THE UNIVERSAL VIRUS DATABASE ICTVDB

The International Committee on Taxonomy of Viruses database is a universally available taxonomic research tool for understanding relationships among all viruses. ICTVdB's fundamental goals are to provide researchers with precise virus identification and to link the agreed taxonomy to sequence databases.

Recent events have catapulted biological weapons to the forefront of public attention. Although government and agencies responded rapidly to many aspects of the threat, people were dismayed to discover the uncertainties involved in identifying different strains of pathogens. Clearly, much diverse information, ranging from soil microbiology to genetically engineered sequences in the genomes of pathogens, was necessary to precisely identify the strains of anthrax bacteria circulating in Fall 2001.

For many microbiologists, this exercise in applying modern systematics under pressure justifies the philosophy adopted a decade ago by the International Committee on Taxonomy of Viruses. Compared with professional groups dealing with more complex organisms, ICTV committed itself to developing a universal database in 1991—one of the earliest initiatives in electronic data management in systematics.[1] One ICTV goal is to use the virus database ICTVdB to describe all viruses of animals (vertebrates, invertebrates, and protozoa), plants (higher plants and algae), bacteria, fungi, and archaea at all taxonomic levels.

This article describes the development and functions of ICTVdB, a database initially developed to facilitate the rapidly changing perception of what is important in characterizing a virus. Because ICTVdB offers precise virus identification and links an agreed taxonomy to sequence databases, it turns sequences into correctly identified viruses. Thus, it serves all areas of agriculture, human and veterinary medicine, and the pharmaceutical industry, and stays abreast of what characterizes a virus. The database uses the Delta system, the world standard for data exchange in taxonomy, which is designed to generate reports from the data matrix in hard copy and HTML format, produce interactive identification protocols online, and calculate distance matrices for cladistic and phylogenetic research to establish hypothetical relationships among organisms. I also describe a few of the data management issues resolved in ICTVdB, including how it is helping sequence databanks turn sequences submitted with incorrect taxonomies into identifiable viruses. ICTVdB has grown in concept and capability to become a major reference resource and research tool.

1521-9615/03/$17.00 © 2003 IEEE
Copublished by the IEEE CS and the AIP

CORNELIA BÜCHEN-OSMOND
Columbia University

Need for a Universal Virus Database

Viruses and viroids are the smallest infectious biological entities that depend on their host for replication. The number of viruses found as pathogens or silent passengers in organisms, from the bacteria to the dominant mammal, is immense. As the international court of experts that rules on names and relationships of all viruses, ICTV now recognizes about 3,000 virus species,[1] but virologists in different fields of biology have tracked some 30,000 virus strains and isolates. Since viruses have evolved many times and infect organisms from all kingdoms of life, researchers want to know more about their origin and evolution, especially in relation to emergence of new diseases and in the context of other threats such as bioterrorism.

As we explore new niches for life, and as the sensitivity and specificity of detection techniques improve, the list of viruses expands and accurate identification becomes paramount. However, we cannot achieve identification without a robust and reliable classification system. Classification is an important identification tool for assessing if the virus on hand is a threat and for starting the decision-making processes that lead to an appropriate course of action.

Taxonomists were among the first to embrace computing as a classification and identification tool.[2–4] With the advent of molecular technologies, such as genomic-sequence analyses, using conventional taxonomic criteria based on morphological and developmental properties was considered old fashioned, less important, and fell into disuse. Unfortunately, sequence databanks accessions are not peer reviewed and contain many sequences of little value because they lack defining information on the organism and other vital specifications. It is a near-impossible task for curators of genomic and protein databanks to sort the wheat from the chaff . Curators have come to realize that the sequence of an isolate of a particular pathogen is most useful if we accurately know the pathogen, its host, the symptoms of its infection, and its geographical distribution. Likewise, laboratories focusing on developing techniques to detect infectious agents, or devising drugs to treat infections, need to know with certainty the pathogens' correctly assigned taxonomies and relationships, informed by the latest information at all levels.

References

1. M.H.V. van Regenmortel et al., *Virus Taxonomy. Classification and Nomenclature of Viruses, Seventh Report of the International Committee on Taxonomy*, Academic Press, 2000, p. 1162.
2. J.M. Dallwitz, "A Flexible Program for Generating Identification Keys," *Systematic Zoology*, vol. 23, no. 1, 1974, pp. 50–57.
3. J.C. Gower, "A General Coefficient of Similarity and Some of Its Properties," *Biometrics*, vol. 27, no. 4, 1971, pp. 857–871.
4. R.W. Payne, "Genkey: A Program for Constructing Diagnostic Keys," *Biological Identification with Computers*, R.J. Pankhurst, ed., Academic Press, 1975, pp. 65–77.

