

ICTVdB - The Universal Virus Database

Status report for ICTV EC meeting ICV San Francisco July 2005

Executive Summary

1. A summary of features of [ICTVdB](#), the copyrighted, interoperable, universal virus database of ICTV approved taxa is given on the poster for ICV 2005 ([Fig. 1](#)). A brief history of ICTVdB is given in [Appendix 1](#).
2. The database is being updated to the 8th Report using the [Master Species List](#) released by the President of ICTV in June 2005. The delay allegedly arose because of publication embargoes surrounding the release of the ICTV 8th Report.
3. ICTVdB already provides ICTV with a
 - web interface to [approved taxonomy](#) to the level of [species](#) including strain and isolate information
 - fully functional and active links of approved taxonomy to PubMed Central, CDC fact sheets, sequence databases, such as GenBank and Swiss-Prot, and other databases (example [Astroviridae](#))
4. ICTVdB potentially provides ICTV with a
 - smooth path to updating and publishing future reports
 - ready made format for researchers to [submit](#), and for study groups to review, proposals for new viruses and revisions of existing taxa.
5. Each description of a taxon in ICTVdB is constructed from a list of some 2,700 ICTV approved [characters](#), from virus genome and particle properties to disease, vector, and host range and is further supported by images. The long list of characters is managed through intrinsic, logical dependencies that interlink related sets of characters thus providing or blocking access to related sections of characters depending on the selected feature at each decision point.

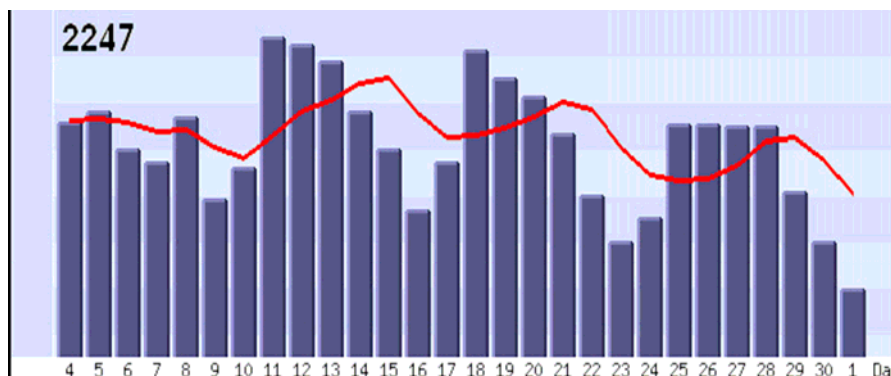


Figure 2: Visits to ICTVdB web pages on the US sites: Web statistic for the month of April 2005.

Table 1. The decimal code applied to taxonomic revisions of Poliovirus and constructed to anticipate the explosion of lower-level data (serotypes, strains, and isolates).

Taxonomic Level	Decimal Code
Order	00. = (not assigned)
Family	00.052. = Picornaviridae
Subfamily	00.052.0. = (no subfamilies)
Genus	00.052.0.01. = Enterovirus
Subgenus (serogroup)	Superseded by species concept
Species (type species)	00.052.0.01.001. = Poliovirus
Species	00.052.0.01.007. = Poliovirus
Subspecies	00.052.0.01.007.00. = (not assigned)
Serotype	00.052.0.01.007.00.001. = Poliovirus 1 00.052.0.01.007.00.002. = Poliovirus 2 00.052.0.01.007.00.003. = Poliovirus 3
Isolate (strain)	00.052.0.01.007.00.001.001. = PV-1 Mahony 00.052.0.01.007.00.001.002. = PV-1 Brunhilde 00.052.0.01.007.00.002.001. = PV-2 Lansing 00.052.0.01.007.00.003.001. = PV-3 Leon/37

