This Word module should be used for all taxonomic proposals.

Please complete **Part 1** and:

either **Part 3** for proposals to create new taxa or change existing taxa

or **Part 2** for proposals of a general nature.

Submit the completed Word module, together with the accompanying Excel module named in Part 3, to the appropriate ICTV Subcommittee Chair.

The Word module explains and justifies your proposal. The Excel module is a critical document that will be used to implement the proposed taxonomic changes once they are approved and ratified. If proposals presented in the Word module are not presented accurately in the Excel module, the taxonomic changes cannot proceed.

For guidance, see the notes written in blue, below, and the Help Notes in file Taxonomic\_Proposals\_Help\_2019.

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

|  |  |  |
| --- | --- | --- |
| **Code assigned:** | ***2019.004G*** |  |
| **Short title:** Create a megataxonomic framework, filling all principal/primary taxonomic ranks, for dsDNA viruses encoding HK97-type major capsid proteins |
|  |
| **Author(s) and email address(es):**  |
| List authors in a single line *Archives of Virology* citation format (e.g. Smith AB, Huang C-L, Santos, F) | Provide email address for each author in a single line separated by semi-colons |
| Koonin EV, Dolja VV, Krupovic M, Varsani A, Wolf YI, Yutin N, Zerbini M, Kuhn JH | koonin@ncbi.nlm.nih.gov; doljav@science.oregonstate.edu; mart.krupovic@pasteur.fr; Arvind.Varsani@asu.edu; wolf@ncbi.nlm.nih.gov; yutin@ncbi.nlm.nih.gov; zerbini@ufv.br; kuhnjens@mail.nih.gov  |
| **Corresponding authors** |
| Koonin, Eugene V.; koonin@ncbi.nlm.nih.govKuhn, Jens H.; kuhnjens@mail.nih.gov |
| **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) | **ICTV *Caudovirales* Study Group Chair; ICTV *Herpesvirales* Study Group Chair; ICTV Bacterial and Archaeal Viruses Subcommittee; ICTV Animal dsRNA and ssRNA- Viruses Subcommittee Chair; ICTV Animal DNA Viruses and Retroviruses Subcommittee Chair; ICTV Plant Viruses Subcommittee Chair; ICTV Fungal and Protist Viruses Subcommittee Chair****This is a direct submission to the entire ICTV Executive Committee** |
| **ICTV Study Group/Author comments (if any) and response of the proposer:** |
| Here we propose a megataxonomic framework for a subset of double-stranded DNA (dsDNA) viruses by assigning ICTV-ratified taxa (i.e., species, genera, subfamilies, families, and orders) to available, but presently unfilled major megataxonomic ranks (classes, phyla, and kingdoms). The goal of this proposal is to provide taxonomic “buckets” or “place holders” that enable ICTV Study Groups to accommodate the close-to-exponentially increasing number of novel viruses that are related to, but distinct from, viruses that constitute the already established taxa. The awareness of these novel viruses often goes hand-in-hand with the realization that current orders might have to be promoted to classes (e.g., *Caudovirales*) or that entire family structures need to be completely re-evaluated (e.g., caudoviral subtaxa). We surmise that the absence of established higher taxa and the absence of ICTV Study Groups for such taxa may have had an adverse effect, leading to large groups of classifiable viruses not becoming classified. Conversely, placing currently established taxa together into higher-rank taxa can be expected to initiate long-overdue, likely intense, discussions between currently non-interacting ICTV Study Groups to examine higher-rank evolutionary relationships of the viruses they are engaged with. The megataxonomy outlined in this proposal is to be seen as an initial step in this direction, and we expect this framework to change substantially over time.We:* aim to bring virus taxonomy into better accordance with other biological taxonomies, which require novel organisms to be classified into all available principle/primary ranks even if this means that certain higher-ranked taxa only include single lower-ranked taxa. For instance, in animal taxonomy, the unranked supergroup Hemimastigophora includes a single class Hemimastigidea, which includes a single order Hemimastigida, which includes a single family Spironem(atelli)idae (which includes 4 genera). Likewise, in prokaryotic taxonomy, the bacterial species *Elusimicrobium minutum* is the only included species in genus *Elusimicrobium*, which is the only genus in family *Elusimicrobiaceae*, which is the only family in order *Elusimicrobiales*—that order is the only order in class *Elusimicrobia*, which is the only class in phylum *Elusimicrobia*. Obviously, taxon demarcation criteria cannot be established for single taxon-including higher-ranked taxa and hence their definitions are identical to those of the higher-ranked taxa for the time being, i.e., until the discovery of novel organisms requires the creation of sister taxa. However, filling all principle ranks provides a sense of “scaling”, i.e. a current assessment of how distant a particular organism is from other classified taxa; this “scaling” argument was used successfully previously in TaxoProps establishing the availability of taxonomic ranks above order and the establishment realm *Riboviria*;
* deliberately propose the creation of higher-ranked taxa that currently include only single lower-ranked taxa, either because we are aware from the literature that an existing lower-rank taxon will have to be promoted to a higher rank in the near future due to overbearing virus diversity (e.g., *Caudovirales*), or because we are aware from the literature of large virus groups for which higher-rank taxa will have to be established shortly; we hope that the created higher-ranked taxa will provide an impetus for the community to classify already known highly divergent virus groups;
* deliberately did not fill any secondary (sub-)ranks as the filling of such ranks is not mandatory in other biological taxonomies;
* deliberately focus this proposal only on already official taxa (rather than, for instance, proposing novel species that could become the founding members of “obvious” novel higher-rank taxa we are certain will need to be established) to keep this proposal relatively simple; and
* emphasize that, although we posit that focusing on a single protein fold is initially sufficient for creation of the dsDNA virus megataxonomy outlined here (from the highest rank, realm, down to approximately class/order), such a focus must be seen as a rough guide for lower-rank taxonomy, so that other methods (e.g., sequence-based methods such as GRAViTy, pairwise genome sequence comparisons, phylogenies of individual ORFs or proteins; structural comparisons of encoded proteins or virions; phenotypic virus characteristics) will have to be used to resolve lower-rank relationships and likely to refine higher-rank relationships.
 |
|  |
| Date first submitted to ICTV: | June 19, 2019 |
| Date of this revision (if different to above): | October 18, 2019 |

