2).

These individual species comprise sequences from similar hosts from either the Old or New

worlds with the exception of *Pegivirus A* which includes sequences derived from New world primates and Old world bats. Two rodent sequences are both included in *Pegivirus I* despite having an ambiguous p-distance for the region 888-1635 (0.303) since they group together on the phylogenetic tree and both are from rodents sampled in the New world. However, if an amino acid p-distance of > 0.36 for the region 2398-2916 is used to demarcate species, the amino acid p-distances between *Pegivirus F, Pegivirus G and Pegivirus J* would all fall below the cutoff. Higher or lower p-distance demarcation points also produce inconsistent assignments. In particular, we could not find demarcation points that divided *Pegivirus A* into exclusively primate or bat-derived groups of sequences.

Pegivirus A includes GBV-A and other isolates from New World monkeys (U22303, U94421, AF023425, AF023424) (Leary *et al.*, 1997; Simons *et al.*, 1995) as well as viruses obtained from African bats (KC796085, KC796082, KC796086, KC796081, KC796075, KC796089) (Quan *et al.*, 2013); *Pegivirus B* includes viruses (GBV-D) derived from bats in Asia (GU566735, GU566734)(Epstein *et al.*, 2010) and Africa (KC796073, KC796083) (Quan *et al.*, 2013). *Pegivirus C* is proposed as a new species to include GBV-C/hepatitis V virus (Leary *et al.*, 1996; Linnen *et al.*, 1996) and related viruses isolated from Old World Primates (Bailey *et al.*, 2015; Birkenmeyer *et al.*, 1998; Kapusinszky *et al.*, 2015; Sibley *et al.*, 2014) . Within this species, the virus phylogeny corresponds closely to that of the host (Bailey *et al.*, 2015; Sharp & Simmonds, 2011; Sibley *et al.*, 2014) with separate lineages for human (78 complete genome sequences), chimpanzee (AF070476), Yellow baboon (KR996153, KR996142, KR996146, KR996144, KR996152, KR996151, KR996150, KR996149, KR996148, KR996147, KR996145, KR996143, KP890673, KP890672), Olive baboon (KF234530), Red tailed guenon (KF234529, KF234528, KF234526, KF234525, KF234527), red colobus (KF234523, KF234524, KF234507, KF234522, KF234521, KF234520, KF234519, KF234518, KF234517, KF234516, KF234515, KF234514, KF234513, KF234512, KF234511, KF234510, KF234509, KF234508, KF234506, KF234505, KF234504, KF234503, KF234502, KF234501, KF234500, KF234499) and African green monkey (KP296858). The proposed species *Pegivirus D* (KC145265) (Chandriani *et al.*, 2013) and *Pegivirus E* (KC410872) (Kapoor *et al.*, 2013b) both include single complete coding region sequences derived from horses. *Pegivirus F, G and I* all include viruses derived from Old and New world bats (Quan *et al.*, 2013), *Pegivirus H includes* viruses described as Human pegivirus 2 and Human hepegivirus (Berg *et al.*, 2015; Kapoor *et al.*, 2015), while *Pegivirus J* includes viruses derived from rodents (Firth *et al.*, 2014; Kapoor *et al.*, 2013a). *Pegivirus K* is a recently described virus isolated from pigs (Baechlein *et al.*, 2016). Some of the proposed species identifiers used will assist association with previous isolate names or designations (*Pegivirus* A: GBV-A, *Pegivirus C*: GBV-C, *Pegivirus E*: Equine pegivirus, *Pegivirus H*: Human pegivirus 2).

MODULE 11: **APPENDIX**: supporting material

additional material in support of this proposal

References:

Baechlein, C., Grundhoff, A., Fischer, N., Alawi, M., Hoeltig, D., Waldmann, K.-H. & Becher, P. (2016). Pegivirus Infection in Domestic Pigs, Germany. *Emerg Infect Dis* **22**, in press.

Bailey, A. L., Lauck, M., Mohns, M., Peterson, E. J., Beheler, K., Brunner, K. G., Crosno, K., Mejia, A., Mutschler, J. & other authors. **(2015).** Durable sequence stability and bone marrow tropism in a macaque model of human pegivirus infection. *Sci Transl Med* **7**, 305ra144.