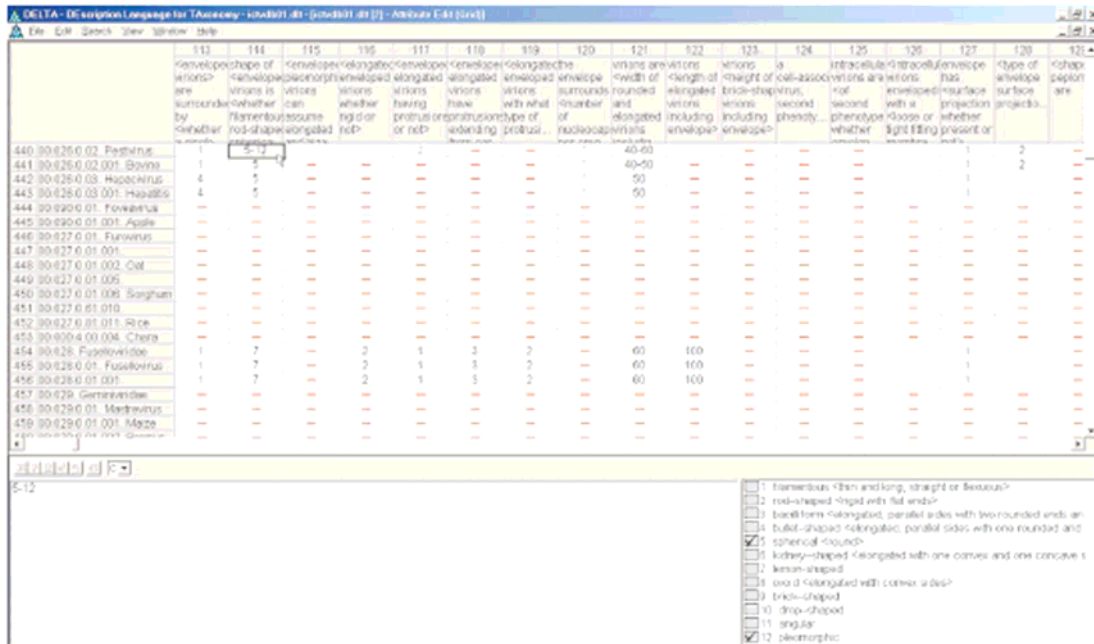


Figure 2. Screen capture of the spreadsheet view in the Delta editor. The red bars indicate that a cell is made inaccessible through dependencies.

```
# 00.005. \i{Astroviridae}\i0{/
83,8 84,1 85,2 87<00.005.> 88<00005FAM> 89<05.> 90<05000000> 91,39733 121,1
122,1 123,5 137,3 185,2 186,2 188,1 189,2 191,27-30 199,1
200,1<star\endash{like with five or six points> 203,1 209,2 210,3 211,2 280,2
285<[[http:]www.vir.gla.ac.uk/staff/bhellad/Astro_510_50k.jpg];
[[http:]www.vir.gla.ac.uk/staff/bhellad/Astro_510_50k.jpg]> 294,1.36-1.39 333,2
336,2 338<presumed to be>,2 339,1 340,2 357<about>,6800-7900 508,45-47 562,1
580,2 584,39733 585,39733 597,1 1133,2 1699,3 1701,3 1730,4 1887,13 1910,3
1911,6/7 1992,8/9/16/18 2656,2 2659,22 2670<[rf_astro.htm]> 2671<Monroe SS,
Carter MJ, Herrmann JE, Kurtz JB, Matsui SM> 2673,39733
```

Box 1: ICTVdB description of *Astroviridae* in DELTA-format, the source code for natural language translation.

6. The database is presented on the web as a [natural language output](#) that is styled to match ICTV Reports. They are NOT scanned copies of ICTV reports, but are based on codes selected from the 2700 characters and generated from the [encoded data](#) (Box 1) in ICTVdB. On average, ICTVdB on the web is consulted every 30s through the US site alone (Fig. 2) and is mirrored in Rothamsted, UK, and Academia Sinica.
7. The database core is a decimal code (Table 1) that serves as a unique identifier of every entry and at the same time indicates the taxonomic status of the entry. This core serves also as the link into ICTVdB from sequence databases such as GenBank, Swiss Prot etc that use the database as the link to ICTV approved taxonomy and nomenclature.
8. IntKey, the interactive data retrieval and identification program of the database, can also constructs descriptions on the fly from available properties on some 3,400 species and isolates that are available to ICTV study groups imported from other databases (Table2), and from on-line entries from researchers via ICTVdB to ICTV study groups. Work is proceeding on some 8,000 other isolates submitted for inclusion ([Table 3](#)).

