VDN
Virology Division News

Identification and classification of viruses that have not been propagated*

J. Maniloff

Department of Microbiology and Immunology, University of Rochester,
Rochester, New York, U.S.A.

Summary

Microorganisms that cannot be grown in the laboratory can now be tentatively identified, by cloning and sequencing particular nucleic acid segments and then carrying out a comparative sequence analysis with an appropriate database. For bacteria, a few universally distributed genes and gene products have enabled comparative sequence analysis to be used for tentative identification and classification of uncultured bacteria. For viruses, there is no universally distributed viral gene or gene product. However, in a few cases, viruses that could not be propagated in the laboratory have been identified and classified. In these cases, either the entire viral genome sequence was determined or partial sequence information was supplemented with additional data. The Executive Committee of the International Committee on Taxonomy of Viruses (ICTV) has reviewed the issue of identification and classification of viruses that have not been propagated. Under the ICTV system, formal review of any taxonomic proposal is carried out by the relevant ICTV Subcommittee or Study Group. The few examples of unpropagated viruses that have arisen thus far have been readily accommodated within existing viral taxonomy, with the international group of experts comprising each Subcommittee and Study Group determining the necessary and sufficient amount of information needed for classification of an unpropagated virus on a case-by-case basis.

Introduction

Recent advances in nucleic acid methodology now enable phylogenetic relationships between organisms to be investigated by analyses of nucleic acid and protein sequence similarities, signature sequences, and genome organization. For bacteria, reconstruction of a fairly detailed phylogenetic tree in the past few years has made it possible to design experiments to identify and classify uncultured bacteria. The general strategy has been to clone (with or without PCR amplification) and sequence specific genes or gene products,

---

like 16S rRNA, and carry out comparative analyses of the sequence data from uncultured bacteria with a bacterial sequence database. A similar strategy has been applied to some unpropagated viruses – a couple of recent well-known cases are hepatitis C virus and Sin Nombre (also referred to as Four Corners) virus.

For both uncultured bacteria and unpropagated viruses, it must be remembered that the adjective "uncultured" or "unpropagated" is an acknowledgment of current experimental limitations. Obviously, uncultured bacteria and unpropagated viruses grow or replicate in nature – the "uncultured" or "unpropagated" modifier simply indicates the inability, thus far, to devise appropriate growth conditions in the laboratory.

The sequence analysis strategy that seems to be working reasonably well for identification and classification of bacteria cannot, in general, simply be applied to viruses, because of basic differences between bacterial and viral molecular biology. However, the experience of microbiologists in using sequence analysis to identify and classify uncultured bacteria provides important lessons in the strengths and weaknesses of this strategy as it may apply to viruses.

For this discussion, the difference between identification and classification must be emphasized. While identification of an uncultured microorganism may be tentative, it can still be important. Knowledge of the properties of phylogenetically related microorganisms points the way for diagnosis, therapeutic and preventive measures, and experimental designs to confirm or question the tentative identification of the uncultured microorganism. However, classification involves taxonomy and nomenclature and has a formal status. Established international scientific organizations are responsible for certification of new taxa, following rules and guidelines meticulously developed over many years and under continual revision in collaboration with the scientific community. For bacteria, the taxonomic organization is the International Committee on Systematic Bacteriology (ICSB) and for viruses it is the International Committee on Taxonomy of Viruses (ICTV).

Identification and classification of uncultured bacteria

Woese and coworkers have defined the properties of a gene or gene product whose sequence is useful for comparative analyses to determine phylogenetic relationships among members of a group of organisms: (i) the gene or gene product must be universally distributed, (ii) it must have functional constancy (to minimize different selective pressure in different organisms), (iii) it must not be laterally transferred, and (iv) its sequence must change in a "clock-like" manner (i.e., accumulate base changes randomly with time) (e.g. [15]). These criteria led to the choice of rRNAs for sequence analysis and phylogenetic tree reconstruction in bacteria, with 16S rRNA being chosen because of its size – 5S rRNA is too small to give good "clock-like" behavior while 16S rRNA is a good chronometer and easier to sequence than the longer 23S rRNA. Therefore, comparison of the 16S rRNA sequence of an uncultured bacterium with sequences in a large rRNA database (the Ribosomal Database Project at the University of Illinois contains over 1,000 bacterial 16S rRNA sequences) allows tentative identification of the uncultured bacterium based on its phylogenetic proximity to bacteria in the database (reviewed in [1, 12]).

It has been found that comparative analyses of 16S rRNA sequences effectively distinguish between bacteria in different genera or well-separated species [5, 14]. However, 16S rRNA sequence analysis cannot always resolve phylogenetic relationships between

closely related or recently diverged bacterial species. This means that organisms with almost identical sequences (>99 % similarity) may belong to different species with significantly different phenotypic properties.