|  |
| --- |
| **ICTV-EC comments and response of the proposer:** |
| Per the minutes of the last ICTV Executive Committee meeting (EC51, July 15–17, 2019, Berlin, Germany), the EC voted for minor revisions of this proposal (Uc) with 18/19 votes in favor. The EC asked for the following steps to be taken prior to submission of this revision:1. Consult all affected Study Group once again for feedback

Response: all relevant Study Group Chairs were contacted for a second time and asked to provide input and criticisms. None of the Study Groups disagreed with the overall taxonomic proposal, i.e., the proposed relationship between officially established taxa. Concerns were voiced about certain proposed taxon names and the Study Groups’ suggestion were mostly followed and names were changed accordingly for this revision.Some members of the ICTV *Herpesvirales* Study Group voiced their dislike over filling all primary/principle ranks despite some of them only including single lower-ranked taxa. We modified this revised proposal to further explain why we chose this path (see first bullet point above).Some members of the ICTV *Herpesvirales* Study Group wondered whether the taxonomic main split between the two major groups of viruses in the proposal ought to occur at higher or lower ranks, but no scientific data were provided justifying the one or the other.Some members of the ICTV *Herpesvirales* Study Group pointed out that the conservation of the terminase large subunit, which is detectable at the primary sequence level, provides additional support to the HK97-MCP fold-based taxonomy proposed here. We agree with this assessment and note so now in the proposal text. Future analyses of these sequences may result in better defined realm inclusion criteria than are currently outlined.1. Provide any feedback from Study Groups to the ICTV President

Response: all relevant Study Group responses (and all TaxoProp author rebuttals or explanations) were forwarded to the ICTV President and the ICTV Executive Committee per email. |

**Part 3:** **PROPOSED TAXONOMY**

|  |
| --- |
| **Name of accompanying Excel module:** 2019.004G.A.v1.Duplodnaviria.xlsx |

**INTRODUCTION**

Recent advances of comparative genomics and metagenomics uncover a close-to-exponentially increasing number of diverse viruses. These discoveries not only vastly improve our understanding of the evolutionary relationships within the virosphere, but also emphasize that the existing taxonomic framework is inadequate to depict the relationships within the virosphere. The currently available dataset of virus genome sequences and increasingly sophisticated methods for analysis beyond “simple” phylogenies (e.g., gene network analysis, iterative and self-optimizing sequence alignments) enable us now to roughly outline the global organization of the virus world in its entirety, including key evolutionary events that resulted in the emergence of major virus clades.