ICTVdB's Development

Initially developed at the Australian National University, and now being maintained and refined at Columbia University's Biosphere 2 Center, ICTVdB owes much to the US National Science Foundation's support and the American Type Culture Collection's sponsorship. The Australian National University realized early on the value of the Web and launched the world's third Web server in 1992 (after CERN and Brazil). Thus, ICTVdB became available on the Web at the beginning of 1993. ICTVdB is now recognized worldwide as a successful prototype biological database and is currently consulted over 30,000 times daily on three mirror sites in the US (www.ncbi.nlm.nih.gov/ICTVdb), the UK (www.ictvdb.iacr.ac.uk), and China (http://ictvdb.mirror.ac.cn).

ICTVdB began as a taxonomic database[2] for understanding relationships among viruses. Virus taxonomy, like any other taxonomy, is in flux because our understanding of relationships between viruses increasingly depends on genomic data that continually challenges earlier decisions based on morphology. Strategies to facilitate communication across semantic boundaries are particularly important in ICTVdB, which deals with data from diverse sources such as bacteriology, agriculture, veterinary, and medical sciences, each of which has evolved a distinctive vocabulary. Although we have standardized terms within ICTVdB, we cannot impose these standards on virologists in all disciplines or impose them retrospectively on the literature.

Taxonomy of most organisms is largely based on morphology. Although scientists have now seen most viruses under the electron microscope, many of today's infectious particles tend to be better known by their chemical and genomic make up, the complex disease symptoms in their hosts, and their vectors and geographical distribution. The human rotavirus, for example, the major cause of diarrhea in young children, is well known as a particle 75 to 80 nm in diameter (see Figure 1), and epidemiologists know that outbreaks of the disease coincide with abrupt

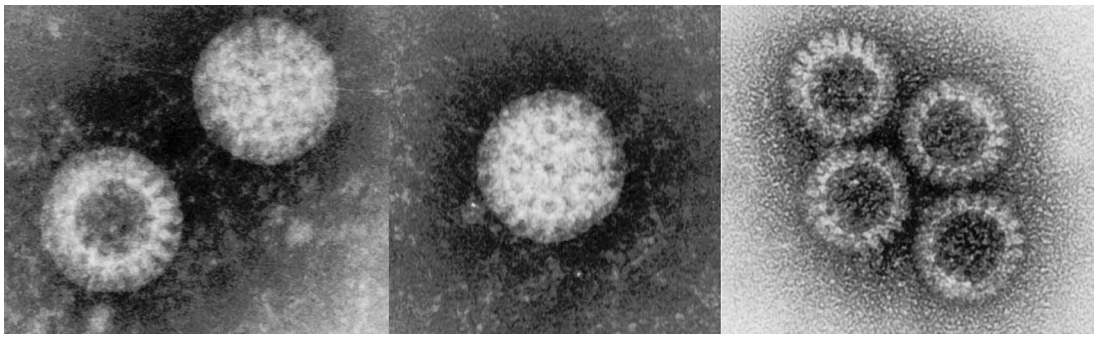

Figure 1. Electron micrographs of a human strain of Rotavirus B. (a,b) A complete virus; (a,c) Stages in the breakdown of virus particles as seen in preparations of stool samples.

