This form should be used for all taxonomic proposals. Please complete all those modules that are applicable.

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.

Part 1: **TITLE, AUTHORS, etc**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Code assigned:** | ***2017.010D*** | | | | (to be completed by ICTV officers) |
| **Short title:** creation of 17 new species, 4 new genera and 2 subfamilies in the family *Papillomaviridae* | | | | | |
| **Modules attached**  (Modules 1, 4 and either 2 or 3 are required. | | **1**  **2  3  4** | | | |
| **Author(s):** | | | | | |
| Robert Burk and Zigui Chen for the Papillomaviridae Study Group | | | | | |
| **Corresponding author with e-mail address:** | | | | | |
| Robert Burk, robert.burk@einstein.yu.edu | | | | | |
| **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 (there are six virus subcommittees: animal DNA and retroviruses, animal ssRNA-, animal ssRNA+, fungal and protist, plant, bacterial and archaeal) | | | Papillomaviridae Study Group (pending) | | |
| **ICTV Study Group comments (if any) and response of the proposer:** | | | | | |
|  | | | | | |
|  | | | | | |
| Date first submitted to ICTV: | | | | June 18, 2017 | |
| Date of this revision (if different to above): | | | | June 21, 2017 | |

|  |
| --- |
| **ICTV-EC comments and response of the proposer:** |
|  |

**Part 2**: **PROPOSED TAXONOMY**

|  |
| --- |
| Present the proposed new taxonomy on accompanying spreadsheet |
| **Name of accompanying spreadsheet:** 2017.010D.N.v1.Papillomaviridae\_2subf |

Please display the taxonomic changes you are proposing on the accompanying spreadsheet module 2017\_TP\_Template\_Excel\_module. Submit both this and the spreadsheet to the appropriate ICTV Subcommittee Chair.

**Part 4:** **APPENDIX**: supporting material

|  |
| --- |
| **References:** |
| 1. **Bernard HU, Burk RD, Chen Z, van Doorslaer K, zur Hausen H, de Villiers EM.** 2010. Classification of papillomaviruses (PVs) based on 189 PV types and proposal of taxonomic amendments. Virology **401:**70-79.  2. **Simmonds P, Adams MJ, Benko M, Breitbart M, Brister JR, Carstens EB, Davison AJ, Delwart E, Gorbalenya AE, Harrach B, Hull R, King AM, Koonin EV, Krupovic M, Kuhn JH, Lefkowitz EJ, Nibert ML, Orton R, Roossinck MJ, Sabanadzovic S, Sullivan MB, Suttle CA, Tesh RB, van der Vlugt RA, Varsani A, Zerbini FM.** 2017. Consensus statement: Virus taxonomy in the age of metagenomics. Nat Rev Microbiol **15:**161-168.  3. **Stamatakis A.** 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics **22:**2688-2690.  4. **Edgar RC.** 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics **5:**113.  5. **Team RC.** 2014. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2013. ISBN 3-900051-07-0. |

|  |
| --- |
| **Annex:**  Please explain the reasons for the taxonomic changes you are proposing and provide evidence to support them. The following information should be provided, where relevant:   * **Species demarcation criteria**: Explain how new species differ from others in the genus and demonstrate that these differences meet the criteria previously established for demarcating between species. If no criteriahave previously been established, and if there will now be more than one species in the genus, please state the demarcation criteria you are proposing. * **Higher taxa**:   + There is no formal requirement to state demarcation criteria when proposing new genera or other higher taxa. However, a similar concept should apply in pursuit of a rational and consistent virus taxonomy.   + Please indicate the **origin of names** assigned to new taxa at genus level and above.   + For each new genus a **type species** must be designated to represent it. Please explain your choice. * **Supporting evidence**: The use of Figures and Tables is strongly recommended (note that copying from publications will require permission from the copyright holder). For phylogenetic analysis, try to provide a tree where branch length is related to genetic distance. |

The current taxonomic classification of papillomaviruses has been based on the nt sequence of the L1 ORF. The rationale for using this region of the genome emerges from the absolute conserved and consistent structure of the viral particle that uses an icosahedral geometric capsule composed of predominantly L1 capsids. For example, the L1 protein self-assembles into virus like particles (VLPs) and is necessary and sufficient for this activity. Thus, there is sufficient amino acid homology to align essentially all papillomavirus L1 protein sequences and back-translate and align the nucleotides. This has served as a valuable tool for taxonomy and phylogeny, but with the rapid expansion of PVs from a large variety of hosts, the L1 is becoming a screening tool with further support from the 4 core ORFs (E1/E2/L1/L2) present in all known PVs to date.

The original criteria distinguishing genera stated, “Most types within a PV genus show less than 60% sequence identity to types of other genera based on global multiple sequence or pairwise alignments of the L1 genes. Nevertheless, the suggested percentage identities that define PV genera have to be taken as general, but not absolute criteria for a number of reasons.” (1). The criteria for defining a new species is around 70%, that is a novel PV genome with complete genome sequence data available and being <70% related to PV’s within the genus would be designated as a new species. As stated above for genera, this is a general rule, but not absolute and needs curation. The current document also includes the current proposal to create subfamilies within the family *Papillomaviridae*. Recently papillomaviruses have been isolated from a variety of fish. These are highly disparate genomes and their L1 ORF sequence have <45% sequence similarity to the known set of PV’s. Although only one fish papillomavirus, Sparus aurata papillomavirus 1 (SaPV1, KX643372) is available in GenBank, we have in hand another 11 fish genomes that provide evidence that SaPV1 is not something “fishy”. Dr. Chris Buck at the National Cancer Institute (NCI), NIH, Bethesda MD who has been in contact with us, discovered these virus sequences. Since the complete genome sequences have not been deposited in GenBank, according to ICTV criteria, we are not proposing these for taxonomic assignment at this time.