Depicting the evolutionary relationships among viruses necessarily depends on the identification of hallmark genes/proteins that connect them. In contrast to cellular organisms, such hallmark genes are not universally shared among all viruses [[9](#_ENREF_9)] and it is therefore currently presumed that viruses have several distinct points of origin, i.e., that they cannot be united under a single highest taxon rank on evolutionary grounds. Nevertheless, extensive analyses of the evolution of large groups of viruses, rather than all of them, have proved productive. The primary approach taken in such studies is the phylogenetic analysis of genes that are conserved across those groups, known as Virus Hallmark Genes (VHGs), which are responsible for the key functions in virus replication and virion morphogenesis [[9](#_ENREF_9)]. The most widely spread VHGs are:

* RNA-directed RNA polymerases (RdRps);
* RNA-directed DNA polymerases/reverse transcriptases (RTs) that are homologous to RdRps;
* superfamily 3 helicases (S3Hs);
* single jelly-roll major capsid proteins (SJR-MCPs);
* double jelly-roll major capsid proteins (DJR-MCPs); and
* rolling-circle replication initiation endonucleases (RCREs) [[9](#_ENREF_9), [11](#_ENREF_11), [12](#_ENREF_12)].

Using these VHGs, megataxonomic scaffolds can be established that are further informed, if necessary, by dissection of bipartite gene-genome networks of viruses into distinct modules [[4-6](#_ENREF_4), [10](#_ENREF_10), [21](#_ENREF_21)]. These analyses indicate that a substantial majority of currently classified viruses can be assigned to one of four, likely, evolutionarily independent virus realms. Because the International Committee on Taxonomy of Viruses (ICTV) has recently formally approved creation of taxa above the rank of order, the door is now open to formalize the megataxonomic scaffolds that resulted from VHG analyses within the official ICTV-supported taxonomy.

Here we propose a megataxonomic structure for one of these groups: dsDNA viruses that encode HK97 major capsid proteins and small terminase subunits required for DNA packaging into capsids.

**MEGATAXONOMY OF HK97 MAJOR CAPSID PROTEIN dsDNA VIRUSES**

The HK97 major capsid protein (MCP) supermodule of viruses, named after Escherichia coli phage HK97 (*Caudovirales*: *Siphoviridae*: *Hendrixvirus*) whose capsid was among the first determined capsid structures of this virus assemblage [[23](#_ENREF_23)], unites tailed prokaryotic double-stranded DNA (dsDNA) viruses in the established order *Caudovirales* and animal dsDNA viruses in the established order *Herpesvirales*. The icosahedral capsids of these viruses are made of homologous proteins with a distinct fold that are unrelated to those of other dsDNA viruses (Figure 1), which share vertical single or double jelly roll (DJR) MCPs [[3](#_ENREF_3), [5](#_ENREF_5), [9](#_ENREF_9), [23](#_ENREF_23)]. In addition to the HK97-MCPs, the viruses in this supermodule are further connected by portal proteins, capsid maturation proteases, and two subunits (large and small) of the terminases [[4](#_ENREF_4), [8](#_ENREF_8), [14](#_ENREF_14), [20](#_ENREF_20)]. The large terminase subunits are a distinct variety of packaging ATPases that share only the general structural design of the P-loop fold with the packaging ATPases of the DJR-MCP viruses and are not directly related to the latter [[13](#_ENREF_13)]. Detailed structural, proteomic, and computation studies continue to uncover additional proteins shared between members of the two virus orders [[10](#_ENREF_10), [21](#_ENREF_21)].