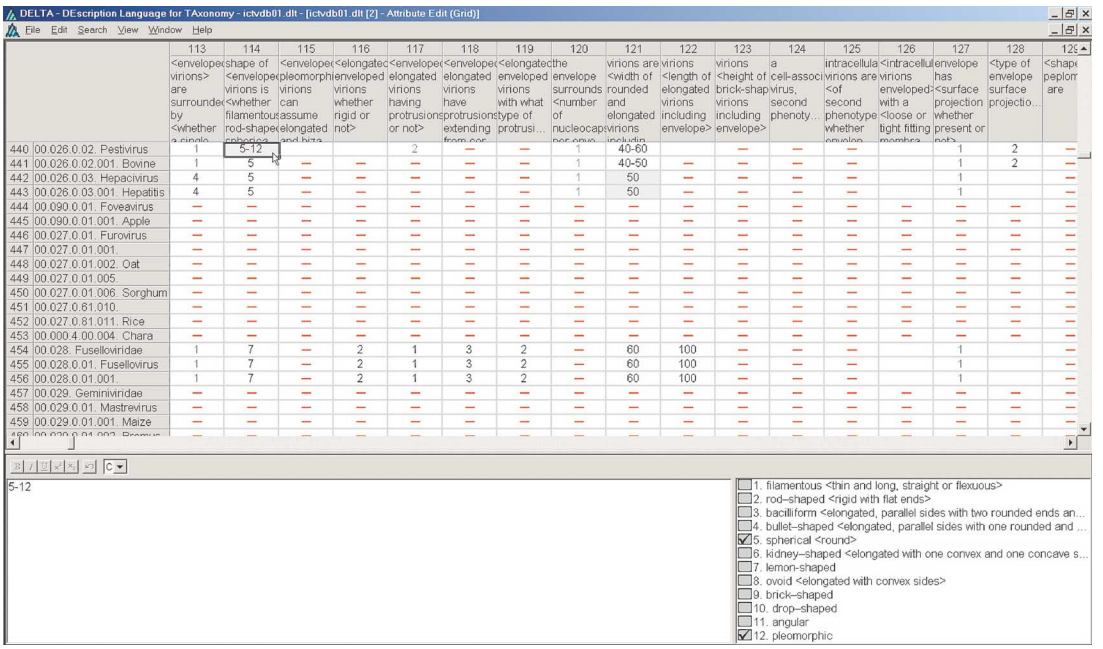

Figure 2. Screen capture of the spreadsheet view in the Delta editor. The red bars indicate that a cell is made inaccessible through dependencies.

temperature changes in winter. These simple statements represent a challenging array of diverse information that, when compared to a sequence's coded data, is difficult to search for in a database. We are addressing a similar challenge as annotation assumes higher priority in sequence databases. In these databases, the problem for annotation is to relate an isolated pathogen's sequence to the correct virus strain and species. Although ICTVdB pioneered many aspects of banking diverse data, a huge task remains to populate the database with historic and contemporary data as researchers recognize and sequence more isolates.

Demands on Delta

ICTVdB uses the Delta (Description Language for Taxonomy) system,[3] which Michael Dallwitz developed at CSIRO Entomology[4] (http://biodiversity.uno.edu/delta/www/delta.htm). Some of Delta's distinctive features include its capacity to store diverse data and to translate these data into natural language for traditional reports and Web publication. Delta's ability to handle large data sets, one item at a time, is ideally suited to the long list of diverse properties (see http://ictvdb.bio2.columbia.edu/chars.htm for a character list) often accompanying the extensive text comments and images we must use to de-

Figure 3. Comparison of phylogenetic analysis of RNA-dependent RNA polymerase sequences of a set of positive-sense single-stranded RNA viruses (tree on the right) and a similar analysis using classical taxonomic characteristics like morphology, physico-chemical properties, and host range of the same set of viruses (tree on the left) The middle column shows ICTV members' consensus taxonomy, which is based on sequence analysis and classical taxonomic characteristics.

scribe an infectious particle. Although only partly populated, ICTVdB already lists more than 2,500 virus descriptions (items) constructed from 2,600 characters, some with up to 2,000 states. By the time we enter all available data on virus isolates and strains, the number of items will be closer to a million.

A key side requirement is that ICTVdB be user friendly, with online data entry available for peer review of new information (http://ictvdb.bio2.columbia.edu/EntVir/index.htm) ranging from a virus' molecular properties to its geographic distribution and host range. Such diverse information, with intrinsic dependencies between genomic data, protein composition, particle structure, and infectivity, places particular demands on Delta's flat-file system. Building a dependency network into the data specification files has accommodated these demands. The Delta editor's spreadsheet display (see Figure 2) is particularly useful for reviewing these dependencies, which is a critical step in developing and working with the ICTVdB data set.

Although Dallwitz designed Delta for taxonomic research, its output formats transcend this specialist interest. Taxonomists can use its translation facilities to construct nearest-neighbor relationships and blend data from diverse sources. Significant events in virology tend to be associated with host jumping, epidemics, and environmental disturbances, all information of which researchers can retrieve from ICTVdB. For example, ICTVdB does not contain sequence data, but conversion of ICTVdB data from Delta into Nexus format was deemed essential by sequence databases users for comprehensive phylogenetic analyses. In this way, we can use classical taxonomic data simultaneously with sequence data, download it from genome databanks in Nexus format, and use it for a phylogenetic data evaluation, as Figure 3 shows. Such capabilities in Delta make it indispensable for monitoring viruses' evolution in relation to emerging diseases, such as the disease associated with the West Nile virus that is unfolding in the US.

ICTVdB's Structural Features

ICTVdB's core infrastructure is its Index of Viruses (http://ictvdb.bio2.columbia.edu/Ictv/ICTVindex.htm), a list of approved virus names sanctioned by ICTV. Because virus names change frequently, contain diverse linguistic and geographical elements, and are usually coupled with a disease or its symptoms, virus nomenclature

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)	Superceded 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

presents challenging semantic problems for a database. Thus, ICTVdB has several distinctive features not usually associated with taxonomy but that we introduced out of necessity. Chief among these is its decimal code.[5] Originally introduced because the peculiar nomenclature used in virology defies direct and systematic interrogation in a database, and because virus taxonomy was changing rapidly, a decimal code (analogous to the code of enzyme nomenclature) seems to offer a simple resolution of diverse problems.