Lastly, there are many new genomes that have been identified through metagenomic data that we are including in this proposal. We have based our criteria on suggestions from the recent review by Simmonds et al., 2017 (2). Thus, a complete genome with sufficient read depth is required and the genome has to basically present no significant problems in the identification of ORFs, etc. Although we would have liked to include 2 new species for human papillomaviruses, we have decided to postpone this until the community comes to consensus on this matter. Currently, Dr. Joakim Dillner runs a human papillomavirus repository (<http://www.nordicehealth.se/hpvcenter/>). This facility provides an important public service and Dr. Dillner is an outstanding investigator and member of Papillomaviridae Study Group. In the past, he has been firm on the receipt of either a cloned PV genome and/or the PCR fragments covering the genome. Alternatively, people discovering novel PV genomes in metagenomic data would have to provide some type of biological material that Dr. Dillner’s facility could replicate the sequence. No one has done that yet, and although some of us (RB) believe people will not take the effort, in deference to Dr. Dillner we are working on this issue.

The evidence to propose 4 new genera and 2 subfamilies are discussed below.

Myotis ricketti papillomavirus 1 (MrPV1, JQ814847) is a genome identified from a metagenomic analysis from a Rickett's big-footed bat. Its position in the L1 tree “L1 nt RAxML Tree” is not well supported, so we go to the 4ORF tree “4ORF nt RAxML Tree” where MrPV1 forms a well-supported clade with *Omegapapillomavirus 1* (UmPV1, EF536349 – isolated from a polar bear), *Dyopipapapillomavirus 1* (PphPV4, GU117623 – isolated from a harbor porpoise) and *Dyodeltapapillomavirus 1* (SsPV1, EF395818 – isolated from a domestic pig). Given the disparate hosts and the diversity of the 4 ORF analysis, it appears MrPV1 represents a new genus ***Treisiotapapillomavirus*** that further supports the creation of a new species ***Treisiotapapillomavirus 1***.

Rusa timorensis papillomavirus type 1 (RtiPV1, KP757765) was isolated from a deer. RtiPV1 is proposed as a new species *Treisthetapapillomavirus 1* forming a new genus ***Treisthetapapillomavirus****.* RtiPV1 forms a clade with Capra hircus papillomavirus 1 (ChPV1, DQ091200) isolated from a domestic goat that is a species type and member of the genus *Phipapillomavirus* as shown in the Tree figure labeled, “L1 nt RAxML Tree”. To the right of the tree is shown a plot of the 1 X 1 sequence comparisons of each proposed virus with all others. To confirm the topology and percent difference from all other species and genera papillomaviruses, we have also displayed a 4 ORF nt tree and similarity for these regions (see 4ORF nt RAxML Tree).

Equus caballus papillomavirus 8 (EcPV8, KU963288, isolated from a horse) is designated as a newspecies *Treiskappapapillomavirus 1* and genus***Treiskappapapillomavirus****.* It is related to a *Zetapapillomavirus 1* (EcPV1, AF498323) in the L1 ORF tree, but is part of a larger clade in the 4ORF tree that includes *Zetapapillomavirus 1*, *Dyoiotapapillomavirus 1/2* (EcPV2, EU503122; EcPV4, JQ031032; EcPV5, JQ031033 – other EcPVs isolated from horses) and *Dyorhopapillomavirus 1* (EcPV3, GU384895). Thus there appears to be a clade with 4 different genera (included the newly proposed *Treiskappapapillomavirus 1*) of related papillomaviruses isolated from horses.

Lastly, the proposition of creating subfamilies within *Papillomaviridae* is well supported as the data provides compelling evidence that the PV isolated from a gilthead sea bream (*Sparus aurata*) (SaPV1, KX643372) fish meets the criteria described below. It has many unique characteristics including a smaller PV genome and a spliced L1 ORF. Both the L1 and 4ORF tree shows a unique position and lower sequence similarity than seen for other PVs. We propose this PV genome as a new species *Alefpapillomavirus 1* of the genus ***Alefpapillomavirus*** constituting a new subfamily ***Secondpapillomavirinae***. We further suggest to call the other papillomaviruses *Firstpapillomavirinae* indicating their characterization preceding the fish PVs.

**Figure 1.** Phylogenetic tree showing papillomavirus species. Maximum likelihood (ML) trees were constructed using RAxML MPI v8.2.9 (3) with GTR substitution model based on the **aligned L1 ORF nucleotide sequences** of 133 papillomavirus types representing species. The nucleotide sequence alignments were based on the aligned amino acid using the program MUSCLE v7.221 (4). Numbers on or near branches indicate ML bootstrap percentages with autoMRE-based bootstopping criterion in RAxML. The bar indicates the nucleotide substitution of 0.5 changes per site. The percent nucleotide identity for new species compared to all current species are shown in the panel to the right of the phylogeny. Values for each comparison (i.e., these are all 1 X 1 comparisons) of a given species are connected by lines and the comparison to self is indicated by the 100% identity point. The values were calculated based on the alignments used for tree construction employing the *dist.dna* function in R’s package “ape” v4.1 (5), with setting of “model” as “raw” and “pairwise.deletion” as “true”. Different colored lines are used to distinguish each isolate.

Analysis/Results/tree_ICTV2017_133PV_L1n.pdf

**Analysis/Results/tree_ICTV2017_133PV_4ORFn.pdf**

Figure 2. Phylogenetic tree showing papillomavirus species. The tree was based on the **alignment of concatenated 4 open reading frames** (ORFs, E1, E2, L2 and L1) of 133 papillomavirus types representing species.