Berg, M. G., Lee, D., Coller, K., Frankel, M., Aronsohn, A., Cheng, K., Forberg, K., Marcinkus, M., Naccache, S. N. & other authors. **(2015).** Discovery of a Novel Human Pegivirus in Blood Associated with Hepatitis C Virus Co-Infection. *PLoS Pathog* **11**, e1005325. Public Library of Science.

Birkenmeyer, L. G., Desai, S. M., Muerhoff, A. S., Leary, T. P., Simons, J. N., Montes, C. C. & Mushahwar, I. K. (1998). Isolation of a GB virus-related genome from a chimpanzee. *J Med Virol* **56**, 44–51.

Chandriani, S., Skewes-Cox, P., Zhong, W., Ganem, D. E., Divers, T. J., Van Blaricum, A. J., Tennant, B. C. & Kistler, A. L. (2013). Identification of a previously undescribed divergent virus from the Flaviviridae family in an outbreak of equine serum hepatitis. *Proc Natl Acad Sci U S A* **110**, E1407–15.

Epstein, J. H., Quan, P.-L., Briese, T., Street, C., Jabado, O., Conlan, S., Ali Khan, S., Verdugo, D., Hossain, M. J. & other authors. **(2010).** Identification of GBV-D, a novel GB-like flavivirus from old world frugivorous bats (Pteropus giganteus) in Bangladesh. *PLoS Pathog* **6**, e1000972. Public Library of Science.

Firth, C., Bhat, M., Firth, M. A., Williams, S. H., Frye, M. J., Simmonds, P., Conte, J. M., Ng, J., Garcia, J. & other authors. **(2014).** Detection of zoonotic pathogens and characterization of novel viruses carried by commensal Rattus norvegicus in New York City. *MBio* **5**, e01933–14.

Kapoor, A., Simmonds, P., Scheel, T., Hjelle, B., Cullen, J., Burbelo, P., Chauhan, L., Duraisamy, R., Sanchez, L. M. & other authors. **(2013a).** Identification of Rodent Homologs of Hepatitis C Virus and Pegiviruses. *MBio* **4**, e00216–13.

Kapoor, A., Simmonds, P., Cullen, J. M., Scheel, T. K. H., Medina, J. L., Giannitti, F., Nishiuchi, E., Brock, K. V, Burbelo, P. D. & other authors. **(2013b).** Identification of a pegivirus (GB virus-like virus) that infects horses. *J Virol* **87**, 7185–90.

Kapoor, A., Kumar, A., Simmonds, P., Bhuva, N., Singh Chauhan, L., Lee, B., Sall, A. A., Jin, Z., Morse, S. S. & other authors. **(2015).** Virome Analysis of Transfusion Recipients Reveals a Novel Human Virus That Shares Genomic Features with

References:

Hepaciviruses and Pegiviruses. *MBio* **6**, e01466–15.

Kapusinszky, B., Mulvaney, U., Jasinska, A. J., Deng, X., Freimer, N. & Delwart, E. (2015). Local Virus Extinctions following a Host Population Bottleneck. *J Virol* **89**, 8152–61.

Leary, T. P., Muerhoff, A. S., Simons, J. N., Pilot-Matias, T. J., Erker, J. C., Chalmers, M. L., Schlauder, G. G., Dawson, G. J., Desai, S. M. & Mushahwar, I. K. (1996). Sequence and genomic organization of GBV-C: a novel member of the flaviviridae associated with human non-A-E hepatitis. *J Med Virol* **48**, 60–7.

Leary, T. P., Desai, S. M., Erker, J. C. & Mushahwar, I. K. (1997). The sequence and genomic organization of a GB virus A variant isolated from captive tamarins. *J Gen Virol* **78 (Pt 9)**, 2307–13. Microbiology Society.

Linnen, J., Wages, J., Zhang-Keck, Z.-Y., Fry, K. E., Krawczynski, K. Z., Alter, H., Koonin, E., Gallagher, M., Alter, M. & other authors. **(1996).** Molecular Cloning and Disease Association of Hepatitis G Virus: A Transfusion-Transmissible Agent. *Science (80-)* **271**, 505–508. American Association for the Advancement of Science.