Decimal Code

I constructed the decimal code for the Index of Viruses, and the code serves as a file name for database outputs as well as an access number for externally linking to ICTVdB. The code was inspired by the recommendations on enzyme nomenclature that the International Union on Biochemistry prepared in 1978. In modern terms, it is analogous to the IP number of a PC. The decimal code affords unequivocal identification of a virus to the level of strain or isolate, and importantly, indicates its taxonomic context.[6] For example, the close relationship between two semantically challenging members of the family *Reoviridae*, *Mal del Rio Cuarto virus* (name includes geographical information) and *Nilaparvata lugens reovirus* (name includes vector–host and taxonomic family information) is immediately evident from inspection of their decimal codes (00.060.0.07.004 and 00.060.0.07.008, respectively). Moreover, the code conveniently accommodated the recent, controversial relegation[7] of such widely used species names as Poliovirus 1, 2, and 3 to serotypes on the basis of pair-wise comparison of genomic data (see Table 1).

The prevalence of zeros in Table 1 is designed to absorb the rapid expansion of virus taxonomy in the direction of ever more detailed subdivision of taxa on the basis of sequence data.[8] ICTV has yet to rule on higher-level taxonomy (families, orders, and above) or on controversial concepts such as subspecies. Increasingly, the most pertinent new data are generated by virologists working at the level of serotypes, strains, and isolates. As ICTV revised the taxonomy in light of new data, the coding system will track the history of the decisions from the old to new code, both of which are valid for search purposes. Now expanded to 19 digits to accommodate lower-level taxonomic information, the code should cope with even the most ambitious splitters in the taxonomic community.

As the database grew, it became clear that the decimal code served as more than an unequivocal identifier for taxonomically correct internal linkages within the database. The code is used as a file name for transposing ICTVdB to the Web, and sequence databases such as GenBank (see Figure 4), European Molecular Biology Laboratory (EMBL), and Swiss-Prot (the original protein sequence database of the Swiss Institute for Bioinformatics) that link to ICTVdB use it as a surrogate access number. Taxonomic databases such as the Springer Index of Viruses[9,10] also use these numbers to link to ICTVdB, as does the Species 2000 project,[9] which is the online version of the catalog of life in Figure 5. Be-

cause the decimal code unequivocally identifies a virus and simultaneously indicates its taxonomic status from order to isolate, researchers should routinely cite it in publications.

If a database is to accept the latest data from all branches of virology and place them into contemporary taxonomic context, it will most commonly deal with information at the isolate level. Descriptions at the low end of the taxonomic hierarchy include data needed for an unambiguous identification of and differentiation between strains and isolates, and are of great importance in medicine and agriculture. These data give insight into evolutionary trends, and precise identification becomes especially important for new, re-emerging, or uncommon viruses.

ICTVdB serves virus taxonomy from the bottom up, facilitating the submission of primary data from researchers who use discipline-specific semantics (which often reflect diverse geographic and linguistic features) as well as internationally accepted terminology at the sequence-data level. Simultaneously, the database must also accept revisions and consolidations from the top down, as the consensus in virus taxonomy reflects new information. These taxonomic decisions taken by ICTV increasingly rely on sequence data.

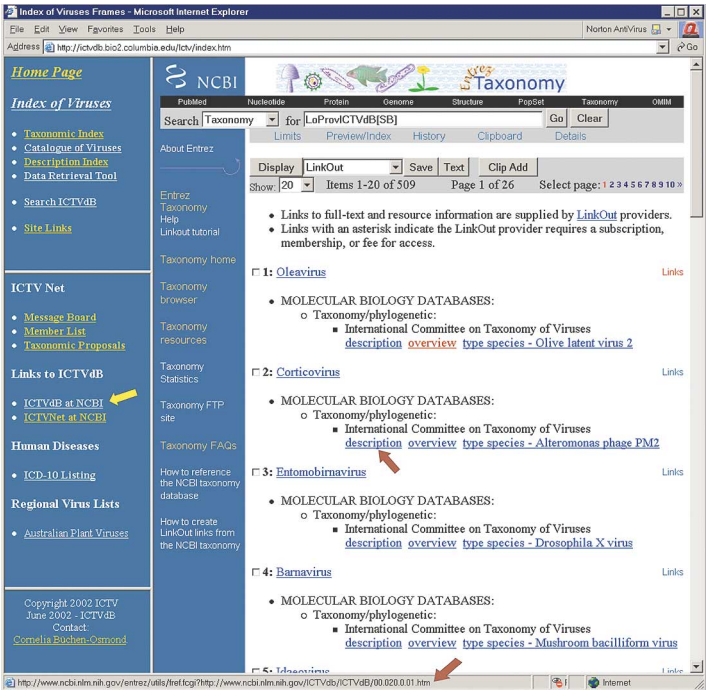

Figure 4. Screen capture showing the LinkOut facilities of National Center for Biotechnology Information's (NCBI)GenBank, which provides links to databases outside NCBI. (LinkOut is an Entrez feature [www.ncbi.nlm.nih.gov/Entrez] that is designed to provide users with links from PubMed and other Entrez databases to a various relevant Web-accessible online resources, including full-text publications, biological databases, consumer health information, research tools, and more. The goal is to facilitate access to relevant online resources beyond the Entrez system to extend, clarify, or supplement information found in the Entrez databases.) The index pages of ICTVdB have a link (yellow arrow) to NCBI, which uses the LinkOut filter set-up for ICTVdB. The list is generated on the fly; only 500 virus descriptions can be linked automatically using the filter because discrepancies between names used in the databanks interfere with the automated alignment. The red arrows point to the link for the virus description on the page and at the address bar to ICTVdB's URL, with the file name using the decimal code.