The shared properties of HK97-MCP viruses and the absence of shared properties between these viruses and any other virus group justifies the establishment of a realm (proposed to be called *Duplodnavira*) that includes the currently established orders *Caudovirales* and *Herpesvirales*. We propose to separate both orders at the phylum rank (proposed phyla *Uroviricota* and *Peploviricota*, respectively) given that the vast order *Caudovirales* is currently undergoing major reorganization and is expected to massively expand in the near future, likely requiring the promotion of subfamilies to families or higher and the order to class or higher [[2](#_ENREF_2), [11](#_ENREF_11), [12](#_ENREF_12), [15](#_ENREF_15)]. In addition, new families of HK97-MCP viruses are being continuously discovered by metagenomics and single-cell genomics. Examples include the recently identified large group of “cross assembly (crAss)-like viruses” [[7](#_ENREF_7), [22](#_ENREF_22), [24](#_ENREF_24)], “magroviruses” that infect mesophilic euryarchaeota [[19](#_ENREF_19)], and “Lak-like megaphages” [[1](#_ENREF_1), [6](#_ENREF_6)], all of which have to be accommodated in the megataxonomy. This proposal is to be considered as an operational move to formalize the obvious evolutionary relationship between *Caudovirales* and *Herpesvirales* based on a common VHG.

**Taxon demarcation criteria:**

The International Code of Virus Classification and Nomenclature (ICVCN) is currently ambiguous regarding the need for taxon demarcation criteria at higher ranks. Three Rules appear to be applicable (emphasis in italics is ours):

“3.5 Taxa will be established only when representative member viruses are sufficiently well characterized and described in the published literature so as to allow them to be identified unambiguously *and the taxon to be distinguished from other similar taxa*”;

“3.22 Every individual virus is a physical entity and treated as belonging to a number of taxa of hierarchical ranks, *some of which may remain undefined*”;

and

“3.24 The classification of a virus at the species and genus ranks is mandatory. *Classification may also encompass any further number of taxa at higher hierarchical ranks*”

Because our proposal only encompasses already established taxa, all viruses affected by our proposal have been “sufficiently well characterized” as otherwise they would not have been classified into these established taxa in the first place. Furthermore, Rule 3.22 permits establishing ranks that for the moment remain undefined; and Rule 3.24 indicates no restriction of ranks to be established.

Nevertheless, for the time being, we suggest the following provisional taxon demarcation criteria while being aware that these may have to be revisited whenever new members of the realm are being proposed:

1. *Duplodnaviria*: a virus is a member of this realm if it has a dsDNA genome encoding a major capsid protein containing the HK97 fold (HK97-MCP). Viruses of this realm also share portal protein and the terminase complex which is unrelated to the genome packaging ATPase conserved in the majority of vertical jelly-roll major capsid protein dsDNA viruses (proposed realm *Varidnaviria*)
2. *Uroviricota*: a *Duplodnaviria* member is a member of the included phylum *Uroviricota* if it infects prokaryotes, but not eukaryotes
3. *Peploviricota*: a *Duplodnaviria* member is a member of the included phylum *Peploviricota* if it infects eukaryotes, but not prokaryotes

If a principle rank taxon includes only a single lower-ranked taxon, then the definition of the lower-ranked taxon is, for now, identical to the definition of the higher-ranked taxon.

Truly useful taxon demarcation criteria will have to be established in the future, likely by incorporating yet-unclassified virus groups into the realm. The conservation of the terminase large subunit, which is detectable at the protein sequence level in all viruses proposed to be classified here and which is also readily identifiable among “crAss-like viruses”, “magroviruses”, and “Lak-like megaphages” could aid in the further delineation of taxon demarcation criteria.