Table 2: Nature and origin of isolate data in ICTVdB

Database Name from which Isolate data derived	Numbers of Isolates
Orbis database from CD Fort Collins, USA	726 (in ICTVdB)
Foot-and-Mouth Disease, IAH Pirbright, UK	8126 (work in progress)
Arbovirus database, NICD, South Africa	1326 (work in progress)
Isolate description from literature and direct submission	1508 (+ 726 orbiviruses)
Total number of serotype/isolate descriptions	11686

Action requested

1. Steps be taken to avoid future delays to incorporation of ICTV decisions into ICTVdB
2. Consideration be given to the use of ICTVdB for
 - generation of future ICTV Reports
 - streamlining study group consideration of new taxa by on-line submissions through ICTVdB
 - maintenance of the inventory of isolates to species in ICTVdB to ensure up to date access to information on research into emerging viral diseases and other issues
 - maintenance of ICTVdB itself in the longer term through income generated by a licensing arrangement with a respected entity such as ASM
3. Clarification of the relationship between ICTV and other apparently unauthorised listings of virus species and linkages carrying the imprimatur of ICTV
 - such as the Virus index 2005 published at the Danforth Centre
 - the location of the ICTVnet and list server presently hosted at the Danforth Centre
 - it would make sense to have all database activities of ICTV under a single IT authority.
4. Clarification of the relationship between ICTV and Columbia University (at present likely to be terminated with the end of Cornelia Büchen-Osmond's appointment on March 31 2006), the present sponsor of all ICTVdB activities.

Ian Lipkin and Yves Lyssier at Columbia University have pledge to further support ICTVdB, but new curators need to be found for the day-to-day work on ICTVdB. John Mackenzie has agreed to take on the intellectual leadership for ICTVdB at Columbia University.

Prior sponsors such as the Australian National University have expressed no recent interest in intellectual property issues.

Figure 1 and Table 3 are too large to send by email, please download from the web as indicated.

Appendix 1

Historical background to ICTVdB (documentation available on all events)

Spring 1987: Initial discussion meeting on the need for an all inclusive virus database in Durham NC involved Lois Blaine (ATCC), Mikhail Krichevsky (NIH) and Cornelia Büchen-Osmond (ANU).

Mar 1990: Lois Blaine invited a broad spectrum of people from different areas in virology, informatics and database experts, representing ICTV, EMBL/EBI, NCBI, NIH, USDA and many universities from around the world, to a think-tank workshop in ATCC Rockville MD, to establish if there was a need for a universal virus database. ([Program and Documentation](#))

It was decided to petition the Executive Committee of ICTV with a recommendation to build a **Universal Virus Database** based on the successful model of the plant virus database VIDE developed by AJ Gibbs at ANU, Canberra, using DELTA (DEscription Language for TAXonomy) developed by MJ Dallwitz, CSIRO Entomology, Canberra. The DELTA system is a versatile taxonomic database that has been accepted as the world standard. Its core data matrix can be used to translate coded data into camera ready printed descriptions, keys, phylogenetic trees and distance matrices.

Aug 1990: The [petition](#) was tabled and accepted at the next ICTV Executive Meeting. Among other things, it was envisaged that the database should serve to generate future ICTV Reports, avoiding the labor intensive assembly and formatting of earlier Reports. During this meeting a new format was proposed for virus descriptions which served as templates for the camera-ready typeset ICTVdB descriptions.

Sept 1991: Initial funding for the development of ICTV was received through a grant to ATCC.
NSF Award #9107464; “Data Integration for Improved Analysis of Microbial Strain Characteristics”.

The first objective was to standardize the descriptors necessary to describe all characteristics of all types of viruses. The standardization effort built on the “ICTV code for the description of virus characters”, published in 1983, a previous ICTV effort which failed on the computer system of early electronic systems, but which was successfully revitalized and improved by AJ Gibbs for the plant virus database VIDE. All other available virus databases were scrutinized for useful descriptors to capture the growing number of characteristics for viruses.