Dependencies

Unlike many other databases that deal with relatively uniform data types and a few fields, ICTVdB is not a relational database; rather, it is a flat-file system, at least at first glance (see Table 2). All key components of ICTVdB in Delta format (character list, specification, and items file) are readable text files, as are the directive files used for data translation and conversion. Database developers can read and manipulate all these files simultaneously in the Delta editor, which displays the character list and items file on one screen including the specific instructions for data and characters defined in the data specification file.

ICTVdB's character list is distinctive in that it must accommodate data of all sorts, from the geometry of virus particles through the chemical composition of components to the host range and geographic distribution. It also supports these data with explanatory commentary and images. Each character is specified in terms of ordered or unordered multistate properties, integer or real numeric properties, text, and images—the later being handled as a special category of text—all of which the data specification file defines. Table 2 unfolds the specification of the general genomic characteristics of viruses (excluding sequences); <> denotes commentary in the character list and in the items file. Table 2 is an example of an individual data entry using all character types defined in Delta (see specification file, column 1). The natural-language translation of the coded entry (see items file, column 4) reads "Genome is (usually) monopartite; contains RNA; is 9128-9738 nucleotides long (depending on isolate) with a weight ranging between (9.0-)9.2-9.5 or 9.8 (for strain Y). Genome organization: *5'-gag-pro-pol-env-3'*. Genome

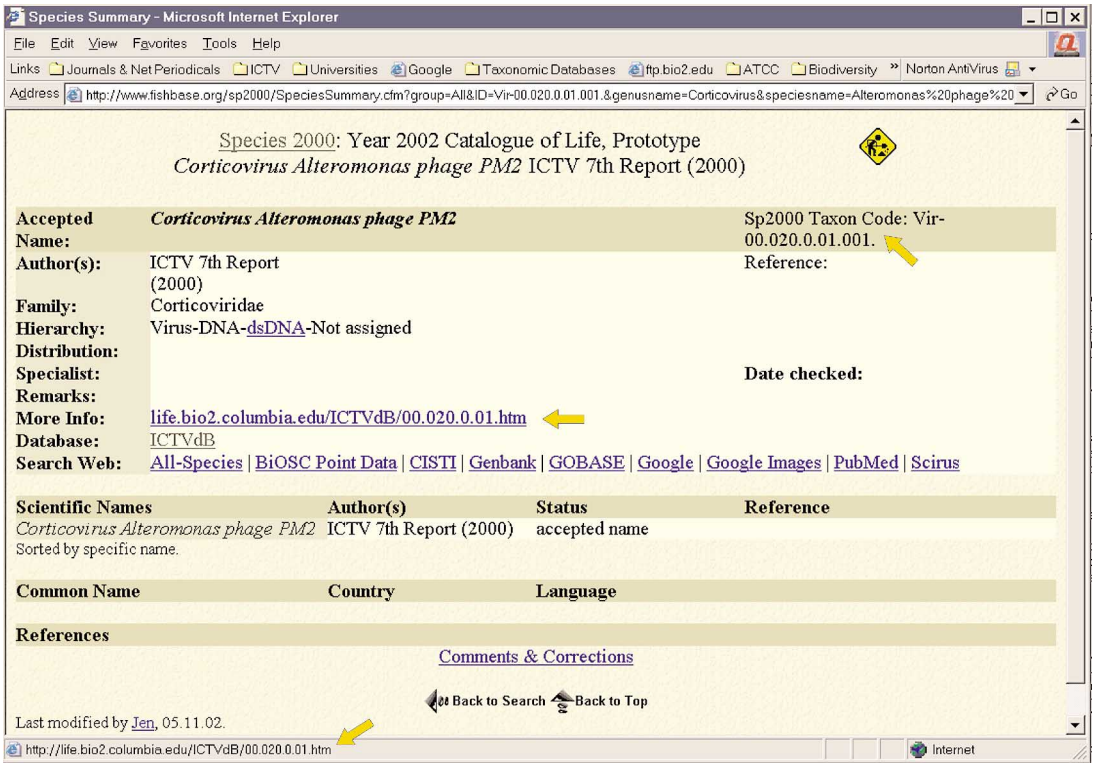

Figure 5. Screen capture of the record for Corticovirus in Species 2000: Catalog of Life 2002. Note at the top right hand corner (yellow arrow) the decimal code that Species 2000 is using as taxon code. Each record is automatically linked, using the decimal code as URL (arrows), to the corresponding ICTVdB description.

map[6] gm_lenti.gif (Figure 6)."