Quan, P.-L., Firth, C., Conte, J. M., Williams, S. H., Zambrana-Torrelio, C. M., Anthony, S. J., Ellison, J. A., Gilbert, A. T., Kuzmin, I. V & other authors. **(2013).** Bats are a major natural reservoir for hepaciviruses and pegiviruses. *Proc Natl Acad Sci U S A* **110**, 8194–8199.

Sharp, P. M. & Simmonds, P. (2011). Evaluating the evidence for virus/host co-evolution. *Curr Opin Virol* **1**, 436–41.

Sibley, S. D., Lauck, M., Bailey, A. L., Hyeroba, D., Tumukunde, A., Weny, G., Chapman, C. A., O'Connor, D. H., Goldberg, T. L. & Friedrich, T. C. (2014). Discovery and characterization of distinct simian pegiviruses in three wild African Old World monkey species. *PLoS One* **9**, e98569. Public Library of Science.

Simons, J. N., Pilot-Matias, T. J., Leary, T. P., Dawson, G. J., Desai, S. M., Schlauder, G. G., Muerhoff, A. S., Erker, J. C., Buijk, S. L. & Chalmers, M. L. (1995). Identification of two flavivirus-like genomes in the GB hepatitis agent. *Proc Natl Acad Sci* **92**, 3401–3405.

Annex:

Include as much information as necessary to support the proposal, including diagrams comparing the old and new taxonomic orders. The use of Figures and Tables is strongly recommended but direct pasting of content from publications will require permission from the copyright holder together with appropriate acknowledgement as this proposal will be placed on a public web site. For phylogenetic analysis, try to provide a tree where branch length is related to genetic distance.