**Etymology of proposed taxa:**

* *Duplodnaviria*; from Latin dūplō, meaning double and DNA (referring to the fact that all founding members of the realm have double-stranded dsDNA genomes); and the suffix -*viria* for realm taxa
* *Heunggongvirae*; from Cantonese 香港 [Hēunggóng], meaning (and approximately pronounced) Hong Kong—a reference to Escherichia coli phage HK97, the founding member of the HK97 [Hong Kong 97]-fold major capsid protein viruses in this taxon; and the suffix -*virae* for kingdom taxa
* *Uroviricota*; to honor the proposed name for a higher-rank taxon for tailed phages proposed by Lwoff and Tournier in 1966 (“Urovirales”; from Greek οὐρά [ourá/uros, meaning tail) for the “LHT system” [[16-18](#_ENREF_16)]; and the suffix -*viricota* for phylum taxa
* *Caudoviricetes*; derived from Latin caudo, meaning tail—a reference to the current order *Caudovirales*, which will likely be reorganized (and its name dissolved) in the near future; and the suffix -*viricetes* for class taxa
* *Peploviricota*; to honor the proposed name for a higher-rank taxon including herpesviruses proposed by Lwoff and Tournier in 1966 (“Peplovirales”; from Greek πέπλος [peplos], meaning garment, a reference to the unique tegument of herpesviruses) for the “LHT system” [[16-18](#_ENREF_16)]; and the suffix -*viricota* for phylum taxa
* *Herviviricetes*; from herpesvirus; and the suffix -*viricetes* for class taxa

None of the proposed taxon names above is derived from a person’s name.



**Figure 1.** Major capsid proteins of HK97-MCP viruses infecting hosts from the three domains of life. The structures are coloured according to the secondary structure elements: α‑helices, red; β‑strands, blue; and random coil, grey. Only the ‘floor’ domain of the human cytomegalovirus (HCMV) capsid protein is shown. The X‑ray structure of the capsid protein of HSTV-1 is not available and is represented with a homology-based model. HK97, Escherichia coli phage HK97; HSTV-1, Haloarcula sinaiiensis tailed virus 1.

**List of official (ss and ds) DNA virus taxa that we propose remain unassigned to any realm, including the one proposed here, until further data become available:**

* *Anelloviridae*
* *Ampullaviridae*
* *Baculoviridae*
* *Bicaudaviridae*
* *Clavaviridae*
* *Dinodnavirus*
* *Finnlakeviridae* [proposed]
* *Fuselloviridae*
* *Globuloviridae*
* *Guttaviridae*
* *Halspiviridae* [proposed] including *Salterprovirus*
* *Hytrosaviridae*
* *Ligamenvirales*
* *Nimaviridae*
* *Nudiviridae*
* *Ovaliviridae*
* *Plasmaviridae*
* *Polydnaviridae*
* *Portogloboviridae*
* *Rhizidiovirus*
* *Spiraviridae*
* *Thaspiviridae* [proposed]
* *Tristromaviridae*