At critical points in the character list, ICTVdB uses binary statements, such as "virus particle with or without envelope," to establish dependencies so that only the subsequently valid characters can be used. These dependencies provide the internal linkages hierarchy in the data, direct the search path during interrogation, and among other things, reveal errors during data entry. The dependencies are important for the decision-making process during identification and data comparison. Some multistate characters in key positions (for example, a plant or animal virus) can control the validity of up to 2,000 characters down the line. The dependencies are automatically indicated in the Delta spreadsheet display (see Figure 2).

Although Delta is considered to be a flat-file system because it is set up in one big table, it is as flexible as a relational database system with many tables since taxonomic relationships are generally linear and not interrelated. In spite of this linear relationship between different taxonomic characteristics, the Delta systems provides dependencies that open or close sets of characters that become applicable or inapplicable depending on the decision-making process. Because all data are in a single table, the dependencies turn Delta into a unique form of relational database—that is, one with many tables within one big table. The tables or blocks of characters become accessible or not depending on the specific relationship of the data they are linked to, as the example for enveloped and not enveloped viruses in Figure 2 shows. These intrinsic relationships are controlled by a dependency network integrated in the data specification file. In contrast to the often cumbersome rebuilding of linkages between tables in a relational database, changing dependencies on the fly is easy, and the built-in error-checking program will alert the user immediately if a logical inconsistency has arisen in the data set as a consequence of the change.

At other points in the character list, ICTVdB uses pseudo-characters to overcome semantic difficulties arising in different subfields of virology and to establish dependencies among blocks of characters. For example, a tailed phage is composed of an isometric head and an elon-

Table 2. Components of a Delta database.

Specification file	Character list		Items file
Type	**Feature**	**Attribute**	**Code**
1,OM ordered multistate	#1. genome is <whether segmented>/	1. monopartite/ 2. bipartite/ 3. tripartite/	1<usually>,1
2,UM unordered multistate	#2. genome contains <nucleic acid type>/	1. DNA/ 2. RNA/	2,2
3,IN integer numeric	#3. genome <length> is/	<number of> nucleotides long/	3,9128-9738 <depending on isolate>
4,RN real numeric	#4. genome with a weight/	kDa/	4<ranging between>,(9.0-) 9.2-9.5/9.8<for strain Y>
5,TE text	#5. genome organization: <order of genes or ORFs>/		5<5′-gag-pro-pol-env-3′>
6,TE image	#6. Genome map <image path to diagram>/		6<gm_lenti.gif>

(see Figure 6)

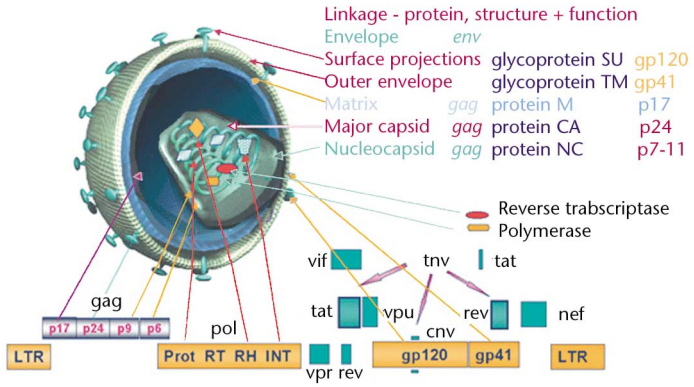

Figure 6. Genome map of human immuno-deficiency virus (HIV) from the ICTVdB database. The genome is transcribed from the 5′-end (left) to 3′-end. The map indicates the positions of the LTR (long terminal repeats) and encoded genes in the sequence. The *gag* gene encodes a precursor of a polyprotein which is translated into MA (matrix), CA (capsid), and NC (nucleocapsid) proteins, *prot* encodes a protease, *pol* a polymerase, and *env* the envelope. The gene order *5′-gag-pro-pol-env-3′* is typical for *Retroviridae*.

gated tail. The head has the same morphological features as the capsid of an isometric virus, and the tail has the same helical features of an unenveloped elongated virus. To avoid having two sets of characters for almost identical morphological structures, pseudo-characters have been introduced which let us choose between the semantic term *head* and *capsid* in the case of isometrics capsid structures, or *tail* and *capsid* in the case of elongated helical capsid structures. Table 3 shows pseudo-character 98, which handles the semantic equivalence of the term tegument, used for some virus description, but is referred to as inner lipid protein membrane in other virus families, or the generally used term capsid, which is called head in the case of a tailed phage. It also shows the dependencies established by the states, so that state 4, as indicated in the right hand column, only opens the character section 646 to 690, whereas all other character sections are made inaccessible. It also handles the semantic equivalence of head and capsid, as previously explained, but excludes the block 691 to 722, because the definition of head does not include the characters describing an inner capsid, as the inner most shell of a double-shelled capsid is called. The dependencies build the database's internal hierarchy and prevent during data entry the prompting of characters that do not apply for a given virus. The dependencies are also important to streamline the data matrix used for interactive interrogation and identification of the data.