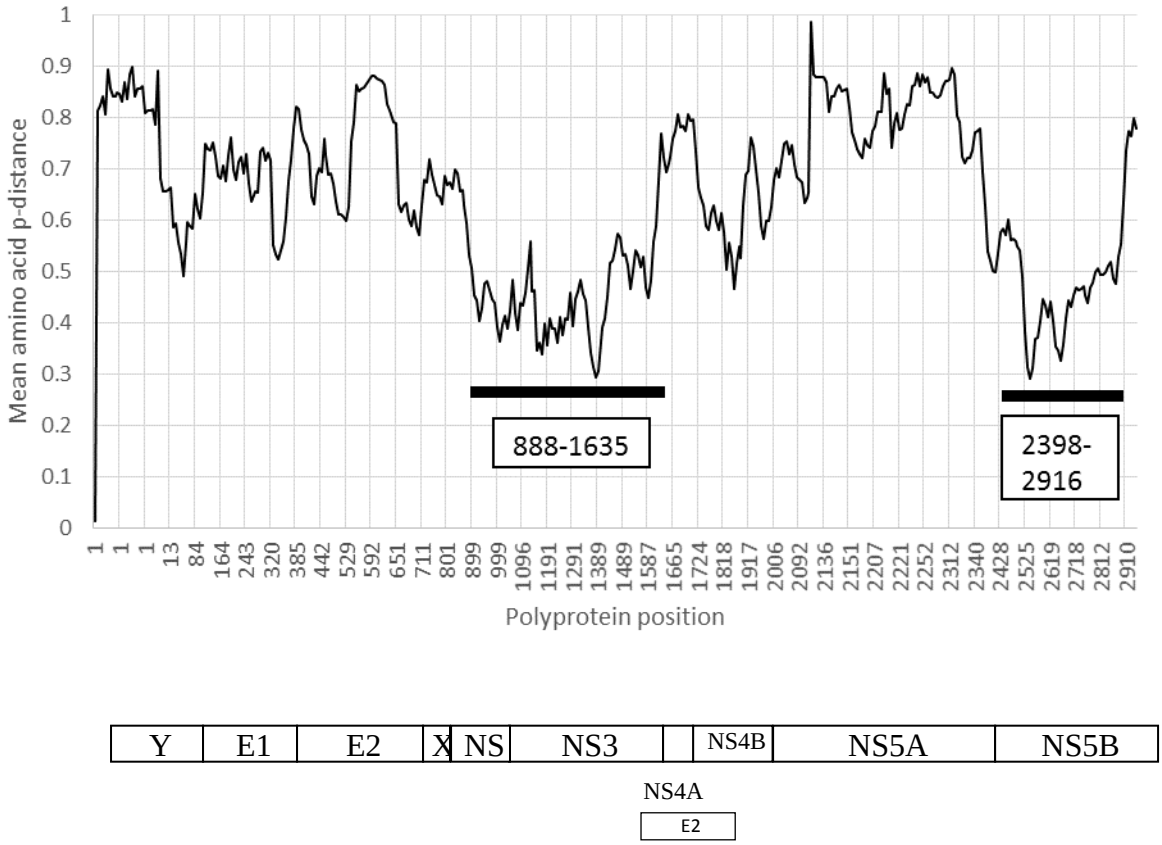
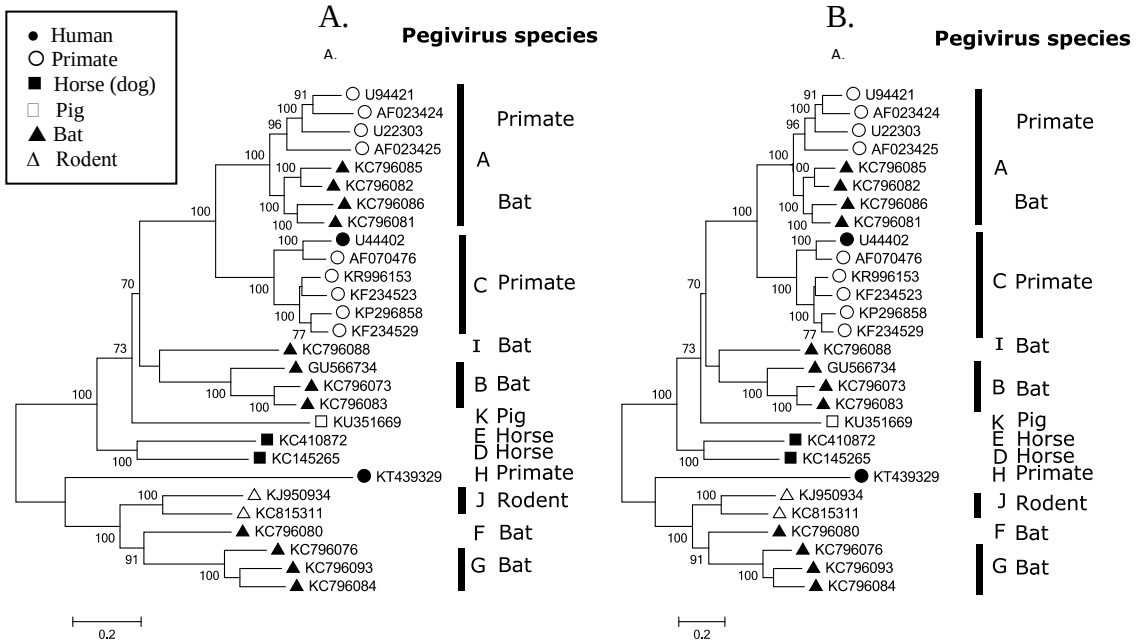
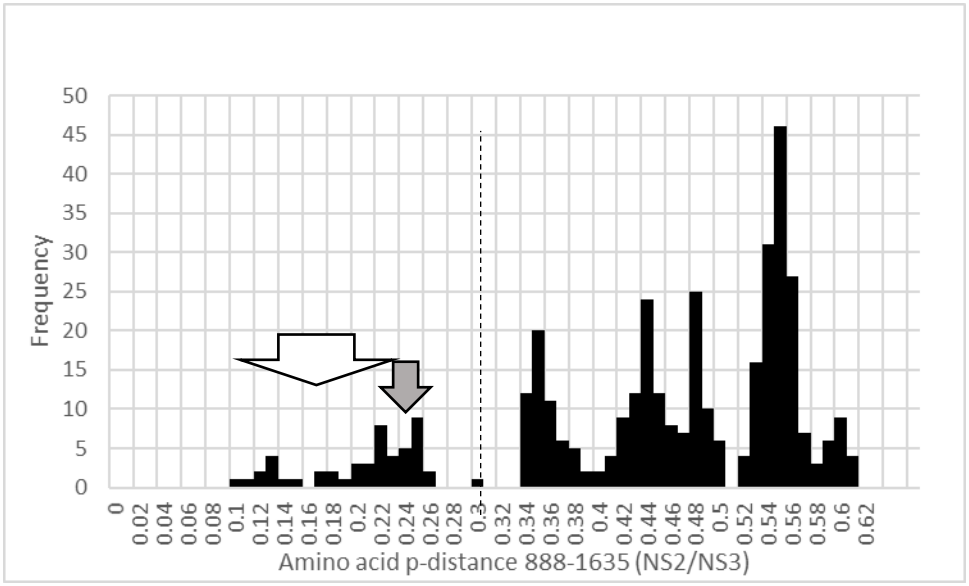