In some cases, comparative analyses of complete 16S rRNA sequences have been insufficient to determine phylogenetic relationships and have had to be supplemented by signature sequence analysis (e.g. [17, 18]). Signature sequences are positions or groups of contiguous positions in a gene or gene product that have compositions characteristic of particular phylogenetic branches, and can be used to define such branches. Signature sequence analysis has been needed to resolve ambiguous phylogenetic branching orders and to identify phylogenetic relationships in rapidly evolving branches.

An uncultured bacterium may yield a novel sequence, and concern has been noted about novel gene sequences obtained from natural samples after PCR amplification [7]. The issue is that, while a novel sequence may indeed represent a novel microorganism, novel sequences also can be generated by PCR sequencing and amplification errors and other experimental artifacts.

This comparative sequence analysis approach has enabled identification of uncultured bacteria in a variety of ecosystems, including two human pathogens (e.g. [11]). In some cases, information available on the growth characteristics of bacteria phylogenetically related to an uncultured bacterium has provided the key for development of media components and conditions for growing the previously uncultured bacterium.

The question of the taxonomic designation of uncultured bacteria has been considered by the ICSB, and the provisional status *Candidatus* established for "procaryotic entities for which more than a mere sequence is available but for which characteristics required for description according to the *International Code of Nomenclature of Bacteria* are lacking" [8]. As noted by Murray and Schleifer [7]: "The whole capability of the organism must be considered in taxonomic arrangements." The designation *Candidatus* indicates a provisional status, not a taxonomic rank.

Identification and classification of unpropagated viruses

No single viral gene or gene product satisfies the criteria (described above) for comparative analysis and determination of viral phylogenetic relationships. In particular, no viral gene or gene product is universally distributed and there are considerable data indicating extensive transfer of viral genes, both between viruses and between viruses and host cells (e.g. [4, 6]). Given the genetic variation that characterizes viruses, particularly RNA viruses, there is also a question of "clock-like" behavior in viral gene sequence changes. Therefore, while there may be signature properties (e.g., presence of reverse transcriptase activity) that narrow identification of an unpropagated virus to a specific family or families, there is no single "universal" viral property or gene sequence, analogous to 16S rRNA in bacteria, whose determination in principle allows identification of all viruses.

Unlikely results can arise from efforts to use sequence data to reconstruct a universal virus phylogenetic tree, because of the lack of a universal viral gene or gene product for comparative sequence analysis and phylogenetic tree reconstruction. This is illustrated by attempts to analyze DNA and RNA polymerase amino acid sequences to determine virus and cell phylogenetic relationships [2, 16]. Some of the conclusions of these studies are that a few bacteriophages (PRD1, M2, and phi 29) are phylogenetically related to a group of

eukaryotic DNA plasmids, and that this phylogenetic branch in turn is related to the adenoviruses. These results are improbable – they postulate a common ancestor for adenoviruses and some bacteriophages, which is implausible in terms of current understanding of the evolution of prokaryote and eukaryote molecular biology and virus-host adaptation.

Phylogenetic trees based on comparative sequence analyses are trees of the gene or gene product whose sequence has been used. On an evolutionary time scale, for prokaryotes and eukaryotes the majority of an organism's essential genes share a common ancestry, in spite of some lateral gene transfer. Therefore, these essential genes can be used to represent organisms, and phylogenetic trees derived from sequences of these genes represent phylogenetic trees of the organisms [10].

The situation in virology is more complicated. The great diversity of viral genome structures and replication strategies means that members of certain taxa (e.g., the Order *Mononegavirales* or Family *Retroviridae*) will have some common essential genes and sequence analysis of these genes can determine phylogenetic relationships among these taxa members. However, members of other virus taxa have different common essential genes and members of some taxa (e.g., the Family *Microviridae*) have no apparent common essential genes.

The recent successful identification and classification of two unpropagated viruses, hepatitis C virus and Sin Nombre virus, provide different examples of the use of sequence analysis. For hepatitis C virus, the entire viral genome was cloned and sequenced. Comparative sequence analyses then showed similarities in signature nucleic acid and protein sequences and genome organization between hepatitis C virus, flaviviruses, and pestiviruses (reviewed in [3]). Based on these data, the ICTV established a third (as yet unnamed) genus to accommodate hepatitis C virus in the family *Flaviviridae*, along with the established genera *Flavivirus* and *Pestivirus*. In contrast, identification of the Sin Nombre virus was based on serological cross-reactivity between patient sera and hantavirus antigens, and PCR amplification and sequence analysis of a 241 base pair hantavirus signature sequence [9]. More recently, the complete Sin Nombre virus genome has been sequenced and its phylogenetic relationship to other members of the genus *Hantavirus* determined [13].

This level of information will continue to be required for classification of unpropagated viruses, because the mobility and variation of viral genes confound efforts to follow the bacterial model and use comparative analysis of a single gene or gene product sequence to determine virus phylogenetic relationships. Also, as observed in the bacterial studies, particular problems arise in trying to use comparative sequence analysis to resolve phylogenetic relationships between closely related, recently diverged, and rapidly evolving organisms, and there is uncertainty about the origin of novel sequences and their appropriateness for phylogenetic considerations. The data needed for classification of an unpropagated virus must be sufficient to address such concerns.