Images

Images of virus particles are used in several

Table 3. Semantic equivalencies and dependencies among the major morphological properties of virus particles.

98. <virion or phage> consists of <components of particle>/	Dependent character blocks for proteins
1. an envelope <including inner and outer envelope>/	Envelope. #563-604
2. a surface membrane/	Surface membrane # 605-645
3. a tegument/	Tegument #862-883
4. a head <of phage treated as isometric capsid>/	Capsid (coat protein) #646-690
5. a capsid <including inner and outer capsid>/	Capsid, inner capsid. #646-722
6. a tail <of phage treated as elongated capsid>/	Tail #764-795
7. a nucleocapsid/	Nucleocapsid #723-763
8. a core <including nucleoid>/	Core. #796-840
9. lateral bodies/	Lateral bodies #881-904
10. a matrix/	Matrix. #884-905
11. inner lipid membrane	Tegument #862-883

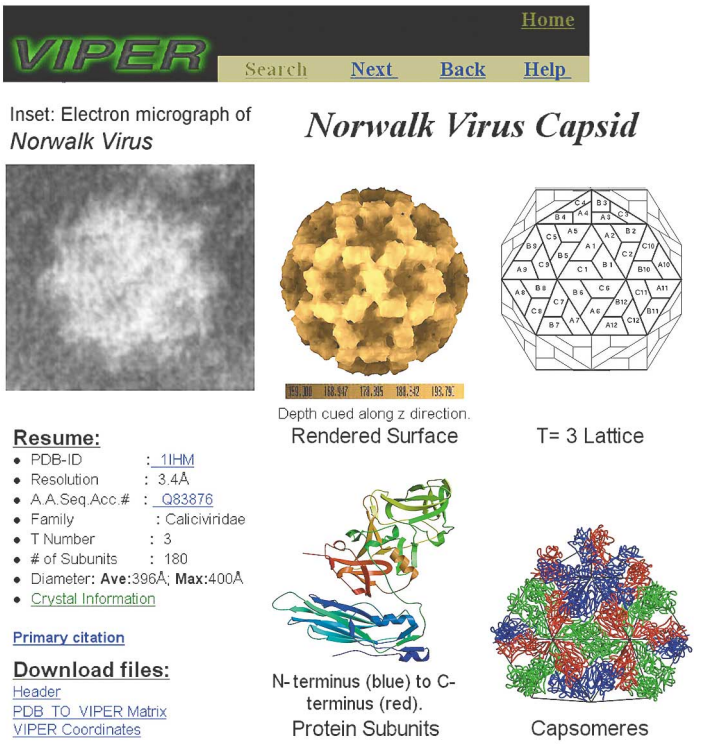

Figure 7. *Norwalk virus.* Electron micrograph and screen capture of a Web page from the Viper database (http://mmtsb.scripps.edu/viper/1ihm.html). The Web page presents the fly generated resume of the protein structure and protein subunits of *Norwalk virus*. I inserted the electron micrograph of a *Norwalk virus* to demonstrate the similarity of the 3D image reconstruction with the native virus preparation.

ways in ICTVdB. For example, text descriptions of key morphological characters become more precise when they are linked in the character list to representative vignettes from electron microscope photographs. Thin-section electron micrograph (EM) images of infected tissues illustrate virus infection cycles and host pathology. EM images of the type species will enhance descriptions of all viruses generated from ICTVdB, regardless of the presentation format selected. Not surprisingly, images of virus particles are among the most frequently accessed files in ICTVdB on the Web. A large image file is more instructive to users, but in the database, it is functionally equivalent to numerous characters, such as "virus 75–80 nm in diameter" in the case of *Rotavirus B* (see Figure 1). File-size considerations and access paths dictate that we store image files outside the main data set, in either local files or files accessed on the Internet.

Figure 7 shows how we can use the outside links to augment the database's content for the example *Norwalk virus*, another diarrhea-causing food-borne virus that is emerging fast as a major cause of severe diarrhea in schools, hospitals, and the hospitality sector (especially aboard cruise ships). We can find a 3D reconstruction of the *Norwalk virus* capsid, lattice information, and structure of the protein subunits forming the capsomers in the Viper database. The Viper Web site (see http://mmtsb.scripps.edu/viper/lihm.html) displays structures of various icosahedral viruses that are stored in the Protein Data Bank (PDB) and provides modeling tools for structural biology. Such detailed protein information is beyond ICTVdB's scope but might be of great value for the database's users.