Figure 1
Amino acid divergence across *Pegivirus* polyproteins. Mean amino acid p-distances were calculated for 26 aligned *Pegivirus* polyprotein sequences that differed by > 0.11 of amino acid positions using a sliding window of 50 amino acids incremented by 10 residues and plotted against the amino acid position of the start of the fragment. Increments on the x-axis scale are uneven because of unnumbered gaps in the reference sequence (U22303). Two regions with distances consistently < 0.6 are indicated by bars. A schematic representation of the *Pegivirus* polyprotein is shown to scale below.



C.



D.

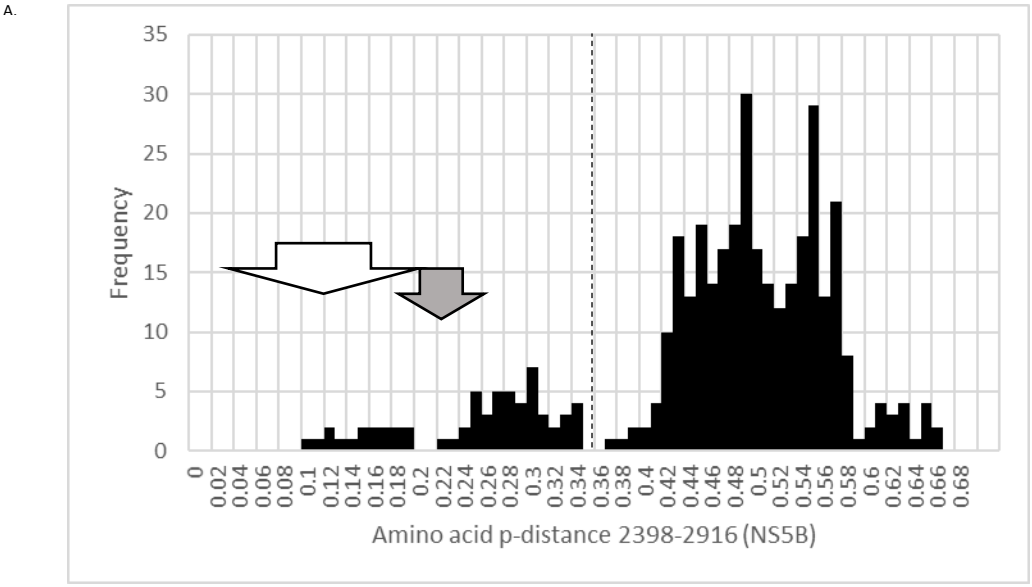


Figure 2

Analysis of *Pegivirus* conserved regions. Maximum likelihood trees were produced using MEGA 6 for (A) amino acid positions 888-1635 using the Le and Gascuel (LG) model with frequencies and a gamma distribution of variation with invariant sites, and for (B) amino acid positions 2398-2916 using the LG model with a gamma distribution of invariant sites. Branches observed in >70% of bootstrap replicates are indicated. Frequency histograms of amino acid p-distance between *Pegivirus* sequences in the regions 888-1635 (C) and the region 2398-2916 (D). Amino acid p-distances between *Pegivirus C* sequences derived from different primate species are indicated by an open arrow, while those between primate and bat-derived *Pegivirus A* sequences are indicated by a shaded arrow. The distance that demarcates different species is indicated by a dotted line.