| **References:** |
| --- |
| 1. Al­Shayeb B, Sachdeva R, Chen L-X, Ward F, Munk P, Devoto A, Castelle CJ, Olm MR, Bouma-Gregson K, Amano Y, He C, Méheust R, Brooks B, Thomas A, Lavy A, Matheus-Carnevali A, Sun C, Goltsman DSA, Borton MA, Nelson TC, Kantor R, Jaffe AL, Keren R, Farag IF, Lei S, Finstad K, Amundson R, Anantharaman K, Zhou J, Probst AJ, Power ME, Tringe SG, Li W-J, Wrighton K, Harrison S, Morowitz M, Relman DA, Doudna JA, Lehours A-C, Warren L, Cate JHD, Santini JM, Banfield JF (2019) Clades of huge phage from across Earth’s ecosystems. bioRxiv:5723622. Barylski J, Enault F, Dutilh BE, Schuller MBP, Edwards RA, Gillis A, Klumpp J, Knezevic P, Krupovic M, Kuhn JH, Lavigne R, Oksanen HM, Sullivan MB, Jang HB, Simmonds P, Aiewsakun P, Wittmann J, Tolstoy I, Brister JR, Kropinski AM, Adriaenssens EM (2019) Analysis of spounaviruses as a case study for the overdue reclassification of tailed phages. Syst Biol:doi: 10.1093/sysbio/syz10363. Bayfield OW, Klimuk E, Winkler DC, Hesketh EL, Chechik M, Cheng N, Dykeman EC, Minakhin L, Ranson NA, Severinov K, Steven AC, Antson AA (2019) Cryo-EM structure and in vitro DNA packaging of a thermophilic virus with supersized T=7 capsids. Proc Natl Acad Sci U S A 116:3556-35614. Cheng H, Shen N, Pei J, Grishin NV (2004) Double-stranded DNA bacteriophage prohead protease is homologous to herpesvirus protease. Protein Sci 13:2260-22695. Dai X, Zhou ZH (2018) Structure of the herpes simplex virus 1 capsid with associated tegument protein complexes. Science 360: eaao72986. Devoto AE, Santini JM, Olm MR, Anantharaman K, Munk P, Tung J, Archie EA, Turnbaugh PJ, Seed KD, Blekhman R, Aarestrup FM, Thomas BC, Banfield JF (2019) Megaphages infect *Prevotella* and variants are widespread in gut microbiomes. Nat Microbiol 4:693-7007. Dutilh BE, Cassman N, McNair K, Sanchez SE, Silva GG, Boling L, Barr JJ, Speth DR, Seguritan V, Aziz RK, Felts B, Dinsdale EA, Mokili JL, Edwards RA (2014) A highly abundant bacteriophage discovered in the unknown sequences of human faecal metagenomes. Nat Commun 5:44988. Feiss M, Rao VB (2012) The bacteriophage DNA packaging machine. Adv Exp Med Biol 726:489-5099. Fokine A, Leiman PG, Shneider MM, Ahvazi B, Boeshans KM, Steven AC, Black LW, Mesyanzhinov VV, Rossmann MG (2005) Structural and functional similarities between the capsid proteins of bacteriophages T4 and HK97 point to a common ancestry. Proc Natl Acad Sci U S A 102:7163-716810. Hernandez Duran A, Greco TM, Vollmer B, Cristea IM, Grunewald K, Topf M (2019) Protein interactions and consensus clustering analysis uncover insights into herpesvirus virion structure and function relationships. PLoS Biol 17:e300031611. Iranzo J, Koonin EV, Prangishvili D, Krupovic M (2016) Bipartite network analysis of the archaeal virosphere: evolutionary connections between viruses and capsidless mobile elements. J Virol 90:11043-1105512. Iranzo J, Krupovic M, Koonin EV (2016) The double-stranded DNA virosphere as a modular hierarchical network of gene sharing. MBio 7:e00978-0091613. Iyer LM, Leipe DD, Koonin EV, Aravind L (2004) Evolutionary history and higher order classification of AAA+ ATPases. J Struct Biol 146:11-3114. Liu J, Mushegian A (2004) Displacements of prohead protease genes in the late operons of double-stranded-DNA bacteriophages. J Bacteriol 186:4369-437515. Low SJ, Dzunkova M, Chaumeil PA, Parks DH, Hugenholtz P (2019) Evaluation of a concatenated protein phylogeny for classification of tailed double-stranded DNA viruses belonging to the order Caudovirales. Nat Microbiol 4:1306-131516. Lwoff A, Horne RW, Tournier P (1962) Un système des virus. C R Hebd Séances Acad Sci 254:4225-422717. Lwoff A, Tournier P (1966) The classification of viruses. Annu Rev Microbiol 20:45-7418. Lwoff A (1967) Principles of classification and nomenclature of viruses. Nature 215:13-1419. Philosof A, Yutin N, Flores-Uribe J, Sharon I, Koonin EV, Beja O (2017) Novel abundant oceanic viruses of uncultured marine group II Euryarchaeota. Curr Biol 27:1362-136820. Rao VB, Feiss M (2008) The bacteriophage DNA packaging motor. Annu Rev Genet 42:647-68121. Rixon FJ, Schmid MF (2014) Structural similarities in DNA packaging and delivery apparatuses in herpesvirus and dsDNA bacteriophages. Curr Opin Virol 5:105-11022. Shkoporov AN, Khokhlova EV, Fitzgerald CB, Stockdale SR, Draper LA, Ross RP, Hill C (2018) PhiCrAss001 represents the most abundant bacteriophage family in the human gut and infects *Bacteroides intestinalis*. Nat Commun 9:478123. Wikoff WR, Liljas L, Duda RL, Tsuruta H, Hendrix RW, Johnson JE (2000) Topologically linked protein rings in the bacteriophage HK97 capsid. Science 289:2129-213324. Yutin N, Makarova KS, Gussow AB, Krupovic M, Segall A, Edwards RA, Koonin EV (2018) Discovery of an expansive bacteriophage family that includes the most abundant viruses from the human gut. Nat Microbiol 3:38-46 |