- Nov 1991: Collaboration began with G. Cameron on management of large data systems, EMBL Data Library, which subsequently led to incorporation of links from ICTVdB to the emerging online accessibility of sequence databanks.
- Feb 1992: A meeting initiated by ICTV members, the newly formed virus data subcommittee which included Marian Horzinek (chair), Adrian Gibbs, Brian Mahy and President Fred Murphy, was held at the European Commission in Brussels to discuss funding options. This meeting was presented with a well prepared a grant proposal with the title “Universal Virus Database”, and although the ICTV EC sought funding for ICTVdB development, it failed to find support for the international endeavor. ICTVdB development continued with support from ANU. ([Meeting Documentation](#))
- Dec 1992: First ICTVdB WebPages were mounted on RSBS-ANU server, establishing an early presence for the Universal Virus Database of ICTV that has now been mirrored in the US, Europe and Asia, and is consulted about 2,000 times daily on the US sites alone. ([1998 hit rate graph](#)).
- Aug 1993: ICTV EC meeting in Glasgow during ICV Congress: ICTVdB Progress Report on the standardization of virus descriptors. ([report](#))
- Feb 1994: Marian Horzinek, the virus data subcommittee (VDSC) chair, runs the first ICTV Virus Data Subcommittee Workshop in Doorn, the Netherlands. Discussion meeting on progress and consolidation of list of descriptors ([progress report](#)).
- Aug 1994: Founding of [Species2000 project](#); ever since ICTVdB has provided the most recent approved virus taxonomy to Species 2000. [First publication on ICTVdB](#).
- Sep/Oct 1994: First ideas for a data entry system, EntVir, in discussion with IT student Peter Sienkowsky. The development of a functional data entry system took several years.
- Jan 1995: Agency funding for the development of ICTVdB commenced with a grant from NSF to ATCC (Lois Blaine), CSIRO (DELTA Mike Dallwitz) and ANU-RSBS (Cornelia Büchen-Osmond). NSF Award #9505661: “Planning and Preliminary Design of a World Virus Database (ICTVdb)”.
- Mar 1995: Automated translation of ICTVdB descriptions directly from the DELTA formatted database matrix into hypertext with active links to sequence databanks utilizing the sequence accession numbers as links.

- Apr 1995: ICTV EC meeting in Bethesda at which Claude Fauquet persuaded the EC not to use ICTVdB to generate natural language hard copy for ICTV publications, preferring instead conventional hard copy publication, CD-ROM production of encyclopedic virus descriptions, periodic ICTV Reports on Nomenclature and Classification of Viruses. ([email exchange between Claude and CBO](#), [ICTV progress report](#))
- May 1995: Start of collaborative consultation with EMBL/EBI/Swiss-Prot and NCBI to improve and facilitate active linkage from ICTVdB species descriptions to sequence and publication database records, an ongoing effort
- After a discussion meeting in Rothamsted the longstanding links to the plant virus group there were revitalized and a collaborative data exchange ensued between plant virus genome project and ICTVdB which is still actively pursued. The genotypic database effort was primarily developed by Mike Adams and John Antoniw.
- Apr 1996: [Rothamsted Mirror of ICTVdB established](#).
- Aug 1996: ICV in Jerusalem, [presentation](#) of ICTVdB development and progress
- Aug 1996: Further agency funding for the development of ICTVdB with a grant from NSF to ATCC (Lois Blaine), CSIRO (DELTA Mike Dallwitz) and ANU-RSBS (Cornelia Büchen-Osmond).
NSF Award # 9631047: "Software Tools for Interoperable Taxonomic Databases".
- Sep 1996: [NCBI Mirror of ICTVdB established](#). Collaboration with Scott Federhen, NCBI taxonomy, to improve NCBI virus taxonomy, an ongoing effort
- Jan 1997: [Joining of the plant virus database VIDEdB with ICTVdB](#). This major effort entailed a complete update of the plant virus taxonomy found in VIDEdB which was different then the now prevailing ICTV approved taxonomy.
- [A meeting at Rothamsted](#) resulted in a manifesto which cemented the coordination for a phenotypic and the genotypic database effort.
- Apr 1997: First demonstration of the data entry system for ICTVdB at the Strasbourg ICTV EC meeting. ([progress report](#))
- Apr 1998: Without consulting the development team, Claude Fauquet invited representatives of Academic Press and Williams and Wilkins (Kluwer) to the ICTV EC meeting in San Diego where he the proposed the sale of ICTVdB to the publishers. The proposal was not accepted, and again, although the ICTV EC was not persuaded to use ICTVdB to generate the 7th

Report, this report included an introduction to the status of the database and its potential was included. ([email exchange](#); [progress report](#))