The original PC-based ICTVdB is available as a natural-language translation on the Web using HTML conversion of the Delta-formatted data. The Web environment is essential for universal access, interactive data entry and interro-

gation, and interoperability with other databases. Currently, a plethora of accessories is available, many of which are standard components of Delta—for example, Web Intkey, an interactive identification program and search facility. The virology research community has developed others, such as the data-entry forms, Java applets, and scripts used to display directory trees, specifically for use in ICTVdB. In the future, XML tagging will vastly improve interoperability, and the Taxonomic Database Working Group is making many efforts to transpose the taxonomic standard developed for Delta into an XML data architecture. Just as we are sure the flow of new information about viruses will not slow, it is certain that new technologies will become available to ICTVdB to handle these data.

Thus far, I have been the only investigator working on ICTVdB's development, with a lot of goodwill and software support from colleagues. The principal impediments to its usefulness and sustainability are common to most biological databases. First, populating the database requires commitment from the virological community for data entry and update. It is essential to have researchers deposit new virus data in ICTVdB at the same time as they deposit sequence data elsewhere. As publication of peer-reviewed data takes new paths, editors can help by insisting that publication be accompanied by assigning or reporting a decimal code of the particular organism. Second, and more important, is the upkeep of the operational system in the face of changing database concepts. All recent attempts to improve the taxonomic standard (www.tdwg.org/nwsltr_p5.html) for data architecture used in the most advanced relational and object-oriented databases systems have come back to the Delta standard, to the same philosophy and algorithms that Mike Dallwitz advanced many years ago. No doubt we can improve the operating system underlying Delta, but not the taxonomic standard itself. Delta has been successful for more than three decades because the developers and users worked together. There is an urgent need to identify mechanisms for long-term support of Delta and ICTVdB and ensure software development to keep the system that has served the world's taxonomists for so many years. CISE

Acknowledgments

The universal virus database project was supported by US NSF grant BIR-9808037 through a consultancy with the American Type Culture Collection, Manassas, Virginia.

References

1. J.G. Atherton, I.R. Holmes, and E.H. Jobbins, "ICTV Code for the Description of Virus Characters," *Monographs in Virology*, vol. 14, 1983, p. 152.
2. C. Büchen-Osmond, L. Blaine, and M.C. Horzinek, "The Universal Virus Database of ICTV (ICTVdB)," *Virus Taxonomy. Classification and Nomenclature of Viruses, Seventh Report Int'l Committee on Taxonomy*, M. H. V. van Regenmortel et al., eds., Academic Press, 2000, pp. 19–24.
3. J.M. Dallwitz, T.A. Paine, and E.J. Zurcher, *User's Guide to the Delta System: A General System for Processing Taxonomic Descriptions*, CSIRO Division of Entomology, 1993, p. 136.
4. J.M. Dallwitz, "A General System for Coding Taxonomic Descriptions," *Taxon*, vol. 29, no. 1, 1980, pp. 41–46.
5. C. Büchen-Osmond and M.J. Dallwitz, "Towards a Universal Virus Database - Progress in the ICTVdB," *Archives of Virology*, vol. 141, no. 2, 1996, pp. 392–399.
6. C. Büchen-Osmond, "Further Progress in ICTVdB, A Universal Virus Database," *Archives of Virology*, vol. 142, no. 8, 1997, pp. 1734–1739.
7. C.R. Pringle, "Virus Taxonomy at the XIth International Congress of Virology, Sydney, Australia 1999," *Archives of Virology*, vol. 144, no. 10, 1999, pp. 2065–2070.
8. C.M. Fauquet, "Taxonomy, Classification and Nomenclature of Viruses," *Encyclopedia of Virology*, vol. 3, A. Granoff, and R.G. Webster, eds., Academic Press, 1999, pp. 1730–1756.
9. C. Büchen-Osmond, "ICTVdB Virus Database (CD-ROM and online) www.fishbase.org/sp2000/search.cfm," *Catalogue of Life: Indexing the World's Known Species*, R. Froese and F.A. Bisby, eds., 2nd ed., Species2000, 2002.
10. C.A. Tidona, G. Darai, and C. Büchen-Osmond, *The Springer Index of Viruses*, Springer-Verlag, 2001, p. 1511.

Cornelia Büchen-Osmond is a virologist and research scientist with the Columbia Earth Institute and is stationed at the Biosphere 2 Center in Arizona. She trained in electron microscopic identification of viruses at the Hygiene Institute, Klinikum JW Goethe-University, Germany. She was invited to develop the universal virus database in 1992 and commenced this work in the Bioinformatics Group at Research School of Biological Sciences, Australian National University. She is a member of the American Society for Microbiology. Contact her at the Columbia Earth Inst., Biosphere 2 Ctr., Columbia Univ., PO Box 689, Oracle, AZ 85623; buchen@bio2.columbia.edu.