Therefore, ICTV consideration of a proposal to classify an unpropagated virus based on sequence analysis must include either analysis of a reasonably complete genome sequence or a significant partial sequence in addition to enough other information (e.g., morphology, antigenic properties, host range, and clinical features) to provide distinguishing features of the virus. Less complete information may be sufficient for tentative viral identification but not classification.

According to ICTV statutes, consideration of any taxonomic proposal, including review of the available data, is carried out by the appropriate ICTV Vertebrate, Invertebrate, Plant, Fungal, or Bacterial Virus Subcommittee or Subcommittee Study Group. Each Subcommittee and Study Group is composed of an international group of experts and will have to determine the necessary and sufficient amount of information needed for classification of an unpropagated virus on a case-by-case basis. Final ratification of virus taxonomic proposals requires a vote of the full ICTV membership.

The unpropagated viruses that have emerged thus far have been readily accommodated within existing viral taxonomy. Therefore, at this time, no special taxon need to be considered for uncultured viruses.

This statement has been reviewed by and expresses the views of the ICTV Executive Committee. Additional input from the virology community would be useful in further ICTV considerations of this issue, and can be submitted to the author for transmission to the ICTV Executive Committee.

Acknowledgement

References

1. Amann RI, Ludwig W, Schleifer K-H (1995) Phylogenetic identification and in situ detection of individual microbial cells without cultivation. Microbiol Rev 59: 143–169
2. Braithwaite DK, Ito J (1993) Compilation, alignment, and phylogenetic relationships of DNA polymerases. Nucleic Acids Res 21: 787–802
3. Choo Q-L, Kuo G, Weiner A, Wang K-S, Overby L, Bradley D, Houghton M (1992) Identification of the major, parenteral non-A, non-B hepatitis agent (hepatitis C virus) using a recombinant cDNA approach. Semin Liver Dis 12: 279–288
4. Casjens S, Hatfull G, Hendrix R (1992) Evolution of dsDNA tailed-bacteriophage genomes. Semin Virol 3: 383–397
5. Fox GE, Wisotzkey JD, Jurtshuk P (1992) How close is close: 16S rRNA sequence identity may not be sufficient to guarantee species identity. Int J Syst Bacteriol 42: 166–170
6. Koonin EV, Dolja VV (1993) Evolution and taxonomy of positive-strand RNA viruses: implications of comparative analysis of amino acid sequences. Crit Rev Biochem Mol Biol 28: 375–430
7. Murray RGE, Schleifer KH (1994) Taxonomic note: a proposal for recording the properties of putative taxa of procaryotes. Int J Syst Bacteriol 44: 174–176
8. Murray RGE, Stackebrandt E (1995) Taxonomic note: implementation of the provisional status *Candidatus* for incompletely described procaryotes. Int J Syst Bacteriol 45: 186–187
9. Nichol ST, Spiropoulou CF, Morzunov S, Rollin PE, Ksiazek TG, Feldmann H, Sanchez A, Childs J, Zaki S, Peters CJ (1993) Genetic identification of a hantavirus associated with an outbreak of acute respiratory illness. Science 262: 914–917
10. Olsen GJ, Woese CR (1993) Ribosomal RNA: a key to phylogeny. FASEB J 7: 113–123
11. Relman DA (1993) The identification of uncultured microbial pathogens. J Infect Dis 168: 1–8
12. Schmidt TM, Relman DA (1994) Phylogenetic identification of uncultured pathogens using ribosomal RNA sequences. Methods Enzymol 235: 205–222
13. Spiropoulou CF, Morzunov S, Feldmann H, Sanchez A, Peters CJ, Nichol ST (1994) Genome structure and variability of a virus causing hantavirus pulmonary syndrome. Virology 200: 715–723
14. Stackebrandt E, Goebel BM (1994) Taxonomic note: a place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. Int J Syst Bacteriol 44: 846–849
15. Stackebrandt E, Woese CR (1981) The evolution of prokaryotes. Symp Soc Gen Microbiol 32: 1–31
16. Ward CW (1993) Progress towards a higher taxonomy of viruses. Res Virol 144: 419–453

17. Woese CR (1985) Why study evolutionary relationships among bacteria? In: Schleifer KH, Stackebrandt E (eds) Evolution of prokaryotes. Academic Press, New York, pp 1–30

18. Woese CR, Maniloff J, Zablen LB 1980) Phylogenetic analysis of the mycoplasmas. Proc Natl Acad Sci USA 77: 494–498

Author's address: Dr. Jack Maniloff, Department of Microbiology and Immunology, University of Rochester, Medical Center Box 672, Rochester, NY 14642-8672, U.S.A.