- Aug 1999: Second ICTV Virus Data Subcommittee Workshop in Canberra is held under the leadership of Marian Horzinek, the outgoing VDSC chair, and Tony Della-Porta, the incoming VDSC chair ([poster](#)).
- Sep 1999: [Claude Fauquet announced](#) signing of contract with Academic Press for Encyclopedia Virology, without discussion or presentation at ICTV EC meeting.
- Jul 2000: [Systems Interoperability and Access to Global Species Databases](#), International Congress of Culture Collections
- Apr 2001: Cornelia Büchen-Osmond moved from RSBS-ANU to Biosphere 2 Center, Columbia University and transferred the web server for ICTVdB to the Medical BioInformatics core at Columbia.
- Oct 2001: Installation of the LinkOut to ICTVdB on NCBI server which enables linkage from the NCBI taxonomy browser to [ICTVdB](#).
- Nov 2001: Testing of the data entry system and ICTVdB by ICTV members: Charlie Calisher starts to test drive the ICTVdB system by entering representative members from data of the arbovirus catalogue. His helpful comments greatly improved the system, particularly the questionnaire, the text used for the virus descriptions.
- Dec 2001: Collaboration with IBIS Therapeutics to correctly identify viral pathogens using ICTVdB as the reference source for approved taxonomy and nomenclature.
- Feb 2002: NLM grant: [1G08LM007198-01](#) "ICTVdB: The Universal Virus Information System" awarded to ATCC.
- July 2002 ICTV EC meeting in Paris (ICTVdB [progress report](#) including Charlie Calisher's arbovirus descriptions, i.e. [St. Louis encephalitis virus](#); [Western equine encephalitis virus](#))
- ICV Congress in Paris: [ICTVdB presentations](#) and [database demonstration](#)
- Aug 2002: The WebPages of ICTVdB, the Universal Virus Database of the International Committee of Taxonomy of Viruses, were copyrighted by ICTV ([email from Jack Maniloff](#)).

- Sep 2002: [Confidentiality agreement](#) between Academic Press and ICTV for publication of the 8th ICTV Report which was signed by ICTV. Agreements of that sort have prevented ICTVdB to update taxonomic information.
- Sep 2002: Collaborative project with Ulrich Melchers and [Virologo](#), Oklahoma State University, and other plant virologists at on the sequence status of plant virus genomes.
- Dec 2002: The Newt project with links to [ICTVdB descriptions](#) is one of the more recent developments by the EBI/SWISS-Prot taxonomy group.
- Apr 2003: Tony Della-Porta (VDSC chair) runs the third ICTV Virus Data Subcommittee Workshop in St Lois ([progress report](#)). This workshop doubles up as a virus data entry training course for ICTV study group chairs.
- Jun 2003: Collaborative project with Judith Brown, UA Tucson, AZ, on Biodiversity & Inventory of New World Begomoviruses. The data entry system of ICTVdB (in [Spanish](#) and English) is used to describe the new virus isolates collected from Mexico and the South West of the US.
- Dec 2003-: Expansion, integration and improved interoperability of ICTVdB commenced with support from Ian Lipkin and Yves Lussier, Department of Epidemiology and BioInformatics Core, Columbia University.
- Mar 2004: Collaborative project with Species 2000 using a XML wrapper kit for "[Annual Checklist](#)" online, a work in progress.
- Apr 2004: Incorporation of 725 original Orbivirus isolate data into ICTVdB received from Dennis L. Knudson, Department of Bioagricultural Sciences and Pest Management, Colorado State University, College of Agricultural Sciences, Fort Collins, CO, USA.
- Sep 2004: Begin of data conversion of isolate data from Foot-and-Mouth Disease Virus database into DELTA format. Peter Mertens, Institute for Animal Health, Pirbright, UK, will make available over 6000 isolate data of picorna- and reoviruses.
- Oct 2004: Collaborative project with Ulrich Melchers and Michael Palmer, Oklahoma State University, on [Plant Virus Biodiversity and Ecology](#). Virus isolates collected during this project will be submitted to ICTVdB.
- Jan 2005: Sara Mullins, ATCC, submits a complete dataset of all 152 SARS isolates described in the literature to ICTVdB.
- Mar 2005: Publication of most recent version of Species 2000 "[Annual Checklist](#)" in CD format and online. ([examples of ICTV virus lists with links to ICTVdB descriptions](#))

Apr 2005: Tony Della-Porta secures virus isolate data from Bob Swanepool, South Africa for ICTVdB. The first batch of about 1,500 isolates records cover around 30 years of arbovirus collection, many of them prototypes of the virus. More data are expected to follow.

July 2005: ICTVdB provides a standardized vocabulary necessary to describe all characteristics of all types of viruses. The questionnaire list contains 2770 characters more than 1500 alone are reserved for host range data. ICTVdB contains presently descriptions of 3 orders; 72 families, 255 genera, 1763 species and 2234 isolates.

Numbers of descriptions	Orders	Families	Genera	Species	Total	isolates	Total
7 th ICTV Report, 2000	3	56	233	1550	1842	n/a	1842
8 th ICTV Report, 2005	3	73	287	1938	2301	n/a	2301
ICTVdB 1999 ICV Sydney	3	64	228	1235	1530	128	1668
ICTVdB 2002 ICV Paris	3	69	246	1364	1682	432	2103
ICTVdB 2005	3	72	255	1433	1763	2234	3997

ICTVdB web sites:

home: <http://phene.cpmc.columbia.edu/>

NCBI: <http://www.ncbi.nlm.nih.gov/ICTVdb/>

Europe: <http://www.ictvdb.rothamsted.ac.uk/>

China: <http://ictvdb.mirror.ac.cn/>

Current research regular collaborators on ICTVdB:

Ian Lipkin, Department of Epidemiology, and Yves Lussier, BioInformatics Core, [Northeastern Biodefense Center](#), a Regional Center of Excellence in Emerging Infectious Diseases and Biodefense, Columbia University

Scott Federhen, NCBI Taxonomy, Bethesda, MD, USA

Rolf Appweiler, EBI/Uniprot, Hinxton Genome Campus, UK

Sandrine Pilbout and her taxonomy group, SWISS-Prot, Geneva, Switzerland

Chris Tidona, Springer Index of Viruses project

John Antinow, Mike Adams, DPV and plant virus databank, Rothamsted Research, UK
Species2000 Project

Judith Brown, Geminiviridae Databank, University of Arizona, Tucson, AZ, USA

Dave Ecker, Ibis Therapeutics, Carlsbad CA, USA

Currently consulting with DARPA, Ibis Therapeutics, Science Applications International Corporation and many others

Print publications describing the development and applications of ICTVdB

- Büchen-Osmond C. 2005. Taxonomy and classification of viruses. In: *Manual of Clinical Microbiology*, 9th edition (in preparation)
- Tidona CA C, Darai G, Büchen-Osmond C. *Springer Index of Viruses*, Springer Verlag, Heidelberg 2nd Edition (in preparation)
- Ecker D J, Sampath R, Willett P, Wyatt JR, Samant V, Massire C, Hall TA, Hari K, McNeil JA, Buchen-Osmond C, Budowle B (2005) The Microbial Rosetta Stone Database: A compilation of global and emerging infectious microorganisms and bioterrorist threat agents *BMC Microbiology*, **5**, 19
doi:10.1186/1471-2180-5-19
<http://www.biomedcentral.com/1471-2180/5/19>
- Büchen-Osmond C. (2003). [The Universal Virus Database ICTVdB](#). *Computing in Science and Engineering* 5 (3), 16-25.
- Büchen-Osmond C. 2002. [Taxonomy and Classification of Viruses](#). In: *Manual of Clinical Microbiology*, 8th edition, Vol 2, Ch 76. ASM Press, Washington DC, p. 1217-1226.
- Tidona CA C, Darai G, (eds) Büchen-Osmond C (special editor) 2002. *Springer Index of Viruses*, Springer Verlag, Heidelberg
- Büchen-Osmond C, Blaine. L., Horzinek MC. (2000). [The universal virus database of ICTV \(ICTVdB\)](#). In *Virus Taxonomy. Classification and Nomenclature of Viruses* F. C. van Regenmortel MHV, Bishop DHL, Carstens E, Estes MK, Lemon S, Maniloff J, Mayo MA, McGeoch DJ, Pringle CR, Wickner R. (eds), Seventh Report of the International Committee on Taxonomy. Academic Press, New York, San Diego
- Büchen-Osmond C. 1997. [Further progress in ICTVdB, a universal virus database](#). *Arch Virol* 142, 1734-1739.
- Büchen-Osmond C and Dallwitz M. 1996. Towards a universal virus database-progress in the ICTVdB. *Arch Virol* 141, 392-399.